Pillars – Pathways to Inclusive Labour Markets

Report

What is the Future of Automation? Using Semantic Analysis to Identify Emerging Technologies

This project receives funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101004703.

What is the Future of Automation? Using Semantic Analysis to Identify Emerging Technologies*

Sugat Chaturvedi, University of Sussex
Ekaterina Prytkova, University of Sussex
Tommaso Ciarli, UNU-MERIT
Önder Nomaler, UNU-MERIT

June 2023

Abstract

Identification of emerging digital automation technologies is critical to understanding the changing patterns of work, firm and industry organisation, and labor demand, and thus formulating policies to mitigate the associated risks while harnessing their potential benefits. In this paper, we analyse a large corpus comprising millions of patents and scientific publications from Derwent, PATSTAT, and OpenAlex databases related to automation technologies across a wide range of domains, including but not limited to industrial robots and artificial intelligence. To identify *emerging* technologies, we propose a methodology which combines machine learning methods with state-of-the-art sentence transformers from the field of computational linguistics. We first identify *radically novel* patents and publications using a novelty detection algorithm and their semantic off-shoots. We then cluster them into cohesive technology groups based on similarity in their content. We validate these clusters based on obtained labels and observe that citation patterns across patents and publications are heavily dependent on semantic similarity. Finally, we construct aggregate indicators of emergence for these technologies and characterize these based on trends in novelty, bibliometric impact, uncertainty, and growth rates during the past decade. We identify six patterns of technological and scientific development, which provide a better understanding of which digital automation techologies are likely to emerge in the near future, and which have matured. The resulting data set of emerging technologies will be useful to practitioners, policymakers, and researchers interested in the implications of these technologies on labour markets and the society.

Keywords: emerging technologies, industry 4.0, text as data, patents, publications

*We are grateful to Deyu Li, Fabien Petit, and participants at the seminar at JRC–European Commission and at the 13th Global Tech Mining and Eu-SPRI 2023 Conferences for helpful feedback. This project received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101004703. The data are made available to PILLARS partners, and will be publicly shared once we publish the working paper.

Chaturvedi: Science Policy Research Unit, University of Sussex, UK, email: sc2057@sussex.ac.uk; Prytkova: Science Policy Research Unit, University of Sussex, UK, email: E.Prytkova@sussex.ac.uk Ciarli: email: ciarli@merit.unu.edu; Nomaler: email: nomaler@merit.unu.edu

1 Introduction

How digital automation technologies will eventually impact societies, depends on the directions in which they will evolve, and what policies are implemented to maximise the gains while minimising the negative impacts for all. For instance, digital automation technologies can contribute to achieving the sustainable development goals (Guenat *et al.*, 2022) or can stand in their way (Ciarli, 2022). Focusing on labor markets implications, studies have provided radically different estimates about the potential impacts of some of these automation technologies on employment and wages (Eloundou *et al.*, 2023; Petit *et al.*, 2023; Hirvonen *et al.*, 2022; Acemoglu & Restrepo, 2019; Nedelkoska & Quintini, 2018; Frey & Osborne, 2017). These studies tend to focus either on non-defined groups of technologies (ICT), or a limited set of automating technologies, such as AI and robots, which have been adopted by a small number of companies (Zolas *et al.*, 2020). The focus on some technologies seems to downplay how digital automation technologies are evolving and combining into several applications within and between firms (Ciarli *et al.*, 2021).

For policies to be able to direct and maximise the benefits of digital automation technologies, policy makers need to be aware of which automation technologies are most likely to emerge in the future, how they combine in which applications, and what is the trajectory they are likely to follow. Keeping with the example of the impacts of digital automation technologies on work, these will also depend on how workers' skills will evolve and which new jobs are created (Autor *et al.*, 2022). For instance, building skills requires time, and is most useful at the beginning of a worker career (Langer & Wiederhold, 2023).

In this paper, we identify emerging automation technologies and areas of scientific advance from patented innovations and scientific publications, respectively. We cover a large spectrum of automation technologies, beyond robots and AI, including data acquisition and data management technologies, computing, networking, additive manufacturing, and user interface (Savona *et al.*, 2022).

Building on a corpus of patents from Derwent, documenting efforts in the private and public sector to develop automation technologies since 2001, we build queries to search the OpenAlex publication repository for scientific publications on digital automating technologies and extract a corpus of publications from 2001. The data set is described in more detail in Prytkova *et al.* (2022) — PILLARS deliverable D3.1. Next, from each corpus we identify novel patents and publications using an anomaly detection algorithm. We also identify the patents and publications that are most similar to those novel patents and publications, i.e. which develop the technology in similar directions to the novel ones — their 'offshoots'. Next, having identified the novel patents and publications and their offshoots, we cluster them into

500 groups that we define as technologies. Finally, for each technology (group of patents or publications) we compute five measures that have been used in the literature to identify attributes of emerging technologies (Rotolo *et al.*, 2015): radical novelty, prominent impact, relatively fast growth, uncertainty, and coherence. These attributes are important to form expectations about the future relevance and evolution of those digital automation technologies. We cluster technologies to identify and characterise different patterns of emergence.

We find over one million patents related to digital automation technologies —mostly from China, US, Japan, and Korea. Among these new technologies, we identify around 90 thousand patents which are novel during the period 2012-2021.[1] We classify those novel patents and their offshoots in 500 technologies and applications, ranging from neural networks and self-driving vehicles, to block-chain and medical monitoring devices. We follow a similar exercise for areas of scientific advance starting with over 4 million scientific publications.

Among those, we groups of technologies and scientific areas that, on average, experienced the quickest increase in the growth rate of patents since 2012 and the fastest increase in novel documents in recent years. The most rapidly emerging and novel broad technologies in patented inventions are neural networks, Natural Language Processing (NLP), augmented reality, blockchain, additive manufacturing (AM), Internet of Things (IoT) and cloud computing. These are followed by robots, workflow automation, control systems and Unmanned Autonomous Vehicles (UAV). The most rapidly emerging applications are energy distribution networks, waste management, clothing recommender systems, secure financial transactions and certifications, recruitment and some applications in health care such as biometric data and health record security. Based on publications the increase in the pace of radically novel advancements is fastest in deep learning (e.g. applied to medical diagnostics) and decentralized finance, IoT based precision agriculture and NLP and dialogue systems.

The future developments of those technologies and applications seem rather uncertain, based on the share of radically novel technologies that emerge yearly.

In patented inventions the majority of digital automation technologies are applications combining several technologies (such as mobile payment, management of energy networks, hospital, schools or buildings, or insurance). There are fewer broad technologies (such as neural networks, additive manufacturing or IoT). However, broad technologies grow at a faster rate in the short run, and evolve rapidly, producing more novel patents. They tend to represent the radical innovations.

We do not observe a unique pattern in the relation between scientific and technological developments. In some cases, developments are more novel and fast growing in science, in others the patented inventions seem to take the lead.

[1]The most novel being deposited by inventors in Germany, Japan, Israel, Sweden and Switzerland.

For both scientific and technological developments in digital automation technologies, it seems that radical novelty goes hand-in-hand with fast growth.

Our results point to the relevance of closing monitoring the wide spectrum of digital automation technologies, to better understand their future developments and what it may mean for the future of work. Our results also provide important indications on which technologies are following a more incremental pattern and where we may expect radical innovations to emerge. We expect more radical change to occur in the technologies, rather than in the applications, therefore associating technological developments with scientific developments seems crucial.

Our work is closely related to the literature on emerging technologies. We contribute to the literature by integrating and operationalising attributes of emerging technologies using specific metrics. We also contribute to the literature focusing on identification of automation technologies and their trajectories (Martinelli *et al.*, 2021; Singh *et al.*, 2021). More generally, we add to the literature that uses natural language processing techniques for mapping and forecasting technologies (Hain *et al.*, 2022). Kelly *et al.* (2021) identify breakthrough patents from 1840 onwards by computing text similarity of a patent with prior and subsequent patents. We instead argue for a more nuanced estimation of "breakthroughness" of a patent by computing novelty relative to the set of patents/publications closest to a given document.

In the rest of the paper, we briefly describe the data (Section 2), and the combination of methods used to identify different patterns of novel emerging technologies (Section 3). Section 4 discusses the resulting patterns of emerging technologies and applications, and Section 5 concludes.

2 Data

We use data on patents and publications related to digital automation technologies in the domains of robotics, data acquisition, data management, computing, artificial intelligence and intelligent information systems, additive manufacturing, networking, and user interface. The dataset comprises 1,143,033 patents from 2000–2021 retrieved from the Derwent Innovation Index (DII). Table 1 reports some examples within these technology families. In addition, we use over 4 million publications retrieved by queries related to digital automation technologies across a wide range of domains from OpenAlex open catalogue—a large open repository for publications. The data is a refined version of that submitted as deliverable D3.1 and is described in a separate paper (Prytkova *et al.*, 2022).

Figure 1 shows the distribution of assignee country for all the patents. We see that China accounts for a large share of patents and is the assignee country for approximately two-thirds of patent families. It is followed by the United States, Japan, and South Korea.

3 Methodology

3.1 Representing Patents and Publications as Vectors

We use state-of-the-art natural language processing (NLP) methods to analyze the text of patents and publications. We first concatenate the patent titles and descriptions. In case of publications, we concatenate their titles and abstracts. We then obtain vector representations (or embeddings) of all the patents and publications using the sentence-transformer architecture proposed by Reimers & Gurevych (2019). Sentence transformers produce semantically meaningful document embeddings so that similar patents or publications are closer to each other in the vector space. In other words, they have shorter euclidean distance and higher cosine similarity between them. This makes sentence transformers suitable for finding similar patents or publications while being orders of magnitude faster compared to other transformer models such as BERT or RoBERTa.

Specifically, we use the pre-trained *all-mpnet-base-v2* sentence transformer model which maps patents and publications to a 768 dimensional vector space by creating a single embedding vector for each document.[2] It fine-tunes the pre-trained *mpnet-base* model trained by Microsoft (Song *et al.*, 2020) for natural language understanding task by applying a siamese triplet network on over a billion sentence pairs from diverse domains such as academic papers, Wikipedia, Reddit comments, and Stack Exchange among others, and has shown state-of-the-art results on semantic search and sentence embedding tasks. The siamese triplet network is trained using a triplet loss function. Given an anchor document A having a corresponding similar document P and a contradictory (dissimilar) document N in the training set, the embeddings are fine-tuned so that the distance between A and P is smaller than that between A and N.

3.2 Finding Novel Patents and Publications

One of the most important parameters for a technology to be considered emerging is *radical novelty*. Therefore, we first identify novel patents and publications. To do this, we use a semi-supervised anomaly detection method. Specifically, we use the Local Outlier Factor (LOF) algorithm proposed by Breunig *et al.* (2000). LOF measures the local density of each patent or publication with respect to k-nearest patents or publications. Thus, the documents

[2]This model is particularly suitable for short paragraphs. It takes input length up to 384 words and truncates longer sequences. For more information, see all-mpnet-base-v2.

that have a substantially lower density than their neighbors are classified as anomalous (or outliers). We use the cumulative set of patents or publications until a given year t-1 for computing the distribution of data and classify new patents or publications in the year t as novel if they are anomalous with respect to the existing ones in the embedding space. We use 258,344 patents from 2000–2011 and 1,294,407 publications from 2001–2011 as the base sample and infer the degree of novelty of patents and publications from 2012–2021. We update the base sample to include patents for an additional year to compute novelty for subsequent years. For example, to compute the novelty for patents and publications in 2013, we also include patents and publications in 2012 in the base sample.

The parameter k defines the local neighborhood with respect to which the relative density is computed. A larger value for k includes more adjacent patents while a smaller k restricts the set of patents to a local neighborhood for density computation which might be useful for detecting minor innovations. In our implementation, we set k = 1000 which is in line with our objective of detecting radically novel technologies. Though this algorithm has previously been used by Lee *et al.* (2015) and Jeon *et al.* (2022) to measure novelty of patents, we extend its usage by incorporating semantic information in patent and publications based on the sentence transformer architecture and include patents from various domains as well as publications.

Appendix Figure A.1 reports the distribution of our novelty measure by year from 2012–2021 for patents and publications. We see that the novelty in case of publications has gradually declined over the past decade. This is consistent with a general decline in disruptive scientific ideas over time as documented by Chu & Evans (2021) and Park *et al.* (2023). However, this declining trend in novelty does not strictly reflect among patents. We define top 10% patents and publications within each year as novel. This gives us a set of 88,413 novel patents and 282,822 novel publications during the period 2012–2021. As a validation check of our novelty measure, we incorporate citation information. In line with our expectations, patents that are among the top 10% in our novelty measure have, on average, 20% more citations than other patents while novel publications have 11% more citations compared to other publications.[3,4]

[3]This relation is robust to (i) different values of nearest neighbors (k = 20 or k = 200) for local outlier factor algorithm, (ii) using a continuous measure of novelty, and (iii) comparing patents and academic publications published in the same year using year fixed-effects.

[4]For computational reasons, we split our sample of publications in 13 clusters based on their citation network using leiden algorithm and then separately run the novelty detection for the set of papers in each cluster. Figure A.2 shows that the proportion of novel publications is roughly same across all the clusters.

3.3 Identifying Technologies

Novel Technologies A single or a small set of novel patents or publications may influence future technological trajectories. For example, the Italian computer vision company VisLab's patent on Vision system for an autonomous vehicle combined multiple cameras with a sensor system to collect information on a vehicle's surroundings and influenced future developments in autonomous navigation technology. Similarly, Kang *et al.* (2016) design an integrated tissue–organ printer for human tissues which represents an important milestone in three-dimensional (3D) bioprinting and scores highly on our novelty measure. Therefore, we enrich our set of novel patents and publications by also including subsequent patents and publications (or *offshoots*) that are close to the novel ones in the semantic space. To identify novel technologies and their offshoots, we start with the set of patents or publications that are among the top 10% of novelty metric within each year from 2012–2021. Given this set, we compute cosine similarity of each patent or publication to the most similar novel patent or publication in each year. The impact set (or offshoot) of novel patents (or publications) is the set having more than 90^{th} percentile of cosine similarity with the nearest novel patent/publication in a given year.[5] This gives us a set of 190,714 novel patents and their offshoots, and 605,932 novel publications and their offshoots.

We then apply k-means clustering on the patent and publication document vectors to arrive at a set of technologies based on semantic similarity. Our choice of using k-means is dictated both by its simplicity and scalability to large corpus. The algorithm partitions the data into K disjoint technologies (or areas of scientific advance) by first iteratively choosing cluster centroids to minimize the within-cluster sum-of-squares (or inertia):

$$\sum_{i=1}^{N} \min_{\mu_k} ||x_i - \mu_k||^2$$

where μ_k corresponds to centroid closest to the patent or publication having vector x_i. It then assigns each point to the nearest centroid. We cluster the novel patents and their (semantic similarity based) offshoots into 500 clusters. We then assign labels to these technologies by extracting top 10 words/phrases within each topic using TF-IDF scores. A critical decision for k-means clustering is identifying the "appropriate" number of clusters. For this we rely on the 4-digit International Patent Classification (IPC) codes. IPC codes hierarchically classify patents into different groups. The full list of codes is available at https://www.wipo.int/classifications/ipc/en/. We, therefore, set the number of clusters as

[5] The 90^{th} percentile of cosine similarity to the most similar novel patent in year t ranges from 80.5% in 2012 to 84.13% in 2020. In case of publications, this ranges from 77.07% in 2012 to 82.07% in 2020. Since 2021 is the final period in our data set, we do not find offshoots of novel patents or publications in 2021.

500 which is approximately the number of distinct IPC categories within a technology domain subject to the IPC code not being rare, i.e. occurring at least 15 times. We also validate this approach using the "elbow-method" heuristic by experimenting with values of clusters from 10 to 1,000 (in increments of 10) and find that the gradient of within-cluster sum-of-squared distance flattens at a similar number of clusters. A visual examination of cluster labels also suggests that having 500 clusters does a good job of separating different technologies while grouping together those that are similar. To maintain comparability with the level of granularity of technologies, we use the same number of clusters for assigning scientific publications to fields of research.

To further validate these clusters, we use citation information. We see that over 22.7% citations for patents come from within the same cluster (or *narrow technology*). This is despite having a small number patents in each individual cluster. To further check the co-citation pattern within and across clusters, we use the Balassa revealed comparative advantage index (RCA) borrowed from the literature on trade (Balassa, 1965). Intuitively, this index measures the overlap in citations across any ordered pair of cited and citing clusters (a_1, a_2) normalized by the number of total number of citing and cited patents in these clusters. Specifically, for cited cluster d and citing cluster l, this index is computed as follows:

$$RCA_{a_1,a_2} = \frac{N_{a_1,a_2}/N_{a_1}}{\sum_p N_{p,a_2}/\sum_p N_p}$$

where N_{a_1,a_2} is the number of forward citations in cluster a_1 from cluster a_2, N_{a_1} is the total number of citations received by cluster a_1, $\sum_p N_{p,a_2}$ is total number of citations across all cited clusters p from citing cluster a_2, and $\sum_p N_p$ is the total number of citations across all cited clusters p. Therefore, larger values of RCA indicate a higher degree of overlap between the cluster pair a_1 and a_2. We indeed find that the median value of RCA is much higher when $a_1 = a_2$ (183.45) compared to when this is not the case (3.95). For over 88% of cited and citing clusters, RCA is the highest when the cited and citing clusters are the same.

We also check the specialization of clusters related to the novel technologies in 15 manually constructed queries related to the core technologies. These queries pertain to (1) neural networks (2) block chain (3) Machine translation, NLP, and text/speech analysis (4) Robotics: parts and operations (5) Advanced intelligent communication/control systems (6) Autonomous navigation (7) Real-time/dynamic systems and embedded systems (8) Machine learning (9) Computing (10) Automated workflow/workflow control, RPA (11) Secure data transmission (12) User interface: software, devices/peripherals, protocols (13) Wireless communication and networking (14) Virtual reality, CAD, 3D modeling/digital simulation (15) Additive manufacturing.[6] The proportion of patents retrieved among the novel and offshoot set is around 32%. Around 9.2% clusters specialize only in a single query. These include RFID (specializing in wireless communication query), additive manufacturing, and cloud storage among others. On the other hand, more applied clusters such as those related to power consumption and biometric authentication specialize in as many as 9 queries. We carry out a similar exercise for publications and find qualitatively similar patterns—14% clusters specialize in a single major query. Figure A.3 shows the distribution of diversity of novel clusters in queries.

Broad Technologies To identify broader technologies associated, we again use k-means clustering. In this case, we further cluster the 500 technologies into 100 clusters based on the centroid for each cluster. In case of patents, we also cluster the entire set into 40 technology clusters.[7] This gives us different clusters related to technologies such as block chain, autonomous vehicles, three dimensional printing, neural networks, wireless communication, cloud computing, augmented reality among others and combinations of these. Figure A.4 shows a representation of these technologies projected on a 2-Dimensional space using UMAP (McInnes *et al.*, 2018) and the top 3 labels associated with each of these. To validate these clusters, we again use citation information and find that over 46% of citations come from within the same technology family.[8] We find that the median RCA is much higher when $a_1 = a_2$ (21.52) compared to when this is not the case (0.27).[9]

We also use RCA to check the specialization of these 40 major clusters in the manually crafted queries described above. We report a heatplot of RCA in queries across these major clusters in Figure A.5. We find that the patterns of specialization are intuitive—management systems cluster specializes in queries related to automated workflow/RPA; blockchain cluster specializes in query related to block chain; mobile based GPS in queries related to navigation and networking; industrial robot in robotics and so on. We see that while some clusters specialize in a single query, others are more diverse and specialize in multiple queries. For example, autonomous vehicle cluster specializes only in queries related to autonomous navigation and three dimensional printing cluster specializes only in queries related to additive manufacturing. On the other hand, healthcare, social network, and authentication system

[6]We are able to retrieve approximately 29% of patents within our original set of patents using these queries.

[7]We try cluster numbers ranging from 2 to 150, in increments of 2, and obtain 40 as a reasonable number based on the elbow method.

[8]This excludes the set of patents that don't have any citations, and therefore, includes 263,430 cited patents and 298,051 citing patents.

[9]In fact, for all citing and cited clusters, RCA is the highest when the cited and citing clusters are the same.

clusters specialize in as many as 8 queries.

3.4 Estimating Emergence

To identify emerging technologies, we follow the characterization given by Rotolo *et al.* (2015) who define emerging technologies as those that have five attributes—*radical novelty* in method or function of the technology, relative *fast growth* (or "clockspeed nature"), *coherence* that persists over time, *prominent impact*, and uncertainty.

Below we describe our proposed metrics to estimate these. To measure novelty of a given technology within each year in our period of interest, we compute the share of novel patents or publications in each cluster, i.e. those among the top 10% in our novelty metric within each year. We measure relatively fast growth by the gradient of growth starting from the birth year of the cluster (i.e. the first time a patent or publication appears in that cluster), i.e. number of patents or publications divided by the age of the technology (in years)—where age is the time period since the first patent/publication related to a technology or area of scientific advance was published. Though, our clustering approach based on semantic similarity guarantees some degree of temporal coherence of technologies/areas of scientific advance, we provide a quantitative measure of coherence by computing the mean of cosine similarity between patents or publications within a cluster. In other words, the clusters having documents that closer to each other in the semantic space on average are defined as more coherent. We measure prominent impact using forward citations for a patent or publication. Specifically, we consider patents that have among the top 10% citations in each year as impactful. We then compute the share of impactful patents or publications within each cluster to measure impact. Finally, uncertainty is conceptualized as the number of novel patents or publications in the most recent year as a proportion of all patents or publications within that cluster. The idea behind this is that if most of the novelty in a cluster appears in the most recent year, this implies that the trajectory of a technology is still uncertain, and therefore, the uncertainty in its potential applications is yet to be fully resolved.

Finally, we identify emerging technologies by categorizing the 500 technologies and 500 areas of scientific advance into 6 groups based on the patterns of novelty, impact, and growth of patents and publications within each cluster in each year from 2012–2020. To do this, we first obtain the aggregate measures of novelty, citations, and number of patents within each cluster for each year. This gives us 27 variables. We then standardize these variables to assign them on equal scale and take a quadratic polynomial transformation for all these variables—including all interactions between them. Finally, we apply k-means clustering to obtain patterns of emergence for all the technologies or scientific areas.

4 Results

4.1 Patented inventions

4.1.1 Geographical distribution

We present the geographical distribution of novel and impactful patents in Figure 2 after purging year fixed-effects. We find that, on average the patents having China as the assignee country are among the least novel while those having assignee country as Turkey, Brazil, India, the United States of America, and Canada are the most novel. However, patents from Turkey, South Korea, Japan, Brazil, Malaysia, Russia, and China receive significantly fewer citations. On the other hand, patents having assignee country as Ireland, United States of America, Canada, Israel, and Sweden receive significantly higher number of citations. This results in a geographic selection of patents in the novel and offshoot set. As depicted in Figure 3, the proportion of patents having China as the assignee country has significantly reduced to around 38% while proportion of patents having assignee country as United States, Japan, and South Korea has increased.[10]

For robustness, we also restrict our sample to 442 clusters for which at least 90% patents were filed in multiple countries and recompute the emergence patterns. This ensures that the patents are of high quality. Figure A.7 shows the geographic distribution of total patents within these clusters while Figure A.8 shows the distribution of novelty and impact across geographic regions. We report the results in Table 4 which suggests a partial overlap in the technologies assigned to each of these patterns. However, the broad patterns of emergence remain consistent across these groups and technologies within patterns 2, 3, and 6 continue to be the emerging technologies.[11]

4.1.2 Patterns of emergence

Figure 4 shows the average number of patents and share of novel and impactful patents for each of the six patterns, sorted from the fastest to the slowest growing. Table 2 reports the average for each of the five attributes over the years, for each of the six patterns. Figure 5 shows word clouds of the technology labels corresponding to these patterns. Figure 6 shows the box plot for the emergence metrics across the six patterns.

In this subsection we discuss the main features of the 6 different patterns. In the following

[10] This also changes the geographic distribution of novelty and impact among patents within the novel and offshoot set (see Figure A.6).

[11] Figure A.9 shows the emergence patterns while Figure A.10 shows the corresponding wordclouds of labels of technologies within each pattern.

subsection we discuss the main broad technologies and applications and how they differ across the six patterns. We recall that in our analysis one technology (or application) is one of the 500 clusters identified in Section 3.3 — in turn clustered in six different patterns — and broad technologies are groups of similar technologies. We list the broad technologies and technological applications (and the pattern in which they where clustered) in Tables A.1 and A.2. To facilitate the interpretation of results, we also list the families and subfamilies of technologies (Table 1) of each broad technology-pattern pair.

The technologies and applications grouped in the first three patterns are particularly fast growing and show increase in the share of novel and impactful patents (at least for part of the past decade). The first pattern of emergence includes a relatively small number of technologies and applications (47). Technologies are largely related to *neural networks*, *augmented reality* and *blockchain*, and to a lesser extent *Natural Language Processing (NLP)*, *additive manufacturing*, *autonomous vehicles*, and the *Internet of Things (IoT)*. Few technological applications where clustered in this pattern, mainly using neural networks (energy and waste management and recommendation systems), blockchain (financial transactions, certificates and health data), sometimes in combination in AM (e.g., dental prosthesis). On average, the number of new patented inventions for these technologies and applications has been increasing with an yearly growth gradient of 18.5, increasing over time at a growing rate. These are the technologies that in terms of expected growth are the fastest emerging. Although the average share of novel patents over the years is below average (47%), this is the only group of technologies for which the share of novel patents has also been increasing at an increasing rate. For this reason, these technologies are also the most uncertain on average (uncertainty score is three time the average across patterns), suggesting that for most of these technologies new trajectories are explored each year. This is also in line with the young age of those technologies that have an average birth rate of 2015 (first patent). Despite the young age and the uncertainty, these technologies are generating the largest share of follow up inventions, showing the highest share of impactful patents (18%).

The second pattern of emergence also includes a relatively small number of technologies and applications (48). Technologies are largely related to *additive manufacturing* with a few technologies related to *neural networks* and *autonomous vehicles.* Just a couple of technological applications where clustered within this pattern, mainly in relation to neural network and NLP (recruitment and product recommendation algorithms). On average, the number of new patented inventions following this second pattern has been increasing with an yearly growth gradient of 15.7, decreasing its pace only since 2018. Differently from the technologies in the first pattern, in the case of additive manufacturing and other technologies following this pattern, the growth is accompanied by a decrease in novelty (which peaks in

2016). The average share of novel patents is below average, and is decreasing. As a result, uncertainty is also relatively low (11) compared to most other patterns, and impact is around average. Although the average birth year of those technologies is similar to those in the first pattern these technologies seem to have found a relatively stable pattern of development along given coherent trajectories (based on their semantic similarity). In sum, technologies within this pattern continue to develop at a fast rate, along trajectories that seem to have established relatively soon, as of 2020.

The third pattern of emergence includes the smallest number of technologies and applications (36). Two groups of technologies characterise this pattern: *IoT* and *cloud computing*, with only a couple of technologies related to *additive manufacturing* and *autonomous vehicles*. Just few technological applications where clustered within this pattern, mainly in relation to neural networks, IoT and data management in cloud servers (recruitment and product recommendation algorithms, event scheduling, remote building management, and peripheral devices). On average, the number of new patented inventions has been increasing with a stable yearly growth gradient of 11.5 since 2013, neither increasing as for technologies in the first pattern nor decreasing, as for technologies in the second pattern. Over the years, these are the technologies with the highest share of novel patents (60%). Because the share of novel do not increase, the uncertainty is lower than for technologies in the first pattern, but is the second highest. This suggests that the IoT and cloud computing technologies in this pattern are still opening into several new trajectories. The share of impactful technologies is also the second highest (15%). In sum, like for technologies in the first pattern, technologies in this pattern are still developing in several direction, at a fast rate, and generating a large number of further innovations. Unlike the first group, these technologies have been around for longer on average (average birth years is 2014) and have evolved into novel technologies at a high rate.

The fourth pattern of emergence includes more than one third of all technologies and applications (187), with similar growth rate, novelty, and impact pattern. Three main groups of technologies, related to industry automation, have followed this pattern: *robots*, *workflow automation* and *intelligent control systems*. Some of the groups of technologies discussed before also follow this pattern: *neural networks*, *IoT*, *cloud computing*, *autonomous vehicles*, and *mobile devices*. These are usually related to applications in industry automation. This pattern also includes the majority of technological applications. Some of the applications combine robotics, data management and networking technologies (maintenance of vehicles, car sharing applications, management of construction sites and projects, logistics and delivery of goods, ordering, cooking and delivering food, management of energy consumption, insurance, disaster management, medical imaging, extraction of biometric information). A second group of applications combine technologies in networking and data managing, without the robotic and RPA components, and in some cases involving user interfaces (management of accommodation, schools, building, and parking spaces, intelligent homes). A third group of applications combine AI and intelligent information system with user interfaces, and in some cases data management technologies (product recommendation, event scheduling, teaching platforms and applications, travel information, health monitoring and medical images, sales scheduling, intelligent homes, gaming). A fourth group are mainly user interfaces (peripheral devices, media players, e-trading). On average, the number of new patented inventions has been increasing at a fast rate for these technologies, with no sign to wane down as of 2020. Like technologies in the third pattern, the growth has been increasing at a stable rate, on average; but unlike them, novelty has decreased since 2012, and stabilised at a lower level (above 50%). On average these technologies have been around for longer (the average birth year is 2012) and are relatively settled, alike technologies in the second pattern (average uncertainty score at 12). Impact is still relatively high and growing (to 20% of the patents in 2020). In sum, these are still emerging technologies, attracting a lot of inventive activity, although in directions that are relatively stable when compared to technologies in patterns 1 and 3.

The technologies and applications grouped in the last two patterns (5 and 6) have been growing at a relatively low rate since 2014, on average, with no sign of increasing, and have experienced a fall in the share of novel and impactful patents (especially pattern six, down to only 20% of novel patents). The fifth pattern of emergence includes a relatively large number of technologies that are mainly applications (91). We find only few technologies, which are not specific to this pattern, related to *control systems*, *autonomous vehicles*, *mobile devices* and *Radio Frequency Identification (RFID)*. This pattern also includes a large number of technological applications, which tend to combine with pattern four. That is different technologies combined in the same applications follow different patterns of emergence. Applications that are particular to pattern five combine networking technologies and user interfaces (call centres, digital media, mobile advertisement, auctions), technologies in networking and data management, in some cases involving user interfaces or AI (locations services, mobile payments, medical images, e-commerce, advertisement), technologies in AI and intelligent information system with user interfaces (social network applications), or are based mainly on networking technologies (power supply, emission of coupons, e-mail). On average, the number of new patented inventions has been increasing at slow rate since 2014. Unlike technologies in the first four patterns novelty keeps decreasing over the years (reaching below 50%, on average). Like technologies in the fourth pattern, these technologies have been around for longer (the average birth year is 2012) and are relatively more settled, with a low average uncertainty score of 6. Impact is also relatively low (8% over the years), suggesting little influence on the development of future technologies. In sum, these technologies show a stable but decreasing pattern of emergence, producing less and less novel patents, at a relatively slow rate, and with little impact over following inventions.

The sixth pattern of emergence also includes a relatively large number of technologies that are almost only applications (90). We find only one technology in this pattern: *Radio Frequency Identification (RFID)*, all other technological developments being applications. Most applications in this pattern are shared with pattern four and five, that is they share some of the technological components that make the applications discussed in the other patterns. The applications that are specific to pattern six, combine networking and data management technologies (parking management, teller machines and health care). Novel patents for these technologies and applications are clustered in the first year (2012), in relation to patents between 2001 and 2011. Since then, novelty has declined rapidly and only 20% of patents are novel in these technologies and applications (on average). This implies that the uncertainty is also lowest (3), on average. Impact is also relatively low (10% over the years), signalling a lower influence on future inventions related to automation, with respect to technologies and applications in other patterns. In sum, these technologies show a stable and low pattern of emergence, producing few novel patents, at a relatively slow rate, and with little impact over following inventions.

4.1.3 Technologies and applications

Table A.1 lists all the broad emerging technologies identified in our sample of digital automation patents across the different patterns. These are: neural networks, Natural Language Processing (NLP), augmented reality, blockchain, additive manufacturing (AM), Internet of Things (IoT), cloud computing, robots, workflow automation, control systems, Unmanned Autonomous Vehicles (UAV), mobile devices and Radio Frequency Identification (RFID). Table A.2 lists technological applications extracted from digital automation patents, clustered in the different patterns of emergence. Each technological application is composed of one or more of the 500 technologies identified in Section 3.3. Some examples of technological applications are: management of energy distribution, supply and demand; clothing recommendations; financial transactions; recruitment; car sharing; health care; teaching; and advertisement (see Table A.2 for a full list). The difference between the list of broad technologies and technological applications is of course not clear-cut. Broad technologies include clusters of patents that share a focus on the main technologies (e.g., additive manufacturing), although they often mention an application (which is why they are patented inventions rather than more general publications). Applications include clusters of patents that share a focus on specific applications (e.g., building management), which usually combine several technologies (e.g., control systems, IoT, NLP and user interfaces in the case of building management).

In the patented inventions, we find that the majority of digital automation technologies are actually applications combining several technologies, rather than broad technologies. We also find that broad technologies are more frequently clustered among the fast growing and novel technologies and applications — only a couple of broad technologies are clustered in patterns five and six. Whereas we find only few applications clustered among the first three fast growing and novel patterns. The novelty measure may be partly influenced by a higher similarity in text describing technological applications than technologies, but the growth patterns seem to suggest that in the most recent years the technological race among inventors to develop radically new digital automations has been focused on the development of technologies, while the applications have followed a more stable pattern of incremental inventions.

We discuss the main properties of broad technologies and applications in turn.

Neural networks, NLP, augmented reality, blockchain, AM, IoT, and cloud computing are the broad technologies that have been growing most in the past 10 years, and which are generating most novel invention, i.e., diverging from past technological trajectories. Therefore the following families of technologies seem to be the most emerging. AI & Intelligent Information Systems, particularly Machine Learning, Computer Vision, NLP; AM, particularly Binder Jetting, Material Extrusion, Powder Bed Fusion, Vat Photopolymerization, Liquid additive manufacturing (LAM), Selective Powder Deposition (SPD), 3D Construction Printing (3DCP), 3D Scanning, Powder Bed Fusion, and Sheet Lamination; Data Management Technologies related to Blockchain and Data Sharing; Networking technologies, particularly related to Monitoring and remote control Applications, Network services and applications, IoT networks, Wireless communication and network infrastructures; Computing technologies, related to Local/real-time processing and Distributed computing; and a few User interface technologies related to Biometric Recognition and Extended Reality (XR).

Next is a group of broad technologies related to industry automation, which grows at increasing rate, but less fast than a group above, and which has been generating increasingly novel technologies since 2016 after a period of increased maturity (reduced novelty). These are Robots, Workflow automation, and Control system. These are all included in the Robotics family of technologies, and in particular the following subfamilies: Mobile robots, Robotic control, Robotic navigation, and Software/virtual robots (RPA).

The two main broad technologies that seem to be relatively mature, and which show relatively slow growth in innovative activity are mobile devices and RFID. These are technologies that cut across different families: Networking, User interface, Robotics, Computing, Data

Management Technologies, and Data Acquisition Technologies.

However, these broad technologies are composed of more specific technologies that have been following different patterns, some emerging more rapidly and creating more novelty than others. UAV is an example of a broad technology that includes technologies that were clustered in all the first five patterns. First, we have drones (ID 27), technologies that follow the first two pattern of emergence, that is experiencing high growth rate of patenting activity, and high and increasing share of novelty. These are technologies that combine mobile robotics, NLP, and extended reality. Mobile robotic components are the ones that follow th mos fast emerging and novelty increasing pattern (pattern 1). Second, we observe technologies and technological components related to autonomous driving (of cars) and traffic control (ID 19) following the second and third pattern. These technologies see fast growing patenting activity, but more stable (high) novelty rate (and increasing impact rate). These are mainly networking technologies, related to monitoring and remote control applications connecting the car to the surrounding environment, which would enable a self-driving car to drive safely on a road. The third broad UAV technology includes technologies and technological components to manage self-driving cars (ID 30). These include vehicle and traffic control technologies to manage and control the vehicle, vehicle identification, operate logistic transportation, run safety diagnostics and monitoring, and provide road assistance. Some of these are also networking technologies for monitoring and remote control applications, IoT networks and wireless communication and network infrastructures. But they also combine data management technologies such as data sharing and encryption technologies for secure data transfer, machine learning (AI & Intelligent Information Systems), distributed computing and cloud servers, and robotic navigation and RPA (e.g. for logistic transportation, such as cold chains). Many of these are relatively more mature technologies and applications, that see a less fast growing patenting activity and which were more novel around 2012 than now.

UAV are no exception in combining technologies with different patterns of emergence. For instance, technologies using neural networks for speech recognitions (ID 94) include fast emerging and novel development of NLP in pattern one as well as more mature development of user interfaces technologies in pattern four. Or IoT technologies for managing and operating warehouses (ID 29) combine fast emerging and novel technologies in pattern one with more mature and less fast growing wireless communication and network infrastructures, and network services and applications in patterns four and five.

4.2 Publications

Figure 7 shows emergence patterns across different areas of scientific advance while the corresponding wordclouds for the cluster labels are reported in Figure 8. Table A.3 reports the detailed list of technologies and their applications in the 6 patterns. Table 3 reports the number of distinct technologies, average number of publications per technology, share of radically novel and impactful papers, their growth gradient, internal coherence, uncertainty, and birth year for each of these patterns. Figure 9 depicts a box-plot showing the heterogeneity in our emergence metrics within each pattern.

Pattern 1 is characterized by 27 technologies exhibiting exponential growth with a growth gradient of 69 publications per year since inception, and sharply increasing novelty and impact especially over the last 5 years of our sample period. The technologies in this pattern are mostly young with an average birth year of 2016, and hence the number of publications per technology is relatively small (493). Overall, approximately 60% of the publications are among the top 10% novel and 22.5% are among the top 10% most cited with respect to papers published in the same year—indicative of high impact. The pattern is also characterized by a high degree of uncertainty in the future direction of the technology as nearly 43% of the novel papers per technology were published in the last year of our sample period. This pattern includes technologies such as blockchain, deep learning, cloud computing, and internet of things (IoT) with advances in specific applications such as cryptocurrencies and decentralized finance, deep learning with radiomics—in particular for cancer detection and diagnosis, and Industrial IoT (IIoT). Notably, this pattern also includes Generative Adversarial Networks (GANs) which have led to major advances in the field of computer vision.

Pattern 2 includes 33 technologies pertaining to areas such as computer vision, wireless communication, and natural language processing with applications related to object and target detection for autonomous navigation (e.g. real-time traffic detection), IoT-based precision agriculture, and question-answering, dialogue systems, and named entity recognition (NER) systems. This pattern is also associated with an exponential growth in the number of publications (with an overall growth gradient of around 30 publications per year per technology) over the past 10 years accompanied by an increase in the number of radically novel publications. However, the increase in the number of novel publications per year has stagnated. Overall, nearly 49% of the papers within this pattern are radically novel as per our definition, of which 20% were published in the most recent year—representing a moderate degree of uncertainty.

The pattern 3 includes 63 technologies, many of which were highly novel and experienced rapid growth in the early 2010s, but have since stagnated. This is accompanied by a decline in the number of radically novel publications per year. The average number of publications per technology over our sample period is nearly 2,000 indicating a relatively mature stage of technological development. This pattern includes technological advances in areas such as Robotics, Electroencephalography (EEG)-based brain-computer interface, computing and encryption (e.g. Silicon Photonics or AES encryption), and cloud computing and data management services. Some specific applications include Prosthetic Arm, Robot-assisted arm training after stroke, Advanced Driver Distraction Warning Systems (ADDW), and Steganography and Digital Watermarking.

Pattern 4 includes 176 steadily growing technologies (with average growth gradient of 20 publications per technology), with a relatively high number of novel papers per year. This includes advances in areas such as Additive Manufacturing, Bio-inspired Reinforcement learning for (intelligent) Robots, big data storage devices, smart material (e.g. Shape-memory polymers, Electronic Skin), Quantum Computing, and gesture recognition and sensors (e.g. tactile sensors) for user interface design. Specific applications include Three-Dimensional Bioprinting (e.g. printing implants and bone and tissue regeneration), deep learning for protein structure prediction, neural- network- based landslide/flood susceptibility mapping, Unmanned Aerial Vehicles (UAV), Augmented Reality/Educational Robots for Classrooms, and VR-based assistive technologies and Humanoid/Social Robots for the elderly and for children with Autism Spectrum Disorders (ASD).

Pattern 5 (59 technologies) has the lowest share of radically novel publications (28%) with a substantial decline in novelty and stagnation in the number of papers per technology over the past 10 years (average growth gradient of only 12 patents per year). Many of these technologies have already seen wide deployment and adoption. Consequently, there is very little uncertainty (2.50%) in their future trajectory. The technologies within this pattern include Radio Frequency Identification (RFID) tags for authentication, Smart Watch and Accelerometer, cache memory (e.g. DRAM), Semantic Web Services, and Electrocardiogram-based Biomedical Signal Processing. Some applications include RFID-enabled supply chain and inventory management system, Ambient Assisted Living Systems, facial recognition, audio-source separation, and time-series/short-term electricity load forecasting.

Finally, Pattern 6 includes 142 technologies with stagnation in growth and an initially declining novelty which stabilized at a relatively high share of radically novel publications. On average 54% of papers within each technology in this pattern are radically novel. This pattern is marked by relatively lower impact (8%) and internal technological coherence (avg. within cluster cosine similarity of 59.7). It includes technologies related to robotic control (e.g. Fuzzy, PID and NMPC controllers), robotic arm and legged robots, high performance computing (HPC), holographic, tactile, and LCD/LED displays, Voice User

Interface and conversational robots, neural machine translation and text classification, and wireless communication infrastructure such as LTE and optical networks. Applications include Remote Laboratory (or Virtual Instruments Systems In Reality) and Virtual Classrooms.

5 Conclusion

In this paper, we develop a new methodology for the identification of emerging technologies (patented inventions) and areas of scientific advance (publications). We apply these on a large corpus of patents and scientific publications to identify emerging digital automation technologies and areas of scientific advance. This can be useful to understand their impacts on the economy and labor markets. The final data set of emerging technologies is likely to be of use to researchers and policymakers. In addition, the novelty measure based on the text of patents and publications itself can be useful in identification of *disruptive* science and innovation in other areas.

We identify a wide spectrum of technologies, applications and research areas related to digital automation technologies. Using indicators of technological emergence that measure fast growth, radical novelty, prominent impact, coherence, and uncertainty, we distinguish different patterns of emergence.

We find that, in patented inventions the majority of digital automation technologies are applications combining several technologies (such as mobile payment, management of energy networks, hospital, schools or buildings, or insurance). There are fewer broad technologies (such as neural networks, additive manufacturing or IoT). This is in line with the nature of patenting innovations that are more likely to be exploited in the market.

However, broad technologies grow at a faster rate in the short run, and evolve rapidly, producing more novel patents. They tend to represent the radical innovations. The technological applications tend to seep for longer and attract more incremental and less novel innovations.

The most rapidly emerging and novel broad technologies in patented inventions are neural networks, Natural Language Processing (NLP), augmented reality, blockchain, additive manufacturing (AM), Internet of Things (IoT) and cloud computing. These are followed by robots, workflow automation, control systems and Unmanned Autonomous Vehicles (UAV). Technologies such as mobile devices (wireless) and RFID seem already pertaining to the past in comparison.

The most rapidly emerging applications are energy distribution networks, waste management, clothing recommender systems, secure financial transactions and certifications, recruitment and some applications in health care such as biometric data and health record security. Many applications including transport, location, advertisement, health, social networks, e-payment, and call centres attract inventive activity, but on rather stable patterns.

We do not observe a unique pattern in the relation between scientific and technological developments. In some cases, such as blockchain and deep learning, scientific and technological developments are both in the fastest emerging patterns that are generating more novel documents. In other cases, such as IoT, although the technology is not amongst the most fast emerging in terms of novelty, related scientific developments are (e.g. in Industrial IoT). In other cases, such as additive manufacturing, it is the scientific development, that although steadily growing, are not developing in new areas as fast as the patented inventions. Results seem to confirm the well known non-linear advances of science, technology and innovation (Kline & Rosenberg, 1986).

Based on publications the increase in the pace of radically novel advancements is fastest in deep learning (e.g. applied to medical diagnostics) and decentralized finance, so we need to carefully understand its potential consequences and barriers to adoption. IoT based precision agriculture seems ready to take-off (pattern 2). We've already seen wide-spread adoption of ChatGPT but we can indeed expect more advances in NLP and dialogue systems (pattern 2). The way in which these will evolve and how they will be used, though, also seems to have a high degree of uncertainty. On the other hand, there seems to be only an incremental progress in case of older technologies such as RFID and Ambient Assisted Living systems (pattern 5) which seems to suggest an end to their development as researchers are increasingly devoting their energy to other areas (based on the growth in publications).

For both scientific and technological developments in digital automation technologies, it seems that radical novelty goes hand-in-hand with fast growth. Perhaps this is because as opportunities to do radically novel research in an area dry up, the research and innovation communities focus their attention to the new hot areas — so there seems to be a lot of dynamism and adjustment.

In terms of policy implications, our results suggest that, first, it is important to consider the full array of digital automation technologies, beyond AI and robots. Combination of data management, data acquisition, network technologies, and user interfaces are contributing to fast emerging applications in markets, sales, logistics, management, financial transactions, hospitality and health, to name a few examples.

Second, the diversity in the patterns of emergence is multidimensional: fast growing is important indicator that some technologies may be taking off. But this is often combined with novelty, which in this paper means that the technologies also change at a fast rate, and it is more difficult to predict how they will evolve and how this will affect other technologies.

Third, because novelty (radical innovations) is generated more in the underlying technologies than in their applications (incremental innovations), it is important to track the scientific and technological developments of the fast growing and most novel technologies to improve our understanding of their future directions.

References

Acemoglu, Daron, & Restrepo, Pascual. 2019 (mar). *The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand.* Tech. rept. National Bureau of Economic Research, Cambridge, MA.

Autor, David, Chin, Caroline, Salomons, Anna M, & Seegmiller, Bryan. 2022. *New Frontiers: The Origins and Content of New Work, 1940–2018.*

Balassa, Bela. 1965. Trade liberalisation and "revealed" comparative advantage 1. *The manchester school*, **33**(2), 99–123.

Breunig, Markus M, Kriegel, Hans-Peter, Ng, Raymond T, & Sander, Jörg. 2000. LOF: identifying density-based local outliers. *Pages 93–104 of: Proceedings of the 2000 ACM SIGMOD international conference on Management of data.*

Chu, Johan SG, & Evans, James A. 2021. Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences*, **118**(41), e2021636118.

Ciarli, Tommaso (ed). 2022. *Changing Directions: Steering science, technology and innovation towards the Sustainable Development Goals.* Brighton, UK: STRINGS, SPRU, University of Susse.

Ciarli, Tommaso, Kenney, Martin, Massini, Silvia, & Piscitello, Lucia. 2021. Digital technologies, innovation, and skills: Emerging trajectories and challenges. *Research Policy*, **50**(7), 104289.

Eloundou, Tyna, Manning, Sam, Mishkin, Pamela, & Rock, Daniel. 2023 (mar). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.*

Frey, Carl Benedikt, & Osborne, Michael A. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, **114**(jan), 254–280.

Guenat, Solène, Purnell, Phil, Davies, Zoe G., Nawrath, Maximilian, Stringer, Lindsay C., Babu, Giridhara Rathnaiah, Balasubramanian, Muniyandi, Ballantyne, Erica E.F., Bylappa, Bhuvana Kolar, Chen, Bei, De Jager, Peta, Del Prete, Andrea, Di Nuovo, Alessandro, Ehi-Eromosele, Cyril O., Eskandari Torbaghan, Mehran, Evans, Karl L., Fraundorfer, Markus, Haouas, Wissem, Izunobi, Josephat U., Jauregui-Correa, Juan Carlos, Kaddouh, Bilal Y., Lewycka, Sonia, MacIntosh, Ana C., Mady, Christine, Maple, Carsten, Mhiret, Worku N., Mohammed-Amin, Rozhen Kamal, Olawole, Olukunle Charles, Oluseyi, Temilola, Orfila, Caroline, Ossola, Alessandro, Pfeifer, Marion, Pridmore, Tony, Rijal, Moti L., Rega-Brodsky, Christine C., Robertson, Ian D., Rogers, Christopher D.F., Rougé, Charles, Rumaney, Maryam B., Seeletso, Mmabaledi K., Shaqura, Mohammed Z., Suresh, L. M., Sweeting, Martin N., Taylor Buck, Nick, Ukwuru, M. U., Verbeek, Thomas, Voss, Hinrich, Wadud, Zia, Wang, Xinjun, Winn, Neil, & Dallimer, Martin. 2022. Meeting sustainable development goals via robotics and autonomous systems. *Nature Communications*, **13**(1), 1–10.

Hain, Daniel, Jurowetzki, Roman, & Squicciarini, Mariagrazia. 2022. Mapping Complex Technologies via Science-Technology Linkages; The Case of Neuroscience–A transformer based keyword extraction approach. *arXiv preprint arXiv:2205.10153*.

Hirvonen, Johannes, Stenhammar, Aapo, & Tuhkuri, Joonas. 2022. New Evidence on the Effect of Technology on Employment and Skill Demand.

Jeon, Daeseong, Ahn, Joon Mo, Kim, Juram, & Lee, Changyong. 2022. A doc2vec and local outlier factor approach to measuring the novelty of patents. *Technological Forecasting and Social Change*, **174**, 121294.

Kang, Hyun-Wook, Lee, Sang Jin, Ko, In Kap, Kengla, Carlos, Yoo, James J, & Atala, Anthony. 2016. A 3D bioprinting system to produce human-scale tissue constructs with structural integrity. *Nature biotechnology*, **34**(3), 312–319.

Kelly, Bryan, Papanikolaou, Dimitris, Seru, Amit, & Taddy, Matt. 2021. Measuring technological innovation over the long run. *American Economic Review: Insights*, **3**(3), 303–20.

Kline, Stephen j., & Rosenberg, Nathan. 1986. An overview of innovation. Landau, r, edn. Washington, DC: National Academies Press.

Langer, Christina, & Wiederhold, Simon. 2023. *The Value of Early-Career Skills*.

Lee, Changyong, Kang, Bokyoung, & Shin, Juneseuk. 2015. Novelty-focused patent mapping for technology opportunity analysis. *Technological Forecasting and Social Change*, **90**, 355–365.

Martinelli, Arianna, Mina, Andrea, & Moggi, Massimo. 2021. The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. *Industrial and Corporate Change*, **30**(1), 161–188.

McInnes, Leland, Healy, John, & Melville, James. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Nedelkoska, Ljubica, & Quintini, Glenda. 2018. *Automation, skills use and training*. Tech. rept. OECD, Paris.

Park, Michael, Leahey, Erin, & Funk, Russell J. 2023. Papers and patents are becoming less disruptive over time. *Nature*, **613**(7942), 138–144.

Petit, F, Jaccoud, F, & Ciarli, T. 2023 (jan). *Heterogeneous adjustments of labor markets to automation technologies*. Tech. rept. Munich.

Prytkova, Ekaterina, Ciarli, Tommaso, & Nomaler, Önder. 2022. *Dataset of emerging automation technologies related to industry 4.0. Technical document with the protocol for data extraction*. Tech. rept.

Reimers, Nils, & Gurevych, Iryna. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Rotolo, Daniele, Hicks, Diana, & Martin, Ben R. 2015. What is an emerging technology? *Research policy*, **44**(10), 1827–1843.

Savona, Maria, Ciarli, Tommaso, Steinmueller, Ed, & Vannuccini, Simone. 2022. The design of digital automation technologies: implications for the future of work. *EconPol Forum*, **23**(5), 4–10.

Singh, Anuraag, Triulzi, Giorgio, & Magee, Christopher L. 2021. Technological improvement rate predictions for all technologies: Use of patent data and an extended domain description. *Research Policy*, **50**(9), 104294.

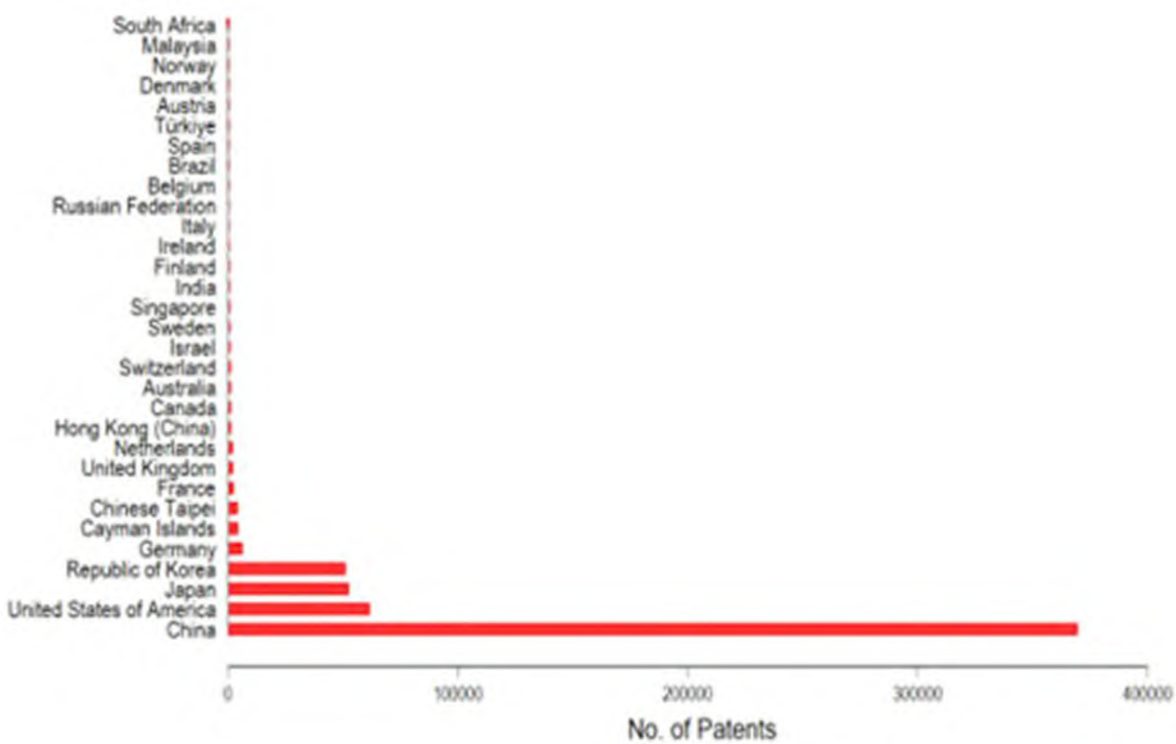
Song, Kaitao, Tan, Xu, Qin, Tao, Lu, Jianfeng, & Liu, Tie-Yan. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, **33**, 16857–16867.

Zolas, Nikolas, Kroff, Zachary, Brynjolfsson, Erik, McElheran, Kristina, Beede, David, Buffington, Cathy, Goldschlag, Nathan, Foster, Lucia, & Dinlersoz, Emin. 2020 (dec).

Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey. Tech. rept. National Bureau of Economic Research, Cambridge, MA.

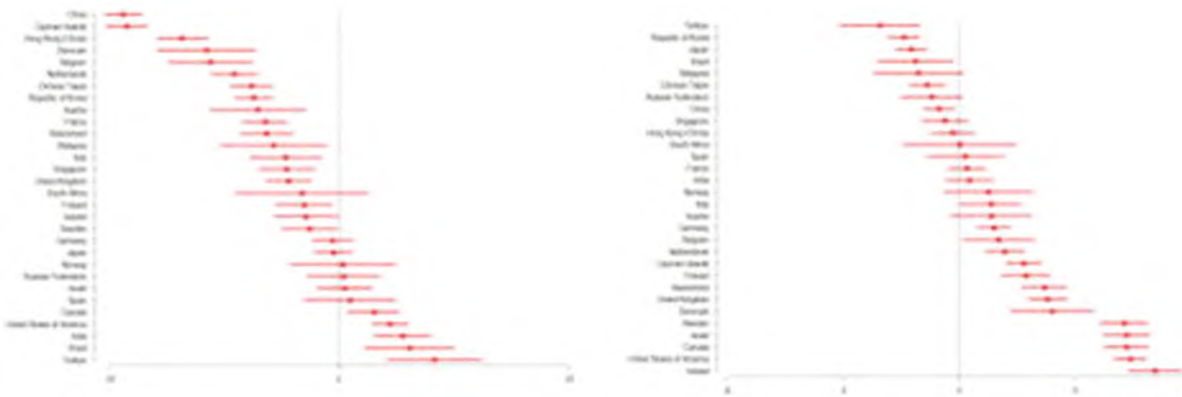
Figures

Figure 1. Number of Patents by Assignee Country for All Clusters



Note: Distribution of assignee country for all the patents. This information is available for 64.55% patents. Out of these, China is the only assignee country for 64.64% patent families.

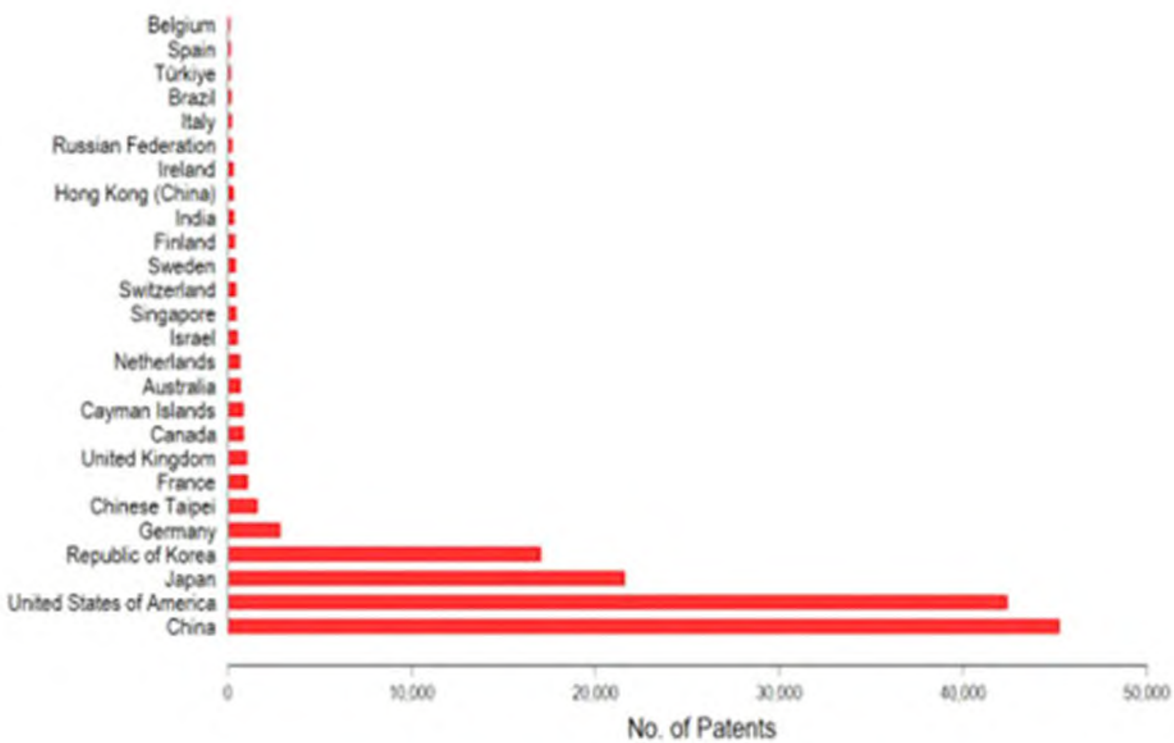
Figure 2. Novelty and Impact by Geography (controlling for year)



(a) Novelty (b) Impact

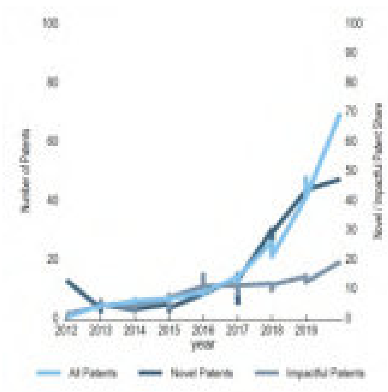
Note: These results correspond to all the patents.

Figure 3. Number of Patents by Assignee Country for Novel Clusters

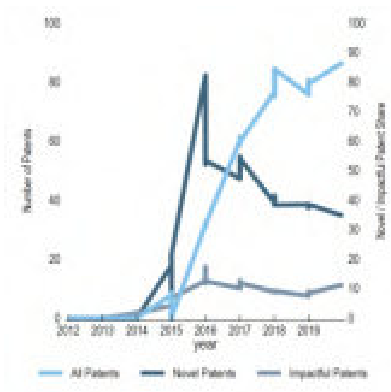


Note: Assignee country for patents in the novel and offshoot set. This information is available for 82.70% of patents. Out of these, the assignee country is exclusively China for 37.78% patents.

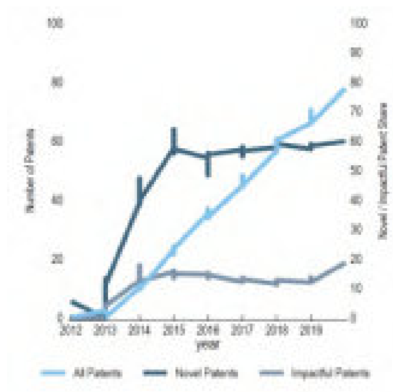
Figure 4. Patterns of Emergence



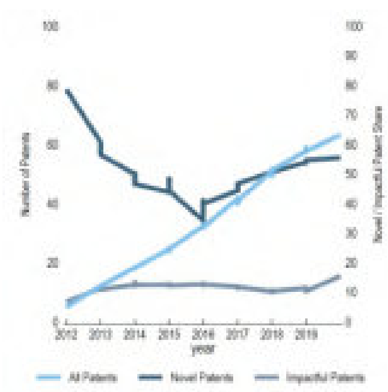
Pattern 1



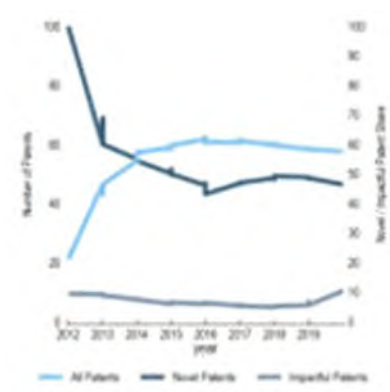
Pattern 2



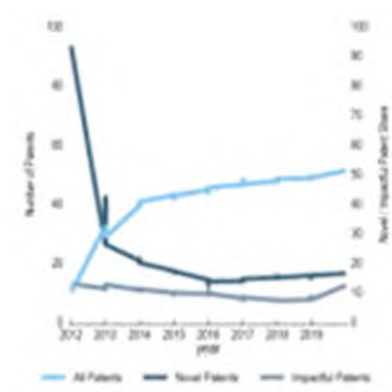
Pattern 3



Pattern 4

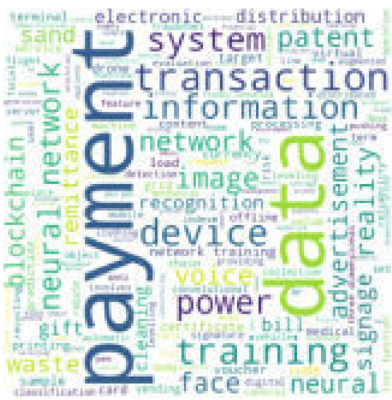


Pattern 5

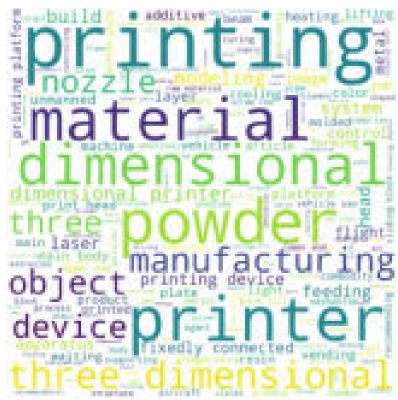


Pattern 6

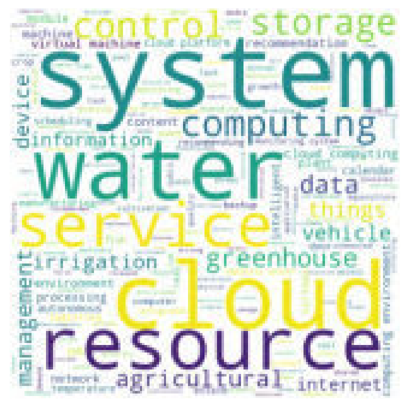
Figure 5. Word Clouds: Emergence Patterns



Pattern 1



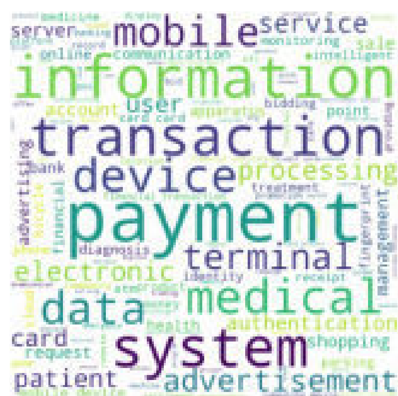
Pattern 2



Pattern 3

Pattern 4

Pattern 5



Pattern 6

Figure 6. Technology Types: Box Plot

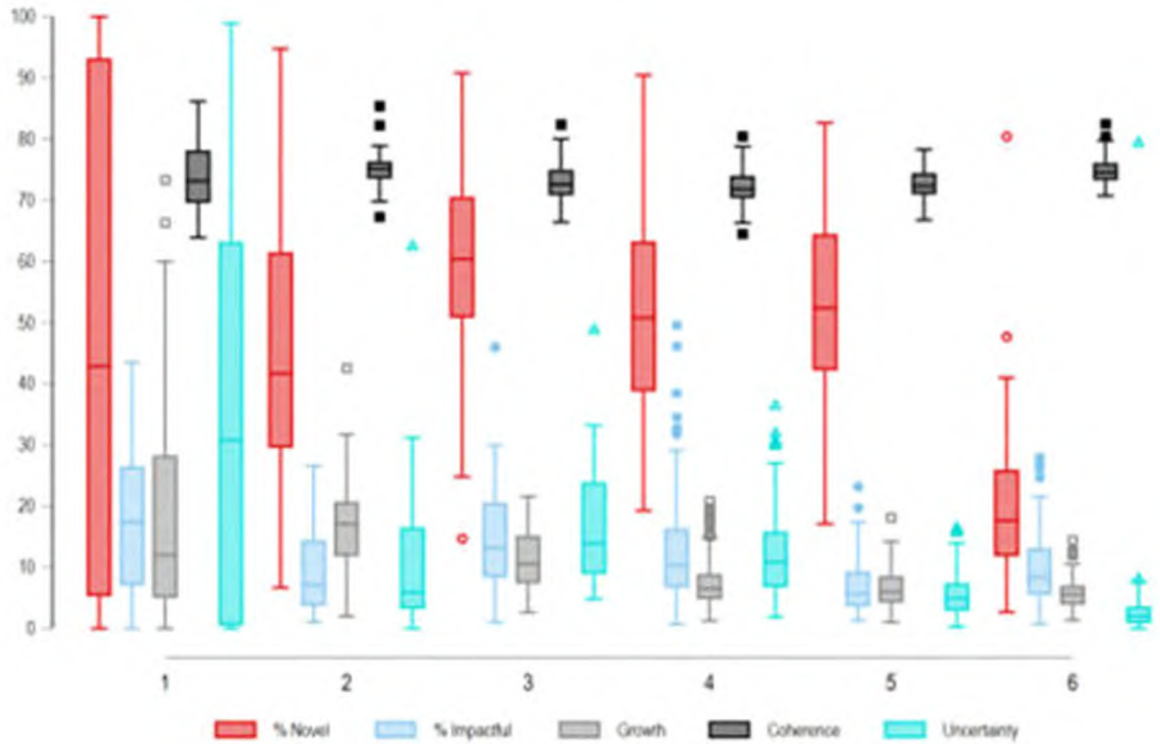
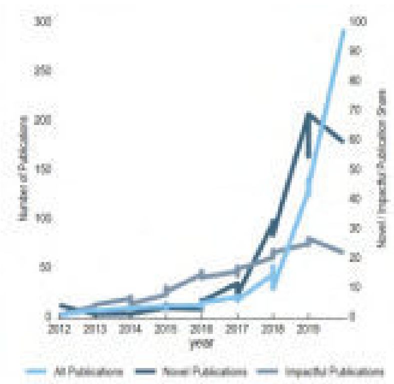
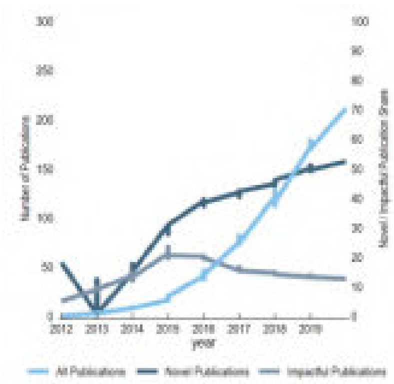


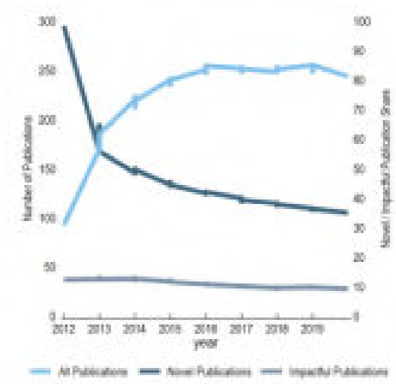
Figure 7. Patterns of Emergence: Publications



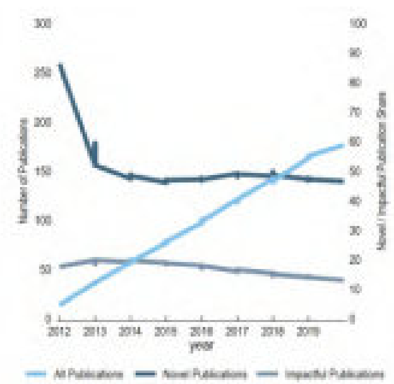
Pattern 1



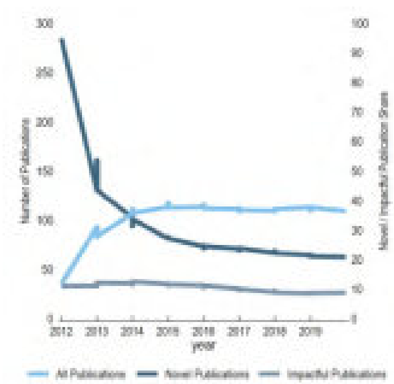
Pattern 2



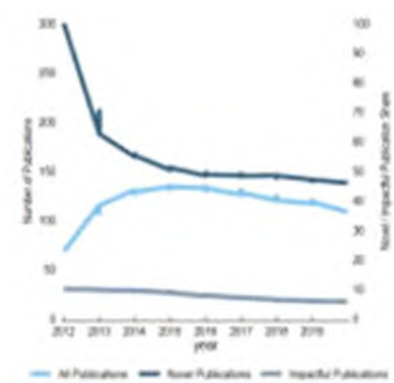
Pattern 3



Pattern 4

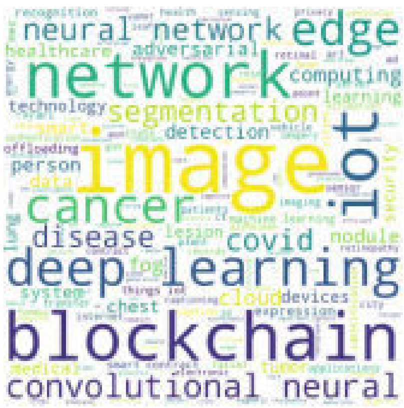


Pattern 5

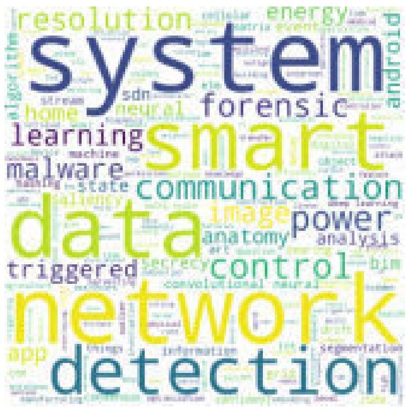


Pattern 6

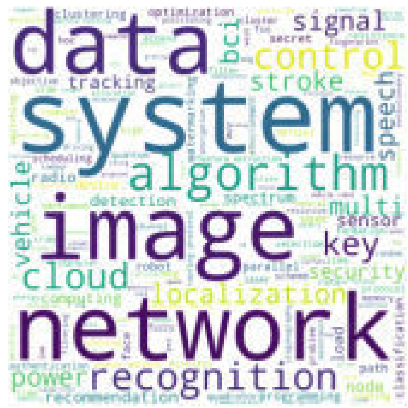
Figure 8. Word Clouds: Emergence Patterns for Publications



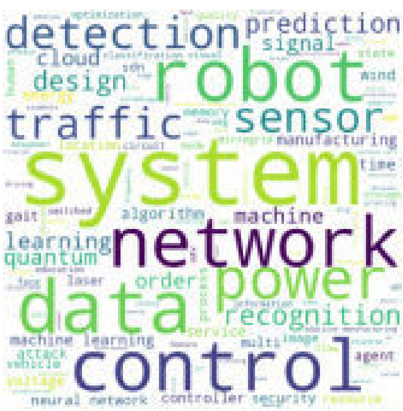
Pattern 1



Pattern 2

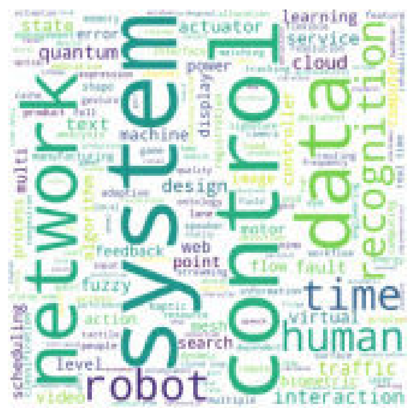


Pattern 3



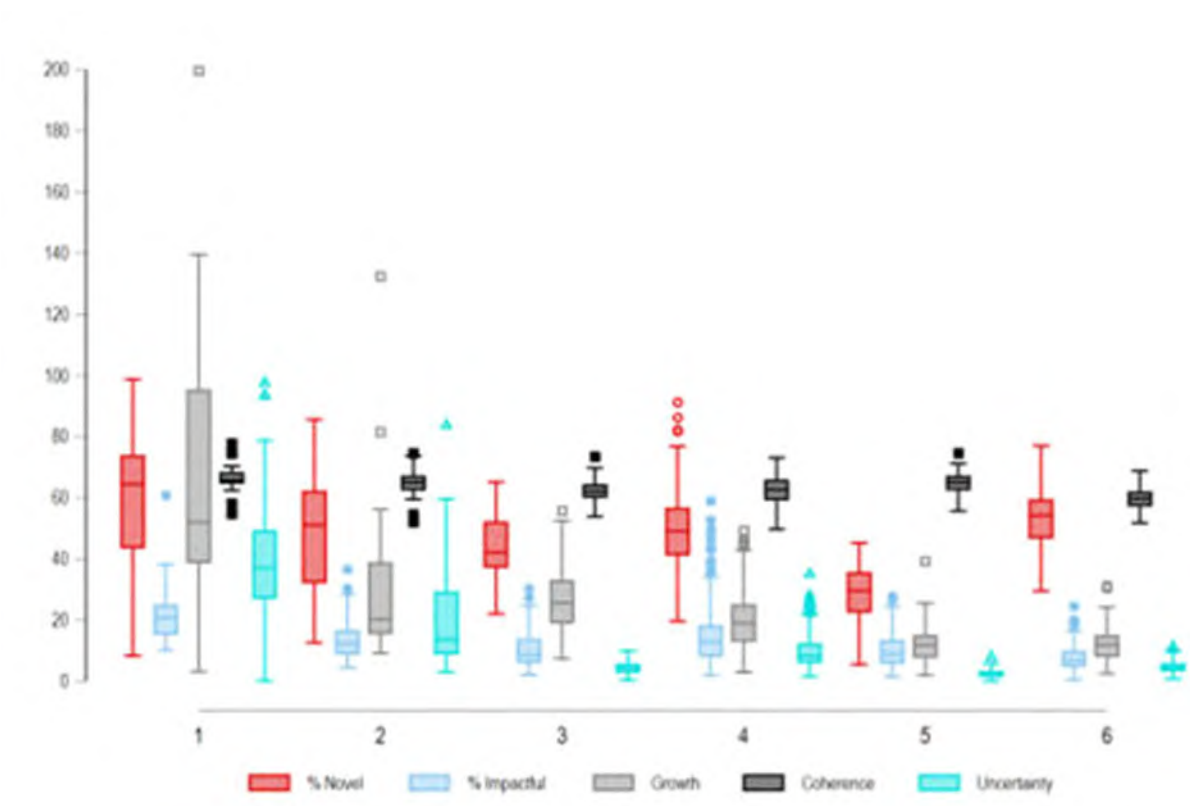
Pattern 4

Pattern 5



Pattern 6

Figure 9. Science Area Types Papers: Box Plot



Tables

Table 1—List of Technology Families

Technology family	Some examples
Robotics	Swarm robotics, robotic vehicles
Data Acquisition Technologies	Remote sensing, IoT, scanners
Data management	Database systems, cryptography, security, blockchain
Computing	computing architectures e.g. cloud, edge, neuromorphic, fog
AI & Intelligent Information System	ML, NLP, multi-modal data processing
Additive manufacturing	CAD, CAM, rapid prototyping, 3D printing
Networking	IoT, wireless communication
User interface	VR/AR, smart personal assistants, interactive holograms

Table 2—Emerging Cluster Patterns: Patents

Pattern	#Technologies	#Patents	% Novel	% Impactful	Growth	Coherence	Uncertainty	Birth Year
1	47	168.83	47.28	17.65	18.47	73.73	32.40	2015.51
2	48	339.44	45.32	9.80	17.15	75.06	10.65	2015.90
3	36	316.55	60.34	14.69	11.47	73.18	16.50	2014.17
4	187	303.92	52.00	11.94	7.24	72.04	11.98	2012.21
5	91	483.08	52.32	6.97	6.41	72.50	5.73	2012.00
6	90	365.46	19.54	9.88	5.72	74.71	3.18	2012.07

Note: The table represents summary of emergence patterns within each pattern. #Patents is the total number of patents within each group, % novel is the share of novel patents, % impactful is the share of impactful patents, growth is the gradient of growth, i.e. number of patents/age of the technology (in years). Coherence is the mean of cosine similarity between patents within a cluster. Uncertainty is conceptualized as the number of novel patents in the most recent year as a proportion of all patents within that cluster.

Table 3—Emerging Cluster Patterns: Publications

Pattern	#Technologies	#Publications	% Novel	% Impactful	Growth	Coherence	Uncertainty	Birth Year
1	27	493.30	59.54	22.56	68.98	66.51	42.75	2016.22
2	33	637.60	49.26	14.48	29.75	64.50	20.08	2013.58
3	63	1994.21	44.33	10.71	27.15	61.95	4.34	2012.02
4	176	882.81	49.45	15.52	19.90	62.59	9.71	2012.14
5	59	913.44	28.38	9.97	12.15	64.71	2.50	2012.05
6	142	1054.00	53.90	7.73	12.04	59.71	4.65	2012.00

Note: The table represents summary of emergence patterns within each pattern. #Publications is the total number of publications within each group, % novel is the share of novel publications, % impactful is the share of impactful publications, growth is the gradient of growth, i.e. number of publications/age of the technology (in years). Coherence is the mean of cosine similarity between publications within a cluster. Uncertainty is conceptualized as the number of novel publications in the most recent year as a proportion of all publications within that cluster.

Table 4—Confusion matrix: Emergence Patterns

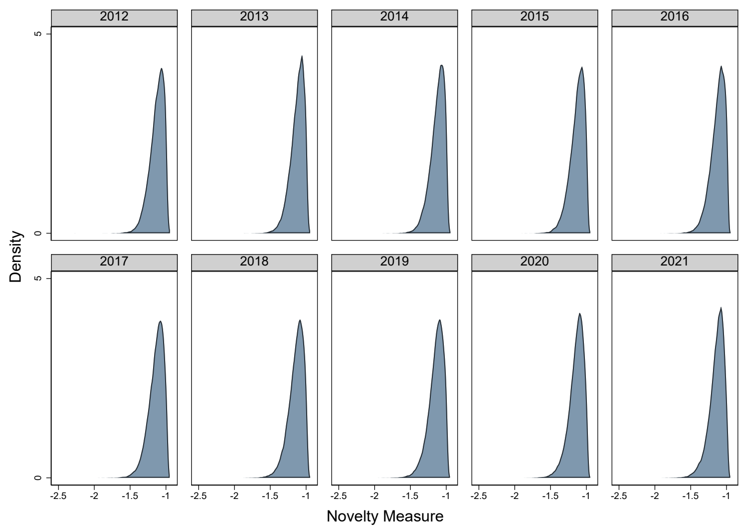
Pattern/Multicountry Pattern	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	64	0	0	22	0	2	88
2	0	34	0	0	0	0	34
3	0	8	20	0	1	0	29
4	0	0	0	42	26	21	89
5	0	12	0	0	22	0	34
6	25	0	14	0	0	129	168
Total	89	54	34	64	49	152	442

Note: Confusion matrix for technology types and technology types for multi-country technologies, i.e. technologies for which at least 90% patents are filed in multiple countries.

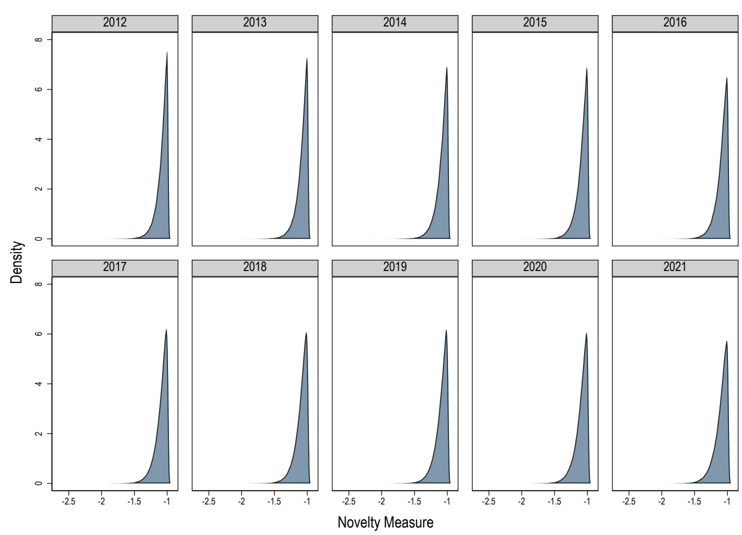
Appendix

Figures

Figure A.1. Novelty Distribution by Year



(a) Patents



(b) Publications

Note: Distribution of the novelty metric for all patents and publications during 2012–2021.

Figure A.2. Distribution of Novel papers across Clusters

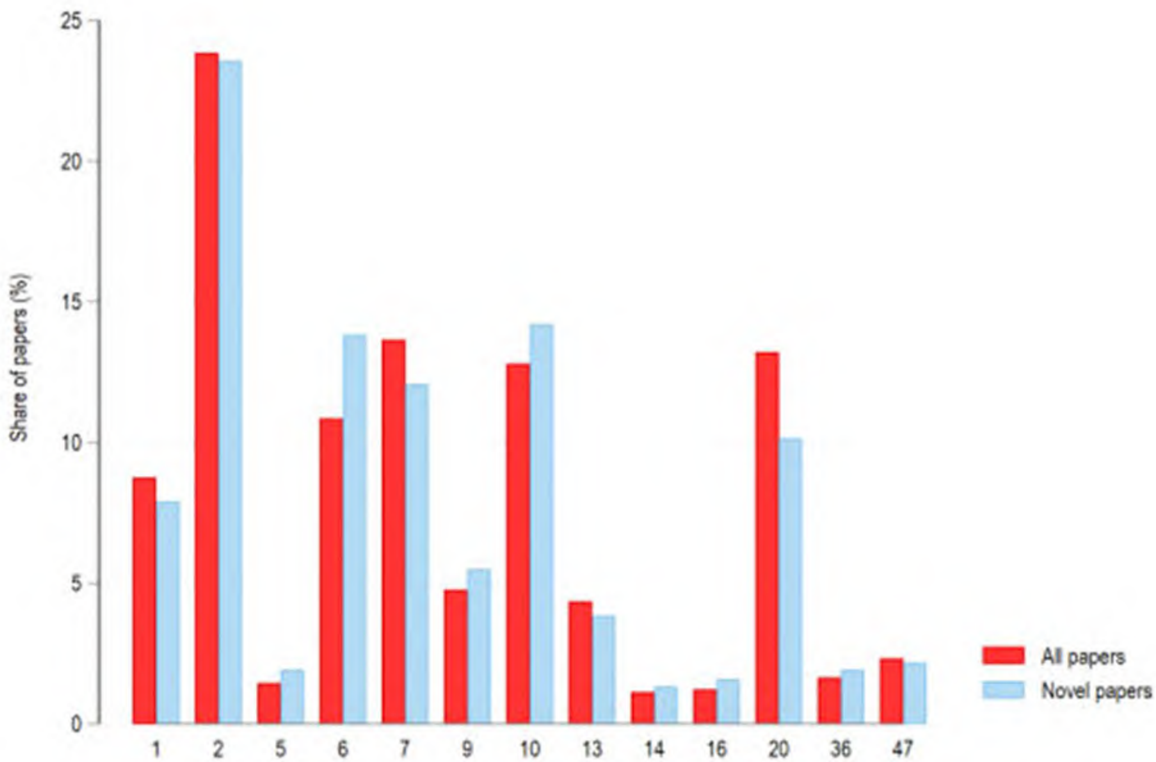
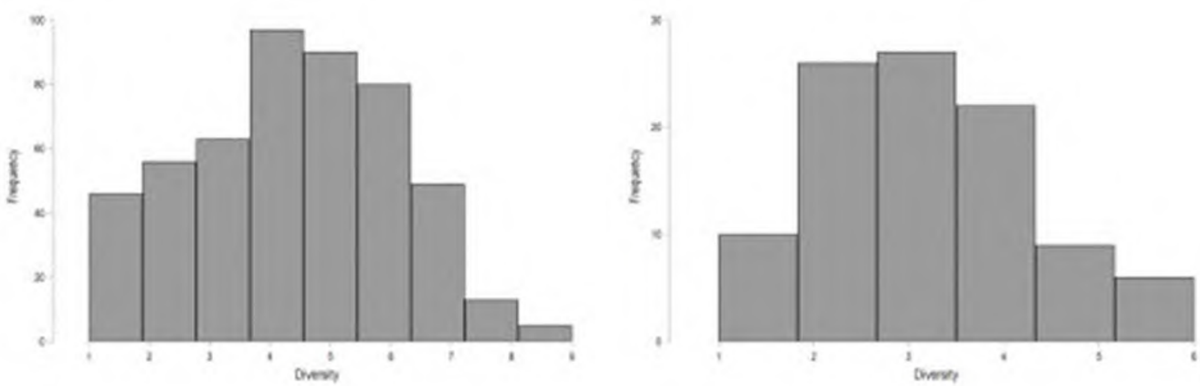


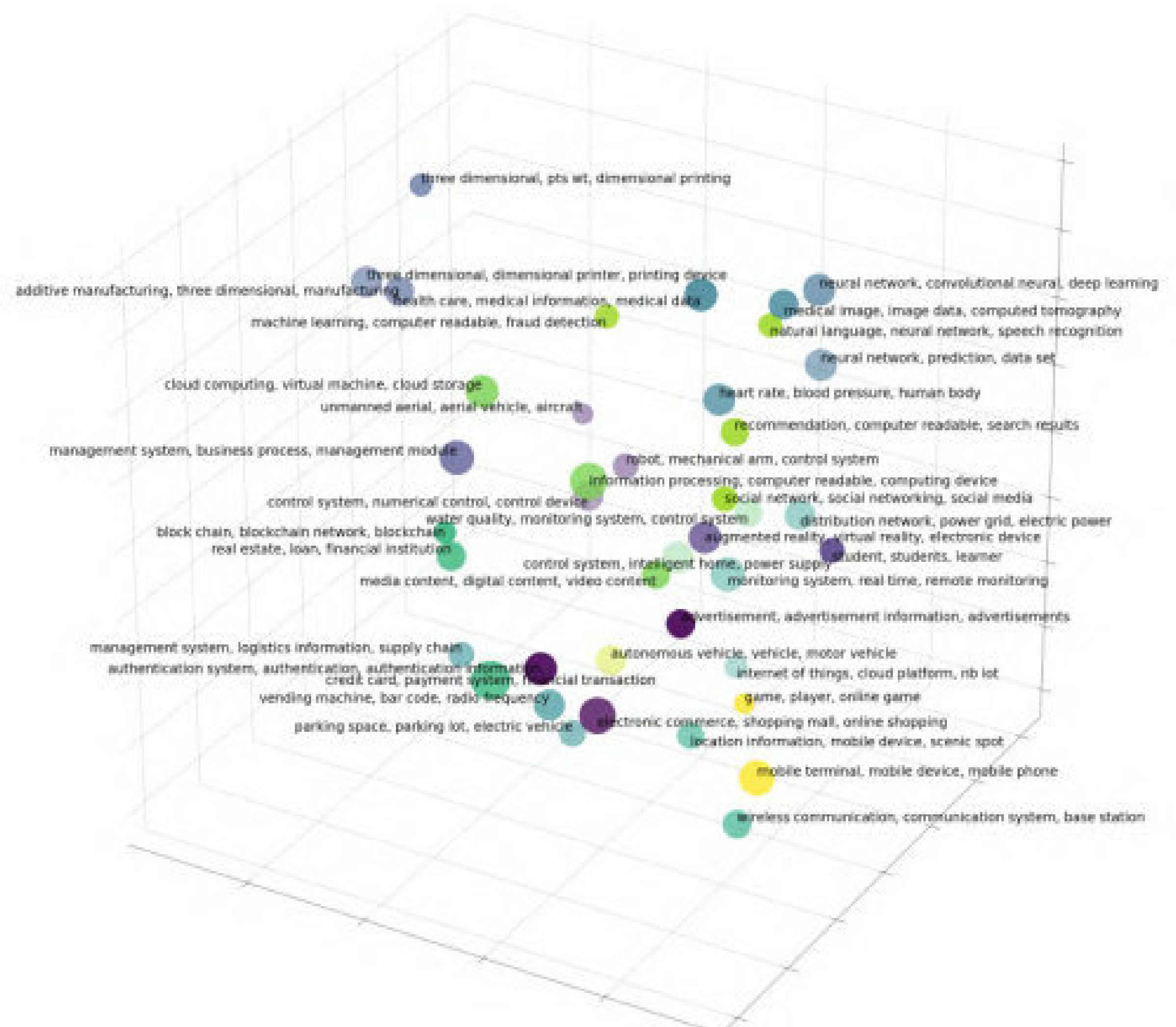
Figure A.3. Diversity of Novel Clusters in Queries



(a) Patents (b) Publications

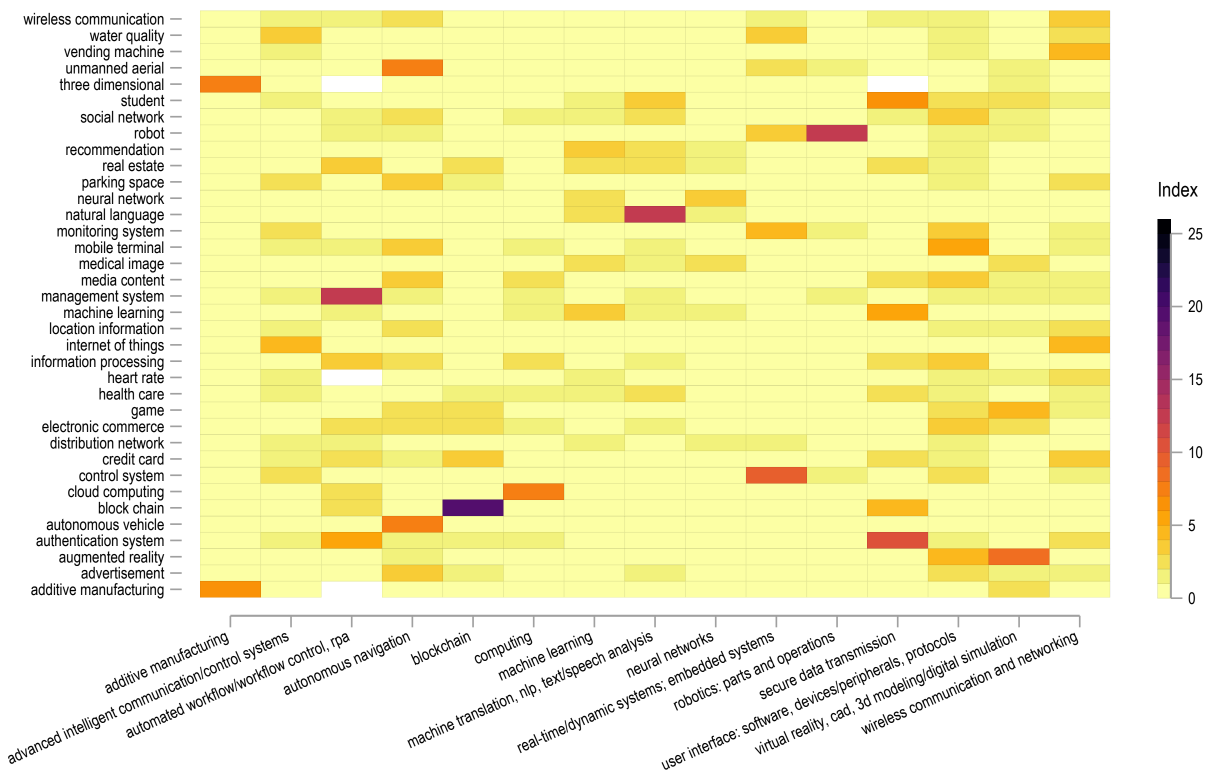
Note: The figure shows histogram for diversity defined as the number of manually constructed queries in which a cluster specializes ($RCA > 1$) for patents and publications.

Figure A.4. Major Technologies



Note: Vector representation of 40 major technologies obtained from the full set of patents related to digital automation technologies. These are projected on a two-dimensional space using UMAP (McInnes *et al.*, 2018). The color coding represents further aggregation based on the cluster centroids.

Figure A.5. Major Technologies and Queries



Note: Heatplot of Balassa's revealed comparative advantage index (RCA) indicating specialization of major technology groups in manually crafted queries.

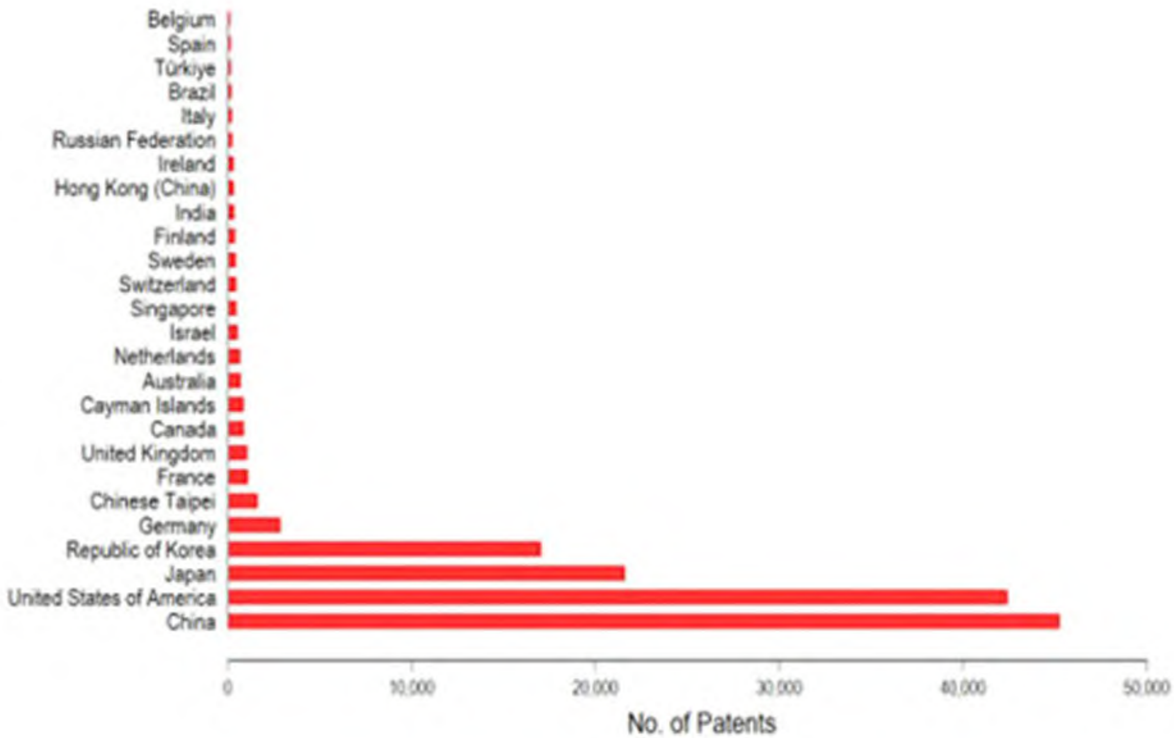
Figure A.6. Novelty and Impact by Geography (controlling for year)

(a) Novelty

(b) Impact

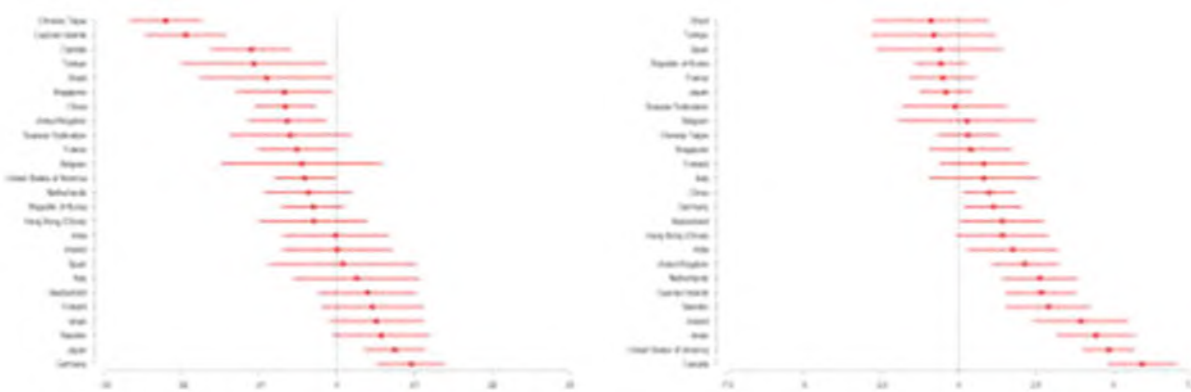
Note: These results correspond to patents in the novel patents and their offshoots set.

Figure A.7. Number of Patents by Country Filed for Novel and Off-shoot Clusters



Note: Patent filing location for novel and offshoot patents s.t. patent filed in multiple countries.

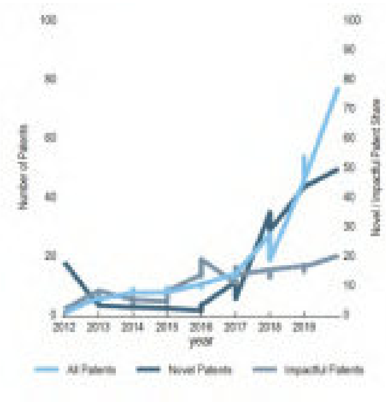
Figure A.8. Novelty and Impact by Geography (controlling for year)



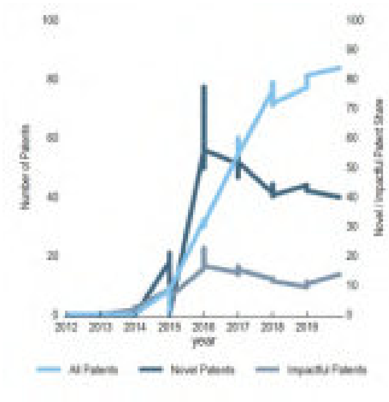
(a) Novelty (b) Impact

Note: These results correspond to novel and offshoot patents filed in multiple countries.

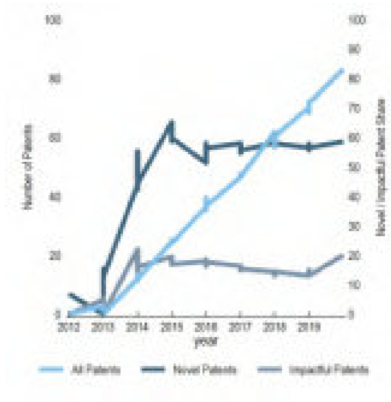
Figure A.9. Patterns of Emergence: Multicountry Clusters



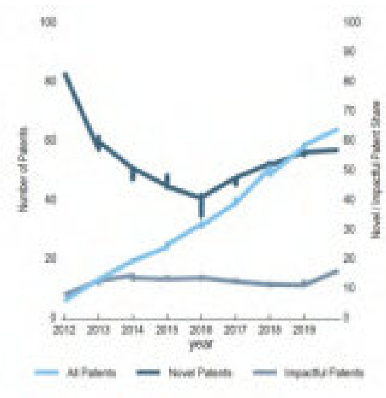
Pattern 1



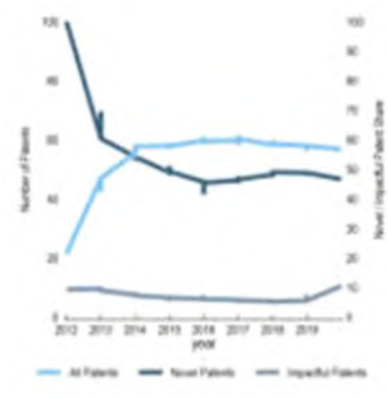
Pattern 2



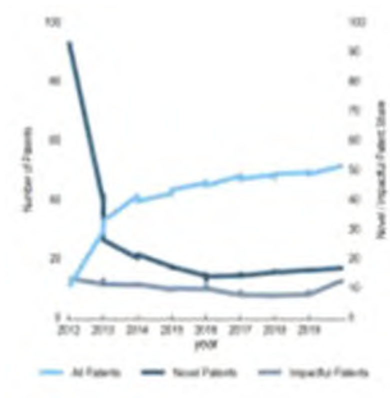
Pattern 3



Pattern 4

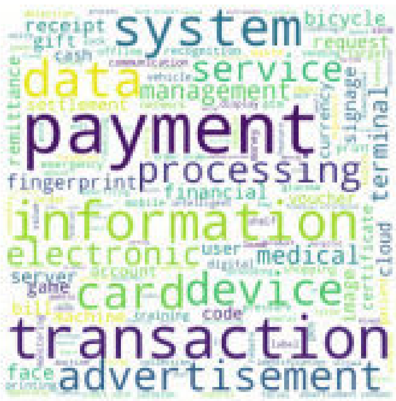


Pattern 5

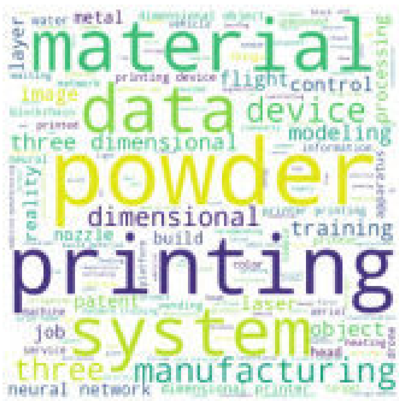


Pattern 6

Figure A.10. Word Clouds: Emergence Patterns for Multicountry Clusters

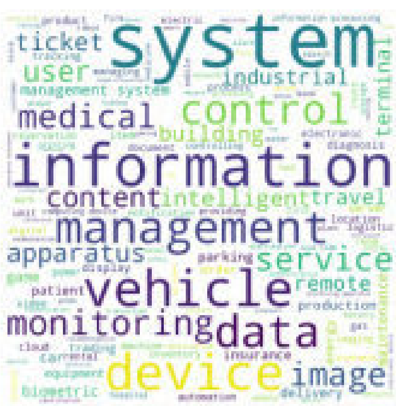


Pattern 1

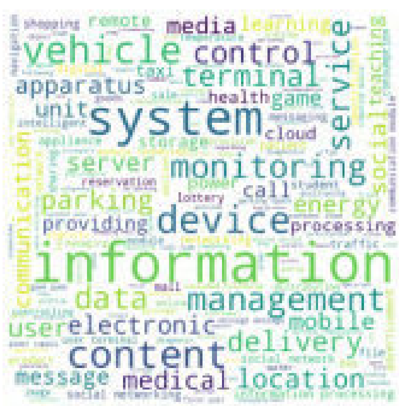


Pattern 2

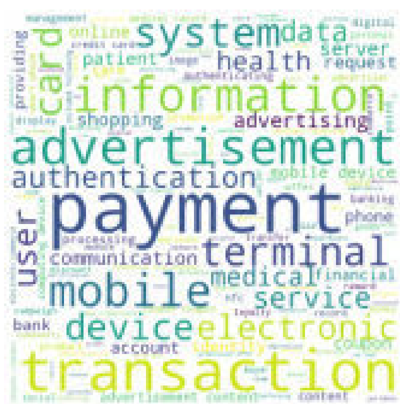
Pattern 3



Pattern 4



Pattern 5



Pattern 6

Tables

Table A.1—Broad technologies by pattern

ID	Technology	Application	Example	Family	Subfamily	Patterns	Similar	Main
14	Neural networks			AI & Intelligent Information Systems	Machine Learning	1		
39	Neural networks	Image recognition		AI & Intelligent Information Systems	Computer Vision	1; 2	No	
39	Neural networks	Image processing		AI & Intelligent Information Systems	Computer Vision	2; 1	No	
94	Neural networks	Voice assistant		User Interface	Biometric Recognition	4; 1	No	
94	Neural networks			AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 1	No	
94	Neural networks			Networking	Monitoring and remote control Applications	4; 1	No	
94	NLP	Speech recognition		AI & Intelligent Information Systems	Natural Language Processing (NLP)	1; 4	No	
75	NLP	Biometric information		AI & Intelligent Information Systems	Computer Vision	1; 4	No	
75	NLP			User Interface	Biometric Recognition	1; 5	No	
51	Augmented reality			User interface	Extended Reality (XR)	1		
82	Blockchain	Security and medical information		Data Management Technologies	Blockchain	1		
70	Blockchain	Anti counterfeiting		Data Management Technologies	Blockchain	1		
61	AM			Additive Manufacturing	Binder Jetting	2		
61	AM			Additive Manufacturing	Material Extrusion	2		
61	AM			Additive Manufacturing	Powder Bed Fusion	2		
61	AM			Additive Manufacturing	Vat Photopolymerization	2		
61	AM			Additive Manufacturing	Liquid additive manufacturing (LAM)	2		

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
61	AM			Additive Manufacturing	Selective Powder Deposition (SPD)	2		
61	AM			Additive Manufacturing	3D Construction Printing (3DCP)	2		
61	AM			Additive Manufacturing	3D Scanning	2		
65	AM			Additive Manufacturing	Powder Bed Fusion	2		
79	AM			Additive Manufacturing	N/A	2		
91	AM			Additive Manufacturing	Powder Bed Fusion	2		
95	AM			Additive Manufacturing	Sheet Lamination	2		
3	AM	printing pen		Additive Manufacturing	N/A	1; 2	No	
3	AM	color printing		Additive Manufacturing	N/A	2; 1	No	
26	AM	levelling		Additive Manufacturing	Binder Jetting	1; 2	No	
26	AM	cleaning		Additive Manufacturing	N/A	1; 2	No	
26	AM	recycling		Additive Manufacturing	N/A	1; 2	No	
26	AM	sanding		Additive Manufacturing	N/A	1; 2	No	
26	AM	Printing		Additive Manufacturing	Material Extrusion	2; 1	No	
26	AM	Extruding		Additive Manufacturing	N/A	2; 1	No	
26	AM	Feeding		Additive Manufacturing	N/A	2; 1	No	
22	AM	Build material		Additive Manufacturing	N/A	2; 3	Yes	
22	AM	Build material		Additive Manufacturing	N/A	3; 2	Yes	
8	IoT	Water monitoring	Fish farming	Data Acquisition Technologies	Control Systems	3		
8	IoT			Computing	Local/real-time processing	3		
8	IoT			Networking	Network services and applications	3		
8	IoT			Networking	Monitoring and remote control Applications	3		
8	IoT			Networking	IoT networks	3		
8	IoT			Networking	Wireless communication and network infrastructures	3		

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
44	IoT	Agriculture	Greenhouse	Networking	Wireless communication and network infrastructures	3		
44	IoT			Data Acquisition Technologies	Control Systems	3		
44	IoT			Networking	IoT networks	3		
44	IoT			Networking	Monitoring and remote control Applications	3		
53	IoT	Irrigation		Networking	Wireless communication and network infrastructures	3		
53	IoT			Data Acquisition Technologies	Control Systems	3		
53	IoT			Robotics	Mobile robots	3		
53	IoT			AI & Intelligent Information Systems	Machine learning	3		
53	IoT			User interface	N/A	3		
29	IoT	Warehouse		Networking	Monitoring and remote control Applications	1; 3; 4	Yes	
29	IoT			Networking	Wireless communication and network infrastructures	1; 3; 5	Yes	
29	IoT	Warehouse		Networking	Monitoring and remote control Applications	3; 1; 4	Yes	
29	IoT			Networking	Wireless communication and network infrastructures	3; 1; 5	Yes	
29	IoT			Networking	Network services and applications	3; 1; 6	Yes	
29	IoT			Networking	Monitoring and remote control Applications	4; 1; 3	Yes	
29	IoT			Networking	Wireless communication and network infrastructures	4; 1; 4	Yes	
29	IoT			Networking	Network services and applications	4; 1; 5	Yes	
49	IoT	Water quality		Networking	IoT networks	3; 4	No	4

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
49	IoT	Water treatment				3; 4	No	4
49	IoT	Water supply		Networking	IoT networks	4; 3	No	4
49	IoT	Water quality		Networking	Monitoring and remote control Applications	4; 3	No	4
49	IoT	Sewage		Networking	Network services and applications	4; 3	No	4
49	IoT	Water purification		Data Acquisition Technologies	Control Systems	4; 3	No	4
94	IoT	Voice assistant		User Interface	Biometric Recognition	2; 1	No	
83	Cloud Computing			Computing	Distributed computing	3		
83	Cloud Computing			User Interface	Extended Reality (XR)	3		
83	Cloud Computing			Robotics	Software/virtual robots	3		
21	Cloud Computing		Task scheduling	Computing	Distributed computing	3; 4	No	
21	Cloud Computing		Platform	Computing	Distributed computing	4; 3	No	
40	Cloud Computing	Cloud storage		Computing	Distributed computing	3; 4	Yes	4
40	Cloud Computing			Data Management Technologies	Data Sharing	3; 4	Yes	4
40	Cloud Computing			Networking	Distributed computing and low-latency	3; 4	Yes	4
40	Cloud Computing	Cloud storage	Encrypt	Computing	Distributed computing	4; 3	Yes	4
40	Cloud Computing		Encode	Data Management Technologies	Data Sharing	4; 3	Yes	4
40	Cloud Computing			Data Management Technologies	Data management Platforms	4; 3	Yes	4
40	Cloud Computing			Data Management Technologies	Encryption technologies/ Data Security	4; 3	Yes	4
35	Robots			Robotics	Mobile robots	4		
35	Robots			Robotics	Robotic control	4		
35	Robots			Robotics	Robotic navigation	4		
35	Robots			Robotics	Software/virtual robots	4		
98	Workflow automation			Robotics	Software/virtual robots	4		

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
9	Control system			Robotics	Robotic control	4		
9	Control system			Robotics	Software/virtual robots	4		
9	Control system			AI & Intelligent Information Systems	Robotics applications	4		
9	Control system			Networking	Network Management and Orchestration	4		
9	Control system			Networking	Monitoring and remote control Applications	4		
64	Control system			Robotics	Robotic control	4; 5	Yes	4
64	Control system			Robotics	Software/virtual robots	4; 5	Yes	4
64	Control system			AI & Intelligent Information Systems	Robotics applications	4; 5	Yes	4
64	Control system			Networking	Monitoring and remote control Applications	4; 5	Yes	4
64	Control system			Data Management Technologies	Encryption technologies/ Data Security	4; 5	Yes	4
64	Control system			Robotics	Robotic control	5; 4	Yes	4
64	Control system			AI & Intelligent Information Systems	Robotics applications	5; 4	Yes	4
27	UAV			Robotics	Mobile robots	1; 2	Yes	
27	UAV			AI & Intelligent Information Systems	Natural Language Processing (NLP)	1; 2	Yes	
27	UAV			Robotics	Mobile robots	2; 1	Yes	
27	UAV			User Interface	Extended Reality (XR)	2; 1	Yes	
19	UAV	Autonomous driving		Networking	Monitoring and remote control Applications	3; 4	Yes	4
19	UAV	Autonomous driving		Networking	Monitoring and remote control Applications	4; 3	Yes	4
19	UAV	Traffic control				4; 3	Yes	4
30	UAV	Vehicle control	Management	Data Management Technologies	Encryption technologies/ Data Security	4; 5	Yes	4

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
30	UAV		vehicle identification	Networking	Monitoring and remote control Applications	4; 5	Yes	4
30	UAV		Logistic transportation	Networking	IoT networks	4; 5	Yes	4
30	UAV		Diagnostic	AI & Intelligent Information Systems	Machine learning	4; 5	Yes	4
30	UAV		Cloud service	Computing	Distributed computing	4; 5	Yes	4
30	UAV		Safety	Data Management Technologies	Data Sharing	4; 5	Yes	4
30	UAV			Robotics	Robotic navigation	4; 5	Yes	4
30	UAV	Vehicle control	Monitoring	Networking	Monitoring and remote control Applications	5; 4	Yes	4
30	UAV		Assistance	Networking	Wireless communication and network infrastructures	5; 4	Yes	4
20	Mobile devices		Communication	Networking	Wireless communication and network infrastructures	4; 5	No	
20	Mobile devices		Computing	User interface	N/A	4; 5	No	
20	Mobile devices			Robotics	Software/virtual robots	4; 5	No	
20	Mobile devices			Computing	N/A	4; 5	No	
20	Mobile devices		Mobile terminals	Data Management Technologies	Encryption technologies/ Data Security	5; 4	No	
20	Mobile devices		UAV	User interface	N/A	5; 4	No	
20	Mobile devices			Networking	Wireless communication and network infrastructures	5; 4	No	
16	RFID	Inventory management		Networking	Wireless communication and network infrastructures	5; 6	No	6
16	RFID	Supply chain				5; 6	No	6
16	RFID	Shopping		Networking	Wireless communication and network infrastructures	6; 5	No	6
16	RFID			User interface	N/A	6; 5	No	6

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
16	RFID			Data Management Technologies	Data Warehousing	6; 5	No	6
16	RFID			Data Acquisition Technologies	Satellite and remote sensing	6; 5	No	

Notes: The table lists broad technologies allocated to the different patterns of emergence. Each broad technology is composed of one or more of the 500 technologies identified in Section 3.3. The table reports technologies only (applications using more than one technology are presented table A.2). *ID*: the unique identifier of the broad technology; *Technology*: the name of the broad technology, where NLP is Natural Language Processing, AM is additive manufacturing, IoT is Internet of Things, UAV is Unmanned Autonomous Vehicle, and rfid is Radio Frequency Identification; *Application*: applications frequently mentioned in the relevant pattern in the patent labels; *Example*: examples of the broad technology/application; *Family* and *Subfamily*: family of technology as described in Table 1; *Patterns*: patterns to which the broad technology pertains (because different technologies fall under one broad technology, the patterns can be multiple, and the technologies within one broad technology can differ, for example in their applications); *Similar*: indication if the technologies across different patterns are similar ("Yes") or not ("No"); *Main*: reports the main pattern to which the broad technology pertains, when there is one (that is when the large majority of technologies and applications where clustered in one pattern).

Table A.2—Technological applications by pattern

ID	Technology	Application	Example	Family	Subfamily	Patterns	Similar	Main
37		Energy distribution network		AI & Intelligent Information Systems	Machine Learning	1		
37		Energy distribution network		Networking	Monitoring and remote control Applications	1		
37		Energy distribution network		Networking	Distributed storage systems	1		
43		Load prediction		AI & Intelligent Information Systems	Neural networks	1; 6	Yes	
43	Neural networks	Load prediction		AI & Intelligent Information Systems	Neural networks	6; 1	Yes	
7		Energy consumption (final)	Smart meters	AI & Intelligent Information Systems	Machine learning	4; 5	No	
7		Energy consumption (final)	Building automation	Data Management Technologies	Encryption technologies/ Data Security	4; 5	No	
7		Energy consumption (final)	Energy storage	Robotics	Software/virtual robots	4; 5	No	
7		Energy consumption (final)		Data Management Technologies	Data Sharing	4; 5	No	
7		Energy consumption (final)		User interface	N/A	4; 5	No	
7		Energy consumption (final)	Energy monitoring	AI & Intelligent Information Systems	Machine learning	5; 4	No	
7		Energy consumption (final)		Networking	Monitoring and remote control Applications	5; 4	No	
7		Energy consumption (final)		Data Acquisition Technologies	Control Systems	5; 4	No	

continued . . .

52

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
54		Power supply		Networking	Monitoring and remote control Applications	5		
59		Waste management		AI & Intelligent Information Systems	Machine Learning	1		
59		Waste management		Networking	Wireless communication and network infrastructures	1		
73		Dental prosthesis		Additive Manufacturing	N/A	1		
73		Dental prosthesis		User Interface	Extended Reality (XR)	1		
73		Dental prosthesis		AI & Intelligent Information Systems	Application: 3D Computer Vision	1		
88		Clothing recommendations		User Interface	Virtual Reality	1		
88		Clothing recommendations		AI & Intelligent Information Systems	Application: Recommender/Recommendation system	1		
88		Clothing recommendations		AI & Intelligent Information Systems	Application: 3D Computer Vision	1		
96	Neural networks	Recommendation	Product recommendation	AI & Intelligent Information Systems	Machine learning	2; 3; 4	No	
96		Recommendation		AI & Intelligent Information Systems	Natural Language Processing (NLP)	2; 3; 4	No	
96		Recommendation	Content recommendation	AI & Intelligent Information Systems	Machine learning	3; 2; 4	No	
96		Recommendation	Multimedia	AI & Intelligent Information Systems	Natural Language Processing (NLP)	3; 2; 4	No	
96		Recommendation	Target user	User interface	N/A	3; 2; 4	No	
96		Recommendation	Product	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 2;3	Yes	
96		Recommendation	Search	AI & Intelligent Information Systems	Machine learning	4; 2;3	Yes	

continued . . .

53

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
66		Vending machines (payment)		Networking	Wireless communication and network infrastructures	1; 2	No	
66		Vending machines (payment)		Data Management Technologies	Blockchain	1; 2	No	
66		Vending machines		User Interface	N/A	2; 1	No	
66		Vending machines		Robotics	N/A	2; 1	No	
66		Vending machines		Networking	Wireless communication and network infrastructures	2; 1	No	
1		Financial transactions		Data Management Technologies	Blockchain	1; 6	No	
1		Financial transactions	cryptocurrency			1; 6	No	
1		Financial transactions	remittances			1; 6	No	
1		Financial transactions				6; 1	Yes	
1		Financial transactions	loans	User interface	N/A	6; 1	No	
1		Financial transactions		AI & Intelligent Information Systems	Natural Language Processing (NLP)	6; 1	No	
1		Financial transactions	financial transactions	Data Management Technologies	Blockchain	6; 1	Yes	
1		Financial transactions		User interface	N/A	6; 1	Yes	
1		Financial transactions	Banking service	Robotics	Workflow automation software	6; 1	No	
1		Financial transactions		Data Management Technologies	Blockchain	6; 1	No	
1		Financial transactions	e-money	Data Management Technologies	Blockchain	6; 1	No	
1		Financial transactions		Networking	Wireless communication and network infrastructures	6; 1	No	
1		Financial transactions		User interface	N/A	6; 1	No	
1		Financial transactions	Money transfer	Data Management Technologies	Blockchain	6; 1	Yes	
1		Financial transactions		User interface	N/A	6; 1	Yes	
1		Financial transactions				6; 1	Yes	
60	Blockchain	Mobile payment		Data Management Technologies	Blockchain	1; 5; 6	No	

continued . . .

54

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
60	Wireless	Mobile payment	public transport	Networking	Wireless communication and network infrastructures	1; 5; 6	No	
60	Automated workflow	Mobile payment	gift cards	Robotics	Software/virtual robots	1; 5; 6	No	
60		Mobile payment	qr codes	User interface	N/A	1; 5; 6	No	
60		Mobile payment	offline payment	Data Acquisition Technologies	Automated tracing and tagging	1; 5; 6	No	
60	Blockchain	Mobile payment		Data Management Technologies	Blockchain	5; 1; 6	No	
60	Wireless	Mobile payment	credit cards	Networking	Wireless communication and network infrastructures	5; 1; 6	No	
60	User interface	Mobile payment	POS	User interface	N/A	5; 1; 6	No	
60		Mobile payment	mobile payment			5; 1; 6	No	
60	NFC	Mobile payment		Networking	Wireless communication and network infrastructures	6; 1; 5	No	
60		Mobile payment	Payment terminal	Data Acquisition Technologies	Automated tracing and tagging	6; 1; 5	No	
60		Mobile payment	mobile device	User interface	N/A	6; 1; 5	No	
60		Mobile payment	mobile banking			6; 1; 5	No	
60		Mobile payment	POS			6; 1; 5	No	
62		E-trading	Financial instruments	User interface	N/A	4; 5; 6	No	
62		E-trading	Online trading	User interface	N/A	5; 4; 6	Yes	
62		E-trading	Online trading	Data Management Technologies	Blockchain	6; 4; 5	No	
45		E-tickets		Data Management Technologies	Encryption technologies/ Data Security	4; 5	No	4
45		E-tickets		Data Management Technologies	Blockchain	4; 5	No	4
45		E-tickets		Robotics	Software/virtual robots	4; 5	No	4
45		E-tickets	Reservations	N/A	N/A	5; 4	No	4

continued . . .

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
31		E-commerce	Online shopping	Data Management Technologies	Encryption technologies/ Data Security	5; 6	Yes	
31		E-commerce	Product information	AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 6	Yes	
31		E-commerce	Sale management	User interface	N/A	5; 6	Yes	
31		E-commerce		Robotics	Software/virtual robots	5; 6	Yes	
31		E-commerce	Mobile terminal	Networking	Wireless communication and network infrastructures	6; 5	Yes	
31		E-commerce	Product information	AI & Intelligent Information Systems	Natural Language Processing (NLP)	6; 5	Yes	
31		E-commerce		User interface	N/A	6; 5	Yes	
4	Blockchain	E-payment	invoicing	Data Management Technologies	Blockchain	1; 6	Yes	
4	Automated workflow	E-payment		Robotics	Software/virtual robots	1; 6	Yes	
4	Blockchain	E-payment	E-receipt	Robotics	Software/virtual robots	6; 1	No	6
4	Automated workflow	E-payment	POS	Data Management Technologies	Blockchain	6; 1	No	6
4		E-payment	Online payment	User interface	N/A	6; 1	No	6
4		E-payment	Payment processing			6; 1	No	6
92	Blockchain	E-sales	vouchers	Data Management Technologies	Blockchain	1; 6	No	
92	Machine learning	E-sales		Data Management Technologies	Encryption technologies/ Data Security	6; 1	No	6
92	Blockchain	E-sales	Promotions	AI & Intelligent Information Systems	Machine learning	6; 1	No	6
92	Secure networks	E-sales	Loyalty data	User interface	N/A	6; 1	No	6
92		E-sales	Discounts			6; 1	No	6

continued . . .

56

... continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
68		Teller machines	Financial transactions	Networking	Wireless communication and network infrastructures	6		
68		Teller machines		Data Management Technologies	Encryption technologies/ Data Security	6		
10		Secure payment		Data Management Technologies	Encryption technologies/ Data Security	6; 1	No	6
10		Mobile transaction		User interface	N/A	6; 1	No	6
10		Secure payment		Data Management Technologies	Blockchain	6; 1	No	6
36		Logistic/delivery		Networking	Monitoring and remote control Applications	4		
36		Logistic/delivery		Robotics	Mobile robots	4		
36		Logistic/delivery		Data Management Technologies	Encryption technologies/ Data Security	4		
36		Logistic/delivery		Data Management Technologies	Data Warehousing	4		
36		Logistic/delivery		Data Acquisition Technologies	Automated tracing and tagging	4		
36		Logistic/delivery		Data Acquisition Technologies	Satellite and remote sensing	4		
89		Sales logistics	Order placement	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 6	No	
89		Sales logistics	Inventories	User interface	N/A	4; 6	No	
89		Sales logistics	Order processing	Robotics	Mobile robots	6; 4	No	
89		Sales logistics		User interface	N/A	6; 4	No	
55	NLP	Recruitment		AI & Intelligent Information Systems	Machine learning	2; 3	No	
55		Recruitment		AI & Intelligent Information Systems	Natural Language Processing (NLP)	2; 3	No	
55	Secure networks	Recruitment		Networking	Intelligent and secure networks	3; 2	No	

continued ...

57

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
55		Recruitment		Data Management Technologies	Encryption technologies/ Data Security	3; 2	No	
55		Recruitment		AI & Intelligent Information Systems	Natural Language Processing (NLP)	3; 2	No	
55		Recruitment		Robotics	Software/virtual robots	3; 2	No	
55		Recruitment		Data Management Technologies	Blockchain	3; 2	No	
55		Recruitment		AI & Intelligent Information Systems	Machine learning	3; 2	No	
97		Network security		AI & Intelligent Information Systems	Machine learning	3		
97		Network security		Networking	Intelligent and secure networks	3		
56		Event scheduling		User interface	N/A	3; 4	Yes	
56		Event scheduling		AI & Intelligent Information Systems	Natural Language Processing (NLP)	3; 4	Yes	
56		Event scheduling		User interface	N/A	4; 3	Yes	
56		Event scheduling		AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 3	Yes	
2		Vehicle maintenance	Maintenance management	Robotics	Mobile robots	4		
2		Vehicle maintenance		AI & Intelligent Information Systems	Robotics applications	4		
13		Car sharing		Data Management Technologies	Blockchain	4		
13		Car sharing		Networking	Monitoring and remote control Applications	4		
13		Car sharing		Data Management Technologies	Encryption technologies/ Data Security	4		
13		Car sharing		Networking	Monitoring and remote control Applications	4		
85		Vehicle allocation	Taxi	Networking	Monitoring and remote control Applications	4; 5	No	
85		Vehicle allocation	Vehicle allocation			4; 5	No	

continued . . .

58

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
85		Vehicle allocation	Taxi	Networking	Monitoring and remote control Applications	5; 4	No	
85		Vehicle allocation	Bus	Networking	Wireless communication and network infrastructures	5; 4	No	
85		Vehicle allocation	Information			5; 4	No	
81		Parking management	Spaces	Networking	Monitoring and remote control Applications	4		
81		Parking management	Payment	Data Management Technologies	Encryption technologies/ Data Security	4		
81		Parking management	Space management	Networking	Network Management and Orchestration	4		
81		Parking management		Networking	Network services and applications	4		
50		Parking	Space	Networking	Wireless communication and network infrastructures	6		
50		Parking	Management	Networking	Network services and applications	6		
50		Parking		Networking	IoT networks	6		
50		Parking		Networking	Network Management and Orchestration	6		
50		Parking		Data Management Technologies	Data Sharing	6		
46		Insurance	Contract	Data Management Technologies	Blockchain	4		
46		Insurance	Self driving	Robotics	Software/virtual robots	4		
46		Insurance	Data acquisition	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4		
46		Insurance	Monitoring	Networking	Monitoring and remote control Applications	4		
47		Medical imaging	ultrasound images	AI & Intelligent Information Systems	Machine learning	4		

continued . . .

59

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
47		Medical imaging	ultrasound diagnostic	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4		
47		Medical imaging	CT	User Interface	Extended Reality (XR)	4		
47		Medical imaging	Surgery	Robotics	Manipulative robots	4		
47		Medical imaging		Robotics	Robotic control	4		
17	Neural networks	Medical images		AI & Intelligent Information Systems	Machine learning	4; 5	No	
17		Medical images		AI & Intelligent Information Systems	Computer Vision	4; 5	No	
17		Medical images		Data Acquisition Technologies	Biometric and body scans	5; 4	No	
6		Medical images	Medical records	Data Management Technologies	Data Sharing	5; 6	Yes	
6		Medical images	Diagnosis	AI & Intelligent Information Systems	Computer Vision	5; 6	Yes	
6		Medical images	Medical records	Data Management Technologies	Data Sharing	6; 5	Yes	
6		Medical images	Diagnosis	AI & Intelligent Information Systems	Computer Vision	6; 5	Yes	
80		Health monitoring devices		AI & Intelligent Information Systems	Machine learning	4; 5; 6	No	
80		Health monitoring devices		Data Acquisition Technologies	Sensing technologies	4; 5; 6	No	
80		Health monitoring devices		Networking	Wireless communication and network infrastructures	4; 5; 6	No	
80		Health monitoring devices		Networking	Intelligent and secure networks	5; 4; 6	No	
80		Health monitoring devices		Networking	Wireless communication and network infrastructures	6; 4; 5	No	
52		Medication	Prescription	Data Management Technologies	Encryption technologies/ Data Security	5; 6	No	
52		Medication	Delivery	AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 6	No	

continued . . .

60

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
52		Medication	Management	Networking	Monitoring and remote control Applications	6; 5	No	
52		Medication		Data Acquisition Technologies	Automated tracing and tagging	6; 5	No	
52		Medication		Networking	IoT networks	6; 5	No	
52		Medication		User interface	N/A	6; 5	No	
87	Blockchain	Health care	Health records	Data Management Technologies	Blockchain	1; 5; 6	No	5
87	Secure networks	Health care		Data Management Technologies	Encryption technologies/ Data Security	1; 5; 6	No	5
87		Health care		AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 1; 6	No	5
87		Health care	Personal health	User interface	N/A	5; 1; 6	No	5
87		Health care	Health records	Data Acquisition Technologies	Biometric and body scans	5; 1; 6	No	5
87		Health care	Hospital management	Data Management Technologies	Encryption technologies/ Data Security	5; 1; 6	No	5
87	Cloud	Health care		Data Management Technologies	Encryption technologies/ Data Security	6; 1; 5	No	5;6
87	ML	Health care	Remote consultation	AI & Intelligent Information Systems	Natural Language Processing (NLP)	6; 1; 5	No	5;6
87		Health care	Treatment plan	AI & Intelligent Information Systems	Machine learning	6; 1; 5	No	5;6
87		Health care	Medical records	Robotics	Software/virtual robots	6; 1; 5	No	5;6
87		Health care	Reservation	Computing	Distributed computing	6; 1; 5	No	5;6
87		Health care		Data Management Technologies	Data management Platforms	6; 1; 5	No	5;6
87		Health care		User interface	N/A	6; 1; 5	No	5;6
24		Health care	Medical device	Networking	IoT networks	6		

continued . . .

61

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
24		Health care	Patient monitoring	Networking	Wireless communication and network infrastructures	6		
24		Health care	Nursing	Networking	Monitoring and remote control Applications	6		
24		Health care	Emergency	User Interface	User Interface Biometric Sensors	6		
24		Health care		User Interface	Extended Reality (XR)	6		
24		Health care		AI & Intelligent Information Systems	Machine learning	6		
24		Health care		AI & Intelligent Information Systems	Natural Language Processing (NLP)	6		
75		Biometric information		Data Management Technologies	Blockchain	4; 1	No	
75		Biometric information		Robotics	Robotic process automation (RPA)	4; 1	No	
75		Biometric information		Robotics	Robotic navigation	4; 1	No	
57		Hospitality: accommodation	Reservation	Data Management Technologies	Encryption technologies/ Data Security	4		
57		Hospitality: accommodation	Intelligent hotel	Data Management Technologies	Blockchain	4		
57		Hospitality: accommodation		Networking	Monitoring and remote control Applications	4		
71		Hospitality: food	Ordering	Networking	Wireless communication and network infrastructures	4		
71		Hospitality: food	Recipe	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4		
71		Hospitality: food	Cooking	Robotics	Manipulative robots	4		
71		Hospitality: food	Delivery	Robotics	Mobile robots	4		
93		School management	Campus management	Networking	Monitoring and remote control Applications	4		

continued . . .

62

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
93		School management	Attendance management	AI & Intelligent Information Systems	Machine learning	4		
23		Construction management	Construction site	User Interface	Extended Reality (XR)	4		
23		Construction management	Project management	Robotics	Software/virtual robots	4		
58		Building/property management	Market transactions	Data Management Technologies	Blockchain	3; 4	No	4
58		Building/property management		User Interface	Extended Reality (XR)	3; 4	No	4
58		Building/property management	Building automation	Data Acquisition Technologies	Control Systems	4; 3	No	4
58		Building/property management	Property management	Networking	Monitoring and remote control Applications	4; 3	No	4
58		Building/property management		Networking	IoT networks	4; 3	No	4
58		Building/property management		User interface	N/A	4; 3	No	4
33		Building management	Door control	Data Acquisition Technologies	Control Systems	3; 4	Yes	4
33		Building management		Networking	Monitoring and remote control Applications	3; 4	Yes	4
33		Building management		Networking	IoT networks	3; 4	Yes	4
33		Building management		Networking	Network Management and Orchestration	3; 4	Yes	4
33		Building management	Intelligent home	Data Acquisition Technologies	Control Systems	4; 3	Yes	4
33		Building management	Security system	Networking	Monitoring and remote control Applications	4; 3	Yes	4
33		Building management	Fire alarm	Networking	IoT networks	4; 3	Yes	4

continued . . .

63

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
33		Building management	Work safety	User interface	N/A	4; 3	Yes	4
33		Building management		AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 3	Yes	4
33		Building management		Networking	Network Management and Orchestration	4; 3	Yes	4
41		Terminal/peripheral devices	Data processing	Data Management Technologies	Data Sharing	3; 4; 5	No	4
41		Terminal/peripheral devices	Collaboration	User interface	N/A	4; 3; 5	No	4
41		Terminal/peripheral devices	Meeting			4; 3; 5	No	4
41		Terminal/peripheral devices	Information display			4; 3; 5	No	4
41		Terminal/peripheral devices	Chat rooms			4; 3; 5	No	4
41		Terminal/peripheral devices	Communication devices			5; 3; 4	No	4
18		Teaching	Online teaching	Data Management Technologies	Data Sharing	4; 5	Yes	4
18		Teaching		Data Management Technologies	Encryption technologies/ Data Security	4; 5	Yes	4
18		Teaching		AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 5	Yes	4
18		Teaching		Data Management Technologies	Real-time streaming	4; 5	Yes	4
18		Teaching		User interface	N/A	4; 5	Yes	4
18		Teaching	Online teaching	Data Management Technologies	Encryption technologies/ Data Security	5; 4	Yes	4
18		Teaching		AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 4	Yes	

continued . . .

64

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
18		Teaching		Data Management Technologies	Real-time streaming	5; 4	Yes	
86		Teaching	Exams	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 5	Yes	4
86		Teaching		Data Management Technologies	Real-time streaming	4; 5	Yes	4
86		Teaching		Data Management Technologies	Encryption technologies/ Data Security	4; 5	Yes	4
86		Teaching	Exams	AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 4	Yes	4
86		Teaching		Data Management Technologies	Real-time streaming	5; 4	Yes	
86		Teaching		Data Management Technologies	Encryption technologies/ Data Security	5; 4	Yes	
99		Travel information	Navigation	User Interface	Extended Reality (XR)	4; 5	Yes	4
99		Travel information	Exploration	Networking	Monitoring and remote control Applications	4; 5	Yes	4
99		Travel information		AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 5	Yes	4
99		Travel information		Networking	Monitoring and remote control Applications	5; 4	Yes	4
77		Media players		User Interface	Extended Reality (XR)	4; 5	Yes	4
77		Media players		User Interface	N/A	4; 5	Yes	4
77		Media players		Data Management Technologies	Encryption technologies/ Data Security	5; 4	Yes	4
77		Media players		User Interface	N/A	5; 4	Yes	
28		Digital media	Digital rights	User interface	N/A	5		
28		Digital media		Networking	Intelligent and secure networks	5		
28		Digital media		Robotics	Software/virtual robots	5		
0		Intelligent home	Voice recognition	User Interface	Conversational UI	1		

continued . . .

65

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
67	IoT	Intelligent home	Temperature	Networking	Monitoring and remote control Applications	4		
67		Intelligent home		Data Acquisition Technologies	Control Systems	4		
67		Intelligent home		Networking	IoT networks	4		
32		Intelligent homes		Data Acquisition Technologies	Control Systems	4; 5	Yes	
32		Intelligent homes		Networking	Monitoring and remote control Applications	4; 5	Yes	
32		Intelligent homes		Networking	Wireless communication and network infrastructures	4; 5	Yes	
32		Intelligent homes		Networking	IoT networks	4; 5	Yes	
32		Intelligent homes		Data Acquisition Technologies	Control Systems	5; 4	Yes	
32		Intelligent homes		Networking	Monitoring and remote control Applications	5; 4	Yes	
32		Intelligent homes		Networking	Wireless communication and network infrastructures	5; 4	Yes	
32		Intelligent homes		Networking	IoT networks	5; 4	Yes	
15		Social networks		AI & Intelligent Information Systems	Machine learning	4; 5	Yes	
15		Social networks		AI & Intelligent Information Systems	Natural Language Processing (NLP)	4; 5	Yes	
15		Social networks		User interface	N/A	4; 5	Yes	
15		Social networks		AI & Intelligent Information Systems	Machine learning	5; 4	Yes	
15		Social networks		AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 4	Yes	
15		Social networks		User interface	N/A	5; 4	Yes	
72		Social Networks		User interface	N/A	5		
72		Social Networks		AI & Intelligent Information Systems	Natural Language Processing (NLP)	5		
42		Social Networks		User interface	N/A	5		

continued . . .

66

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
42		Social Networks		AI & Intelligent Information Systems	Natural Language Processing (NLP)	5		
42		Social Networks		Robotics	Software/virtual robots	5		
11		Games		Data Management Technologies	Blockchain	4; 5	Yes	
11		Games		User interface	N/A	4; 5	Yes	
11		Games		User Interface	Extended Reality (XR)	4; 5	Yes	
11		Games		User interface	N/A	5; 4	Yes	
11		Games		User Interface	Extended Reality (XR)	5; 4	Yes	
76		Gaming	Gambling	Data Management Technologies	Blockchain	4; 5	Yes	
76		Gaming	Gaming machines	User interface	N/A	4; 5	Yes	
76		Gaming	Lotteries	Robotics	Software/virtual robots	4; 5	Yes	
76		Gaming	Gambling	Data Management Technologies	Blockchain	5; 4	Yes	
76		Gaming	Gaming machines	User interface	N/A	5; 4	Yes	
76		Gaming	Lotteries	Robotics	Software/virtual robots	5; 4	Yes	
90		Disaster management	Alert	Robotics	Mobile robots	4		
90		Disaster management	Evacuation	AI & Intelligent Information Systems	Natural Language Processing (NLP)	4		
90		Disaster management	Prevention	Robotics	Robotic navigation	4		
48		Printing		Computing	N/A	4		
48		Printing		Networking	Wireless communication and network infrastructures	4		
74		Coupons		Networking	Wireless communication and network infrastructures	5		
63		Location services		Data Acquisition Technologies	Satellite and remote sensing	5		
63		Location services		Networking	Wireless communication and network infrastructures	5		
63		Location services		Robotics	Robotic navigation	5		

continued . . .

67

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
63		Location services		User interface	N/A	5		
78		Call centres		Networking	Wireless communication and network infrastructures	5		
78		Call centres		User interface	N/A	5		
78		Call centres		AI & Intelligent Information Systems	Natural Language Processing (NLP)	5		
78		Call centres		Robotics	Software/virtual robots	5		
84		E-mail		User interface	N/A	5		
84		E-mail		Networking	Intelligent and secure networks	5		
12		Mobile advertisement		User interface	N/A	5; 6	Yes	
12		Mobile advertisement		Networking	Wireless communication and network infrastructures	5; 6	Yes	
12		Mobile advertisement		Data Acquisition Technologies	Automated tracing and tagging	5; 6	Yes	
12		Mobile advertisement		Data Acquisition Technologies	Satellite and remote sensing	5; 6	Yes	
12		Mobile advertisement		User interface	N/A	6; 5	Yes	
12		Mobile advertisement		Networking	Wireless communication and network infrastructures	6; 5	Yes	
12		Mobile advertisement		Data Acquisition Technologies	Automated tracing and tagging	6; 5	Yes	
12		Mobile advertisement		Data Acquisition Technologies	Satellite and remote sensing	6; 5	Yes	
25		Advertisement	Online Ads	AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 6	No	
25		Advertisement		User interface	N/A	5; 6	No	
25		Advertisement	Online Ads	AI & Intelligent Information Systems	Natural Language Processing (NLP)	6; 5	No	
25		Advertisement	Targeting	User interface	N/A	6; 5	No	
25		Advertisement	social networks	AI & Intelligent Information Systems	Machine learning	6; 5	No	
25		Advertisement	Games	User Interface	Extended Reality (XR)	6; 5	No	

continued . . .

68

. . . continued

ID	Technology	Application	Details	Family	Subfamily	Patterns	Similar	Main
25		Advertisement	Bidding			6; 5	No	
69		Auctions		Data Management Technologies	Blockchain	5; 6	Yes	
69		Auctions		User interface	N/A	5; 6	Yes	
69		Auctions		Data Management Technologies	Blockchain	6; 5	Yes	
69		Auctions		User interface	N/A	6; 5	Yes	
69		Auctions		Networking	Wireless communication and network infrastructures	6; 5	Yes	
5		Information processing	patent classification	AI & Intelligent Information Systems	Machine learning	1; 5	No	5
5		Information processing	patent retrieval	AI & Intelligent Information Systems	Natural Language Processing (NLP)	1; 5	No	5
5		Information processing				1; 5	No	5
5		Information processing		Computing	N/A	5; 1	No	5
5		Information processing	user information	User Interface	Extended Reality (XR)	5; 1	No	5
5		Information processing	information storage	Robotics	Robotic control	5; 1	No	5
5		Information processing	Industry/merchandise process	Data Management Technologies	Blockchain	5; 1	No	5
5		Information processing	Selling	AI & Intelligent Information Systems	Natural Language Processing (NLP)	5; 1	No	5
10	Blockchain	E-certificate		Data Management Technologies	Encryption technologies/ Data Security	1; 6	No	
10		E-certificate		Data Management Technologies	Blockchain	1; 6	No	
10		E-certificate		Robotics	Software/virtual robots	6; 1	No	6

Notes: The table lists technological applications extracted from digital automation patents, clustered in the different patterns of emergence. Each technological application is composed of one or more of the 500 technologies identified in Section 3.3. The table reports technological applications only (broad technologies are presented in Table A.1). *ID*: the unique identifier of the broad technology; *Technology*: the name of the technology (when the technological application can be attributed to one specific technology) where NFC is near field communication, NLP is Natural Language Processing, ML is machine learning, and IoT is Internet of Things. *Application*: technological applications frequently men-

69

tioned in the patent labels; *Example*: examples of technological applications; *Family* and *Subfamily*: family of technology as described in Table 1; *Patterns*: patterns to which the technological application pertains (because different technologies and components fall under one technological application, the patterns can be multiple, and the components within one technological application can differ); *Similar*: indication if the technological applications across different patterns are similar ("Yes") or not ("No"); *Main*: reports the main pattern to which the technological application pertains, when there is one (that is when the large majority of technologies of a technological application where clustered in one pattern).

70

Table A.3—Broad technologies by pattern (publications)

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
1	96	Blockchain-enabled Supply Chain and IoT Integration		Networking; Data Management Technologies	Blockchain; IoT networks	
			Blockchain: Cryptocurrencies, E-voting, and Decentralized Finance	Data Management Technologies	Blockchain; Application: Smart Contracts; Application: Decentralized Finance; Application: E-voting	
			Blockchain-enabled Internet of Vehicles (IoV)	Data Management Technologies	Blockchain	
			Blockchain for Electronic Health Records	Data Management Technologies	Blockchain	
1	12		Medical Imaging Analysis using Deep Learning (Radiomics)			
			Deep Learning for Cancer Detection (cervical, breast cancer, melanoma, brain tumor segmentation, lung cancer)	AI & Intelligent Information Systems	Machine learning; Application: Medical diagnosis (including self-diagnostic)	
			Deep Learning for Occular Disease Recognition	AI & Intelligent Information Systems	Machine learning; Application: Medical diagnosis (including self-diagnostic)	
			Deep Learning for Alzheimer's Disease Diagnosis	AI & Intelligent Information Systems	Machine learning; Application: Medical diagnosis (including self-diagnostic)	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
			Deep Learning for Covid Detection	AI & Intelligent Information Systems	Machine learning; Application: Medical diagnosis (including self-diagnostic)	
			Machine Learning for Mortality Prediction	AI & Intelligent Information Systems	Machine learning	
1	16	Deep Learning for Object Detection, Semantic Segmentation, and Remote Sensing				2
			Person Identification	AI & Intelligent Information Systems	Computer Vision	
			Object Detection and Semantic Segmentation: Autonomous Navigation	AI & Intelligent Information Systems	Computer Vision; Instance and Semantic segmentation; Application: Object Detection; Application: Scene Understanding	
			Remote Sensing for Land Cover Classification	AI & Intelligent Information Systems; Data Acquisition Technologies	Satellite and remote sensing; Application: Camera-based crop monitoring systems	
		Generative Adversarial Networks (GANs)		AI & Intelligent Information Systems	Machine learning; Generative Adversarial Networks (GANs); Computer Vision; Neural Style Transfer	
		Image Captioning and Visual Question Answering		AI & Intelligent Information Systems	Computer Vision; Image captioning; Convolutional Encoder-Decoder	
1	27		Internet of Things (IoT) for Smart Cities, Smart Homes, and Real-Time Monitoring			2

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Fog, Edge, and Cloud computing		Computing	Distributed Computing; Fog Computing; Cloud Computing; Local/real-time processing; Edge Computing	
		Internet of Things (IoT)	Industrial Internet of Things (IIoT)	Networking	IoT networks; Network services and applications; Application: Industrial Internet of Things (IIoT)	
			Smart City	Networking	Network services and applications; Application: Smart City Applications	
1	84		Hyperspectral Imaging and Deep Learning for Plant Disease Detection and Classification			4
			Deep learning for Plant Disease Detection	AI & Intelligent Information Systems	Computer Vision; Application: Camera-based crop monitoring systems	
1	2	Cloud Computing (Edge computing)				4,6
1	17	Wireless Sensor Networks (WSNs)				3,5
1	22		Facial Emotion Recognition System			3,4
2	16	Deep Learning for Object Detection, Semantic Segmentation, and Remote Sensing				1

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Object Detection and Semantic Segmentation	Autonomous Navigation; urban scene understanding	AI & Intelligent Information Systems	Computer Vision; Application: Panoptic segmentation; Instance and Semantic segmentation; Application: Object Detection; Application: Scene Understanding	
			Target Detection (Pedestrian and real-time traffic detection)	AI & Intelligent Information Systems	Machine learning; Application: Target detection/tracking	
2	27		Internet of Things (IoT) for Smart Cities, Smart Homes, and Real-Time Monitoring			1
			Smart Home and Building	Networking	Wireless communication and network infrastructures; Zigbee; Network Management and Orchestration; Smart Building	
			Smart Agriculture; Precision Agriculture	Networking	Monitoring and remote control Applications; Application: Intelligent Agriculture	
2	83		Malware detection (also uses Machine Learning)			4
2	81	Face Recognition				5
			Image Retrieval	AI & Intelligent Information Systems	Computer Vision; Application: Image Retrieval	
		Non-negative Matrix Factorization and Dimensionality Reducation		AI & Intelligent Information Systems	-	

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
2	48	Fuzzy PID Controller				6
			Maximum Power Point Tracking Photovoltaic Fuzzy Logic Controller	Data Acquisition Technologies	Control Systems	
2	29	Software-Defined Networking (SDN) and Network Function Virtualization (NFV)				4
		Software-Defined Networking (SDN)		Networking	Network Management and Orchestration; Software-Defined Networking (SDN)	
2	1		Smart Grids and Countermeasures against side-channel attacks			3
			Smart Grids and Meters	Networking	Network Management and Orchestration	
2	56	Natural Language Processing and information retrieval				6
		Question answering, dialogue systems, named entity recognition, and relation extraction		AI & Intelligent Information Systems	Natural Language Processing (NLP); Question Answering; Application: Dialogue Systems/chatbot; Application: Named Entity Recognition (NER)	

continued ...

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
2	4		Process Monitoring and Fault Diagnosis			5,6
2	8	Cognitive Radio Networks				3,5
2	34		Smart Manufacturing			4,6
2	40		Cerebral Palsy Rehabilitation			5,6
2	43	Multi-Agent Formation Control				3,4
2	58	Resource and Power Allocation for Cellular Networks				4,6
2	61	Saliency Detection and Eye Tracking				5,6
2	62		Minimally Invasive Surgery with Virtual Reality and Robotic Assistance			3,5
2	63	Simultaneous Localization and Mapping (SLAM)				3,6
2	69	Multi-objective Topology Optimization using Evolutionary Algorithms				3,4
2	71	Visible Light Communication (VLC)				3,6
2	75	Super Resolution and Image Fusion				5,6
2	76	Machine Learning				3,4
2	85		Health Monitoring			4,5
2	28	Neuromorphic Computing				4,5,6
2	30	Process Control of Time Varying System				4,5,6

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
2	50		Sentiment Analysis and Opinion Mining			3,4,5
3	73	Resistive Random Access Memory (RRAM)		-	-	
3	92	Brain-Computer Interface (BCI) Systems (electroencephalography)		User Interface; Data Acquisition Technologies	Brain-computer Interface; Electroencephalography (EEG)-based brain-computer interfaces	
3	46	Autonomous Underwater Vehicles (AUVs) and Underwater Target Tracking		Robotics	Mobile robots; Underwater robot/vehicle	
			Target Tracking	AI & Intelligent Information Systems	Machine Learning; Application: Target detection/tracking	
3	78	Silicon Photonic Integrated Circuits		Computing	Optical/Photonic Computing; Silicon photonics	
3	1		Smart Grids and Countermeasures against side-channel attacks			2
			Countermeasures against side-channel attacks	Data Management Technologies	Encryption technologies/ Data Security; Advanced Encryption Standard (AES)	
3	42		Data hiding, digital watermarking, and steganography			5
			Steganography; digital watermarking	Data Management Technologies	Encryption technologies/ Data Security	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
3	14	Chaotic maps for Image encryption		Data Management Technologies	Encryption technologies/ Data Security	5
3	80	Indoor Positioning and Localization System		Data Acquisition Technologies	Satellite and remote sensing; Indoor positioning systems; Global Navigation Satellite Systems (GNSS) (GPS, GLONASS, Galileo)	4
		Bluetooth low energy (BLE) beacons		Networking	Wireless communication and network infrastructures; Bluetooth low energy (BLE) beacons	
			Wireless Sensor Networks Localization Algorithm	Networking	Wireless communication and network infrastructures; Wireless sensor networks (WSN)	
3	70	Ad hoc network (MANET and WSN)		Networking	Wireless communication and network infrastructures; Wireless sensor networks (WSN)	5
3	66		Recommender Systems (Collaborative Filtering e.g. Matrix Factorization)	AI & Intelligent Information Systems	Machine Learning; Application: Recommender/Recommendation system	5
3	51		Upper Limb Rehabilitation After Stroke (sEMG and Exoskeletons)	Robotics		6
			Prosthetic arm; Surface electromyography (sEMG)	Robotics	Manipulative robots; Application: Robotic Arm	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
3	87	Optical Character Recognition (OCR)		AI & Intelligent Information Systems	Computer Vision; Application: Optical Character Recognition (OCR)	6
3	77	Radio Frequency Identification (RFID)		Networking	Wireless communication and network infrastructures; Radio Frequency Identification (RFID)	5
3	93	Cloud Computing		Computing	Distributed computing; Cloud Computing; Application: Multicloud	5
		Cloud Storage		Data Management Technologies	Data Warehousing; Cloud data warehousing	
		Multi-cloud Data Management		Data Management Technologies	Data management Platforms; Hybrid and multi-cloud data management	
		Cloud Sharing and Services		Data Management Technologies	Data Sharing; Cloud services	
		Cloud Security and Privacy		Data Management Technologies	Encryption technologies/ Data Security; Homomorphic encryption; Cloud encryption, context-aware security, extended detection and response (XDR)	
3	97	Computer Stereo Vision for Autonomous Navigation		AI & Intelligent Information Systems	Computer Vision; Application: Object Detection; Application: Object Tracking	6

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
3	31	Neural Networks for Water Quality and landslide susceptibility				4
		Neural Networks for Water Resource Management		AI & Intelligent Information Systems	Machine Learning; Application: Remote sensing	
3	7	Parallel Manipulators				6
		Parallel Robots		Robotics	Manipulative robots; Parallel robot	
3	68		Driver Assistance Systems/Automated Driving			4
			Advanced Driver Distraction Warning (ADDW) System	User Interface	Immersive/Interactive Displays; Head-up display (HUD); Motion Sensing and Control; Eye-tracking	
3	49		Hybrid Electric Vehicle: Regenerative Braking/Predictive Control			5
			Control Systems for Autonomous Vehicles (steering and tires)	Robotics	Robotic control	
3	90		Computational Thinking			3
3	44	Vehicular Ad hoc Networks/Intelligent Transportation		Networking	Monitoring and remote control Applications; Application: Autonomous vehicles	5

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
3	88	Obstacle Avoidance for Robots & Unmanned Aerial Vehicles		Robotics	Mobile Robots; Wheeled robots; Robotic Navigation; Algorithm-based navigation (grid,sample,geometric,interval,etc)	4
3	0		Image Retrieval	AI & Intelligent Information Systems	Computer Vision; Application: Image Retrieval	6
3	3		Wearable Sensors for Assisted Living; human activity recognition	AI & Intelligent Information Systems; User Interface	Motion Sensing and Control; Wearable Computer;	5
				Data Acquisition Technologies	Sensing technologies; Body sensors	
3	67		Forecasting (time series including load, stock market)			5
			Short-term Electricity Load Forecasting	AI & Intelligent Information Systems; User Interface		Machine Learning
3	38	Quantum Communication/Quantum Key Distribution		Networking; Data Management Technologies	Intelligent and secure networks; Quantum networking; Encryption technologies/ Data Security; Quantum cryptography	6
3	60		Biomedical Signal Processing (electrocardiogram)	AI & Intelligent Information Systems; Data Acquisition Technologies	Machine Learning; Application: Medical diagnosis (including self-diagnostic); Biometric and body scans	5
3	5	Virtual Reality				4,6
3	8	Cognitive Radio Networks				2,5

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
3	17	Wireless Sensor Networks (WSNs)				1,5
3	22		Facial Emotion Recognition			1,4
3	32		Intrusion Detection Systems (IDS)			4,5
3	37	Hyperspectral Remote Sensing				4,5
3	43	Multi-Agent Formation Control				2,4
3	57	Biometric Authentication				5,6
3	62		Minimally Invasive Surgery with Virtual Reality and Robotic Assistance			2,5
3	63	Simultaneous Localization and Mapping (SLAM)				2,6
3	69	Multi-objective Topology Optimization using Evolutionary Algorithms				2,4
3	71	Visible Light Communication (VLC)				2,6
3	76	Machine Learning				2,4
3	50		Sentiment Analysis and Opinion Mining			2,4,5
4	10	Three-Dimensional Bioprinting Hydrogel-based Bioink		Additive Manufacturing	Three Dimensional Bioprinting	
			Implants; Bone and Tissue Regeneration			

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
4	35	Millimeter-Wave Communication (Mitigation of Rain Attenuation)		Networking	Wireless communication and network infrastructures; mmWave networks; 5G mobile networks	
4	41	Bio-inspired Reinforcement learning (Neural Circuits)		AI & Intelligent Information Systems	Machine learning; Reinforcement Learning	
				Robotics	Robotic control; Deep Reinforcement Learning	
4	11	Control strategies for microgrids (e.g. droop control, virtual synchronous generators)		Robotics	Robotic Control; Distributed control	
			Renewable energy (wind) generation, storage, and transmission			
4	33	Additive Manufacturing				
		Powder Bed Fusion		Additive Manufacturing	Powder Bed Fusion; Selective Laser Melting (SLM); Electron beam melting (EBM); Selective Laser Sintering (SLS)	
		Directed Energy Deposition (DED); Gas Metal Arc Welding (GMAW)		Additive Manufacturing	Directed Energy Deposition (DED); Wire Arc Additive Manufacturing (WAAM); Laser Metal Deposition-wire (LMD-w); Direct Metal Deposition (DMD)	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Fused Deposition Modeling		Additive Manufacturing	Material Extrusion; Fused Deposition Modeling/Fused Filament Fabrication (FFF)	
		Rapid Prototyping; Design for Additive Manufacturing		Additive Manufacturing	Rapid Prototyping	
4	55	SSDs for Big Data Storage (including hadoop, map reduce, Non-volatile memory)				
		Distributed Storage Systems		Data Management Technologies	Data Storage; Distributed Storage Systems	
		Non-volatile memory		Computing	Integrated circuit design; Non-volatile memory	
		NVMe; Solid State Drive; NAND Flash Memory		Data Management Technologies	Data Storage; 3D NAND Flash Memory	
		Distributed Databases		Data Management Technologies	Distributed Databases	
4	52	Machine Learning for Drug Discovery and Gene Regulatory Networks				
		Support Vector Machine; random forest	Drug Discovery	AI & Intelligent Information Systems	Machine Learning	
		Deep Learning	Protein Structure Prediction	AI & Intelligent Information Systems	Machine Learning; Application: Protein structure prediction (e.g. AlphaFold)	

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
4	84		Hyperspectral Imaging and Deep Learning for Fruit Disease Detection and Classification			1
			Deep learning for Fruit Classification and Disease Detection	AI & Intelligent Information Systems	Computer Vision; Application: Camera-based crop monitoring systems	
4	83		Malware detection (also uses Machine Learning)			2
			Control-Flow Integrity and SDX			
4	29	Software-Defined Networking (SDN) and Network Function Virtualization (NFV)				2
		Network Function Virtualization (NFV) and Software-Defined Networking (SDN) Controller		Networking	Network Management and Orchestration; Network Function Virtualization (NFV)	
4	80	Indoor Positioning and Localization System				3
		WiFi Fingerprinting Based Indoor Positioning		Data Acquisition Technologies	Satellite and remote sensing; Indoor positioning systems; Global Navigation Satellite Systems (GNSS) (GPS, GLONASS, Galileo)	

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
4	31		Neural Networks for Water Quality and landslide susceptibility			3
			Landslide and Flood Susceptibility Mapping	AI & Intelligent Information Systems	Machine Learning; Application: Remote sensing	
4	68		Driver Assistance Systems/Automated Driving			3
		Autonomous/semi-Autonomous Navigation (Interface and Simulator)		User Interface	Immersive/Interactive Displays	
			Driver Drowsiness Detection System; AI & Intelligent Information Systems	Data Acquisition Technologies	Biometric and body scans; Electroencephalography (EEG) and Magnetoencephalography (MEG); Automated tracing and tagging; Eye Tracking; Machine learning; Neural networks	
4	88	Obstacle Avoidance for Robots & Unmanned Aerial Vehicles				3
		Path Planning for Unmanned Aerial Vehicles (UAV)		Robotics	Mobile Robots; Aerial robot/vehicle; Robotic Navigation; Algorithm-based navigation (grid,sample,geometric,interval,etc)	

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Multi-Robot Systems; Swarm Intelligence		AI & Intelligent Information Systems	Robotics applications; Application: Multi-Robot Systems; AI-based computing; Swarm Intelligence	
		Robotic Learning; Reinforcement Learning		Robotics; AI & Intelligent Information Systems	Robotic Control; (Deep) Reinforcement learning; Imitation learning; Robotics applications; Application: Robotic Learning from Demonstration	
4	20		Social Media Analytics			6
			Point of Interest Recommendations in Location-based Social Networks; Topic Modeling	AI & Intelligent Information Systems	Machine Learning; Application: Recommender/Recommendation system; Natural Language Processing; Application: Text classification and Text Categorization	
4	36	Smart Materials and Soft Robotics				6
		Shape-memory polymers (SMPs)		Robotics	Smart Materials	
		Electronic Skin; Flexible pressure/strain sensors for wearable electronics		Data Acquisition Technologies	Sensing technologies; Tactile sensors; Soft robotics sensors	
		Soft Actuators for Soft Robotics; Dielectric Elastomers; Artificial Muscles		Robotics	Smart Materials; Electroactive Polymers (EAPs); Silicone elastomers	

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
4	9	Quantum Computing and Quantum Dots		Computing	Quantum computing; Quantum dot quantum computer; Ion trap quantum computer; Superconducting quantum computer; Adiabatic quantum computer; Quantum error-correcting turbo codes; Nuclear magnetic resonance quantum computer	6
4	98	Augmented Reality and Gamification	Augmented Reality for Teaching and Tourism	User Interface	Extended Reality (XR); Augmented Reality; Virtual Reality; Immersive/Interactive Displays; Head-mounted Display (HMD); Wearable Computer; Smart Glasses	6
4	82	Gallium Nitride (GaN) High Electron Mobility Transistors (HEMTs)				6
4	53	Fractional Order Systems and Control (Fractional Order PID Controller)		Data Acquisition Technologies; AI & Intelligent Information Systems	Control Systems; Programmable logic controller (PLC) and PID Controller; Robotics Applications; Application: Adaptive Control	5
4	94		Information Security and Enterprise Resource Planning Systems (ERP)			6
			Information Security Governance	Data Management Technologies	Encryption technologies/ Data Security	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
4	72	Human Action and Gesture Recognition				6
			Human Activity Recognition	User Interface; AI & Intelligent Information Systems	Wearable Computer; Machine Learning; Neural Networks	
		Sign Language and Gesture Recognition		User Interface	Motion Sensing and Control; Hand gesture recognition	
		Skeleton Action Recognition		User Interface	Motion Sensing and Control; Human Pose Estimation (HPE)	
4	54		Peak-to-Average Power Ratio (PAPR) Reduction in orthogonal frequency-division multiplexing (OFDM) Systems	Networking		6
4	74	Tactile Sensors and Haptic Feedback				6
		Tactile Sensor		Data Acquisition Technologies	Sensing technologies; Tactile sensors; Soft robotics sensors	
		Rubber Hand Illusion (RHI)		Robotics	Smart Materials; Electroactive Polymers (EAPs)	
4	25		Remote Laboratory and Educational Robots			6
			Educational Robots	Robotics	Manipulative robots	
4	79	Social/Humanoid Robots				6

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
			Virtual Reality-based Assistive Technology for Children with Autism Spectrum Disorder	User Interface; AI & Intelligent Information Systems	Extended Reality (XR); Social and web-based VR; Application; Natural Language Processing (NLP); Speech-generating device	
			Humanoid and Socially Assistive Robots for Older Adults and Children with ASD	AI & Intelligent Information Systems; Robotics	Robotic control; Telepresence; Robotics applications; Application: Human-like Robotics; Application: Robot personality	
4	24		Participatory Design for Human Computer Interaction (HCI)	User Interface	Spatial computing; Tangible User interface	6
4	2	Cloud Computing				1,6
4	5	Virtual Reality				3,6
4	6		Traffic Flow Management			5,6
4	19		Privacy-Preserving Data Mining			5,6
4	22		Facial Emotion Recognition			1,3
4	32		Intrusion Detection Systems (IDS)			3,5
4	34		Smart Manufacturing			2,6
4	37	Hyperspectral Remote Sensing				3,5
4	43	Multi-Agent Formation Control				2,3

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
4	58	Resource and Power Allocation for Cellular Networks				2,6
4	69	Multi-objective Topology Optimization using Evolutionary Algorithms				2,3
4	76	Machine Learning				2,3
4	85		Health Monitoring			2,5
4	30	Process Control of Time Varying System				2,5,6
4	28	Neuromorphic Computing				2,5,6
4	50		Sentiment Analysis and Opinion Mining			2,3,5
5	81		Face Recognition			2
			Audio Source Separation using Non-negative Matrix Factorization	AI & Intelligent Information Systems	-	
5	42		Data hiding, digital watermarking, and steganography			3
			Steganalysis; Reversible Data Hiding and Watermarking	Data Management Technologies	Encryption technologies/ Data Security	
5	14	Chaotic maps for Image encryption				3
		Chaos communications		Data Management Technologies	Encryption technologies/ Data Security	

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
5	70	MANET Routing Protocol (AODV, DSR)		Networking	Wireless communication and network infrastructures; Wireless sensor networks (WSN)	3
5	66		Data Mining Algorithms (e.g. Association rule and frequent item set)	AI & Intelligent Information Systems	Machine Learning	3
5	77	Radio Frequency Identification (RFID)				3
		RFID tags for authentication		Networking	Wireless communication and network infrastructures; Radio Frequency Identification (RFID)	
			RFID-enabled Supply Chain and Inventory Management System	Networking	Wireless communication and network infrastructures; Radio Frequency Identification (RFID); IoT networks	
5	93	Cloud Computing				3
		Cloud Trust Management; Cloud Services		Data Management Technologies	Data Sharing; Cloud services	
		Mobile Cloud Computing; Cloudlets		Computing	Distributed computing; Application: Cloudlet	
5	49	Hybrid Electric Vehicle: Regenerative Braking				3

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
5	44	Vehicular Ad hoc Networks/Intelligent Transportation/Internet of Vehicles (IoV)		Networking	Monitoring and remote control Applications; Application: Autonomous vehicles	3
5	3	Wearable Sensors for Assisted Living				3
			Wearable sensors for Gait Analysis of Elderly	User Interface; AI & Intelligent Information Systems; Data Acquisition Technologies	Wearable Computer; Smart Watch; Sensing technologies; Body sensors; Accelerometers	
			Ambient Assisted Living Systems	Networking	Network Management and Orchestration; Application: Smart Buildings	
5	67		Forecasting (time series including load, stock market)			3
			Forecasting: Time series; Short-term Electricity Load; Solar Radiation	AI & Intelligent Information Systems	Machine learning; Neural networks	
5	60	Biomedical Signal Processing (electrocardiogram)	Atrial Fibrillation; Arythmia Classification	Data Acquisition Technologies; AI & Intelligent Information Systems; Data Acquisition Technologies	Biometric and body scans; Machine Learning; Application: Medical diagnosis (including self-diagnostic); Neural networks	3
5	53	Fractional Order Systems and Control				4

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Synchronization of Complex Dynamical Networks and Chaotic Systems		Data Acquisition Technologies; AI & Intelligent Information Systems	Control Systems	
5	13	Semantic Web Service Composition; Web service Discovery		Computing; Robotics	Software/virtual robots; Workflow automation software	6
5	99		Vehicle Routing	Computing	AI-based computing; Swarm Intelligence; Evolutionary Computing	6
5	26	Optical Networking (Optical Burst Switching)		Networking	Wireless communication and network infrastructures; Wireless optical communication (e.g., Li-Fi, FSO)	6
5	18	Cache Memory				6
		DRAM, Magnetoresistive RAM (MRAM), Non-volatile Memory (NVM), NVRAM		Computing; Data Management Technologies	Integrated circuit design; Non-volatile memory; Data Storage	
5	91	Mixed Criticality Real-time Systems		Computing	Local/real-time processing	6
5	89	Network Mobility Management (e.g. PMIPV6)		Networking	Network Management and Orchestration	6
5	4	Process Monitoring and Fault Diagnosis				2,6
5	6		Traffic Flow Management			4,6
5	8	Cognitive Radio Networks				2,3
5	17	Wireless Sensor Networks (WSNs)				1,3

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
5	19	Privacy-Preserving Data Mining				4,6
5	32		Intrusion Detection Systems (IDS)			3,4
5	37	Hyperspectral Remote Sensing				3,4
5	40		Cerebral Palsy Rehabilitation			2,6
5	57	Biometric Authentication				3,6
5	61	Saliency Detection and Eye Tracking				2,6
5	62		Minimally Invasive Surgery with Virtual Reality and Robotic Assistance			2,3
5	75	Super Resolution and Image Fusion				2,6
5	85		Health Monitoring			2,4
5	30	Process Control of Time Varying System				2,4,6
5	28	Neuromorphic Computing				2,4,6
5	50		Sentiment Analysis and Opinion Mining			2,3,4
6	59	Piezoelectric Actuators for Micro/Nano Positioning and Atomic Force Microscopy (AFM)		Robotics	Smart Materials; Piezoelectric Materials	
6	15	Genetic Algorithm for job-shop scheduling (ML)		Computing	AI-based computing; Evolutionary Computing	
6	47	Adaptive Video Streaming		Data Management Technologies	Real-time streaming	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
6	86	Predictive Control and PID Controllers for Manufacturing Systems				
		Non-Linear Model Predictive Control; Distributed control		Robotics	Robotic control; Nonlinear Model Predictive Control (NMPC); Distributed control	
		PID Controller		Data Acquisition Technologies	Control Systems; Programmable logic controller (PLC) and PID Controller	
6	65	Energy-Efficient High-Performance Computing with Multi-Core Processors and GPUs		Computing	Local/real-time processing; Low-power computing	
			High-Performance Computing	Computing	Distributed computing; Application: High performance clusters	
		Embedded Computing		Computing	Local/real-time processing; Embedded Computing	
		Parallel Computing		Computing	Distributed computing; Parallel computing	
			Application: Field programmable gate array (FPGA)	Computing	Integrated circuit design; Application: Field programmable gate array (FPGA)	
6	95	Legged and Humanoid Robots (Bipedal and Quadrupedal)				
		Legged Robot		Robotics	Mobile robots; Legged robots	

continued . . .

... continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Humanoid Robot		AI & Intelligent Information Systems	Robotics applications; Application: Human-like Robotics	
6	45	Fringe Projection and Holography for Three-Dimensional Measurement and Display				
		Computer Generated Holographic Display (CGH); Autostereoscopic Display; Light Field Display (LFD)		User Interface	Immersive/Interactive Displays; Holographic user interfaces (e.g. Light Field Displays or LFDs)	
		Liquid Crystal Display (LCD)		User Interface	Immersive/Interactive Displays; Transparent Displays (OLED, LED, LCD, microLED, AMOLED)	
6	64	Assistive Technology for People with Disabilities				
		Tactile Display		User Interface	Immersive/Interactive Displays; Tactile Display	
			Voice User Interface	User Interface; AI & Intelligent Information Systems	Conversational UI; Application: Voice User Interface (e.g., Alexa); Application: Virtual assistants	
		Vibrotactile Feedback				
6	23	Machine Translation and Speech Recognition				
		Automatic Speech Recognition and Synthesis		User Interface; AI & Intelligent Information Systems	Conversational UI; Speech Recognition; Speech Synthesis	

continued ...

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Speaker Recognition and Verification		User Interface	Biometric Recognition; Voice recognition	
			Music Genre Classification			
		Transcription		AI & Intelligent Information Systems	Natural Language Processing (NLP); Application: Text-to-Speech (TTS) and Speech-to-Text (STT); Application: Neural Machine Translation	
6	21	Sliding Mode Control for Unmanned Aerial Vehicle				
		Unmanned Aerial Vehicle (UAV)		Robotics	Mobile robots; Aerial robot/vehicle	
		Fuzzy Sliding Mode Control		Robotics	Robotic control; Fuzzy Control	
6	48	Fuzzy PID Controller				2
		Fuzzy PID Controller	Induction motor Fuzzy Logic PI Controller	Data Acquisition Technologies; Robotics	Control Systems; Programmable logic controller (PLC) and PID Controller; Robotic control; Fuzzy control	
6	56	Natural Language Processing and information retrieval				2
		Text Classification (e.g. Naïve Baye's, SVM)		AI & Intelligent Information Systems	Natural Language Processing (NLP); Application: Text classification and Text Categorization	
		Information Retrieval and Personalized Search		AI & Intelligent Information Systems	Natural Language Processing (NLP)	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Text Mining (e.g. Biomedical); Word Sense Disambiguation; Web Ontology Language (OWL); NER, Question-Answering		AI & Intelligent Information Systems	Natural Language Processing (NLP); Question Answering; Application: Named Entity Recognition (NER)	
6	51		Upper Limb Rehabilitation After Stroke (sEMG and Exoskeletons)			3
			Robotic exoskeletons for post-stroke upper limb Rehabilitation; Surface electromyography (sEMG) to detect muscle fatigue	Robotics	Manipulative robots; Application: Robotic Arm	
6	87	Text Detection and Segmentation for Optical Character Recognition (OCR)		Data Acquisition Technologies; AI & Intelligent Information Systems	Scanning; Computer Vision; Application: Optical Character Recognition (OCR)	3
6	97	Computer Stereo Vision for Autonomous Navigation	Lane Detection	AI & Intelligent Information Systems	Computer Vision; Application: Object Detection; Application: Object Tracking	3
		Stereo Matching; Disparity Map for Stereo Vision; Depth Map		AI & Intelligent Information Systems	Computer Vision; Stereo matching; Depth map/image	
6	7	Parallel Manipulators				3
		Flexible Joint Robotic ManipulatorParallel Robots		Robotics	Manipulative robots; Application: Robotic arm	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
		Haptic Bilateral Teleoperation		Robotics	Robotic control	
6	90		Engineering Education			3
6	0	Image Retrieval		AI & Intelligent Information Systems	Computer Vision; Application: Image Retrieval	3
		Ray Tracing; Facial Motion Capture for Animation; Shape Retrieval		AI & Intelligent Information Systems; User Interface	Motion Sensing and Control	
		Image Registration	Remote Sensing	Data Acquisition Technologies		
6	38	Quantum Computing and Quantum Communication		Networking; Computing	Intelligent and secure networks; Quantum networking; Quantum computing	3
6	20	Social Media Analytics (data visualization)				4
6	36	Smart Materials and Soft Robotics				4
		Ionic Polymer Metal Composite Actuators (IPMCs); Dielectric Elastomers (DEA)	Artificial Muscles	Robotics	Smart Materials; Electroactive Polymers (EAPs); Carbon Nanotubes (CNTs)	
6	9	Quantum Dots		Computing; Data Acquisition Technologies	Quantum computing; Quantum dot quantum computer; Sensing technologies; Application: Quantum dot biosensors	4

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
6	98	Educational Games and Virtual Reality (e.g. Second Life)	Virtual Classroom	User Interface	Extended Reality (XR); Virtual Reality; Mixed Reality	4
6	82	Gallium Nitride (GaN) Light Emiting Diodes (AlGAN/InGan)		User Interface	Immersive/Interactive Displays; Transparent Displays (OLED, LED, LCD, microLED, AMOLED)	4
6	94	Information Security and Enterprise Resource Planning Systems (ERP)				4
			Enterprise Resource Planning (ERP) Systems	Data Management Technologies; Robotics	Software	
6	72	Human Action and Gesture Recognition				4
		Human Action/Activity Recognition		AI & Intelligent Information Systems	Computer Vision; Application: Action Recognition	
		Sign Language and Gesture Recognition		User Interface	Motion Sensing and Control; Hand gesture recognition; Hand and finger tracking	
6	54		Peak-to-Average Power Ratio (PAPR) Reduction in orthogonal frequency-division multiplexing (OFDM) Systems	Networking		4
		MIMO-OFDM and CDMA communication systems		Networking	Wireless communication and network infrastructures	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
6	74	Tactile Sensors and Haptic Feedback				4
		Tactile Display and Feedback		User Interface	Immersive/Interactive Displays; Tactile Display; Haptic technology; Haptic Feedback	
		Touch Screen and Gesture Interaction		User Interface	Immersive/Interactive Displays; Motion Sensing and Control; Hand gesture recognition	
6	25		Remote Laboratory and Educational Robots			4
			Virtual Instruments Systems In Reality (Remote Laboratory)	User Interface	Extended Reality (XR); Virtual Reality	
6	79	Social/Humanoid Robots				4
		Human Robot Interaction; Humanoid and Social Robots; Robot Personality: Emotion Recognition and Empathy		User Interface; AI & Intelligent Information Systems; Robotics	Software/virtual robots; Conversational robots/Virtual assistants; Manipulative robots; Application: Cobot (collaborative robot) Robotic control; Telepresence; Robotics applications; Application: Human-like Robotics; Application: Robot personality; Motion Sensing and Control; Facial expression recognition	

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
6	24	Participatory Design for Human Computer Interaction (HCI)	User Interfaces for Elderly	User Interface	Spatial computing; Tangible User interface	4
6	13	Semantic Web Service Composition		Computing; Robotics	Software/virtual robots; Workflow automation software	5
6	99	Vehicle Routing		Computing	AI-based computing; Evolutionary Computing	5
6	26	Optical Networking (elastic optical networks)		Networking	Wireless communication and network infrastructures; Wireless optical communication (e.g., Li-Fi, FSO)	5
6	18	Cache Memory (Cache prefetching)		Computing	-	5
6	91	Fixed Priority Real-time Systems		Computing	Local/real-time processing	5
6	89	Wireless Mesh Networks		Networking	Network Management and Orchestration	5
		Medium Access Control for Wireless Local Area Networks (WLANs) (e.g. EDCA, DCF)		Networking	Wireless communication and network infrastructures	
		Long-term Evolution (LTE) Network; Vertical Handoff in Heterogeneous Wireless Networks		Networking	Wireless communication and network infrastructures	
6	2	Cloud Computing				1,4
6	4		Process Monitoring and Fault Diagnosis			2,5

continued . . .

. . . continued

Pattern	ID	Technology	Application	Family	Subfamily	Patterns
6	5	Virtual Reality				3,4
6	6		Traffic Flow Management			4,5
6	19	Privacy-Preserving Data Mining				4,5
6	34		Smart Manufacturing			2,4
6	40		Cerebral Palsy Rehabilitation			2,5
6	57	Biometric Authentication				3,5
6	58		Resource and Power Allocation for Cellular Networks			2,4
6	61	Saliency Detection and Eye Tracking				2,5
6	63	Simultaneous Localization and Mapping (SLAM)				2,3
6	71	Visible Light Communication (VLC)				2,3
6	75	Super Resolution and Image Fusion				2,5
6	30	Process Control of Time Varying System				2,4,5
6	28	Neuromorphic Computing				2,4,5

Notes: The table lists broad technologies allocated to the different patterns of emergence. Each broad technology is composed of one or more of the 500 technologies identified in Sections 3.3. *ID*: the unique identifier of the broad technology; *Technology*: the name of the broad technology; *Application*: applications frequently mentioned in the relevant pattern in the patent labels; family of technology; Patterns: patterns to which the broad technology pertains (because different technologies fall under one broad technology, the patterns can be multiple, and the technologies within one broad technology can differ, for example in their applications).

Revealing Semantics: Exposure of Industries and Occupations to Emerging STI Areas*

Ekaterina Prytkova† Fabien Petit‡ Deyu Li§

Sugat Chaturvedi¶ Tommaso Ciarli‖

June 2023

Abstract

This paper proposes a novel methodology to estimate occupation and industry exposure to digital automation technologies and research areas. We measure the semantic connection between patents and publications documenting emerging digital automation technologies and descriptions of industries (NACE) and occupations (ISCO) using Natural Language Processing (NLP). We distinguish industries exposure between technology users and producers, leveraging firm's patent portfolios. We find that, besides machine operators, technicians and professionals (non-routine occupations) are also highly exposed, and managers are in the middle of the exposure distribution. Whereas the most exposed industries are those producing automation technologies. We next estimate if highly exposed occupations-industries in Europe are associated with changes in employment between 2011 and 2019. We find that highly exposed occupations are associated with employment growth in sectors producing automation technologies and employment decline in sectors using automation technologies. Using more granular German data we show that the nature of user industries matters, as tasks across industries differ for the same occupation.

Keywords: Industry exposure; Occupation exposure; Text as data; Natural language processing; Emerging technologies; Automation
JEL Codes: O33, O25, J24, O52

---

*We are grateful to Önder Nomaler as well as to audiences at the ESCoE Conference on Economic Measurement 2023, 13th Global Tech Mining conference, Eu-SPRI Conference 2023, Venice Summer Institute 2023, and seminar participants at JRC–European Commission for useful comments and suggestions. This project received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101004703. The data are made available to PILLARS partners and funder, and will be publicly shared once we publish the working paper.

†University of Sussex, SPRU. Email: e.prytkova@sussex.ac.uk.

‡University of Sussex, SPRU. Email: fabien.petit@sussex.ac.uk. Website: fabienpetit.com.

§Utrecht University.

¶University of Sussex, SPRU. Email: sc2057@sussex.ac.uk

‖UNU-MERIT.

1 Introduction

Emerging automation technologies are likely transform industries and occupations across countries and regions (Nedelkoska and Quintini, 2018; Mitchell and Brynjolfsson, 2017; Frank et al., 2019; Ciarli et al., 2021). Although knowledge and technological progress is serendipitous, scientists and engineers, contribute to developing technologies with a purpose, within more or less defined technological paradigms (Dosi, 1982). This is no different for digital automation technologies, were developers and producers aim to develop technologies that can perform specific tasks in specific industries (Leopold et al., 2018; Cirillo et al., 2020) — e.g., picking ripe strawberries or developing mobile robots (Cebollada et al., 2021).

Understanding the implications of digital automation on work, therefore requires and understanding of which are the tasks, occupations and industries for which technologies are developed. Looking ahead, to understand the future potential impacts of automation technologies, requires an understanding of which technologies are more likely to emerge (Chaturvedi et al., 2023).

This paper proposes a novel methodology to estimate occupational and industrial exposure to Sciences, Technologies, and Innovations (STIs) as described in documents such as patents and publications. We use this methodology to estimate the potential exposure of industries and occupations to emerging digital automation technologies. We consider the 1,000 emerging automation technologies and research areas identified in Chaturvedi et al. (2023), which include 500 technologies and 500 scientific areas. These latter correspond to clusters of, respectively, patents and publications. We use industry and occupation classifications to represent economic activities and jobs in order to estimate their exposure to emerging STIs.[1]

We construct the semantic connection between patents and publications describing emerging digital automation technologies with the descriptions of the 3-digit NACE Rev.2 industries and of the 4-digit ISCO-08 occupations. The semantic links between emerging STIs and industries/occupations are established based on their respective textual descriptions. For patents, we use titles that include information on the *essence* of the technology in its first part (e.g. "Energy usage managing method use in data center"), and information on the function of the technology in its second part (e.g. "involves automatically change power state of predetermined system based on predicted non-zero future power requirement for performing expected workload"). For industries, we rely on the description provided in the NACE Rev.2 classification at the 3-digit level by concatenating descriptions of sub-level industries. For occupations, we rely on the task descriptions provided in the ISCO-08 classification at

[1] This version presents results using patented inventions (technologies). A follow-up version will also include the research areas (publications).

the 4-digit level.

Building upon those descriptions as text data, we derive the embeddings by using pre-trained sentence transformers, namely, MPNet v2 which maps sentences from a description into a 768-dimensional vector space (Reimers and Gurevych 2019). For each industry-patent and occupation-patent combinations, we compute the cosine similarity whose magnitude reflects the strength of similarity between a pair of documents.

To improve the matching quality, we introduce a redundancy filtering procedure in the calculation of cosine similarity for both types of combinations. Once filtered, we aggregate from patents to technologies since the latter are defined as clusters of the former. Lastly, we perform a second filtering procedure at the technology level that considers only industry–technology and occupation–technology pairs that matter to the technology.

Because some of the industries that are likely to have a description similar to that of automation technologies are those that produce the technology — e.g. machine manufacturers — we next distinguish between producing and using industries. To identify industries that produce emerging automation technologies, we leverage patent portfolios of the firms that filed patents that constitute our emerging technologies. We identify as producers the firms that are specialised in an automation technology. We then distinguish the exposure measure for producing and using industries.

We find that, besides the machine operators the most exposed 1–digit ISCO-08 categories are 'Clerical support workers' and 'Technicians and Associate Professionals', followed by 'Professionals' and 'Managers', which are in the middle of the distribution. These are occupations that involve also several non-routine and non-manual tasks. That is, automation technologies that we identify as emerging in the most recent decades seem to be designed to perform more and more tasks that are performed by non-routine workers. We also find that technologies that involve tangible capital are more relevant to perform tasks related to medium–high skilled workers, whereas intangible technologies (digital, information processing) are more relevant to perform tasks related to both high and low skilled workers.

Second, we find that the most exposed industries are those that produce the automation technologies, such as 'Manufacture of computer, electronic and optical products', 'Computer programming, consultancy and related activities' (62), or 'Data processing, hosting and related activities; web portals'. These are also the industries that are connected to most technologies. If the most exposed industries are those that produce the technologies, do exposed occupations in these industries complement or are replaced by the technologies?

To address this question, we next estimate the relation between industry-occupation exposure to emerging automation technologies and the change in industry-occupation employment for 28 European countries between 2011-19. We distinguish between producing and using industries (using 1-digit NACE classification). We also analyse the same relation for Germany, where we can distinguish between users and producers at a more granular level (2-digits NACE), and also distinguish between different types of users.

Model estimates at the 1-digit industry level indicate that employment has increased with occupational exposure for producing sectors but decreasing for using sectors. The negative relation for using sector, though is small, even for highly exposed occupations. This is because there is substantial heterogeneity between using industries: some benefit from exposure (complement technologies), others have seen a substantial decrease in employment in Europe across occupations.

We see this when we dig at the two digit-level for Germany. We still find a positive relation for producing industries. In fact, also industries that are intensive users (in the middle of the industry exposure distribution) and non-IT deep users (among the most exposed industries, which are not IT based) see employment in the most exposed occupations. Instead, the IT deep users (among the most exposed industries, which are IT based) see a substantial reduction in employment of up to 20%, for most occupations. The light users (the least exposed industries) see only small negative (non significant) changes in employment.

The contribution of this paper is twofold. First, we propose a novel methodology that relies on state-of-the-art Natural Language Processing (NLP) to estimate industrial and occupation exposure scores to emerging automation technologies. This creates a dynamic mapping that shows the relevance of emerging automation technologies to: (i) industries based on the nature of their economic activity, (ii) occupations based on the tasks and functions they perform, and the ontological domain of knowledge (e.g. law, medicine, engineering, etc.). Thus, we outline a reproducible and scalable method that allows construction of dynamic links between sectors/occupations and automation technologies.

Second, we estimate the employment change in occupations, based on their exposure to emerging STIs, while differentiating this effect between user and producer sectors of technologies, and different user industries.

The paper is organized as follows. Section 2 presents the data. Section 3 describes the methodology. Section 4 shows the exposure of occupations and industries to emerging STIs. Section 5 estimates the effect of exposure on employment growth over the last decade. Section 6 concludes.

2 Data

Emerging technologies. We measure the exposure of industries and occupations to a set of 500 technologies identified in Chaturvedi et al. (2023).The patent sample $\mathcal{P}$ includes

190,714 patents clustered into 500 distinct technologies $k = [0; 499]$ with average size of 381.4 patents ($\sigma = 157.9$). In this paper, we use only *patent titles* for matching with industrial and occupational descriptions.

Industry and occupation taxonomies. We use industry and occupation classifications to represent economic activities and jobs respectively in order to estimate their exposure to emerging STIs. We adopt the European statistical classification of economic activities, Nomenclature of Economic Activities or NACE (Rev. 2), using both (i) industrial titles and descriptions, and (ii) hierarchical structure of the classification for matching with STI texts. Analogously, the International Standard Classification of Occupations or ISCO (in particular, the latest version ISCO-08) serves as a textual and hierarchical representation of occupation groups.

Producer industries. We use information from the European Patent Office Worldwide Patent Statistical Database PATSTAT (2023 Spring) and ORBIS Intellectual Property Database (ORBIS IP) to identify industries that produce each technology. We proceed in three steps.

First, we search for Derwent patents in PATSTAT based on the publication number.[2] We match 99% of Derwent patents in our sample to patents in PATSTAT.

Second, using the matched publication number from PATSTAT, we query the ORBIS IP database to retrieve information on companies that filed the patents. For each company, we record their ID, company name, and 2-digit NACE Rev.2 code.[3]

Third, we extend the set of industrial codes of companies collected from ORBIS IP with those recorded in PATSTAT.

As a result, the final set of industries assigned to emerging STI patents p includes the core, primary and secondary NACE Rev.2 codes and is a combination of those found in both ORBIS IP and PATSTAT.

Employment. We use data on employment from the European Labor Force Survey (EU-LFS) and the German Labor Force Survey to estimate the effect of emerging STIs on employment over the last decade.

The EU-LFS data include information on the level of employment from 2011 to 2019

[2] Derwent and PATSTAT process the original publication number from national and regional patent offices differently. Thus, we manually created a concordance table between the publication number of Derwent patents (i.e. column 'PD') and the publication number of PATSTAT patents (i.e. column 'PUBLN_NR' in table 'TLS211_PAT_PUBLN'). The concordance table is available upon request.

[3] Each company in the ORBIS IP database has a unique BvD ID number, created from its national company number.

at the 1-digit industry and 3-digit occupation levels.[4] We consider both employed and self-employed workers between the age of 15 and 89 in 28 European countries.[5]

For what concerns industries, we observe changes in employment for broad 1-digit NACE sectors in response to identified emerging automation technologies. This level of analysis is still too aggregated to disentangle dynamics within large sectors, such as Manufacturing (C). Since EU-LFS data are not available at a more disaggregated level for industries, we rely on the German LFS that includes information at the 2-digit NACE industry level as well as the 3-digit ISCO occupation level. In Section 5, we use both EU-LFS and German LFS data to construct the empirical model and compare the results.

3 Methodology

In this section, we present the methodology to compute the exposure of industries and occupations to emerging STIs. It's crucial to stress already here that we posit that the computed exposure scores reflect the *relevance* of each technology to an industry/occupation rather their de facto adoption. In the case of industries, the relevance link is based on whether or not a technology has something to do with the way an industry produces its output or the technology is the improved output of an industry. For what concerns occupations, the link reflects relevance of a technology to tasks and functions performed within an occupation. As the technologies in our STI sample are emerging, their de facto impact on employment is unfolding, with different pace across countries, industries and occupations. This will have implications on the interpretation of results in Section 5.

We start with a brief description of textual data and its processing before proceeding with the calculation of document embeddings. Then we explain the rationale and calculations of semantic-based exposure scores. Finally, leveraging established semantic links specifically between technologies and industries, we employ data from PATSTAT and ORBIS IP databases to identify an industry's producer (source) or user (recipient) status.

3.1 Text Corpora: STI, Industries, and Occupations

The semantic links between emerging STIs and industries or occupations are established based on their respective textual descriptions. We introduce some notations. Let $i \in \mathcal{I}$ be

[4] Although the EU LFS provides data until 2021, we prefer to restrict our analysis to the period between 2011 and 2019, hence, not considering the employment changes over the COVID-19 pandemic period.

[5] This list of countries includes (in alphabetical order): Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Spain, Sweden, Switzerland, and United-Kingdom.

an industry, $o \in \mathcal{O}$ an occupation, and $p \in \mathcal{P}$ a patent.

Patent. Patents consist of a *title* and an *abstract* which includes labeled topical blocks such as novelty, independent claims, use, etc.[6]

Titles are comprised of two parts: the first part summarises the *essence of the technology* in a phrase or short sentence (e.g. "Energy usage managing method use in data center"), we denote this part $p_1 \in p$; whereas the second one describes the *function of the technology*, or equivalently, its intended use (e.g. "involves automatically change power state of predetermined system based on predicted non-zero future power requirement for performing expected workload"), we denote this part $p_2 \in p$.

Compared to the abstract, the title has two significant advantages for the purpose of our analysis. First, the title is always present and compact and follows the described structure (i.e. stable meaning). Second, the language used to represent the technology—and its function—conveys technical information through detailed descriptions rather than highly technical jargon. These two properties of the patent title map well onto characteristics of industrial and occupation descriptions making the title the most suitable candidate to use for semantic matching.

Industry. We choose the 3-digit NACE Rev.2 as the most fine-grained level at which we consider industries. The choice is based on, first, the fact that at the 4-digit level subsets of industries under the same 3-digit parent might not be substantially different with regard to their connections to emerging STIs, hence, could be merged with no loss of information. Second, the inclusion of titles and descriptions of nested 4-digit industries as part of the 3-digit industry description provides more text to use for matching.

Eventually, we break each industrial description (3- and 4-digit) into sentences and concatenate each sentence with its corresponding title. This loosely resembles the structure of patent title used for matching.

This yields 271 industries at the 3-digit level with an average of 11 composite sentences representing an industry.[7] These sentences are used for matching with patent titles.

Occupation. We choose the 4-digit ISCO-08 as the most fine-grained level at which we consider occupations. Differently from industries, the 4-digit level of ISCO-08 contains a set of distinct occupations, informative for our analysis. Moreover, each ISCO-08 occupation has

[6] Note that not all patents contain all blocks in their abstracts.

[7] We exclude industry with NACE 3-digit code 18.1 from the analysis because of the persistent conflation of its real meaning i.e. 'Printing and service activities related to printing (newspapers, magazines, periodicals, etc.)' with 3D printing.

a set of tasks performed in this job, although some tasks are shared by several occupations. Similar to industries, for each 4-digit occupation, we use the occupation title along with each task listed for this occupation that again attempts to replicate the patent title's structure.

This yields 433 occupations at the 4-digit level with an average of 7.5 composite sentences representing an occupation.[8]

3.2 Semantic-based Exposure Scores

We create a dynamic mapping that shows the relevance of emerging STIs to industries, based on the nature of their economic activity, and occupations, based on the tasks and functions they perform, as well as the ontological domain of knowledge (e.g. law, medicine, engineering, etc.). We again draw reader's attention to the interpretation of the exposure scores as *relevance* of a technology (i) to the economic activity being process or product innovation an industry *can* implement and (ii) to a set of functions (i.e. tasks) within an occupation that *can* be performed using a technology (either substituting or augmenting labor).

Embeddings. To establish industry–patent or occupation–patent links, we use a pre-trained sentence transformer for text similarity tasks, namely MPNet v2. This sentence transformer maps sentences from a description of an industry i, an occupation o, or a patent p, into a 768-dimensional vector space, which is called its embedding. We denote these embeddings, respectively, Emb_i, Emb_o and Emb_p.

Cosine similarity. For each patent title $p \in \mathcal{P}$, we compute the cosine similarity:

$$C_p^d = \frac{Emb_d \cdot Emb_p}{||Emb_d|| \; ||Emb_p||}, \text{ with } d = \{i, o\}, \tag{1}$$

which expresses the idea of the semantic connection between p and $d = \{i, o\}$.

Yet, similarity can be established based on various shades of meaning; in our case, it can be, for instance, an application, a technical domain, or some mentioned functions (either core or peripheral). Eventually, all this information is compressed in a scalar (i.e. C_p^d) whose magnitude *approximately* reflects the strength of similarity but says nothing about the thematic premise.

[8] We exclude the 4-digit occupations under the 3-digit occupation 'Printing trades workers' (732) from the analysis for the same faulty connection with 3D printing.

Redundancy. To address this issue and improve the matching quality, we introduce *redundancy* in the calculation of cosine similarity of industry–patent combinations (i, p) and occupation–patent combinations (o, p). The implementation of redundancy differs for the industry–patent and occupation–patent combinations.

For industry–patent combinations (i, p), we compute two cosine similarity scores that correspond to each part of the title, namely, $C^i_{p_1}$ for the essence of the technology and $C^i_{p_2}$ for its function. We rank sub-combinations $(i, p)_1$ and $(i, p)_2$ separately according to their respective cosine similarity scores $C^i_{p_1}$ and $C^i_{p_2}$. We select combinations (i, p) for which *both* sub-combinations $(i, p)_1$ and $(i, p)_2$ belong to the top 10 of their respective ranking as relevant industry–patent pairs $(i, p)^\star$. This causes some pairs to be removed because they are retained in one but not the other top 10.

For occupation–patent combinations (o, p), we consider the patent title as a whole and we instead divide the occupation description into two parts, one being the occupation title o_1 and the other being the tasks' description o_2. We follow the same procedure as for industries. We compute both cosine similarity scores $C^{o_1}_p$ and $C^{o_2}_p$ and select combinations (o, p) for which *both* sub-combinations $(o, p)_1$ and $(o, p)_2$ belong to the top 10 of their respective ranking as relevant occupation–patent pairs $(o, p)^\star$.

For the relevant pairs, we compute the harmonic mean with both cosine similarity scores. This gives the cosine similarity score for, respectively, industry–patent pairs and occupation–patents pairs:

$$C^i_p = 2\left(\frac{1}{C^i_{p_1}} + \frac{1}{C^i_{p_2}}\right)^{-1}, \tag{2}$$

$$C^o_p = 2\left(\frac{1}{C^{o_1}_p} + \frac{1}{C^{o_2}_p}\right)^{-1}. \tag{3}$$

As a result of calculations expressed in Equation (2) and (3), we establish connection between a technology identified in a single patent $p \in \mathcal{P}$ and a set of relevant industries and occupations. Each emerging STI includes hundreds to thousands of individual patents that share technological underpinnings and applications.

Aggregation from patents to technologies. The last step towards the exposure score is the aggregation of cosine similarity C^d_p, with $d = \{i, o\}$, to emerging technology k defined as $\mathcal{P}_k \subset \mathcal{P}, \forall\ \ k = [0, 499]$. More specifically,

$$C^d_k = \sum_{p^\star \in \mathcal{P}_k} C^{d^\star}_{p^\star}, \text{ with } d = \{i, o\}, \tag{4}$$

where $\star$ superscript refers to relevant pairs.

Focusing on a given industry $i = I$, we obtain a set Θ^I which is composed of industry–technology pairs (I, k) with a vector of cosine similarity scores $C^I_{k\in\Theta^I}$. The latter is the vector of industry I exposure scores to potentially relevant subset of emerging technologies $k \in \Theta^I$.

Selection of relevant STI pairs. We implement a filtering procedure to select the relevant industry–technology pairs. We provide a step–by–step explanation of the procedure. The procedure is identical to obtain the occupation–technology ones.

Let Θ be the set of industry–technology combinations (i, k). Each combination has a cosine similarity $C^i_k > 0$ which measures to which extent the industry i is semantically related to the technology k. We denote $\Theta_k \subset \Theta$ the set of pairs that are candidates to technology k.

We define the *cosine similarity share of a combination* (i, k) as the share of its cosine similarity within the total cosine similarity for technology k (i.e. aggregated across all industries that are linked to k). We denote this share as

$$\alpha^i_k = \frac{C^i_k}{\sum_{i'\in\Theta_k} C^{i'}_k} \tag{5}$$

We need to identify $\alpha^\star$ that split industry–technology pairs (i, k) into two groups: those with a cosine similarity share above the threshold (i.e. $\alpha^s_k \geq \alpha^\star$) and those with a share below the threshold (i.e. $\alpha^s_k < \alpha^\star$).

We denote Θ^a_k (resp. Θ^b_k) the subset of candidate pairs to technology k which belong to the group above (resp. below) the threshold. The size of both subsets depends on the threshold. The higher the threshold, the smaller (larger) the number of candidate pairs above (below) the threshold.

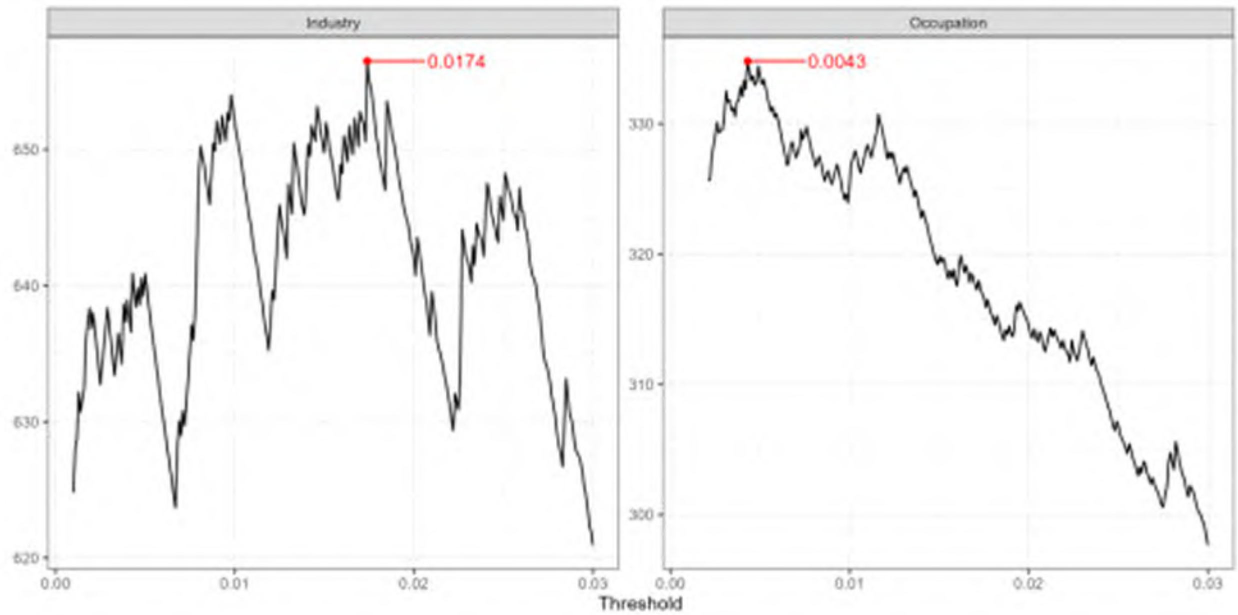
Our *optimal technology-share threshold* is:

$$\alpha^\star = \underset{\alpha\in(0,1)}{\text{argmax}} \left\{ \frac{1}{\#\mathcal{K}(\alpha)} \sum_{k\in\mathcal{K}(\alpha)} \left(\sum_{i\in\Theta^a_k(\alpha)} C^i_k - \sum_{i\in\Theta^b_k(\alpha)} C^i_k \right) \right\} \tag{6}$$

This threshold maximizes the cross-technology average difference between aggregated cosine similarities for pairs that belong to the group above and below the technology-share threshold.

In our data, the optimal technology-share thresholds are 1.74% and 0.44% for, respectively, industry– and occupation–technology pairs as depicted in Figure 1 which shows the

Figure 1: Average difference in aggregated cosine similarity between groups below and above the threshold



Notes: This figure presents the optimal technology-share thresholds for both technology–industry and technology–occupation pairs. These optimal thresholds correspond to the maximum of the average difference in aggregated cosine similarity between groups that lie below and above the threshold.

global maxima.

For industries, we filter out 20,696 industry-technology pairs out of the 25,360 initial combinations (i.e. 81.6%), which gives us a total of 4,664 relevant industry–technology pairs.

For occupations, we filter out 13,207 occupation-technology pairs out of the 24,147 initial pairs (i.e. 54.7%). This gives us a total of 10,940 relevant occupation–technology pairs.

3.3 User or Producer: Identifying Industrial Status

The effects of emerging technologies are likely to differ between sectors that use them (recipients) from those that produce them (sources). Existing literature mostly relies on the industry activities of companies to identify the producer of automation technologies, e.g. artificial intelligence (Dechezleprêtre et al. 2021, Calvino et al. 2022).

Following this stream of literature, we use the NACE Rev.2 codes of companies that filed patents and extend this set of primary industrial codes by including industries in which these companies have a relative specialization based on their patenting activity. Overall, we retrieved companies' industrial information for 137,424 patents or 72.1% of our patent sample by matching ORBIS IP and PATSTAT company identifier.[9]

We introduce a Relative Industry Specialisation (RIS) index which allows us to refine the companies' industrial profiles as companies might produce technologies beyond their primary economic activities to maintain competitiveness (Miller, 2006; Alkemade et al., 2015).

Let p_f^i be the number of patents of a company $f \in \mathcal{F}$ in a given industry $i \in \mathcal{I}$. The RIS index is

$$RIS_f^i = \frac{p_f^i / p_f^{\mathcal{I}}}{p_{\mathcal{F}}^i / p_{\mathcal{F}}^{\mathcal{I}}}, \tag{7}$$

where $p_f^{\mathcal{I}} = \sum_{\mathcal{I}} p_f^i$ is the total number of patents of company f across all the industries, $p_{\mathcal{F}}^i = \sum_{\mathcal{F}} p_f^i$ is the total number of patents in industry i, and $p_{\mathcal{F}}^{\mathcal{I}} = \sum_{\mathcal{I}} \sum_{\mathcal{F}} p_f^i$ is the total number of patents across all industries and companies. We compute the RIS index based on companies' patent portfolios from the last three years prior to the filing year of the patent. The results are robust if we consider a five-year window.

We consider that a company f has a relative specialization in industry i if its share of patents in this industry is higher than the average share of patents in this industry worldwide, that is, when $RIS_f^i > 1$.

We consider that a patent p, which is relevant to industry i, belongs to a producer industry if the company f has a relative specialization in this industry i, otherwise, this patent belongs to a user industry. Formally, a patent $p \in \mathcal{P}^{i,prod}$ if $RIS_f^i \geq 1$ and $p \in \mathcal{P}^{i,user}$ otherwise; where $\mathcal{P}^{i,prod}$ (resp. $\mathcal{P}^{i,user}$) is the set of patents which are relevant to industry i as a producer industry (resp. user industry).[10]

We define the *producer score share* $s_k^i \in [0,1]$ as the share of the aggregated exposure score of patents filed by producers over the total exposure score in the industry for this technology. Formally,

$$s_k^i = \frac{\sum_{p \in \mathcal{P}_k^{i,prod}} C_p^i}{\sum_{p \in \mathcal{P}_k^{i,prod}} C_p^i + \sum_{p \in \mathcal{P}_k^{i,user}} C_p^i}. \tag{8}$$

Thus, the producer score share indicates how much the industry i is a producer of technology k.

Lastly, one can derive the *producer* score $S_k^{i,prod}$ and the *user* score $S_k^{i,user}$ of an industry

[9]For each patent, we match the BvD ID number of the company from ORBIS IP with the applicant identifier (PERSON_ID) in PATSTAT. The PATSTAT column PERSON_ID is not available in ORBIS IP. We download the applicant name of patents (Applicant(s) Name(s)) from ORBIS IP and match it with the applicant name (PERSON_NAME) in PATSTAT to create the link between BvD ID number with PERSON_ID. Some patent applicants cannot be matched with a company in ORBIS IP. For these companies, we use the ID for PATSTAT Standardized Name (PSN_ID) produced by Peeters et al. (2010).

[10]Fractional counting is applied if a patent is assigned to more than one company or the company is associated with more than one industry activity.

i in technology k based on the producer share derived in Equation (8). Specifically,

$$S_k^{i,prod} = s_k^i \times C_k^i, \tag{9}$$
$$S_k^{i,user} = (1 - s_k^i) \times C_k^i, \tag{10}$$

where C_k^i is the exposure score of industry i to technology k as defined in Equation (4). Note that we only derived these scores for relevant industry–technology pairs as selected in Section 3.2.

4 Exposure to the Emerging STIs

In this section, we present results of occupation and industry exposure estimation following the methodology laid out in Section 3.

4.1 Occupation Exposure

4–digit ISCO-08 level. The occupation exposure score is computed at the 4–digit level of ISCO-08 classification; the semantic matching has been done task–by–task, meaning occupations are matched with technologies by *functions* that these technologies perform. This permits a rather wide set of technologies linked to each occupation, on average 30.67 technologies per 4–digit occupation. Table 1 presents the top 20 exposed occupations at the 4-digit level.

'Information and communications technology operations technicians (3511)' is the most connected occupation with 310 technologies linked to it. In general, ICT–related occupations are characterized by a larger number of connections (all above 200); this fact is not surprising as all emerging STIs are digital technologies by construction. Nevertheless, the most exposed occupation is 'Bank tellers and related clerks (4211)' with 104 relevant technologies.

There are many occupations in the top 20 list whose link count does not exceed 100 technologies, for instance, 'Plastic products and machine operators (8142)' and 'Advertising and marketing professionals (2431)'. Together, the number of connections and overall exposure score provide insights about the 'technology composition' of an occupation. More domain–specific occupations (such as 3211, 2431, 4323, 8132) tend to have fewer but stronger technology links while more transversal occupations (such as 3511, 2523, 3513, 1330, 3512) are connected to an array of digital emerging technologies with *relatively* lower strength.

Let us consider an example to illustrate the outlined regularity at work using three occupations from Table 1: 'Photographic products machine operators (8132)' — high exposure

Table 1: Top 20 exposed 4-digit ISCO-08 occupations

Code	ISCO Occupation	C_k^o	$\log(1+C_k^o)$	N_k
4211	Bank tellers and related clerks	4856.60	8.49	104
3252	Medical records and health information technicians	4756.80	8.47	103
8132	Photographic products machine operators	4121.04	8.32	198
8322	Car, taxi and van drivers	3726.06	8.22	106
3133	Chemical processing plant controllers	3535.26	8.17	206
3311	Securities and finance dealers and brokers	3300.29	8.10	146
8142	Plastic products machine operators	2863.96	7.96	55
2523	Computer network professionals	2786.57	7.93	288
8141	Rubber products machine operators	2739.88	7.92	60
3511	Information and communications technology operations technicians	2690.91	7.90	310
4323	Transport clerks	2591.88	7.86	111
3513	Computer network and systems technicians	2425.05	7.79	284
1330	Information and communications technology service managers	2335.20	7.76	211
2431	Advertising and marketing professionals	2141.20	7.67	61
9621	Messengers, package deliverers and luggage porters	1932.38	7.57	148
9623	Meter readers and vending-machine collectors	1862.71	7.53	140
3211	Medical imaging and therapeutic equipment technicians	1798.91	7.50	59
3512	Information and communications technology user support technicians	1753.22	7.47	232
3131	Power production plant operators	1669.62	7.42	79
8143	Paper products machine operators	1527.14	7.33	89

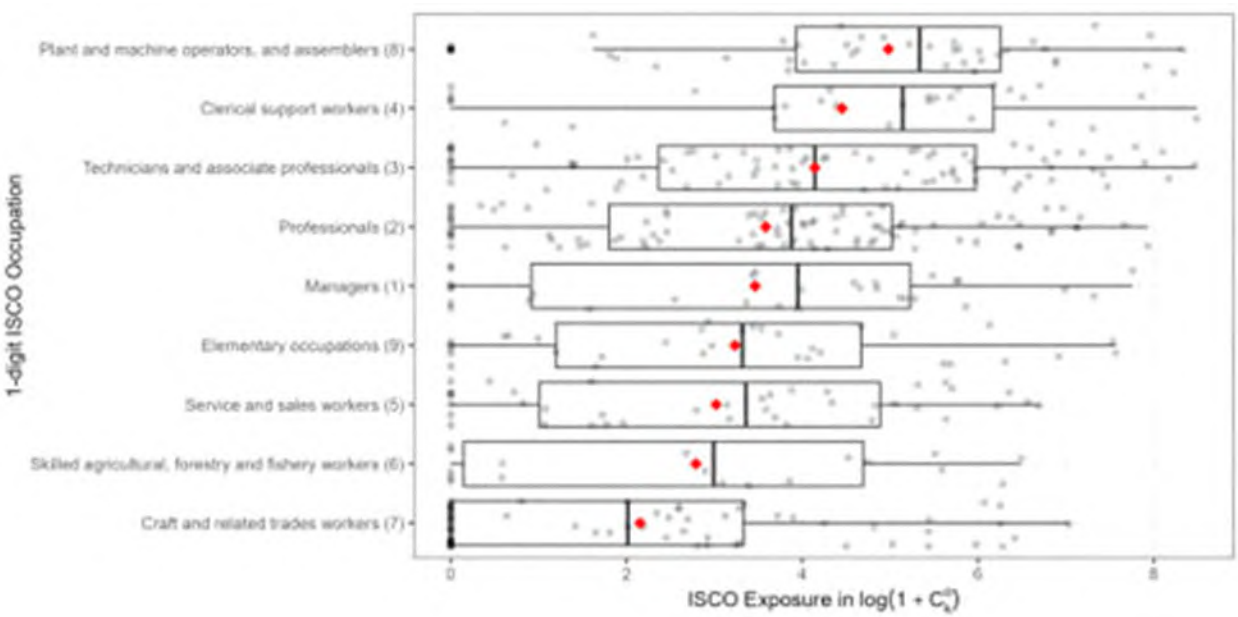
Notes: This table presents the top 20 4–digit ISCO-08 occupations ranked by pooled exposure score (to all technologies). Columns (from left to right) correspond to occupation code, occupation title, exposure score, exposure score expressed in $log(1+x)$ transform, and the number of relevant technologies.

(top 3) and large number of technologies (198); 'Plastic products machine operators (8142)' and 'Medical imaging and therapeutic equipment technicians (3211)' — both with smaller numbers of technologies (55 and 59 respectively) but still relatively high exposure (top 7 and top 17 respectively). All three occupations work with optical equipment however 'Medical imaging technicians (3211)' and 'Plastic products machine operators (8132)' relate to different types of optical equipment in their jobs. The former is related to medical imaging techniques such as tomography, MRI scanning, ultrasound, radiography, etc. while the latter is mainly related to optical equipment in 3D printing systems. In fact, these two occupations share only 5 technologies (hence Jaccard similarity is equal $\frac{5}{55+59-5} = 0.05$). In turn, 'Photographic equipment operators (8132)' occupation has a wider range in which it can relate to optical equipment making this occupation more transversal and less domain–specific; Jaccard similarity for 'Medical imaging technicians (3211)' and 'Plastic products machine operators (8132)' are respectively 0.12 and 0.28.

Finally, one has to take into account the representation of technologies in our emerging STI set. In fact, 500 technology clusters pertain to broader technology families such as Artificial Intelligence, Computing, Robotics, User Interface, and so on (see Appendix A.1). Accounting for the emergence property, families like Artificial Intelligence are relatively more represented than, for example, User Interface among 500 emerging STIs. Thus, some occupations despite being also rather transversal have lower total exposure scores and/or smaller numbers of links compared to domain-specific ones because a substantial share of technologies these occupations relate to are not characterized as emerging and hence are not in our set. For example, the occupation 'Web technicians (3514)' has a large number of technology links (172) but rather a medium total exposure score (641.99) because apparently technologies relevant to this occupation are not in our emerging set.

1–digit ISCO-08 level. We group 4–digit occupations into 1–digit occupations to compare the distribution of exposure scores. Figure 2 presents the results. We observe an interesting finding: the upper half of the boxplot is populated by highly skilled labor groups[11] among which 'Technicians and associate professionals (3)', 'Professionals (2)', with 'Managers (1)' in the middle.

Figure 2: Distribution of occupation exposure by 1-digit ISCO-08 occupation



Notes: This figure presents the distribution of occupational exposure for 4-digit ISCO-08 occupations, expressed in $\log(1+x)$ transform, for each 1-digit occupation separately as boxplots. Diamond points represent the mean value.

Two remarks are in order here: (i) according to the interpretation of exposure, this represents the relevance of our emerging STIs to occupations and not *de facto* adoption, (ii) the nature of the effect of exposure — whether it is augmentation or substitution — is not identified. With regard to (i), given that we are dealing with emerging technologies, we might observe an effect of these technologies on employment in the near future (i.e.

[11] The lower the ISCO code, the higher skilled group it is.

adoption lags). Concerning (ii), this depends on the relation between labor and capital in a given occupation–industry cell; we approximate this relation via the identification of the technology user/producer status of an industry in a given country.

We first comment on the two most exposed 1–digit ISCO groups because they include *relatively* more routinized and/or lower-skilled occupations among the top exposed. Then we proceed to the next two most exposed categories 'Technicians and associate professionals (3)' and 'Professionals (2)' which represent highly skilled labor performing less routinized tasks.

The top exposed category is 'Plant and machine operators (8)'. It includes operators of some capital or equipment with varying degrees of sophistication and domain specificities: from 'Laundry machine operators (8157)' and various drivers (e.g. 8322, 8331, 8332, etc.) to operators of industrial machinery (e.g. 8132, 8141, 8142, etc.). Looking into the set of technologies related to this 1–digit category, we can find unsurprisingly a mixture of Additive manufacturing, (industrial) Internet of Things, Autonomous driving, and Robotics. In sum, due to emerging technologies capital becoming more and more autonomous (from supervised to semi-/unsupervised) hence these occupations are getting more exposed.

'Clerical support workers (4)' is the second most exposed 1–digit ISCO category. We divide this group into 3 subgroups:

1. *Symbol handlers and operators* which include, for instance, 'Data entry clerks (4132)', 'Typists and word processing operators (4131)', 'Coding and proofreading clerks (4413)', and 'Accounting and bookkeeping clerks (4311)';

2. *Document handlers and organizers* which include occupations such as 'Mail carriers and Sorting clerks (4412)', 'Stock clerks (4321)', 'Filing and copying clerks (4415)', 'Library clerks (4411)'.

3. *Information retrieval and provision clerks* which consist of 'Travel consultants and clerks (4221)', 'Contact centre information clerks (4222)', 'Inquiry Clerks (4225)', 'Receptionists (4224, 4226)', 'Survey and Market Research Interviewers (4227)'.

Despite being relatively low-skill, these occupations are less routinized facing a wide range of contingencies (e.g. inquiries) that require domain-specific knowledge (finance, travel, accommodation, etc.) and judgment on how to handle the situation given context and/or available resources (e.g. available hotel rooms). In other words, these are context–dependent information handling jobs.

Generalizing functions that these occupations perform (regardless of their domain) can be summarised as resource record keeping, monitoring, querying and matching, and allocating.

This is a set of functions that separately are rather simple. In fact, delving deeper into the set of technologies related to 'Clerical support workers (4)', one can summarise their nature as various *commercial* digital platforms that perform aforementioned functions in an integrated manner. The variety comes from different application domains such as vehicle/parking allocation, food delivery, inventory management, job search, and ticket reservation, among others. The Word *commercial* is of significance here because a large share of related technologies pertains to financial information operations: authentication, verification, record keeping (including blockchain), processing, security, etc.

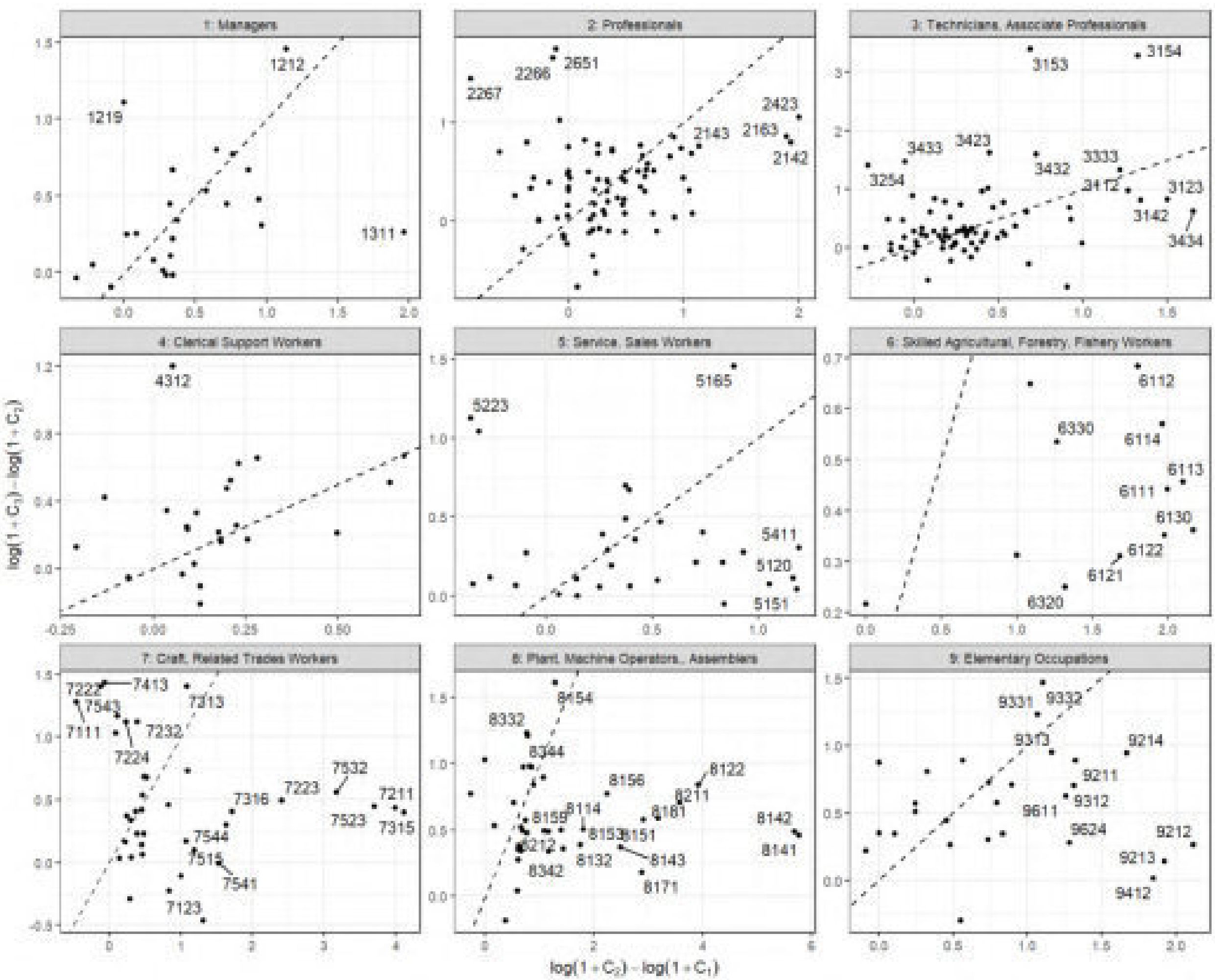
Lastly, only a few technologies explicitly represent an intelligent interface such as smart virtual assistants. There are two reasons for that: (i) being intangible, the software is usually protected with other forms of IPR than patents simply not releasing the source code, (ii) the emergence of transformer models in Natural Language Processing (NLP) is one of the most recent trends with the seminal paper Vaswani et al. (2017).

The discussion above highlights another insight: the more tasks/functions an occupation performs, the less routinized an occupation is. One might say that what constitutes the intelligence in a technology or a worker is the number of tasks/functions they can handle with a substantial degree of autonomy based on the growing breadth of underlying knowledge space. This being said it is getting increasingly hard to generalize both which functions and technologies are relevant for 'Technicians and associate professionals (3)' and 'Professionals (2)' categories because both increase in variety. For technicians, we can certainly say that the majority of top relevant technologies come from the medical domain (medical imaging, medical data management, health monitoring, examination, diagnostics, etc.); this indicates the main domain of application where technological development is affecting jobs. Concerning professionals, the pattern is even less distinct but the most relevant technologies address the function of customized content provision: the first block is about education and the second is about online content including advertising and recommendation.

Dynamics of occupation exposure . In the previous paragraphs, we discussed occupations with the highest absolute exposure. In this part, we want to understand if there is a pattern of exposure growth over time that characterizes each 1–digit ISCO group; in other words, if a 1–digit group of occupations has a spike of exposure in a certain time period or it increases exposure evenly over time.

We consider the period between 2014 and 2019 where the latter aims to exclude the effects of COVID-19 while the former is motivated by the relatively young age of our emerging STIs which were either nascent or not yet born in the period 2012–2013; thus, cumulative exposure over this period is negligible by definition. Eventually, we divide the period into three equal intervals $t_i = \{2014 - 2015, 2016 - 2017, 2018 - 2019\}$ and obtain three snapshots of total exposure score by occupation, i.e. $C_t = \sum_{t \in t_i, k} C_{t,k}$. Hereinafter, we refer to the two percentage changes: $\Delta_1 = log(1 + C_2) - log(1 + C_1)$ and $\Delta_2 = log(1 + C_3) - log(1 + C_2)$. Figure 3 shows the results with each observation being a 4–digit occupation, and the 45-degree line indicating equality of percentage changes, i.e. $\Delta_1 = \Delta_2$.

Figure 3: Percentage change of exposure by 1–digit ISCO occupation



Notes: We compute a two–year exposure score for the following intervals: $t_1 = \{2014, 2015\}$, $t_2 = \{2016, 2017\}$, $t_3 = \{2018, 2019\}$. Thus, we have $C_1 = \sum_{t \in t_1, k} C_{t,k}$. The exposure scores are expressed in $log(1 + x)$ transform. The dashed 45-degree line indicates the equality of percentage changes between both periods.

A large share of occupations in 'Plant and machine operators (8)' increased exposure during Δ_1 more than in Δ_2. As mentioned earlier, a mixture of Additive manufacturing, (industrial) Internet of Things, Autonomous driving, and Robotics is identified as relevant to this 1–digit ISCO group. Noteworthy, all these technologies necessarily involve control over tangible capital (e.g. 3D printing system, vehicle, robot, or industrial machinery). The larger increase in exposure in Δ_1 implies that the bulk of these technologies was developed earlier. If we examine the rightmost observations, i.e. occupations 8142, 8141, 8122, 8181, they represent machine operators that vary with regard to material — plastic, rubber, metal — but they all pertain to additive manufacturing, a more mature technology compared to the other relevant ones in this group.

Occupations located above the dashed line exhibit a larger increase in Δ_2. These are heavy vehicle drivers, i.e. Heavy truck and lorry drivers (8332) and Lifting truck operators (8344). Indeed, the automation of heavy vehicles (for example, for autonomous cargo transportation) is a surging application of autonomous driving in the latest years (compared to passenger cars).

In the case of additive manufacturing, it is likely that the technology itself had a surge of development which increased the exposure of relevant occupations in an earlier period. Contrary, the case of autonomous driving for heavy vehicles can be characterized as rather a surge of application or a niche technology (i.e. particular type of vehicle).

The other two groups that showed top–level exposure are 'Professionals (2)' and 'Technicians and associate professionals (3)' here exhibit a rather even split between prevalently growing in Δ_1 and vice versa in Δ_2. Perhaps, it originates from a heterogeneous set of technologies these occupations are exposed to.

A notable case within 'Technicians and associate professionals (3)', both outliers, i.e. 'Aircraft pilots (3153)' and 'Air traffic controllers (3154)', pertain to aerial navigation and control. The set of technologies these occupations are exposed to unsurprisingly contains unmanned aerial vehicles, drones, remote aerial control, and navigation. This is certainly the case of technology emergence causing these dynamics.

Overall, if an occupation had a surge of exposure in a certain period it can happen for two reasons: (i) a surge in the emergence of relevant technology, (ii) a surge of a particular application of that technology.

4.2 Industrial Exposure

3–digit NACE level. The industry exposure score is computed at 3–digit level of the NACE Rev.2 classification. Industries represent distinct economic activities hence the matching with technology is established rather along the application dimension, with the exception of industries–producers whose description must be the most similar to the description of a technology itself. For example, if a patent describes a remote sensing technology with an application in vital sign monitoring, the matching should be established with both precision measurement (producer) and medical care (user) industries. This leads to a smaller number of technologies per industry 24.42 on average compared to 30.67 per occupation. Table 2

Table 2: Top 10 exposed 3-digit NACE industries

Code	NACE Industry	C_k^i	$\log(1+C_k^i)$	N_k
63.1	Data processing, hosting and related activities; web portals	19629.29	9.88	435
26.3	Manufacture of communication equipment	11796.94	9.38	341
26.2	Manufacture of computers and peripheral equipment	6401.91	8.76	276
73.1	Advertising	5924.81	8.69	64
26.5	Manufacture of instruments and appliances for measuring, testing and navigat [...]	5381.26	8.59	191
64.1	Monetary intermediation	4517.41	8.42	61
47.4	Retail sale of information and communication equipment in specialised stores	4333.50	8.37	95
63.9	Other information service activities	4049.59	8.31	158
61.2	Wireless telecommunications activities	3841.64	8.25	176
62.0	Computer programming, consultancy and related activities	3453.14	8.15	172

Notes: This table presents the top 10 3-digit NACE (Rev. 2) industries ranked by exposure to all technologies. Columns (from left to right) correspond to industry code, industry label, exposure score, exposure score expressed in $\log(1+x)$ transform, and the number of relevant technologies.

presents the top 10 most exposed industries at the 3–digit level.

We referred to occupations with a lower (higher) number of technology links as domain-specific (transversal). In the case of industries, the same idea goes in line with user (producer) status. The emerging STI set consists of digital automation technologies by construction hence only a subset of industries ensures the production of these technologies. Table 2 clearly shows that Advertising (73.1), Monetary intermediation (64.1), and Retail of ICT equipment (47.4) are industries–recipients of technologies, while the rest of the top 10 are producers.

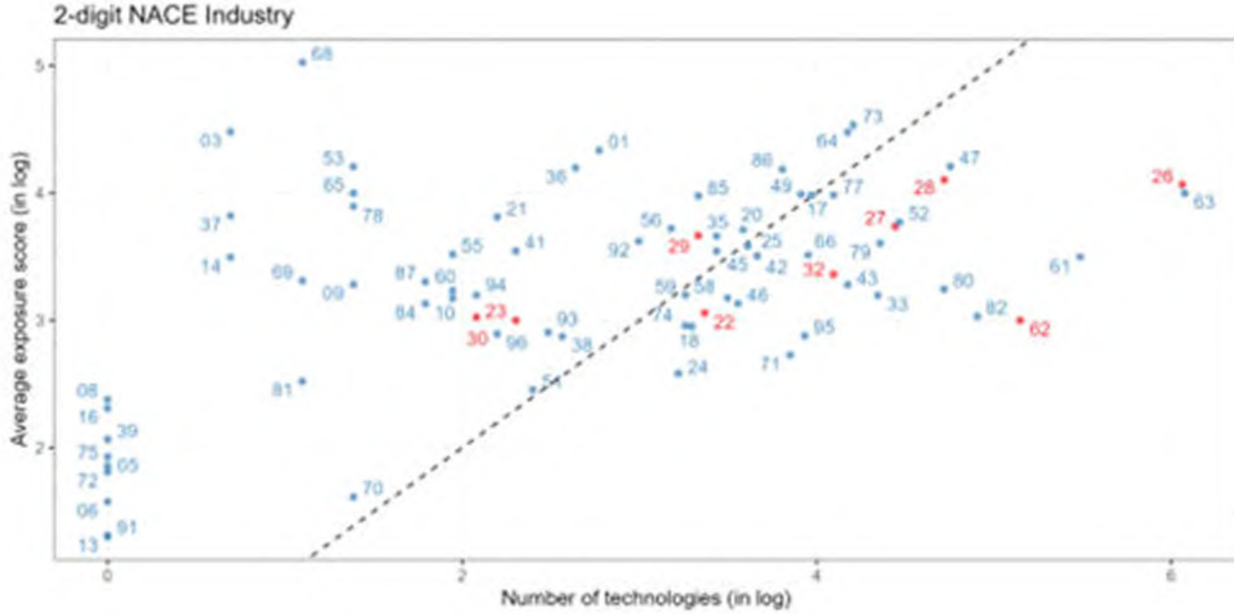
2–digit User/Producer. This regularity is reinforced once we include the RIS index described in Section 3.3. To identify producers, we consider industries in which the producer share is above $\frac{1}{2}$, i.e. $S_i^p/(S_i^p + S_i^u) > \frac{1}{2}$. The producer group includes several industries from Manufacturing (C), such as 'Manufacture of computer, electronic and optical products (26)' and 'Manufacture of electrical equipment (27)', as well as the 'Computer programming, consultancy and related activities (62)' industry from the Information and Communication (J) sector.

For each 2–digit industry, Figure 4 plots the number of technology links versus average exposure; the dashed 45-degree line separates widely connected industries (below) from intensely connected industries (above). Producers tend to be of the widely connected type.

Aggregation and associated loss of information. It is worth stressing that industries already at 3–digit level represent distinct categories.[12] Further aggregation merges industries with different levels of exposure scores or even user/producer status. For example, aggregation from 3– to 2–digit code, pools together industries Manufacture of irradiation, electromedical and electrotherapeutic equipment (26.6) and Manufacture of communication

[12] In NACE Rev.2, 4–digit is the most fine–grained level.

Figure 4: Average industrial exposure score and number of exposed technologies



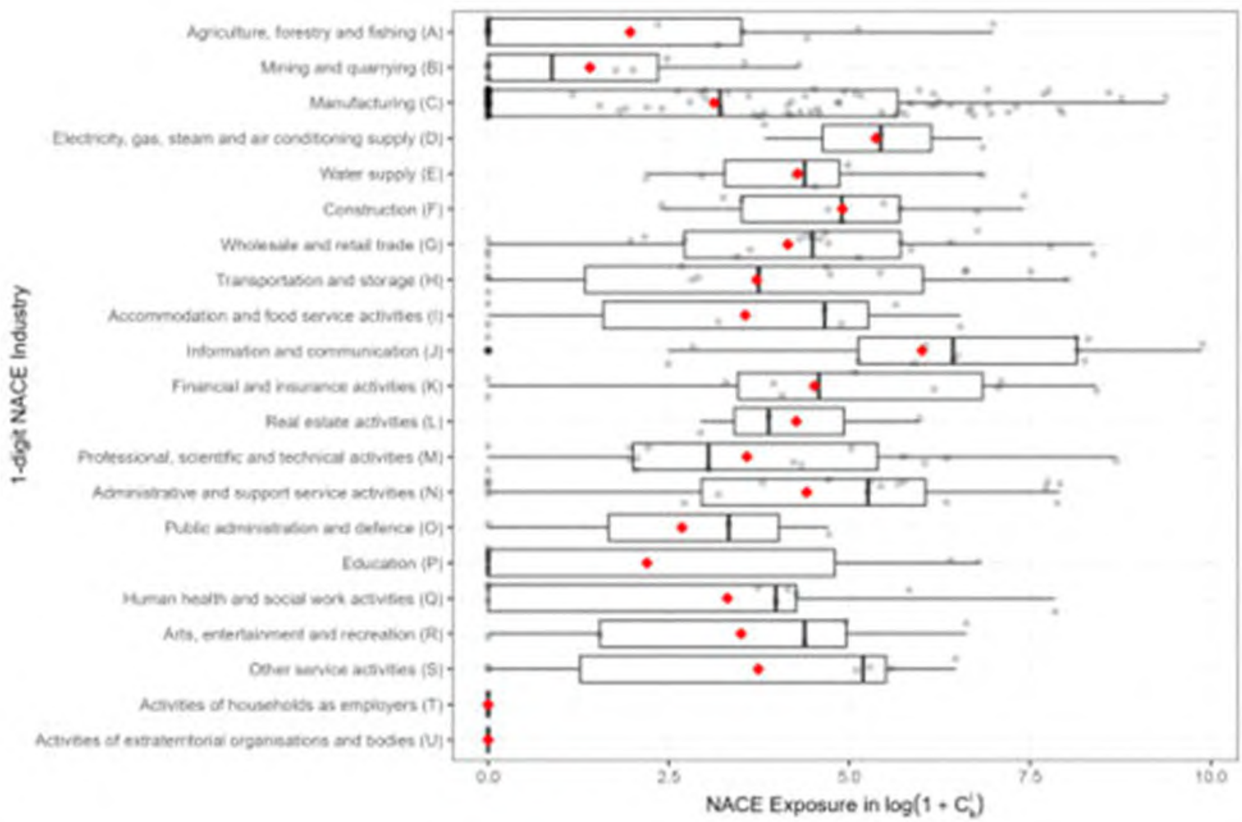
Notes: This figure shows the number of technologies on the x-axis, expressed in log, and the average exposure score on the y-axis, expressed in log, for each 2-digit NACE Rev.2 industries. The dashed line is the 45-degree line.

equipment (26.3) both being rather producers but of different technologies. Going higher, from 2–digit to 1–digit, we end up with broad sectors of the economy such as Manufacturing (C) combine extremely heterogeneous industries; at this level, only coarse technology–sector links remain. Figure 5 illustrates this point, especially for Manufacturing (C). In Section 5 we will further demonstrate the effect of aggregation in the context of the employment change.

Summarizing this Section, we can reiterate the following: (i) high–skilled and non–routinized jobs are exposed to the set of emerging STIs; (ii) one limitation originates from the coverage of the emerging STI set including *novel* technologies; (iii) technologies that involve tangible capital are linked medium–high skilled labor while intangible technologies (digital, information processing) are relevant to both high and low skilled labor.

However, the magnitude of the exposure score alone does not disambiguate the nature of the effect on employment between substitution and augmentation; as the following sections will show, it will depend on the both industry *and* occupation as well as producer or user status.

Figure 5: Average industry exposure across 1-digit NACE industry



Notes: This figure presents the distribution of industrial exposure for 3-digit NACE (Rev. 2) industries, expressed in $\log(1+x)$ transform, for each 1-digit industries separately as boxplots. Diamond points represent the mean value.

5 The employment effect of exposure

In this section, we look at the effects of exposure to emerging STIs on employment over the last decade. We proceed in two steps. First, we estimate the employment change in occupations, based on their exposure to emerging STIs, while differentiating this effect between user and producer sectors of technologies at the 1-digit NACE level for European countries. Second, we focus on the case of Germany for which we can disaggregate sectors into 2-digit NACE industries.

5.1 Emerging STIs and European employment (2011–2019)

We perform our analysis at 3-digit ISCO-08 for occupations and 1-digit NACE Rev.2 for industries.[13] Since our exposure scores are defined at the 4-digit ISCO-08 for occupations

[13] We are limited in the level of analysis using the EU-LFS data as only the 1-digit NACE industries are available (see Section 2 for more details). In the next subsection, we focus on Germany for which we can disaggregate industries into two digits.

and 3-digit NACE for industries, we aggregate industry and occupation exposure scores to the higher levels in their respective classifications.

We estimate the effect of emerging STI exposure on the change in employment over the whole decade (2011–2019). We focus on the change in the number of employees in the country–industry–occupation cell to measure the net employment creation (or destruction) in occupations and industries that are exposed to STIs.

We start with the following identity:

$$L_{cio,2019} \equiv g_{cio} \times L_{cio,2011},$$

where $L_{cio,t}$ is the number of employees in country c industry i occupation o in year t and g_{cit} is the corresponding growth of employment between 2011 and 2019.

We assume that employment growth is given by the following functional form:

$$g_{cio} = A_{cio}(1 + C_o)^{\beta_{j(i)}}, \tag{11}$$

where C_o is the average-yearly exposure of occupation o aggregated across all technologies k and A_{cio} is a scale factor that encompasses all employment changes that are orthogonal to the exposure, i.e. $A_{cio} \perp C_o$. For simplicity, we impose an exponential form to A_{cio} such that

$$A_{cio} = \exp\left(\alpha + \delta_c + \phi_i + X\gamma + u_{cio}\right),$$

where δ_c captures employment changes that are specific to countries, X is a set of control variables and u_{cio} is the error term.

The parameter $\beta_{j(i)}$ characterizes the relation between occupational exposure and employment change that may differ across sectors, e.g. software developers in technology–producer versus technology–user sectors. We allow this parameter to differ across types of sectors $j(i)$ where j distinguishes between user and producer sectors.

We consider two types of sectors: producers and users. Technologies are produced by two sectors: Manufacturing (C) and Information and Communication (J), which account respectively for 72% and 14%, hence 87% in total, of the cumulative producer exposure (see Table B.1 in the Appendix). We assign these two sectors to the producer group and the remaining sectors to users.[14]

Taking the logs in Equation (11), we obtain our baseline specification as a fixed effect

[14]The distributions of occupational exposure across the two types of sectors are presented in Figure B.1 in the appendix. The density of occupations with low exposure to technologies is more important in user sectors compared to producer sectors.

Table 3: Baseline estimate

	Dep. var.: Employment (in log-change)		
	(1)	(2)	(3)
Intercept	-0.061^{**}	-0.051^{**}	-0.107^{***}
	(0.024)	(0.024)	(0.025)
Producer		-0.064^{***}	-0.148^{***}
		(0.014)	(0.019)
ISCO Exposure (in log)	0.012^{**}	-0.003	-0.009^{*}
	(0.005)	(0.005)	(0.006)
ISCO Exposure (in log) × Producer		0.074^{***}	0.082^{***}
		(0.012)	(0.012)
NACE Exposure (in log)			0.030^{***}
			(0.004)
Country FE	Yes	Yes	Yes
R^2	0.359	0.360	0.362
Adj. R^2	0.359	0.360	0.361
Num. obs.	26109	26109	26109

Notes: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses. All regressions are weighted by the level of employment in 2011. ISCO Exposure is the cumulative exposure between 2011 and 2019 in $\log(1 + x)$ transform and employment in log-change between 2011 and 2019. Column (1) presents the raw correlation with country fixed effects, where the baseline country is the United Kingdom. Column (2) adds the interaction of occupation exposure with the type of sector, where the baseline category for sector type is 'User'. Column (3) adds the NACE Exposure between 2011 and 2019 in $\log(1 + x)$ transform as a control variable. Table B.3 presents the same estimates when excluding occupations with null exposure. These estimates are robust to the use of the inverse hyperbolic sine transform instead of the $\log(1+x)$ transform on ISCO exposure (see Table B.4 in Appendix B).

model:

$$\log g_{cio} = \alpha + \beta_{j(i)} \log(1 + C_o) + X\gamma + \delta_c + u_{cio}. \tag{12}$$

Table 3 reports the estimated coefficients and delivers three results.

First, occupations that are more (less) exposed to emerging STIs are associated with increased (decreased) levels of employment between 2011 and 2019. The positive and significant ISCO Exposure coefficient in Column (1) indicates a positive relation, while the negative and significant intercept means that employment change is negative in occupations that are less exposed to emerging STIs.

Second, the positive relation between occupational exposure and employment is driven by Producer sectors and is negative in User sectors. When differentiating the effect of occupational exposure by sector type, Column (2) shows that the positive association between occupational exposure and employment holds only for Producers. The coefficient of the referent group (i.e. Users) is negative but not significant.[15] This pattern holds in Column (3)

[15]This coefficient becomes negative and significant when excluding occupations with null ISCO exposure (see Table B.3 in Appendix A).

when we include the NACE exposure as a control variable.[16]

Third, occupations with high exposure in Producer sectors experience job creation while those with low exposure experience job destruction. The Producer coefficient in Column (2) is negative and significant which indicates that employment in low-exposed occupations in this sector decreased. Thus, the positive employment effect of occupational exposure only holds for those in highly exposed occupations in Producer sectors.

Figure 6 emphasizes the differences in the marginal effect of exposure to technologies between occupations from the two sector types.[17]

Highly-exposed occupations, such as 'Other health associate professionals (325)' and 'Tellers, money collectors and related clerks (421)', experience positive employment growth in producer sectors (i.e. respectively +13% and +14%) but small and negative employment growth in user sectors (i.e -1%).

Low-exposed occupations, such as 'General office clerks (411)' and 'Shop salespersons (522)', experience virtually no employment growth in the user sectors. However, 1-digit occupations 'Craft and related trades workers (7)', 'Plant and machine operators, and assemblers (8)', and 'Elementary occupations (9)' in the producer sector experience a negative employment growth rate. For instance, the employment growth rate for 'Building and finishers and related trades workers' (712) and 'Mining and construction labourers' (931) are about, respectively, -3% and -5% between 2011 and 2019.

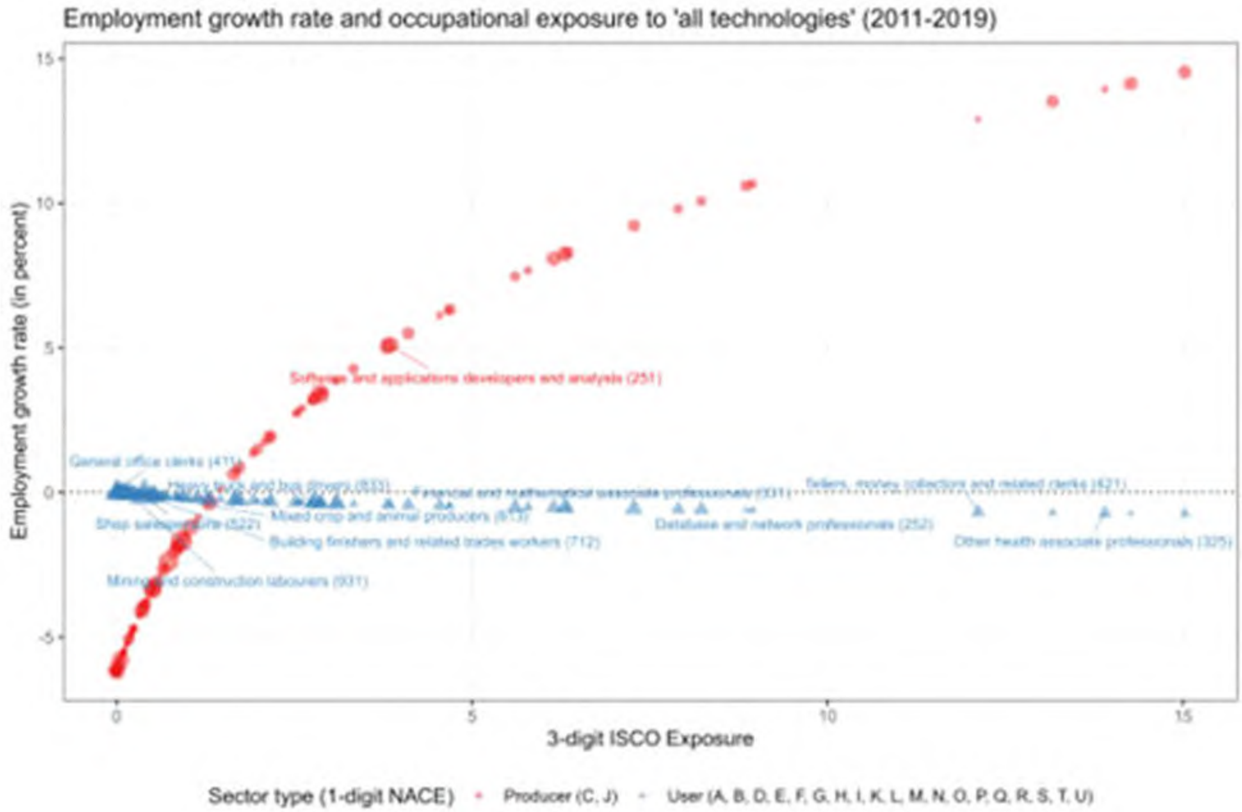
5.2 User–Producer divide and German employment (2012–2019)

The unit of analysis in the previous subsection, i.e. 1-digit NACE, allows us to identify employment patterns that relate to occupational exposure and that hold across all European countries. However, due to data limitations in the EU-LFS, we cannot perform the analysis at a more disaggregated level. This is an issue since some 1-digit NACE sectors consist of 2- and 3-digit industries that differ substantially with respect to the use of emerging technologies. Therefore, we would expect more refined patterns in the relation between occupational exposure and employment. For instance, the 1-digit NACE sector 'Manufacturing' (C) includes both the 2-digit NACE industry 'Manufacture of computer, electronic and optical products' (26) and 'Manufacture of beverages' (11). However, the former is a producer of

[16] Yet, the User group remains a residual category in which all non-producer sectors are included. This is likely to conceal large heterogeneity between 1-digit industries. In the next subsection, we address this issue by disaggregating to the 2-digit NACE level with the case of Germany.

[17] The predicted employment growth rate is derived from the predicted log-change in employment $\widehat{\log g_{cio}}$, estimated from Equation (12), and according to the functional form specified in Equation (11). More specifically, the y-axis corresponds to $\exp(\widehat{\log g_{cio}}) - 1$, where $\widehat{\log g_{cio}}$) is derived from the estimated model in Column (2) from Table 3.

Figure 6: Marginal effect of occupation exposure on employment growth by sector type



Notes: This figure shows the model predictions from the estimated baseline model reported in column (2) from Table 3. The x-axis is the cumulative 3-digit ISCO exposure between 2011 and 2019 to emerging technologies and y-axis is the growth rate of employment (expressed in percent). Sectors are grouped into three types: Producer, Intensive User, and Light User sectors. The size of each data point is proportional to its initial level of employment in 2011 within the sector type. The marginal effect of occupation exposure on employment growth is normalized to the one with a null occupation exposure in the Light User sector.

technology while the latter is a user.

Using German Labor Force Survey data, we can disaggregate to the 2-digit level for industries and, as before, to the 3-digit level for occupations. Our period of analysis starts in 2012 and ends in 2019.[18] We compute both the cumulative producer industry exposure S_i^p and the cumulative user industry exposure S_i^u for every 2-digit industry over the decade.

We classify 2-digit NACE industries into five groups: Producer, Light Users, Intensive and Deep Users (either IT or not IT). Table 4 summarizes the allocation of 2-digit NACE sectors across types of sectors.

To identify producers, we consider industries in which the producer share is above $\frac{1}{2}$, i.e. $S_i^p/(S_i^p + S_i^u) > \frac{1}{2}$. The producer group includes several industries from Manufacturing (C),

[18] Although the analysis with the EU-LFS data starts in 2011, we have to start our analysis in 2012 for Germany as they changed classification between 2011 and 2012. We also run the analysis starting in 2012 with the 1-digit EU-LFS data as a robustness check. This difference in the starting year does not affect the results.

Table 4: Sector types at the 2-digit NACE level

Sector type	2-digit NACE industry
Producer	22, 23, 26, 27, 28, 29, 30, 32, 62
Light User	All other industries not elsewhere classified
Intensive User	17, 49, 52, 77, 79, 80, 82, 86
Deep User - Not IT	47, 73
Deep User - IT	61, 63, 64

Notes: This table summarizes the sector types and the 2-digit NACE industries that belong to each group. Producers are industries in which the producer-user share is above 1/2. User groups are defined based on breaks in the industry exposure distribution (of remaining industries). Deep Users have industry exposure above 9.38, Intensive Users between 9.38 and 4.31, and Light Users include all industries with exposure below 4.31.

such as 'Manufacture of computer, electronic and optical products (26)' and 'Manufacture of electrical equipment (27)', as well as the 'Computer programming, consultancy and related activities (62)' industry from the Information and Communication (J) sector—which is driving the identification of that sector as a producer in the previous analysis at the 1-digit NACE level.

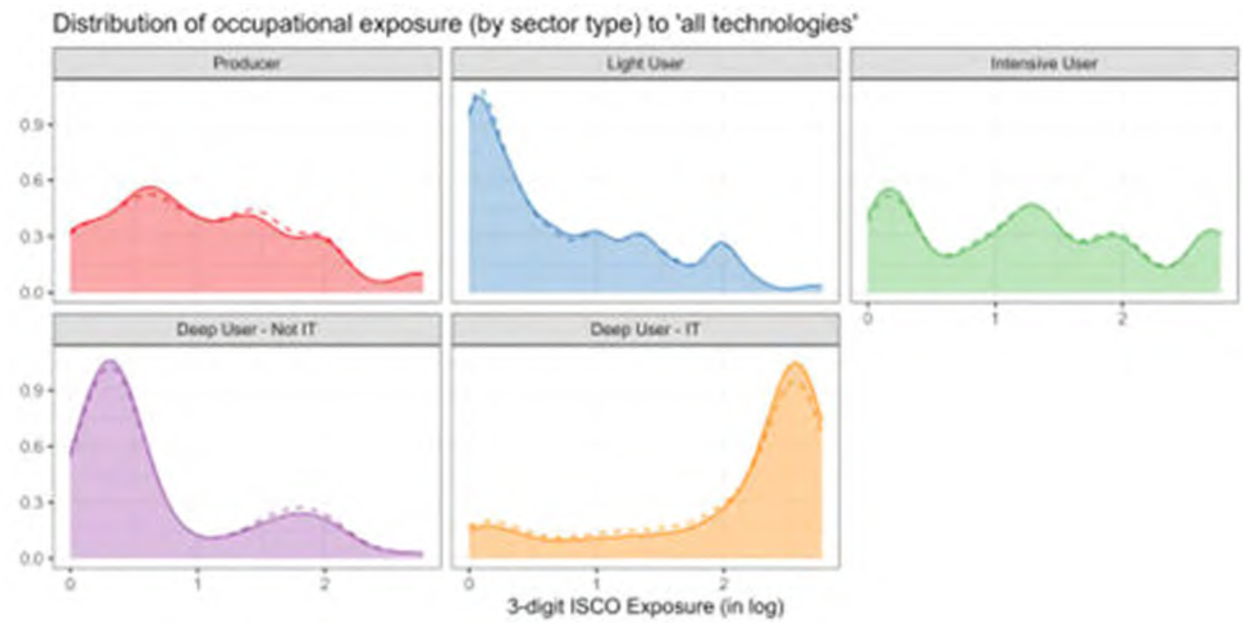
To distinguish between users, we identify breaks in the industry exposure distribution (of remaining industries) to distinguish between Light, Intensive and Deep users (see Figure B.4 in Appendix B for more details).

The deep user category consists of 5 industries (47, 61, 63, 64 and 73). We regroup 'Telecommunications (61)', 'Information service activities (63)', and 'Financial service activities, except insurance and pension funding (64)' as Deep user sectors specialized in IT; and 'Retail trade, except of motor vehicles and motorcycles (47)' and 'Advertising and market research (73)' as Deep user sectors not-specialized in IT. This separation is straightforward when looking at the occupational exposure distributions in Figure 7 as described below.

The Intensive user category includes 2-digit NACE sectors mostly from service sectors, such as 'Office administrative, office support and other business support activities (82)' or 'Human health activities (86)' as well as the industry of 'Manufacture of paper and paper products (17)'. The light user group includes all other industries not elsewhere classified.

Figure 7 presents the distributions of occupational exposure across the five types of sectors. The Deep User (IT) sectors are those with the highest occupational exposure as the left-skewed distribution in the fifth panel indicates. Conversely, the industries which are among the most exposed ones, however not specialized in IT, are mostly composed of low-exposed occupations as shown in the fourth panel. The occupational exposure distribution of this latter group is similar to the one of the Light User industries in the second panel. Lastly, both Producers and Intensive Users have rather spread occupational exposure distributions.

Figure 7: Distribution of occupational exposure by industry types to all emerging technologies in Germany between 2012 and 2019



Notes: This figure shows the distributions of occupational exposure to all technologies by sector type at the 2-digit industry level. The x-axis is the 3-digit ISCO exposure in $\log(1 + x)$ transform and the y-axis is the density. Panels refer to the five types of sectors: Producer (with NACE 2-digit industry codes 22, 23, 26, 27, 28, 29, 30, 32, 62), Intensive User (17, 49, 52, 77, 79, 80, 82, 86); Deep User - not IT (47, 73); Deep User - IT (61, 63, 64); and Light User (all other industries not elsewhere classified). The solid line corresponds to the distribution in 2011 and the dashed line to the distribution in 2019.

We estimate the same specification as in Equation (12) with 2-digit NACE industries in the case of Germany between 2012 and 2019. Thus, we estimate the employment growth as a function of occupational exposure to all emerging technologies and allow the effect to differ between sector types. Table 5 reports the estimated coefficients.

We observe that the three results obtained at the 1-digit level hold when we disaggregate to the 2-digit NACE level. First, Column (1) indicates that there is still a positive effect of occupational exposure to emerging technologies, although only significant at the 10% level.

Second, Column (2) indicates that this effect entails large heterogeneity between industries with the Producer, Intensive Users and Deep Users (Not IT) driving the positive overall relationship, while there is no significant effect for occupations in Light User industries (the referent group), and large negative effect in Deep User specialized in IT.

Third, for Producer sectors, as well as Deep User (Not IT), occupations with high exposure experience job creation, while those with low exposure experience job destruction. This is more concerning in the case of Deep Users (Not IT) as most of the jobs tend to lie at the bottom of the exposure distribution as indicated in Figure 7.

Figure 7 shows the marginal effect of occupation exposure by sector type on the predicted

Table 5: Baseline estimate

	Dep. var.: Employment (in log-change)		
	(1)	(2)	(3)
Intercept	0.119^{***}	0.128^{***}	0.144^{***}
	(0.004)	(0.005)	(0.005)
Producer		-0.055^{***}	-0.024^{*}
		(0.012)	(0.014)
Deep User (IT)		-0.025	0.030
		(0.040)	(0.041)
Deep User (Not IT)		-0.071^{***}	-0.018
		(0.013)	(0.016)
Intensive User		-0.005	0.027^{**}
		(0.012)	(0.013)
ISCO Exposure in log	0.006^{*}	-0.004	-0.004
	(0.003)	(0.005)	(0.005)
ISCO Exposure (in log) × Producer		0.044^{***}	0.044^{***}
		(0.010)	(0.010)
ISCO Exposure (in log) × Deep User (IT)		-0.081^{***}	-0.085^{***}
		(0.018)	(0.018)
ISCO Exposure (in log) × Deep User (Not IT)		0.084^{***}	0.083^{***}
		(0.014)	(0.014)
ISCO Exposure (in log) × Intensive User		0.044^{***}	0.043^{***}
		(0.009)	(0.009)
NACE Exposure (in log)			-0.025^{***}
			(0.004)
R^2	0.000	0.045	0.049
Adj. R^2	0.000	0.044	0.048
Num. obs.	6921	6921	6921

Notes: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses. All regressions are weighted by the level of employment in 2011. ISCO Exposure is the cumulative exposure between 2011 and 2019 in $\log(1 + x)$ transform and employment in log-change between 2012 and 2019. Column (1) presents the raw correlation with country fixed effects. Column (2) adds the interaction of occupation exposure with the type of sector, where the baseline category for sector type is 'Light User'. Column (3) adds the NACE Exposure between 2011 and 2019 in $\log(1 + x)$ transform as a control variable.

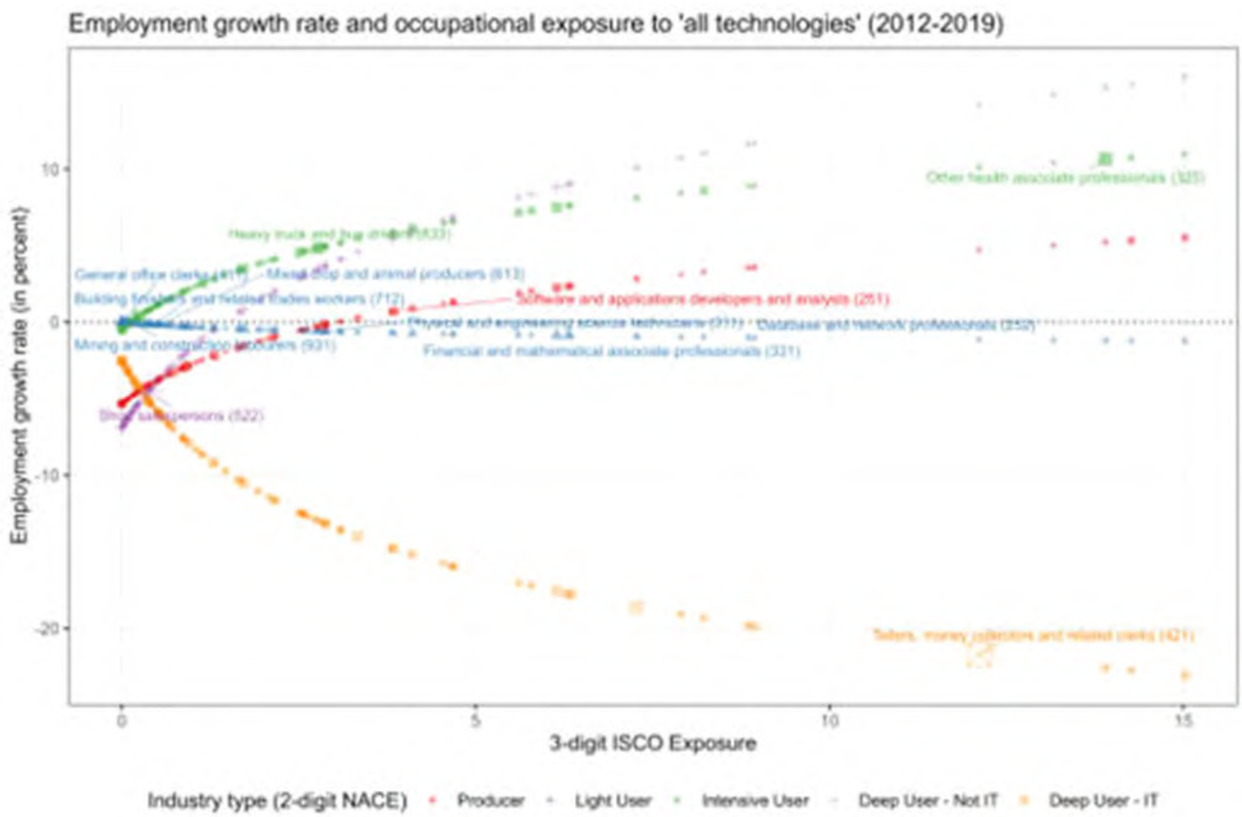
employment growth rate with German data.[19]

We observe that highly-exposed occupations experience employment growth in the Producer, Intensive and Deep User (Not IT) sectors, no growth in the Light user sectors, and negative employment growth in the Deep User (IT) sectors.

Low-exposed occupations, however, experience negative employment growth rates in Producer and Deep User sectors.

[19] The predicted employment growth rate is derived from the predicted log-change in employment $\widehat{\log g_{cio}}$, estimated from Equation (12), and according to the functional form specified in Equation (11). More specifically, the y-axis corresponds to $\exp(\widehat{\log g_{cio}}) - 1$.

Figure 8: Marginal effect of occupational exposure on employment growth rate by sector type in Germany (2012–2019)



Notes: This figure shows the model predictions from the estimated baseline model reported in column (2) from Table 5. The x-axis is the cumulated 3-digit ISCO exposure between 2011 and 2019 to emerging technologies and y-axis is the growth rate of employment (expressed in percent). Sectors are grouped into five types: Producer, Light User, Intensive User, Deep User (Not IT) and Deep User (IT) sectors. The size of each data point is proportional to its initial level of employment in 2012 within the sector type. The marginal effect of occupation exposure on employment growth is normalized to the one with a null occupation exposure in the Light User sector.

Overall, the opposed relationships in both Deep User sector types reveal that the core occupations in these industries, that is low-exposed occupations in Not IT Deep User and high-exposed occupations in IT Deep User, experience negative employment growth.

6 Conclusion

This paper provides a measure of the (yearly) exposure of both industries and occupations to the tasks performed by a wide set of digital automation technologies identified from patents (and publications). These goes beyond AI and robotisation, and include technologies related to data acquisition and management, user interfaces, additive manufacturing, computing and network technologies. That is, technologies that are used across a variety of non manufacturing sectors, and in a variety of occupations.

The method to link technologies to occupations and industries is reproducible and scalable. We also produce a panel database that contains yearly exposure scores to to automation technologies as well as exposure projections.

We find that exposure to automation technology is not only a problem of low skilled or routine tasks and occupations. Besides the machine operators, the most exposed 1–digit ISCO-08 categories are 'Clerical support workers' and 'Technicians and Associate Professionals', followed by 'Professionals' and 'Managers', which are in the middle of the distribution. These are occupations that involve also several non-routine and non-manual tasks.

However, exposure measures based on semantic similarity (Webb, 2019) are likely to capture also the industries and occupations that produce the automation technologies. A patent describing a robot is likely to associated to a robots producing industries. To address this issues, we distinguish between industries that produce the technologies from all the others – assuming that they mainly use them. This allows us to make one further contribution: distinguishing between industries where technology may complement workers, from those where they may substitute workers, for the same occupation.

To assess our exposure measures, we bring them to the employment data for 28 European countries and for Germany for which we have more granular data. Findings confirm expectations. Occupations that are highly exposed to emerging automation technologies in producing industries see an increase in employment between 2011-19. However, those occupations that are not exposed tho those technologies, instead, have seen a reduction in employment. In other words, the workers that gain most are those that work in occupations that are related to the technology (for producing industries). Across all using industries the effect balance out between different types of industries. When we break up industries by their level of exposure (for Germany), we observe that highly exposed occupations see workers replaced by the automation technologies in industries that are highly exposed and that are IT-related.

One implication is that European countries should focus on industries producing automation technologies, to be on the receiving end of the employment growth. This means producing the technologies that will substitute workers in some of the using sectors (the IT-related ones). Importantly, some of the using industries, also experienced increased employment in occupations highly exposed to emerging automation technologies. So these are also strategies industries when thinking about complementing employment with automation technologies. Of course, another possibility is to direct technological change in automation technologies in direction that create more jobs than they replace. In this paper we have taken those technological trajectories for given.

The produced estimations of industry and occupation exposure to emerging automation technologies in conjunction with data on industries, trade, and skills across European regions are very important for forecasting the effects of technological change on the future labour markets and designing effective industrial, education and training policies to address these changes. They can also be used to understand which aspects of these technologies imply a job replacing exposure, and which imply a job-creating exposure.

References

Alkemade, F., Heimeriks, G., Schoen, A., Villard, L., and Laurens, P. (2015). Tracking the internationalization of multinational corporate inventive activity: national and sectoral characteristics. *Research Policy*, 44(9):1763–1772.

Calvino, F., Samek, L., Squicciarini, M., and Morris, C. (2022). Identifying and characterising AI adopters.

Cebollada, S., Payá, L., Flores, M., Peidró, A., and Reinoso, O. (2021). A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data.

Chaturvedi, S., Ciarli, T., Nomaler, Ö., and Prytkova, E. (2023). What's the Future of Automation? Using Semantic Analysis to Identify Emerging Technologies. Technical report.

Ciarli, T., Kenney, M., Massini, S., and Piscitello, L. (2021). Digital technologies, innovation, and skills: Emerging trajectories and challenges. *Research Policy*, 50(7):104289.

Cirillo, V., Evangelista, R., Guarascio, D., and Sostero, M. (2020). Digitalization, routineness and employment: An exploration on Italian task-based data. *Research Policy*, page 104079.

Dechezleprêtre, A., Hémous, D., Olsen, M., and Zanella, C. (2021). Induced Automation: Evidence from Firm-level Patent Data. *SSRN Electronic Journal*.

Dosi, G. (1982). Technological paradigms and technological trajectories : A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3):147–162.

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., and Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences of the United States of America*, 116(14):6531–6539.

Leopold, H., van der Aa, H., and Reijers, H. A. (2018). Identifying candidate tasks for robotic process automation in textual process descriptions. In *Lecture Notes in Business Information Processing*, volume 318, pages 67–81. Springer Verlag.

Miller, D. J. (2006). Technological diversity, related diversification, and firm performance. *Strategic Management Journal*, 27(7):601–619.

Mitchell, T. and Brynjolfsson, E. (2017). Track how technology is transforming work. *Nature*, 544(7650):290–292.

Nedelkoska, L. and Quintini, G. (2018). Automation, skills use and training. Technical report, OECD, Paris.

Peeters, B., Song, X., Callaert, J., Grouwels, J., and Looy, B. V. (2010). Harmonizing harmonized patentee names: An exploratory assessment of top patentees.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market.

Table A.1: Technology families

Technology family	**Some examples**
Robotics	Swarm robotics, robotic vehicles
Data Acquisition Technologies	Remote sensing, IoT, scanners
Data management	Database systems, cryptography, security, blockchain
Computing	Computing architectures e.g. cloud, edge, neuromorphic, fog
AI & Intelligent Information System	Machine learning, NLP, IoT
Additive manufacturing	CAD, CAM, rapid prototyping, 3D printing
Networking	IoT, wireless communication
User interface	VR/AR, smart personal assistants, interactive holograms

Notes: This table presents the list of technology families considered as families of digital automation technologies and provides examples.

Table A.2: Top 10 exposed 2-digit NACE industries

Code	NACE Industry	C_k^i	$\log(1 + C_k^i)$
26	Manufacture of computer, electronic and optical products	25006.79	10.13
63	Information service activities	23678.88	10.07
61	Telecommunications	7950.46	8.98
47	Retail trade, except of motor vehicles and motorcycles	7792.49	8.96
28	Manufacture of machinery and equipment n.e.c.	6770.95	8.82
73	Advertising and market research	6238.41	8.74
64	Financial service activities, except insurance and pension funding	5710.85	8.65
52	Warehousing and support activities for transportation	3778.36	8.24
27	Manufacture of electrical equipment	3567.23	8.18
62	Computer programming, consultancy and related activities	3453.14	8.15

Notes: This table presents the top 10 2-digit NACE (Rev. 2) industries ranked by exposure to all technologies. Columns (from left to right) correspond to industry code, industry label, exposure score, and the exposure score expressed in $\log(1 + x)$ transform.

Appendices

A Appendix A

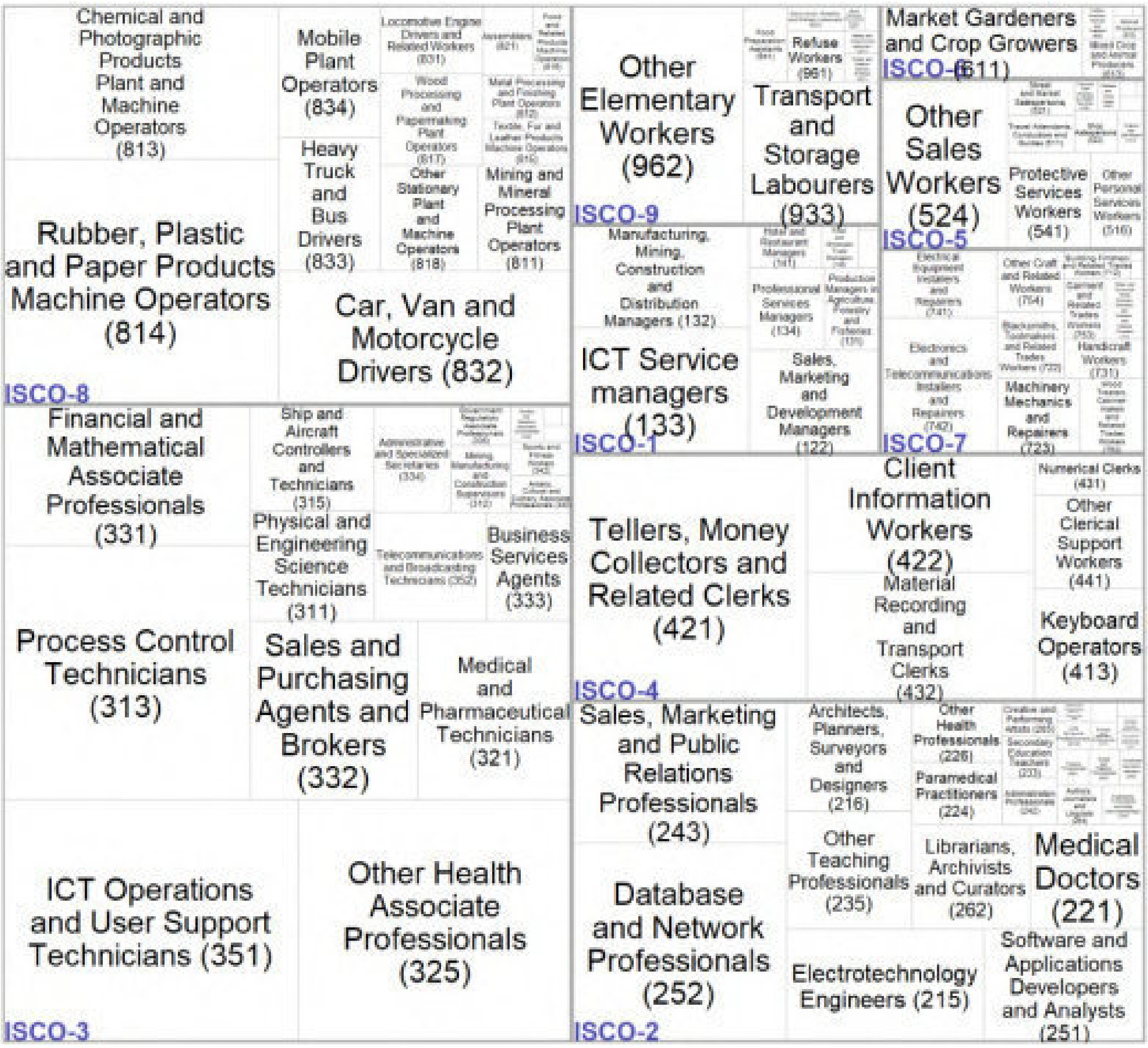
Table A.1 presents the list of technology families with some examples.

Figure A.1 presents the cumulative occupation exposure to all technologies between 2011 and 2021 for 3-digit ISCO-08 occupations.

Table A.2 presents the top 10 exposed 2-digit NACE industries.

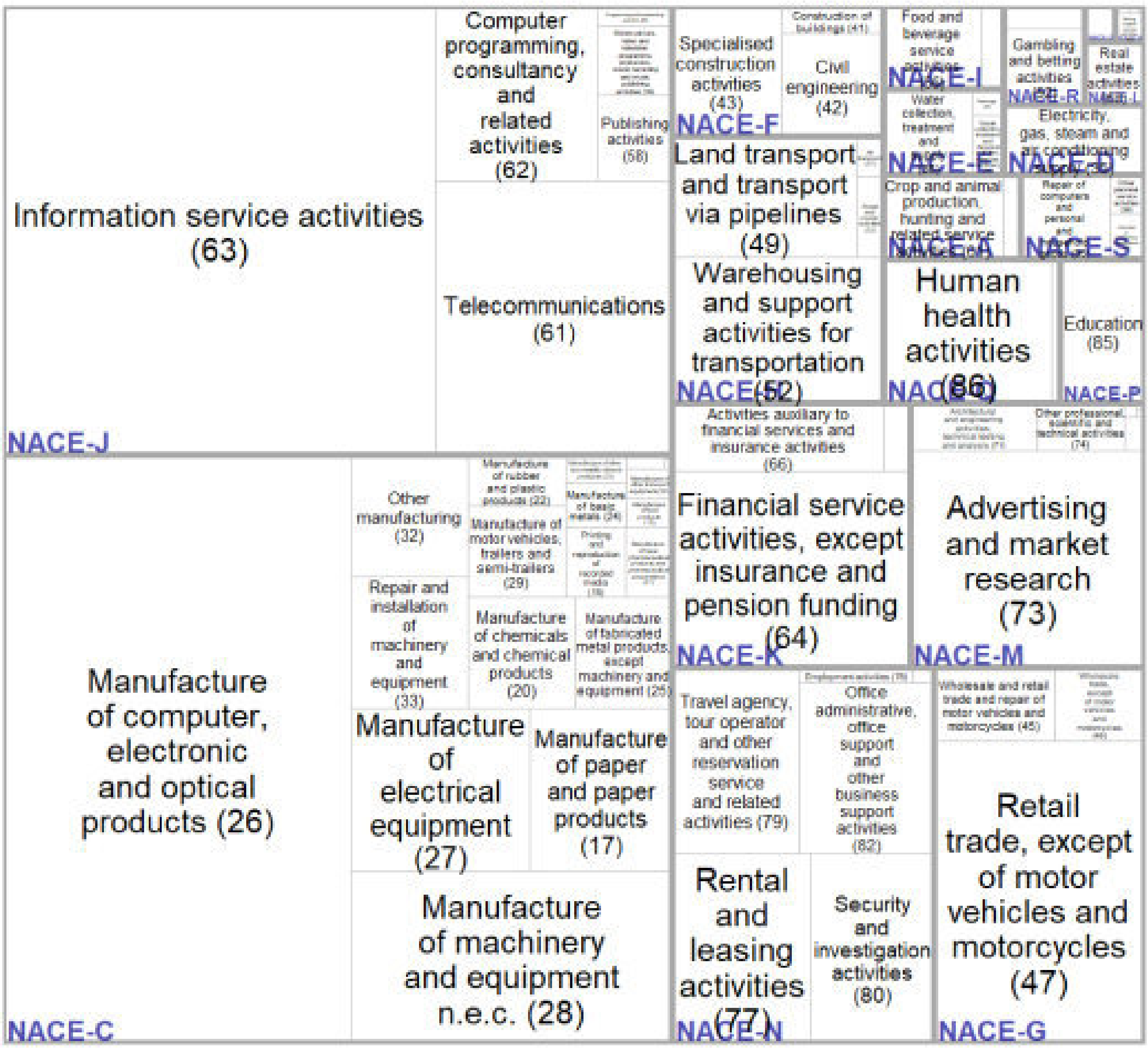
Figure A.2 presents the cumulative industry exposure to all technologies between 2011 and 2021 for 2-digit NACE Rev.2 industries.

Figure A.1: Cumulative Occupation exposure C^o 2011—2021 (average technology)



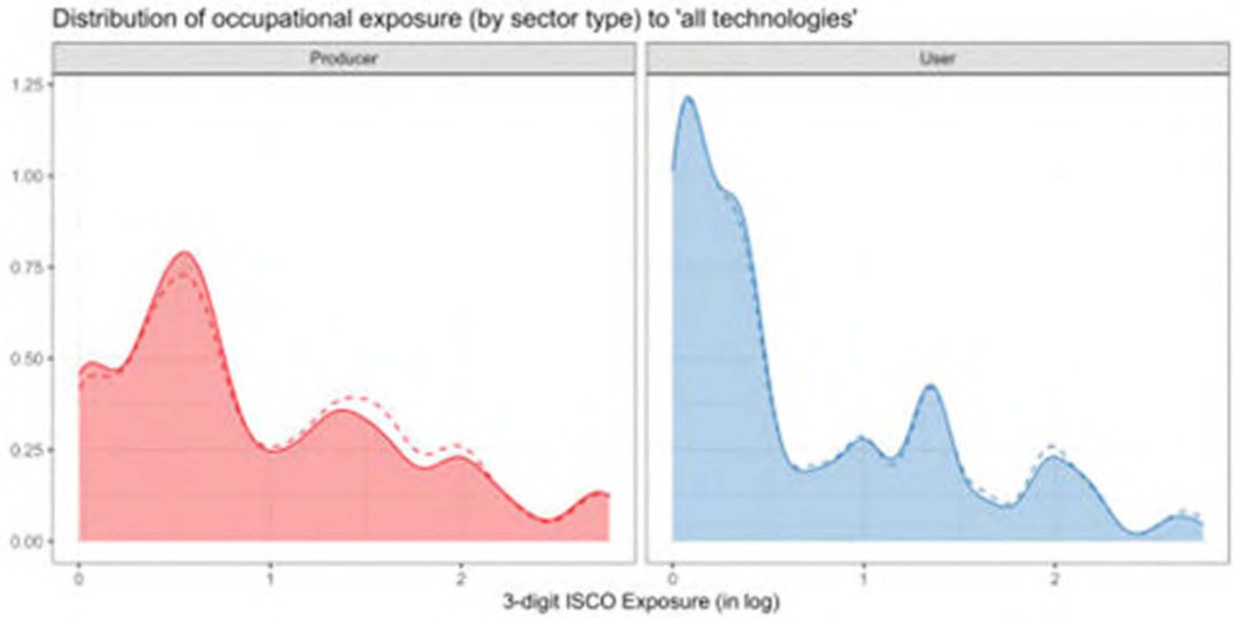
Notes: Smaller tiles represent exposure scores of 3–digit ISCO-08 occupations. Larger tiles group 3–digit occupations under their parent 1–digit ISCO-08 occupation: ISCO-1 — Managers; ISCO-2 Professionals; ISCO-3 — Technicians, Associate Professionals; ISCO-4 —Clerical Support Workers; ISCO-5 —Service, Sales Workers; ISCO-6 — Agricultural, Forestry, Fishery Workers; ISCO-7 — Craft, Related Trades Workers; ISCO-8 — Plant, Machine Operators, Assemblers; ISCO-9 — Elementary Occupations

Figure A.2: Cumulative Industry exposure C^i 2011—2021 (average technology)



Notes: Smaller tiles represent exposure scores of 2–digit NACE Rev.2 industries. Larger tiles group 2–digit industries under their parent 1–digit NACE Rev.2 industry. A — Agriculture, forestry and fishing; B — Mining and quarrying; C — Manufacturing; D — Electricity, gas, steam and air conditioning supply; E — Water supply; sewerage, waste management and remediation activities; F — Construction; F — Wholesale and retail trade; repair of motor vehicles and motorcycles; H — Transportation and storage; I — Accommodation and food service activities; J — Information and communication; K — Financial and insurance activities; L — Real estate activities; M — Professional, scientific and technical activities; N — administrative and support service activities; O — Public administration and defence, compulsory social security; P — education; Q — Human health and social work activities; R — Arts, entertainment and recreation; S — Other service activities; T — Activities of households as employers, undifferentiated goods- and services-producing activities of households for own use; U — Activities of extraterritorial organisations and bodies

Figure B.1: Distribution of occupational exposure by sector type



Notes: This figure shows the distribution of occupational exposure to all technologies by sector type. The x-axis is the 3-digit ISCO exposure in $\log(1+x)$ transform and the y-axis is the density. Panels refer to the two types of sectors: Producer and User. The solid line corresponds to the distribution in 2011 and the dashed line to the distribution in 2019. The producer sectors are Manufacturing (C) and Information and Communication (J), whereas all other sectors are considered user sectors.

B Appendix B

Figure B.1 presents the distribution of occupational exposure across the two types of sectors. The density of occupations with low exposure to technologies is more important in user sectors compared to producer sectors. The average (median) occupational exposure, expressed in $\log(1+x)$ transform, is about 0.690 (0.391) in user sectors and 0.931 (0.650) in producer sectors.

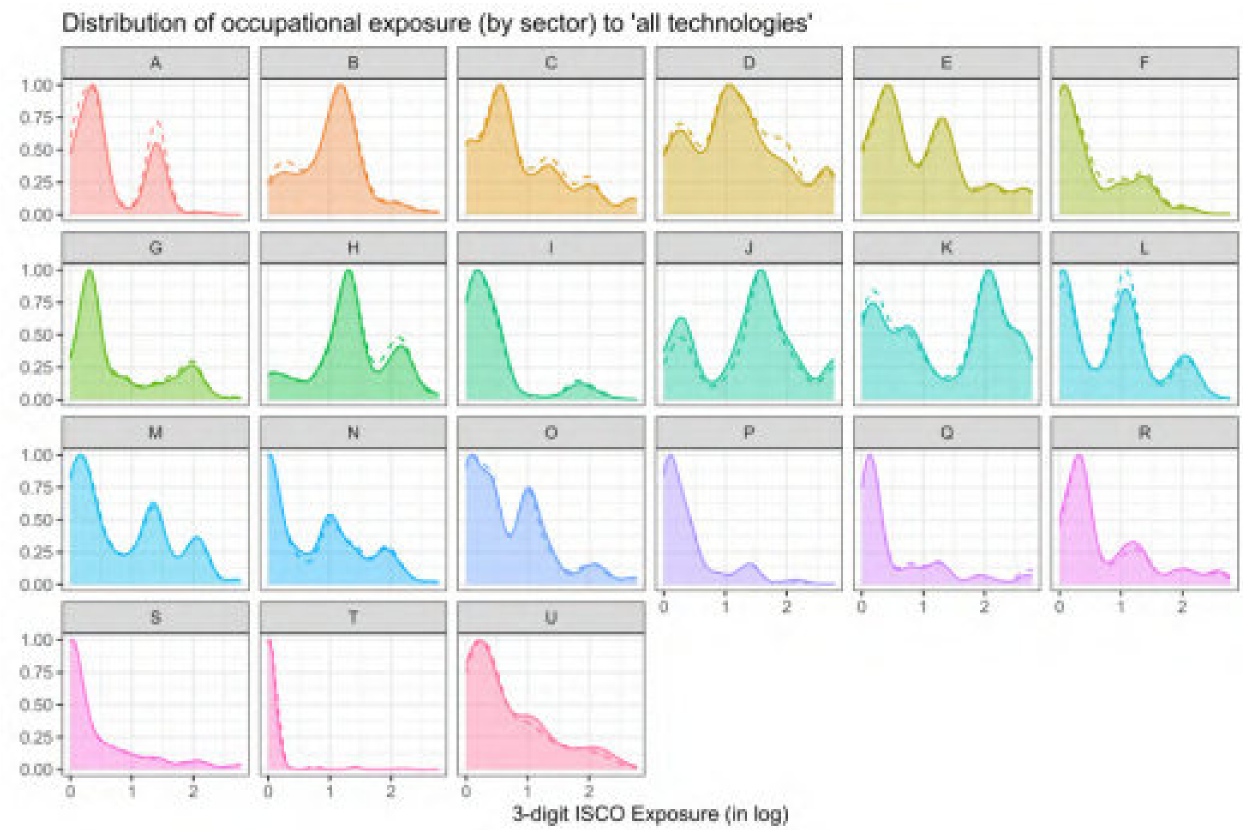
Figure B.2 presents the distribution of occupational exposure by 1-digit sector from the EU-LFS data.

Figure B.3 presents the marginal effect of occupational exposure on employment growth, as in Equation (12) but estimated separately, by 1-digit sector from the EU-LFS data.

Figure B.4 presents the distribution of the 2-digit NACE exposure of User industries and the cutoffs used to distinguish between the several types of Users. Note that the trough between 0 and 1 can also be considered as an additional cutoff to distinguish between Light Users. Yet, we decide to not include it as the results with respect to employment growth do not differ for industries below and above that threshold.

Table B.1 presents the producer exposure to all technologies by 1-digit NACE sectors. Sectors 'Manufacturing (C)' and 'Information and Communication (J)' account for 87% of the total producer exposure.

Figure B.2: Distribution of occupational exposure by 1-digit sector



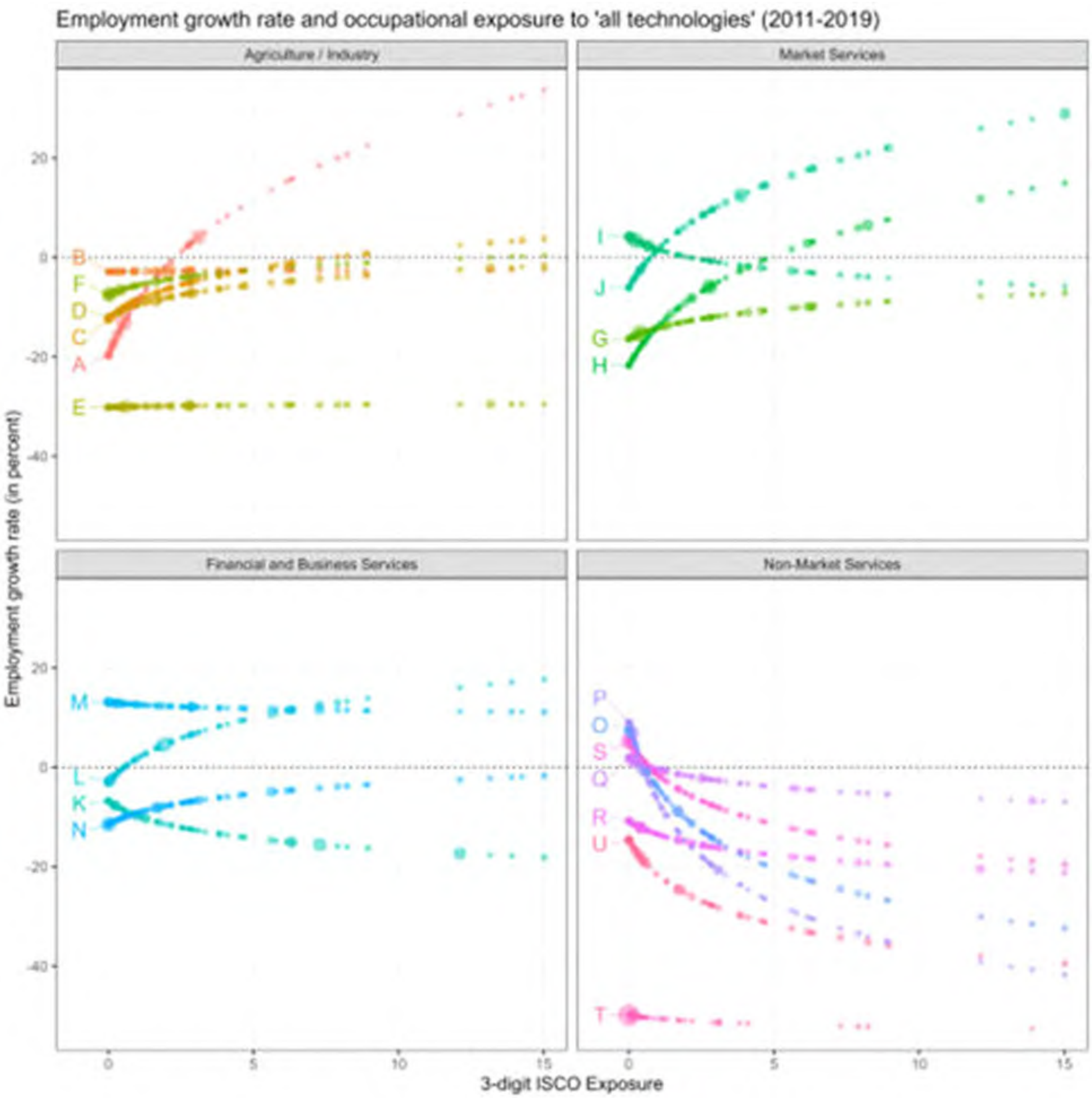
Notes: This figure shows the occupation exposure distribution to all emerging technologies for all 1-digit NACE sectors. The x-axis is the 3-digit ISCO exposure in $\log(1 + x)$ transform and the y-axis is the density. The solid line corresponds to the distribution in 2011 and the dashed line to the distribution in 2019.

Table B.2 presents the summary statistics of the occupational exposure distribution for sector types at the 2-digit NACE level using German data.

Table B.3 reports the estimated coefficients when excluding occupations with null exposure. Compared to the baseline estimate, the coefficient of ISCO exposure in Column (1) becomes insignificant whereas the relationship is now negative for the User sector in Column (2). This indicates that occupations with null ISCO exposure, which are here excluded, tend to experience positive employment growth. Thus, increasing the coefficient of interest in both columns.

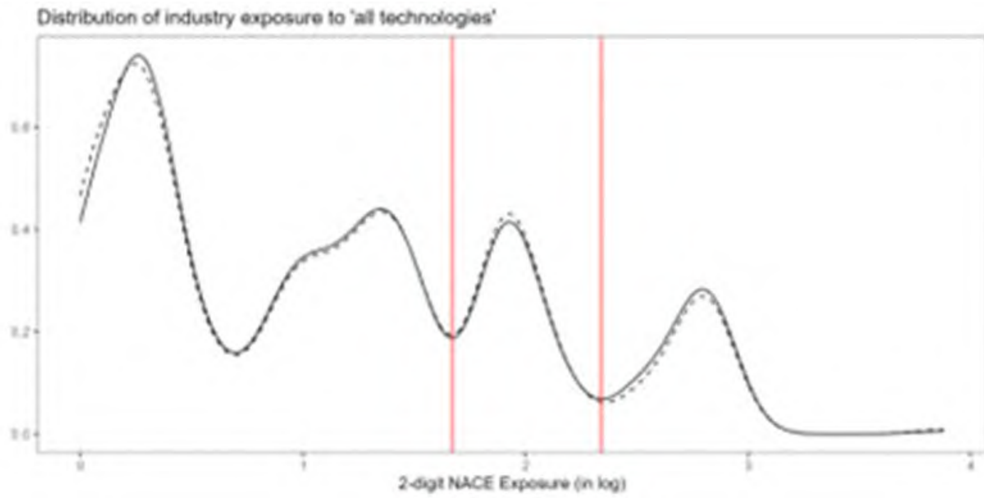
Table B.4 reports the estimated coefficients when using the inverse hyperbolic sine transform instead of the $\log(1+x)$ transform on ISCO exposure. This table accounts for potential bias that could be due to the use of the $\log(1+x)$ transform when computing the exposure for occupations that are close to zero. The results do not differ from the baseline specification. Thus, our results are robust to the use of the inverse hyperbolic sine transform.

Figure B.3: Marginal effect of occupation exposure on employment growth by 1-digit sector



Notes: This figure shows the marginal effect of occupational exposure to all emerging technologies on employment growth when estimated separately for each 1-digit NACE sector.

Figure B.4: Distribution of the 2-digit industry exposure of user industries in Germany



Notes: This figure shows the weighted distribution of industry exposure to all emerging technologies of User sectors. The x-axis is the 2-digit NACE exposure in $\log(1+x)$ transform and the y-axis is the density. The solid line corresponds to the distribution in 2011 and the dashed line to the distribution in 2019. The right-hand side threshold corresponds to the cutoff between Deep User and Intensive User industries, while the left-hand side one to the cutoff between Intensive User and Light User industries.

Table B.1: Producer exposure to all technologies at the 1-digit NACE level

Sector	C^i	s^i	S^i	Share	Cum. share
C	96.36	0.56	54.28	0.72	0.72
J	73.37	0.15	10.82	0.14	0.87
K	15.34	0.26	3.98	0.05	0.92
G	19.33	0.11	2.11	0.03	0.95
F	6.74	0.13	0.89	0.01	0.96
N	23.98	0.03	0.81	0.01	0.97
M	15.15	0.05	0.70	0.01	0.98
D	2.41	0.17	0.41	0.01	0.99
H	13.75	0.03	0.40	0.01	0.99
Q	6.24	0.04	0.25	0.00	0.99
P	2.98	0.05	0.16	0.00	1.00
R	1.94	0.05	0.09	0.00	1.00
L	0.91	0.10	0.09	0.00	1.00
E	2.52	0.01	0.03	0.00	1.00
I	2.46	0.01	0.03	0.00	1.00
S	2.53	0.00	0.00	0.00	1.00
O	0.27	0.01	0.00	0.00	1.00
A	2.79				
B	0.26				

Notes: This table shows the producer exposure to all technologies at the 1-digit NACE level. Columns indicate (from left to right) the 1-digit sector code, the industry exposure score C^i, the share of producer in the industry s^i, the producer exposure score S^i, the share in total producer exposure score, and the cumulative share.

Table B.2: Summary statistics of the occupational exposure distribution for sector types at the 2-digit NACE level

Sector type	Mean	Median	SD	N	SE
Producer	1.04	0.97	0.73	880	0.02
Light User	0.70	0.42	0.70	4928	0.01
Intensive User	1.24	1.29	0.90	698	0.03
Deep User - Not IT	0.68	0.33	0.67	187	0.05
Deep User - IT	2.03	2.57	0.85	228	0.06

Notes: This table provides the summary statistics of the occupational exposure distribution from Figure 7. Columns indicate (from left to right) the sector type, the mean, the median, the standard deviation (SD), the number of observations (N) and the standard error (SE).

Table B.3: Baseline estimate (excluding null ISCO exposure)

	Dep. var.: Employment (in log-change)		
	(1)	(2)	(3)
Intercept	-0.047^{**}	-0.039^{*}	-0.110^{***}
	(0.024)	(0.024)	(0.025)
Producer		-0.050^{***}	-0.156^{***}
		(0.016)	(0.020)
ISCO Exposure (in log)	-0.001	-0.014^{***}	-0.023^{***}
	(0.005)	(0.006)	(0.006)
ISCO Exposure (in log) $\times$ Producer		0.066^{***}	0.076^{***}
		(0.013)	(0.013)
NACE Exposure (in log)			0.038^{***}
			(0.004)
Country FE	Yes	Yes	
Country $\times$ Industry FE			Yes
R^2	0.369	0.369	0.371
Adj. R^2	0.368	0.369	0.371
Num. obs.	24246	24246	24246

Notes: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses. All regressions are weighted by the level of employment in 2011. ISCO Exposure is the cumulative exposure between 2011 and 2019 in $\log(1 + x)$ transform and employment in log-change between 2011 and 2019. Column (1) presents the raw correlation with country fixed effects. Column (2) adds the interaction of occupation exposure with the type of sector, where the baseline category for sector type is 'User'. Column (3) adds the NACE Exposure between 2011 and 2019 in $\log(1 + x)$ transform as a control variable. Table 3 presents the baseline specification.

Table B.4: Baseline estimate (with inverse hyperbolic sine transform)

	Dep. var.: Employment (in log-change)		
	(1)	(2)	(3)
Intercept	-0.061^{**}	-0.051^{**}	-0.111^{***}
	(0.024)	(0.024)	(0.025)
Producer		-0.064^{***}	-0.141^{***}
		(0.014)	(0.018)
ISCO Exposure (in IHS)	0.009^{**}	-0.002	-0.007^{*}
	(0.004)	(0.004)	(0.004)
ISCO Exposure (in IHS) × Producer		0.059^{***}	0.065^{***}
		(0.010)	(0.010)
NACE Exposure (in log)			0.026^{***}
			(0.004)
Country FE	Yes	Yes	
Country × Industry FE			Yes
R^2	0.359	0.360	0.362
Adj. R^2	0.359	0.360	0.361
Num. obs.	26109	26109	26109

Notes: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses. All regressions are weighted by the level of employment in 2011. ISCO Exposure is the cumulative exposure between 2011 and 2019 in inverse hyperbolic sine transform and employment in log-change between 2011 and 2019. Column (1) presents the raw correlation with country fixed effects. Column (2) adds the interaction of occupation exposure with the type of sector, where the baseline category for sector type is 'User'. Column (3) adds the NACE Exposure between 2011 and 2019 in $\log(1 + x)$ transform as a control variable. Table 3 presents the baseline specification.

Online Appendix

Title

Authors

Reviewer Appendix

Title

Authors