The Employment Impact of Emerging Digital Technologies[*]

Ekaterina Prytkova[†] Fabien Petit[‡] Deyu Li[§]
Sugat Chaturvedi[¶] Tommaso Ciarli[‖]

September 2024

Abstract

This paper measures the exposure of industries and occupations to a large set of emerging digital technologies (including robots and AI) and estimates their impact on European employment. Using a novel approach that leverages sentence transformers, we calculate exposure scores based on the semantic similarity between patents and ISCO-08/NACE Rev.2 classifications to construct an open–access database, 'TechXposure'. Using a shift–share approach, we instrument the regional exposure to emerging digital technologies to estimate their impact on employment across European regions. We find an overall positive effect of emerging digital technologies which is driven by an increase in low- and high-skilled employment at the cost of middle-skilled employment, suggesting continued job polarization. Upon examining the individual effects of these technologies, we observe significant heterogeneity in their impact on employment. Notably, we find that robots and machine learning have negative impacts on employment, except for high-skilled workers. Our work suggests that the excessive focus on these specific technologies could potentially overshadow the positive impacts of other emerging digital technologies on employment.

Keywords: Occupation Exposure; Industry Exposure; Text as Data; Natural Language Processing; Sentence Transformers; Emerging Digital Technologies; Automation; Employment
JEL Codes: C81, O31, O33, O34, J24, O52, R23

*We are grateful to Cecilia García-Peñalosa, Florencia Jaccoud, Guy Michaels, Önder Nomaler, Morgan Raux, Maria Savona and John Van Reenen, as well as to audiences at the ESCoE 2023, GTM Conference, Eu-SPRI 2023, VSI 2023, AI and the Economy Conference, JRD 2023 Autumn, EMAEE Workshop, Untangled Final Conference, UNU-MERIT Internal Conference 2023, CESifo Big Data Workshop, JRD 2024 Spring, AMAeA 2024, EEA 2024, and seminar participants at JRC–European Commission, IFS-UCL-IoE HKLM Seminar, UK DfE, CEP, University of Barcelona, and all PILLARS events 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.

†Côte d'Azur University. Email: ekaterina.prytkova@univ-cotedazur.fr.
‡University College London, CEPEO. Email: f.petit@ucl.ac.uk. Website: fabienpetit.com.
§Utrecht University. Email: d.li1@uu.nl
¶Ahmedabad University. Email: sugat.chaturvedi@ahduni.edu.in.
‖United Nations University, UNU-MERIT and University of Sussex, SPRU. Email: ciarli@merit.unu.edu

1 Introduction

The past decade has witnessed rapid advancements in digital automation technologies, such as artificial intelligence, augmented and virtual reality, electric vehicles, self-driving cars, drones, mobile robots, the Internet of Things, 3D printing, and blockchain. While substantial evidence exists on the labor market impact of more established digital technologies, such as Information and Communication Technologies (ICT) and industrial robots,[1] little is known about the employment impact of this diverse array of new digital technologies.

This gap in the literature results from the limited number of available metrics measuring workers' and industries' exposure to emerging digital technologies, which stems from the challenge of identifying the relevance of a technology to an occupation or industry. Prior work, which provide measures of workers' and industries' exposure to more recent technology, focus either on specific technologies such as some applications of artificial intelligence, or provides a catch-all of automation technologies,[2] and only focuses on the US context.[3]

This paper measures the exposure of industries and occupations to a broad set of digital technologies that emerged over the past decade and estimates their impacts on regional employment in Europe. Using state-of-the-art Natural Language Processing (NLP) tools, such as sentence transformers, we introduce an innovative methodology to measure the exposure of industries and occupations to emerging digital technologies. Our approach, based on semantic similarity between patents and industry/occupation descriptions (obtained from standard classification systems), is scalable and reproducible for any type of technology, any period, and any classification system.

The outcome of this methodology is the 'TechXposure' database, a pioneering resource that we have made publicly available. This database stands out as the first of its kind, offering an unprecedented level of granularity in measuring the exposure of NACE industries (up to the 3-digit level) and ISCO-08 occupations (up to the 4-digit level) to a comprehensive and extensive set of technologies.

Using an IV shift-share approach, we estimate the employment impact of a broad set of digital technologies that emerged over the past decade across several demographic groups.

[1] See, for instance, Autor et al. (2003), Autor et al. (2006), Goos and Manning (2007), Goos et al. (2009, 2014), Michaels et al. (2014), Akerman et al. (2015) for the labor market consequences of technological change related to ICT; and Graetz and Michaels (2018), Acemoglu and Restrepo (2020), Vries et al. (2020), Dauth et al. (2021), Aghion et al. (2023) for the labor market effects of industrial automation and industrial robots.

[2] See Felten et al. (2018, 2021), Webb (2019), Alekseeva et al. (2021), Acemoglu et al. (2022b) for studies focusing on AI exposure metrics; Kogan et al. (2019, 2021), Mann and Püttmann (2023), Autor et al. (2024) for studies measuring exposure to a catch-all of automation technologies.

[3] A notable exception is Albanesi et al. (2023). By combining Felten et al. (2018) and Webb (2019) exposure metrics, they study the relationship between labor markets and exposure to AI and software in 16 European countries.

We leverage industry exposures from our database and the baseline employment shares of these industries in each European region to provide valuable insights into the labor market consequences of regional exposure to these technologies.

We start our analysis by grouping patents into technologies based on semantic similarity in their titles. We use the sample of patents identified as core emerging digital technologies in Chaturvedi et al. (2023). This sample includes the digital innovations filed between 2012 and 2021 that are central to the development of digital technologies in this decade. We convert the text of patent titles into vector representations, or *embeddings*,[4] using the pre-trained sentence transformer model *all-mpnet-base-v2* (Reimers and Gurevych 2019).[5] We apply k-means clustering on these embeddings, resulting in the identification of 40 emerging digital technologies, each defined as a group of patents.

We compute the exposure of industries and occupations to these technologies based on the semantic connection between patents and the descriptions of industries and occupations. For each industry-patent and occupation-patent combination, we calculate the cosine similarity score, which reflects the degree of similarity between the documents. To enhance the matching quality, we introduce a filtering procedure that retains only the most relevant pairs. Once filtered, we aggregate the cosine similarity scores from individual patents to the technologies under which they were clustered by taking the citation-weighted sum.

Our exposure metric reflects the degree to which a specific technology is *relevant* to an industry or occupation. For industries, relevance is determined by the integration of technology into the production process and/or if the technology enhances the output of an industry. For occupations, relevance measures the importance of a technology in performing tasks and functions inherent to an occupation. These exposure scores indicate the contextual relevance of each technology to a given industry or occupation, which is a proxy for adoption.[6]

We estimate the causal effect of these digital technologies on European regional employment. To address endogeneity issues, we instrument the regional exposure to these technologies with a shift-share design, in which the industry exposure scores are the *shocks* and the

[4]Text embedding is a Natural Language Processing (NLP) technique used to transform text (words, sentences, documents) into a numerical representation, i.e., high-dimensional numerical vectors, commonly referred to as embeddings. See Gentzkow et al. (2019) for a comprehensive review of NLP applications in the economic literature.

[5]A sentence transformer is a specific architecture of a deep neural network. The features of this architecture enable the model to capture the contextual significance of words in a text and leverage the ensemble effect to produce embeddings. The sentence transformer model *all-mpnet-base-v2* is fine-tuned on over a billion sentence or paragraph pairs from academic papers, Wikipedia, and Stack Exchange, among others, and has shown state-of-the-art results on sentence similarity tasks (Reimers and Gurevych 2019).

[6]However, exposure scores are neutral regarding the nature of the relationship between technology and workers in this industry or occupation, meaning that they do not specify whether the technology and labor are complementary or substitute in producing output.

baseline employment shares of these industries are the *shares*. Our identification strategy relies on the quasi-random assignment of shocks, allowing the employment shares to be endogenous. We argue that the development of emerging digital technologies is predominantly a global phenomenon and therefore not driven by local employment changes in Europe. In addition, we exclude patents originating from Europe. Thus, our industry exposure scores (i.e., the shocks) are assumed to be quasi-exogenous to regional employment changes in Europe. Our approach leverages the equivalence proposed by Borusyak et al. (2021), and we implement the AKM0 inference procedure in line with the literature (Adão et al. 2019).

Our estimation proceeds in two steps. First, we analyze the overall impact of emerging digital technologies on the regional employment-to-population ratio from 2012 to 2019 for several demographic groups. Our sample includes 320 NUTS-2 regions from 32 European countries. Second, we conduct a more detailed analysis to disentangle the individual effects of robotics technologies and data-intensive technologies (including several AI applications). In addition to the quasi-random assignment of shocks, we assume that regions with greater exposure to emerging digital technologies are not disproportionately affected by other labor market shocks or trends.

Our work reveals several new findings. First, we document new insights regarding the exposure of industries and occupations to emerging digital technologies. For occupations, we find that clerical support workers, plant/machine operators, and assemblers are the most exposed to emerging digital technologies, closely followed by high-paying and qualified occupations such as managers, professionals, technicians, and associate professionals. Additionally, we observe that manual occupations are more exposed to *tangible* technology families, such as 3D Printing, Embedded Systems, and Smart Mobility, while cognitive occupations are more exposed to *intangible* technology families, such as Computer Vision, E-Commerce, Payment Systems, HealthTech, and Digital Services. We find a similar divide for industries, with agriculture, manufacturing industries, and services operating physical infrastructures, such as transportation and storage, being more exposed to tangible technologies as compared to other services which are more exposed to intangible technologies.

Second, the overall impact of emerging digital technologies on regional employment is positive; however, we observe a job polarization pattern. We find that a one-standard-deviation increase in regional exposure leads to a 1.042 percentage point (pp.) change, corresponding to 2.08%, in the employment-to-population ratio from 2012 to 2019. When decomposing this effect into skill groups, proxied by education levels, we observe that only low- and high-skilled employment increases due to emerging digital technologies, with respective changes of 0.697 pp. (5.87%) and 0.777 pp. (5.18%) in their employment-to-population ratios, while middle-skilled employment decreases by 0.393 pp. (-1.7%). Additionally, we find that the positive effects are relatively stronger for female and young (aged 15-24) workers compared to male and mature (aged 25-64) workers.

Third, we find significant heterogeneity in the impact of individual technologies. Greater regional exposure to industrial automation (including industrial robots), intelligent logistics (including mobile robots), and machine learning increases the employment of high-skilled workers while decreasing it for both low- and middle-skilled workers. Conversely, some AI applications related to information processing and workflow management display positive impacts on total employment, driven by the employment of low-skilled workers for information processing and shared across the entire skill distribution for workflow management.

Our work contributes to the literature on the labor market consequences of technological change in several ways. First, while our results align with existing literature regarding the negative impact of some automation technologies on employment (e.g. industrial robots and AI), our work suggests that the excessive focus on these specific technologies could potentially overshadow the positive impacts of other emerging digital technologies on employment. Consistent with our results, Mann and Püttmann (2023) and Autor et al. (2024), who use broader definitions of automation technology compared to solely industrial robots as in Acemoglu and Restrepo (2020) or AI and software as in Webb (2019), also find positive employment impacts in local US labor markets. This is particularly relevant when considering the crucial role of complementarities among these technologies in determining their effects on employment.

Second, our work uniquely addresses a gap in the literature regarding exposure metrics. While most existing metrics focus on US classifications and are limited to specific technologies,[7] our work is the first to provide detailed exposure scores for international classifications, specifically NACE Rev. 2 and ISCO-08, at a highly granular level and for a large set of digital technologies. This contribution extends the applicability of exposure metrics beyond the US context and fashionable technologies, offering valuable insights for future research. Additionally, our exposure scores are based on worldwide patents, thereby considering global advances in technologies that extend beyond the US and Europe.

Third, we also provide a methodological contribution by introducing a scalable and ad-

[7]See Jurkat et al. (2022) for international distribution of industrial robots by country and industry, Frey and Osborne (2017) for occupational exposure to computerization, Webb (2019) and Felten et al. (2021) for exposure to AI, and Felten et al. (2023) for exposure to recent advances in AI language modeling capabilities related to Large Language Models (LLM).

vanced methodology using state-of-the-art NLP techniques with sentence transformers.[8] Our methodology is universally applicable, bypassing the need for identifying specific keywords or *tokens*, as it leverages text similarity, thereby requiring only a relevant set of patents. In addition, our approach innovatively uses patents in this context. Although the use of patents to measure technical change is increasingly common, we are the first to define technologies as groups of patents clustered based on semantic distance. This novel method enables us to identify all digital technologies, not limited to general AI or robots, and provides a more precise and interpretable categorization of these technologies.

The paper is organized as follows. Section 2 outlines our methodology for deriving our set of emerging digital technologies from patent data. Section 3 introduces our state-of-the-art NLP-based method for calculating industry and occupation exposure scores to these technologies. Section 4 provides descriptive statistics regarding the exposure of industries and occupations to emerging digital technologies. Section 5 estimates the causal impact of these technologies on regional employment, using an IV shift-share approach. Section 6 concludes.

2 Emerging Digital Technologies

In this section, we derive and describe the emerging digital technologies that we identify, where each technology is a group of patents from the Derwent Innovation Index (DII) database.[9] For simplicity, we refer to 'patent' instead of 'patent family', which is the set of patents in various patent offices about a single invention, in the remainder of the paper. First, we describe the different parts of a patent's text and the properties of our patent sample. Second, we explain our methodology to cluster patents based on semantic similarity and obtain our set of emerging technologies. Finally, we describe the technologies.

[8] Kogan et al. (2021) identify breakthrough innovations with patents from 1850, applying Kelly et al. (2021)'s methodology, to estimate occupational exposure to these innovations via a TFIDF token-based approach. Dechezleprêtre et al. (2021) develop a measure of automation innovation in machinery by analyzing the frequency of specific keywords in patent texts from 1997. Mann and Püttmann (2023) distinguish US patents filed from 1976 to 2014 into automation and non-automation categories using tokens.

[9] DII covers over 120 million global patent publications from 59 worldwide patent-issuing authorities and creates a unique patent family for every invention. Each patent family is represented by a title and abstract translated into English and structured by experts into themed blocks such as novelty, use, advantage, and claims, among others, using standardized terms to harmonize textual descriptions and aid querying. In addition to international CPC and IPC patent classifications, Derwent maintains a custom hierarchical indexing system, Derwent Manual Codes, designed to reflect both the technical and application content of an invention, improving patent retrieval capabilities.

We use a set $\mathcal{P}$ of 190,714 Derwent patents, filed between 2012 and 2021.[10] This patent set constructed by Chaturvedi et al. (2023) represents the core emerging digital technologies and applications invented since 2011. Appendix A.1 provides detailed explanation of the patent corpus construction.

We use the patent title to encode semantic content of a patent in a numerical representation referred to as *embedding* using sentence transformers.[11] Derwent Database provides expertly curated titles and abstracts as available textual data per patent. The choice of patent title is dictated by its construction curated by the DII; all patent titles are comprised of two parts. The first part provides a concise description of the technology in a phrase or short sentence; we denote this part $p_1 \in p$. The second part describes *how the technology functions*; we denote this part $p_2 \in p$. The division between the two parts is marked by the first **comma–verb combination**.[12] This structure ensures a well-balanced representation of an invention in terms of generality vs specificity of description for the purpose of our analysis, i.e. detecting individual technologies. The inclusion of additional information present in abstracts such as independent claims, novelty, or description of drawings, would increase the textual content but provide no precision gains tilting the text's signal-to-noise ratio in favor of the latter.

For the purpose of semantic matching, this concise representation of an invention, combining its essence and intended function or use, can be mirrored using industrial and occupational descriptions. In essence, we represent an industry and an occupation with a set of sentences, each constructed following the same principle: the essence (represented by the industry/occupation title) concatenated with the function (a task for occupation and an activity/process for industry). Later, in Section 3, we provide a more detailed description of the treatment applied to industrial and occupational texts. Mirroring the structure of documents' textual content improves matching between different text corpora, i.e., patents and industrial/occupational taxonomies. The matching is further facilitated due to term standardization in patents and text harmonization performed by the DII; the language used to represent the invention and its function in the title conveys technical information through comprehensive descriptions rather than highly technical jargon. Lastly, unlike the abstract, the title is

[10] Each patent is a document that describes the invention, and how it differs from existing inventions. The information provided for each patent includes a title, an abstract, and additional metadata such as the list of applicants and inventors (i.e., companies or individuals), filing year and authority, citations, and codified technical areas according to various classifications (such as the International Patent Classification or IPC), among others. In turn, the abstract is divided into labeled topical blocks such as novelty, use, independent claims, description of drawings, etc.

[11] Prior work mainly relies on bag-of-words (BoW) approach, i.e. operating with tokens and their weighted frequencies (Kogan et al. 2019, Webb 2019, Arts et al. 2021, Dechezleprêtre et al. 2021, Mann and Püttmann 2023).

[12] Using Part-of-Speech (POS) tagging, we identify that this pattern appears in 87.3% of our patent sample, represented by the following combinations: ', has', ', includes', ', involves', and ', comprises'. For the remaining patents, we divide the patent title at the word space closest to the middle of the document.

always present for all patents.

We provide three examples of patent titles present in our sample:

1. Method for targeting television advertisement based on profile linked to online device**, involves** *selecting television advertisement to be directed to set-top box based on profile information pertaining to user or online activity.* (Patent ID 2013B87254, 2013)
2. Vehicle intelligent logistics control device**, has** *GPS locating module for obtaining position information of transport vehicle through main control chip, RFID reader for reading RFID tag information, and 4G module connected with server.* (Patent ID 201713859U, 2017)
3. System for recognizing training speech**, has** *process or which is configured to increment counter associated with word sequences, and train language model of automatic transcription system using word sequences and counter.* (Patent ID 202048118D, 2020)

For each patent title from our sample $\mathcal{P}$, we obtain its *embeddings* Emb_p, i.e., numerical representation of text in the form of 768-dimensional vector, using the pre-trained sentence transformer model all-mpnet-base-v2 (Reimers and Gurevych 2019). Sentence transformers encode the word's meaning in relation to its surrounding context, providing a significant advantage over the bag-of-words approach that treats text as an unordered collection of words. In particular, the *all-mpnet-base-v2* model is specifically trained for sentence similarity and clustering tasks.[13]

Then, we cluster the embeddings using the k-means algorithm and obtain 40 clusters, each of which we denote as our set of emerging digital technologies $k \in \mathcal{K}$. First, we compute all partitions between 5 and 100 clusters and record the respective Davies-Bouldin Index (DBI) (Davies and Bouldin 1979). We observe that the interval between 30 and 45 clusters contains the best partitions (the DBI is the lowest), i.e., simultaneously high within-cluster and low between-cluster similarity. We inspect partitions from this interval leveraging the most characteristic phrases per cluster (i.e., c-TF-IDF).[14] We find partitioning into 40 clusters to be optimal for further analysis because it consists of technologies commonly found in existing liter-

[13]While other models are available, we prefer the *all-mpnet-base-v2* model because of its high performance in Semantic Textual Similarity (STS) benchmarks (see https://huggingface.co/spaces/mteb/leaderboard), computational efficiency, open-source availability, and convenience of use via SentenceTransformers library. Specialization in text similarity tasks is achieved due to the contrastive loss function used in training: given sentence pairs or triplets, the model adjusts its weights to produce document embeddings that are closer to each other when documents are labeled as similar and are far apart when documents are irrelevant. The training data comprises 1.17 billion sentence pairs or triplets from a wide range of sources: WikiAnswers, Reddit, Stack Exchange, Semantic Scholar academic papers, etc.

[14]A term frequency inverted document frequency (TF-IDF) measure that was modified to compute the most relevant terms based on clusters of documents rather than individual documents. TF-IDF identifies the terms that are most representative of a text corpus because they appear frequently, and are also specific to a subset of documents (or clusters).

ature on digital technologies and automation (Acemoglu and Restrepo 2019, Acemoglu et al. 2022a, Foster-McGregor et al. 2019, Martinelli et al. 2021, Maslej et al. 2024, Ménière et al. 2017, Zolas et al. 2020, [CITE MORE]).

Table 1 lists our set of emerging digital technologies grouped by technology families. We provide a short description for each technology in Tables A.1 to A.3 in Appendix A.2. The grouping of these 40 technologies into 9 families is based on the correlation between the technologies' co-occurrence in occupations (more in the next section). Thus, a family comprises technologies whose occupation structure of semantic links is highly correlated; see Appendix A.6 for a more detailed discussion. Figure A.2, in the appendix, presents the distribution of patents across emerging digital technologies.

3 Semantic–based Exposure

In this section, we present the methodology for computing the exposure scores of industries and occupations to emerging digital technologies. First, we compute the cosine similarity scores of industries and occupations with patents, using textual data and filtering for relevant pairs. Then, we aggregate these similarity scores from the patent to the technology level to obtain the semantic-based exposure scores.

Exposure scores denote the *relevance* of each technology to a given industry or occupation, which we consider later as a proxy for adoption. For industries, the relevance is determined by whether a technology is integrated into the production process or if the technology itself constitutes an enhanced output of an industry. Regarding occupations, the relevance pertains to the significance of technology in the execution of tasks and functions inherent to an occupation.

3.1 Industry Cosine Similarity Scores

Industry Descriptions. We select the 3-digit NACE Rev.2 classification as the most detailed level at which to consider industries' descriptions. This selection is based on two primary considerations. First, this allows us to incorporate titles and descriptions from the 4-digit into the 3-digit industry descriptions—providing a more extensive text corpus for matching. Second, industry subsets under the same 3-digit category do not exhibit substantial differences in their connections to patents, allowing for a merger without significant loss of information.

In Section 2, we described the structure of the patent title and argued that mirroring this structure in the textual representation of industries (and occupations) facilitates matching. Thus, for each industry $i \in \mathcal{I}$, we break the industrial descriptions (both the 3-digit and their

Table 1: List of Emerging Digital Technologies

	Family		Emerging Digital Technology
F1	3D Printing	01	3D Printer Hardware
		02	3D Printing
		03	Additive Manufacturing
F2	Embedded Systems	04	Smart Agriculture & Water Management
		05	Internet of Things (IoT)
		06	Predictive Energy Management and Distribution
		07	Industrial Automation & Robot Control
		08	Remote Monitoring & Control Systems
		09	Smart Home & Intelligent Household Control
F3	Smart Mobility	10	Intelligent Logistics
		11	Autonomous Vehicles & UAVs
		12	Parking and Vehicle Space Management
		13	Vehicle Telematics & Electric Vehicle Management
		14	Passenger Transportation
F4	Food Services	15	Food Ordering & Vending Systems
F5	E-Commerce	16	Digital Advertising
		17	Electronic Trading and Auctions
		18	Online Shopping Platforms
		19	E-Coupons & Promotion Management
F6	Payment Systems	20	Electronic Payments & Financial Transactions
		21	Mobile Payments
		22	Gaming & Wagering Systems
F7	Digital Services	23	Digital Authentication
		24	E-Learning
		25	Location-Based Services & Tracking
		26	Voice Communication
		27	Electronic Messaging
		28	Workflow Management
		29	Cloud Storage & Data Security
		30	Information Processing
		31	Cloud Computing
		32	Recommender Systems
		33	Social Networking & Media Platforms
		34	Digital Media Content
F8	Computer Vision	35	Augmented and Virtual Reality (AR/VR)
		36	Machine Learning & Neural Networks
		37	Medical Imaging & Image Processing
F9	HealthTech	38	Health Monitoring
		39	Medical Information
		40	E-Healthcare

Notes: This table lists the 40 emerging digital technologies along with their respective emerging technology families. Emerging digital technologies are obtained by clustering the embeddings using the k–means algorithm, where the embeddings are derived with the sentence transformer all-mpnet-base-v2. For a short description of these technologies, refer to Tables A.1 to A.3 in Appendix A.2. Technologies are grouped by families, where a family comprises technologies whose occupation structure of semantic links is highly correlated.

nested 4-digit children) into individual sentences and concatenate each sentence with its corresponding title. We represent these composite sentences as $s \in S_i \subset \mathcal{S}_{\mathcal{I}}$, where S_i denotes the set of composite sentences (i.e., title combined with one description sentence) corresponding to industry i. This results in 271 industries at the 3-digit level, each represented by an average of 11 composite sentences.

Embeddings. We produce the embeddings of these composite sentences using the same pre-trained sentence transformer as in Section 2, namely all-mpnet-base-v2. The embedding of a composite sentence s for an industry i is denoted as $Emb_{s,i}$.

Cosine Similarity. For each patent $p \in \mathcal{P}$, we compute the cosine similarity of all composite sentences $s \in \mathcal{S}_{\mathcal{I}}$ with both parts of the patent titles, namely p_1 (representing the invention's description) and p_2 (representing its function). Specifically, the cosine similarities are computed as:

$$C_{s,i}^{p_1} = \frac{Emb_{s,i} \cdot Emb_{p_1}}{||Emb_{s,i}||\ ||Emb_{p_1}||}, \tag{1}$$

$$C_{s,i}^{p_2} = \frac{Emb_{s,i} \cdot Emb_{p_2}}{||Emb_{s,i}||\ ||Emb_{p_2}||}, \tag{2}$$

which quantify the semantic relationship between p_1, respectively p_2, and s. Nevertheless, similarity can be discerned through different nuances of meaning. In our context, this could pertain to aspects such as an application, a technical domain, or specified functions, whether central or ancillary. This data is encapsulated into a scalar, whose magnitude *approximates* the degree of similarity between an aspect of the industry (as described in its NACE 4-digit nomenclature) and an aspect of the invention (as described in the patent).

To reduce the noise and capture the most relevant meaning of the similarity between an invention and an industry, for each (i, p_1) and (i, p_2) combinations, we retain the composite sentence s that exhibits the highest cosine similarity score. Formally,

$$C_i^{p_1} := \arg\max_{s \in S_i} C_{s,i}^{p_1}, \tag{3}$$

$$C_i^{p_2} := \arg\max_{s \in S_i} C_{s,i}^{p_2}, \tag{4}$$

where $C_{s,i}^{p_1}$ and $C_{s,i}^{p_2}$ are, respectively, given by Equations (1) and (2). These scalars summarize the quality of the semantic match between an industry i and the description (p_1) or the function (p_2) of the patent.

Redundancy. To enhance the quality of the matching and filter out irrelevant matches, we incorporate *redundancy* in the calculation of cosine similarity of industry–patent pairs (i, p). For industry–patent combinations (i, p), we separately rank the sub-pairs (i, p_1) and (i, p_2) based on their respective cosine similarity scores $C_i^{p_1}$ and $C_i^{p_2}$. We then identify the industry-patent combinations (i, p) as relevant (denoted as $(i, p)^\star$) if *both* sub-pairs (i, p_1) and (i, p_2) are within the top 10 of their respective rankings. This methodology results in the exclusion of certain pairs that do not rank simultaneously in the top 10 for both components.[15] Thus, we retain inventions for which both the description of the invention and its function are relevant to the industry.

For the identified relevant pairs, we calculate the harmonic mean with both cosine similarity scores, for the description of the invention and its function. This yields the composite cosine similarity score for industry–patent pairs $(i, p)^\star$ as follows:

$$C_i^p = 2\left(\frac{1}{C_i^{p_1}} + \frac{1}{C_i^{p_2}}\right)^{-1}, \tag{5}$$

where $C_i^{p_1}$ and $C_i^{p_2}$ are, respectively, given by Equations (3) and (4). As a result of the calculation presented in Equation (5), we establish a connection between an invention identified in a single patent $p \in \mathcal{P}$ and a set of relevant industries in which that patent can be used to improve the process product or organization.

Table 2 illustrates the redundancy principle at work, considering the first patent example provided in Section 2. This example describes a targeted TV advertising method based on user profile information. For this patent, redundancy helps filter out industries irrelevant to the innovation. The redundancy filtering for the other two patent examples (mentioned in Section 2) is presented in Tables A.4 and A.5 in the appendix.

3.2 Occupation Cosine Similarity Scores

Occupation Descriptions. We choose the 4-digit ISCO-08 as the most detailed level at which to consider the textual description of occupations. Unlike industries, the 4-digit level comprises a set of distinct occupations that are informative for our analysis. Each ISCO-08 occupation is associated with a specific set of tasks, although some tasks may overlap across different occupations.

For each occupation $o \in \mathcal{O}$, we consider two components of the occupation description: the occupation title o_1 and the task description o_2. We divide the task description into indi-

[15] In addition, we manually exclude three very specific connections to improve our exposure scores; see Appendix A.3 for more details.

Table 2: Example of Redundancy Filtering of Industries for Targeted TV Advertising

Code	NACE Industry	Cosine Similarity		
		$C_i^{p_1}$	$C_i^{p_2}$	C_i^{p}
60.2	Television programming and broadcasting activities	0.391	0.445	0.416
73.1	Advertising	0.458	0.373	0.411
73.2	Market research and public opinion polling	0.295	0.272	0.283
59.1	Motion picture, video and television programme activities	0.271	0.263	0.267
61.2	Wireless telecommunications activities	0.290	0.229	0.256
26.3	Manufacture of communication equipment	0.257	0.240	0.249
78.1	Activities of employment placement agencies	0.265		
47.9	Retail trade not in stores, stalls or markets	0.263		
56.3	Beverage serving activities	0.261		
80.1	Private security activities	0.253		
61.3	Satellite telecommunications activities		0.294	
61.1	Wired telecommunications activities		0.237	
97.0	Activities of households as employers of domestic personnel		0.231	
58.1	Publishing of books, periodicals and other publishing activities		0.223	

Notes: This table presents the redundancy filtering of industries for the Patent ID 2013B87254. It displays the cosine similarity of distinct 3-digit NACE Rev.2 industry descriptions with the patent description "Method for targeting television advertisement based on profile linked to online device" (Column 3) and the function principle "selecting television advertisement to be directed to set-top box based on profile information pertaining to the user or online activity" (Column 4). Industries are ranked according to Column 3 in decreasing order. Cosine similarity scores in Columns 3 and 4 are displayed only for sub-pairs belonging to their respective top 10. Column 5 shows the composite patent-industry cosine similarity score, which corresponds to the harmonic mean of Columns 3 and 4. Cosine similarity scores in Column 5 are displayed only for pairs that rank simultaneously in both top 10.

vidual tasks $s \in S_o \subset \mathcal{S}_{\mathcal{O}}$, where S_o is the set of tasks for occupation o. This results in 433 occupations at the 4-digit level, each represented by one title and an average of 7.5 tasks.

Embeddings. Similar to industries, we produce the embeddings using the same sentence transformer model. We represent the embeddings of the occupation title as Emb_{o_1} and the embeddings of a task s as Emb_{s,o_2}.

Cosine Similarity. For each patent $p \in \mathcal{P}$, we compute the cosine similarity of the patent title (in its entirety) with both parts that describe the occupations, namely, the title o_1 and all the tasks separately $o_{s,2}$. More specifically, we compute the cosine similarities as:

$$C^p_{o_1} = \frac{Emb_{o_1} \cdot Emb_p}{||Emb_{o_1}||\,||Emb_p||}, \tag{6}$$

$$C^p_{s,o_2} = \frac{Emb_{s,o_2} \cdot Emb_p}{||Emb_{s,o_2}||\,||Emb_p||}, \tag{7}$$

which express the idea of the semantic connection between o_1, respectively s, and p.

For each (o_2, p) combination, as above for industries, we retain the composite sentence with the highest cosine similarity score. More formally,

$$C^p_{o_2} := \underset{s \in S_o}{\arg\max}\, C^p_{s,o_2}, \tag{8}$$

where C^p_{s,o_2} is the cosine similarity between patent p and task s given by Equation (7). There is no need to aggregate in the case of the title part o_1 as each occupation has only one title. The quality of the semantic match between an occupation and a patent is summarised in both of these scalars, either through the title of the occupation or the tasks performed in that latter.

Redundancy. We employ the same methodology as with industries, designating the occupation–patent combinations (o, p) as relevant (denoted as $(o, p)^\star$) if *both* sub-combinations $(o, p)_1$ and $(o, p)_2$ rank within the top 10 of their respective rankings. Thus, we retain inventions that are relevant to the occupation.[16]

For the identified relevant pairs, we calculate the harmonic mean with both cosine similarity scores. This yields the composite cosine similarity score for industry–patent pairs $(o, p)^\star$ as follows:

$$C^p_o = 2\left(\frac{1}{C^p_{o_1}} + \frac{1}{C^p_{o_2}}\right)^{-1}, \tag{9}$$

where $C^p_{o_1}$ and $C^p_{o_2}$ are, respectively, given by Equations (6) and (8). As a result of the calculation presented in Equation (9), we establish a connection between an invention identified in a single patent $p \in \mathcal{P}$ and a set of occupations to which that patent is relevant. Tables A.6 to A.8, in the appendix, illustrate redundancy filtering of occupations for our patent examples from Section 2.

3.3 Aggregation by Technology

We aggregate cosine similarity scores C^p_i and C^p_o obtained at the patent level in Equations (5) and (9), to the technology level. To this end, we implement a weighting scheme based on the number of citations that a patent receives from other patents to proxy for the relevance of each patent and the likelihood that it is used in industries and occupations. Given the heterogeneity in patent impact, it is pertinent that their weighting reflects this (Hall et al. 2005, OECD 2009).

We assign a weight to the cosine similarity score of each relevant patent–industry/occupation pair. This weight given to the pair is proportional to the number of citations the patent has received relative to the total number of citations accrued by all relevant patents that are associ-

[16] Similar to industries, we manually exclude three very specific connections to improve our exposure scores; see Appendix A.3 for more details.

ated with the same occupation/industry and belong to the same technology within the same year.[17] The specific weight assigned to a relevant pair $(d,p)^\star$ is computed as:

$$\omega_d^p = \frac{m_p}{\sum_{p\in\mathcal{P}_{dt}^k} m_p}, \tag{10}$$

where m_p is the number of citations received by patent p, $\mathcal{P}_{dt}^k$ represents the set of patents associated with emerging digital technology k, filed in year t, and relevant to industry/occupation $d = \{i,o\}$.

We implement this weighting scheme to aggregate the cosine similarity scores at the patent level to the technology level. The cosine similarity of a technology k to an industry/occupation is then computed as:

$$C_{dt}^k = |\mathcal{P}_{dt}^k| \times \sum_{p\in\mathcal{P}_{dt}^k} \omega_d^p C_d^p, \tag{11}$$

where C_d^p denotes the cosine similarity score of the pair (d,p) as derived from Equations (5) and (9), ω_d^p represents the weight from Equation (10), and $|\mathcal{P}_{dt}^k|$ is the total number of patents assigned to industry/occupation–technology pair (d,k) for $d = \{i,o\}$ in year t. This results in the cosine similarity score of industry/occupation i/o with technology k for the year t.[18]

Lastly, we aggregate cosine similarity scores across all years to obtain a cumulative measure for the period 2012–2021. The equation for this aggregation is as follows:

$$C_d^k = \sum_t C_{dt}^k, \text{ with } d = \{i,o\}, \tag{12}$$

where C_{dt}^k is given by Equation (11).

3.4 Exposure Scores

To obtain our final measure of the exposure of 3-digit NACE Rev.2 industries and 4-digit ISCO-08 occupations to emerging digital technologies X_d^k, we apply inverse hyperbolic sine transformation, which helps address the right skewness in the distribution of cosine similarity scores.

[17] Approximately 41% of patents in our sample have not received any citations. This includes 1,733 patents, or 0.91%, which had an indeterminable citation count and are treated as having zero citations. Similarly, there are 77,307 patents, or 40.54%, patents with no citations. Figure A.4 in the appendix shows the distribution of patents of undetermined-count and non-cited patents across technologies. Figure A.3 in the appendix shows the distribution of patent citations across technologies.

[18] Note that aggregating without weighting by citations results in yearly cosine similarity scores very similar to those obtained with the weighting scheme. Figure A.6, in the appendix, displays the correlation between the weighted and unweighted yearly cosine similarity scores. The Pearson correlations between scores derived from both methods are approximately 0.99 for both industries and occupations. The Spearman rank correlation yields a value of about 0.89.

Formally,

$$X_d^k = \sinh^{-1}\left(C_d^k\right), \tag{13}$$

where C_d^k is the cosine similarity score for industry/occupation–technology pair (d,k) between 2012 and 2021 as described in Equation (12).

Our exposure scores necessitate two clarifications regarding their interpretation. Firstly, although related, exposure scores indicate the (contextual) relevance of each technology to a given industry or occupation, which is a proxy for their actual adoption. Secondly, they are neutral regarding the nature of the relationship between a technology and an industry or occupation. This means they do not specify whether the technology and industry or occupation are complementary or substitutes in producing output. This is a deliberate decision in constructing our measure to avoid making ex-ante assumptions about the relationship.

We deliver these data as an open–access database, the 'TechXposure' database. In this database, we also provide measures of exposure at higher levels of aggregation, such as the 1-digit and 2-digit levels for industries, and from the 1-digit to the 3-digit levels for occupations. For details on the derivation of these measures, see Appendix A.7.

Our exposure scores align with existing metrics in the literature. However, our data encompass additional dimensions of these technologies that prior work did not capture, either due to the nonexistence of these technological features or the overly specific focus of their approaches. For example, the AI exposure scores in Webb (2019) are limited to core dimensions of AI (i.e., industrial automation, workflow management systems, cloud computing, and machine learning). In contrast, the AI exposure scores in Felten et al. (2021) cover a broader scope but only address *intangible* AI applications. Thus, they do not account for AI embedded in *tangible* technologies such as industrial and mobile robots, and IoT. For details on the methodology and comparisons, see Appendix A.8.

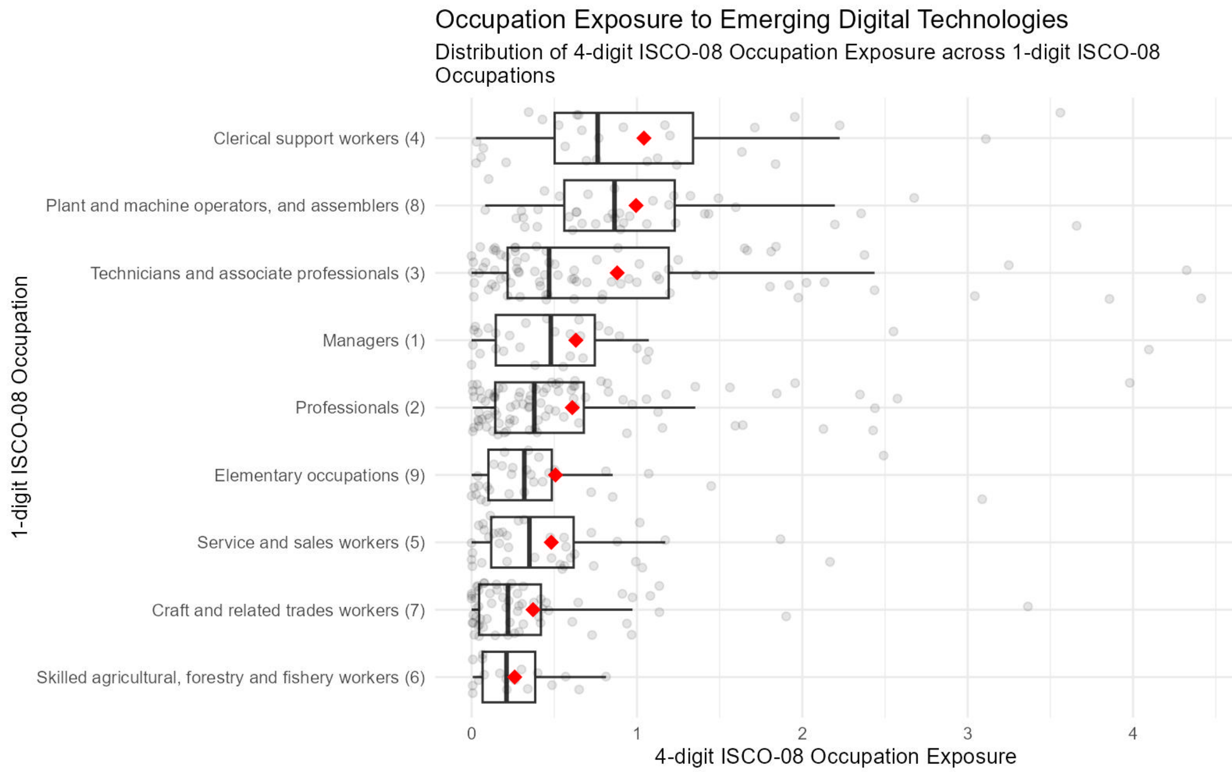
4 Descriptive Analysis

In this section, we describe the exposure of both occupations and industries to emerging digital technologies. We start with occupations and then look at industries.

4.1 Occupation Exposure to Emerging Digital Technologies

We start by examining the overall exposure of occupations, which we define as the average exposure across all technologies. This corresponds to $X_o = \frac{1}{40}\sum_k X_o^k$, where X_o^k is defined by Equation (13). Figure 1 presents the distribution of overall exposure to emerging digital technologies across ISCO-08 occupations. In this figure, 4-digit occupations are grouped into

Figure 1: Overall Occupation Exposure by 1-digit ISCO-08 Occupation



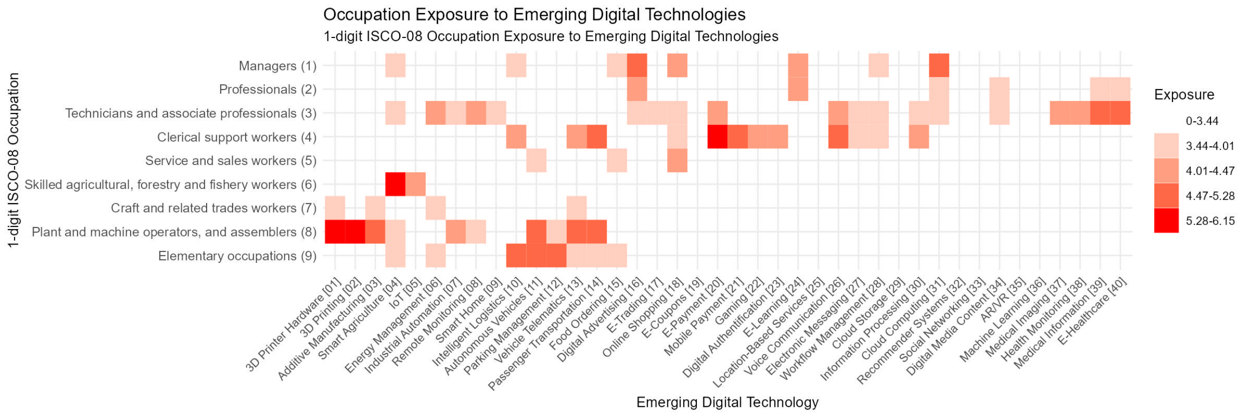
Notes: This figure presents the distribution of exposure to emerging digital technologies across 4-digit ISCO-08 occupations, with each 1-digit occupation displayed separately in boxplots. Vertical bars indicate the median exposure for all 4-digit occupations within the same 1-digit occupation, and diamond points represent the average exposure for these 4-digit occupations.

their respective 1-digit categories, and their distribution is presented as a boxplot. Occupation groups are ranked by their average overall exposure to emerging digital technologies, indicated by the diamond point.

We observe that Clerical Support Workers (ISCO-08 Group 4) and Plant and Machine Operators, and Assemblers (8) are the most exposed to emerging digital technologies. The occupations in these ISCO groups typically involve a higher proportion of routine tasks associated with information handling and production equipment supervision, respectively. Despite having already experienced a significant impact from earlier waves of ICT development (Goos and Manning 2007, Goos et al. 2009, Goos et al. 2014), these middle-paying jobs continue to be strongly related to newer ICT vintages, especially emerging digital technologies that enable handling of information and production equipment in semi- or unsupervised manner.

High-paying occupations, including Managers (1), Professionals (2), and Technicians and Associate Professionals (3), are the next most exposed to the emerging digital technologies. The tasks performed in these occupations are predominantly non-routine and cognitive, often involving the use of a variety of digital technologies. As technologies advance and new

Figure 2: Occupation Exposure by Emerging Digital Technologies (1-digit ISCO-08)



Notes: Each cell shows the exposure of a 1-digit ISCO-08 occupation (row) to a given emerging digital technology (column). Exposure scores below the 80th percentile (0-3.44) are transparent, whereas the four other groups represent respectively the 80th (3.44-4.01), 90th (4.01-4.47), 95th (4.47-5.28), and 99th (5.28-6.15) percentile of the distribution. Figure B.1, in the appendix, presents the same figure at the 2-digit level.

vintages appear, new tasks may also emerge, leading to changes in the task structure of these occupations.

Conversely, we observe that low-paying occupations, such as Service and Sales Workers (5), Skilled Agricultural, Forestry and Fishery Workers (6), Craft and Related Trades Workers (7), and Elementary Occupations (9), are less exposed to emerging digital technologies. These occupations involve more interactive and non-routine tasks, which are less reliant on these technologies.

Lastly, we observe greater heterogeneity in exposure to emerging digital technologies within high-paying occupations (1, 2, and 3) compared to middling occupations (4 and 8). This suggests that only a subset of the former group is related to emerging technologies, while the latter group exhibits more generalized exposure.

We break down the overall exposure of 1-digit ISCO Groups by examining their exposure to each of the 40 emerging digital technologies. Figure 2 presents 1-digit occupation exposure as a heatmap, where the exposure levels are indicated at the intersections of 1-digit occupations (rows) and emerging digital technologies (columns). This visualization reveals two distinct patterns.

First, we observe a distinct divide between *tangible* and *intangible* technologies in terms of their relevance to different occupations. On the one hand, *tangible* technology families, such as 3D Printing, Embedded Systems, and Smart Mobility, are more relevant to manual occupations within ISCO Groups 6 to 9. On the other hand, *intangible* technology families, such as E-Commerce, Payment Systems, Digital Services, Computer Vision, and HealthTech, are more relevant to cognitive occupations, specifically within ISCO Groups 1 to 4.

Second, we note that both Technicians and Associate Professionals (3) and Clerical Support Workers (4) exhibit exposure to a wide range of emerging digital technologies. In contrast, Managers (1) and Professionals (2) appear to have a more limited scope of relevant technologies, primarily concentrated in the realm of intangible technologies. Similarly, exposure of ISCO Groups 6 to 9 is exclusively focused on tangible technologies. It is important to note that this aggregated mapping conceals some heterogeneity in exposure within 1-digit ISCO-08 occupations due to aggregation; see Figure B.1 in the appendix for a more detailed mapping at the 2-digit level.

Due to the granularity of the exposure data, we can track the link with a technology down to the individual task of a 4-digit ISCO occupation. We leverage this information and find the most exposed tasks to the emerging digital technologies, both pooled and individually.

Figure A.5 presents the most exposed tasks by 1-digit ISCO-08 group, providing insights about exact functions responsible for the link between occupations and the emerging technologies.

The most exposed tasks to the individual emerging technologies can be found in Appendix ref to ref.

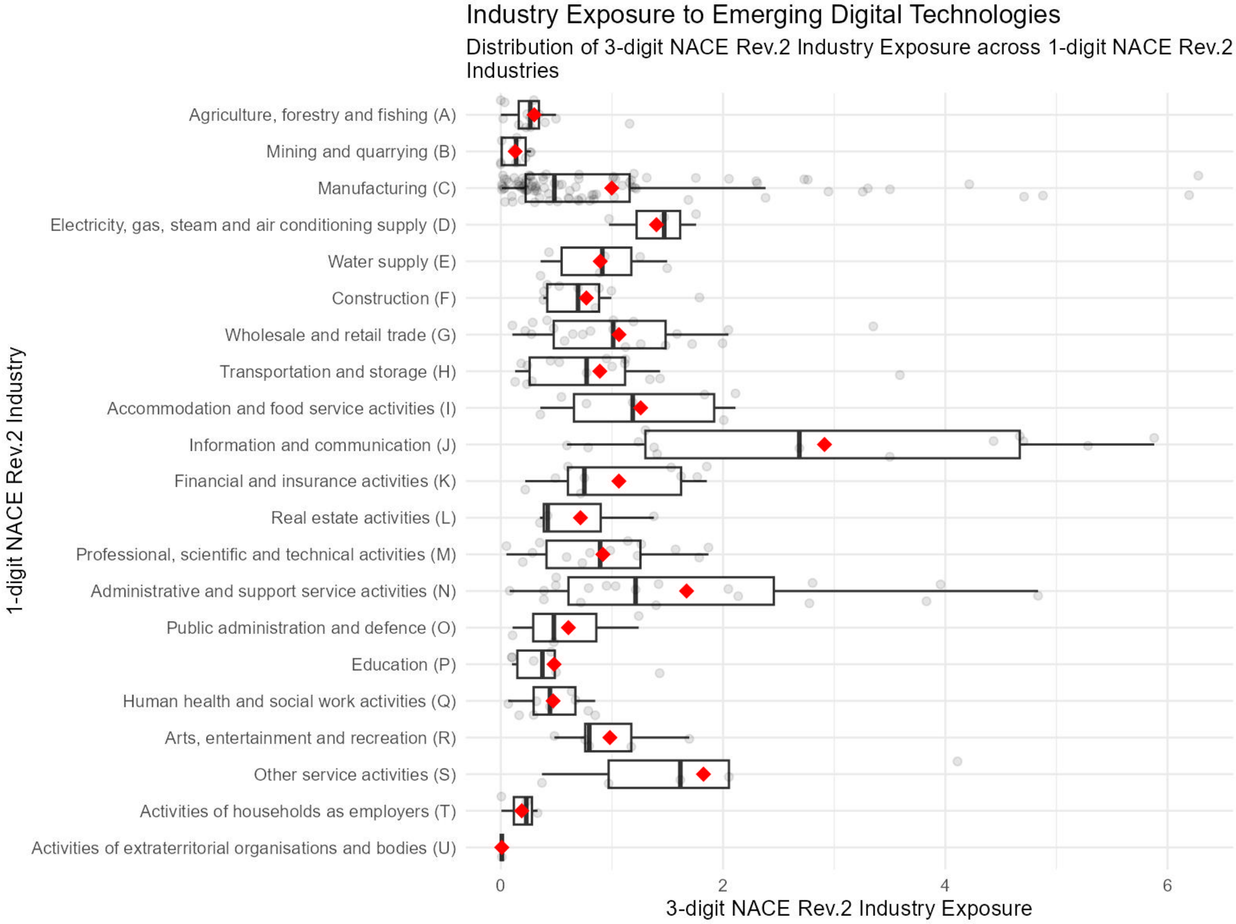
4.2 Industry Exposure to Emerging Digital Technologies

For industries, we examine their overall exposure, which we define as the average exposure across all technologies. This corresponds to $X_i = \frac{1}{40}\sum_k X_i^k$, where X_i^k is defined by Equation (13). Figure 3 presents the distribution of overall exposure to emerging digital technologies across NACE Rev.2 industries. In this figure, 3-digit industries are grouped into their respective 1-digit sectors, and their distribution is presented as a boxplot.

We observe that the Information and Communication (J) and Manufacturing (C) sectors host the most exposed 3-digit industries. This finding is notable due to the significant heterogeneity of industry exposure within these 1-digit industries. Such differences in exposure may indicate the industries' roles as either producers or intensive users, as opposed to light users, of emerging digital technologies. More specifically, industries within the Information and Communication (J) sector are likely to produce intangible technologies, while a specific subset of the Manufacturing (C) sector is likely to produce tangible technologies.

The Administrative and Support Service Activities (N) sector also exhibits a high average level of exposure to emerging digital technologies. Several 3-digit industries within this sector achieve overall exposure levels comparable to those in Sectors C and J. This observation is consistent with the findings presented in Section 4.1, as Sector N is a significant employer of Clerical Support Workers (ISCO Group 4), identified as the most exposed 1-digit ISCO Group (see Fig. 1).

Figure 3: Overall Industry Exposure by 1-digit NACE Rev.2 Industry

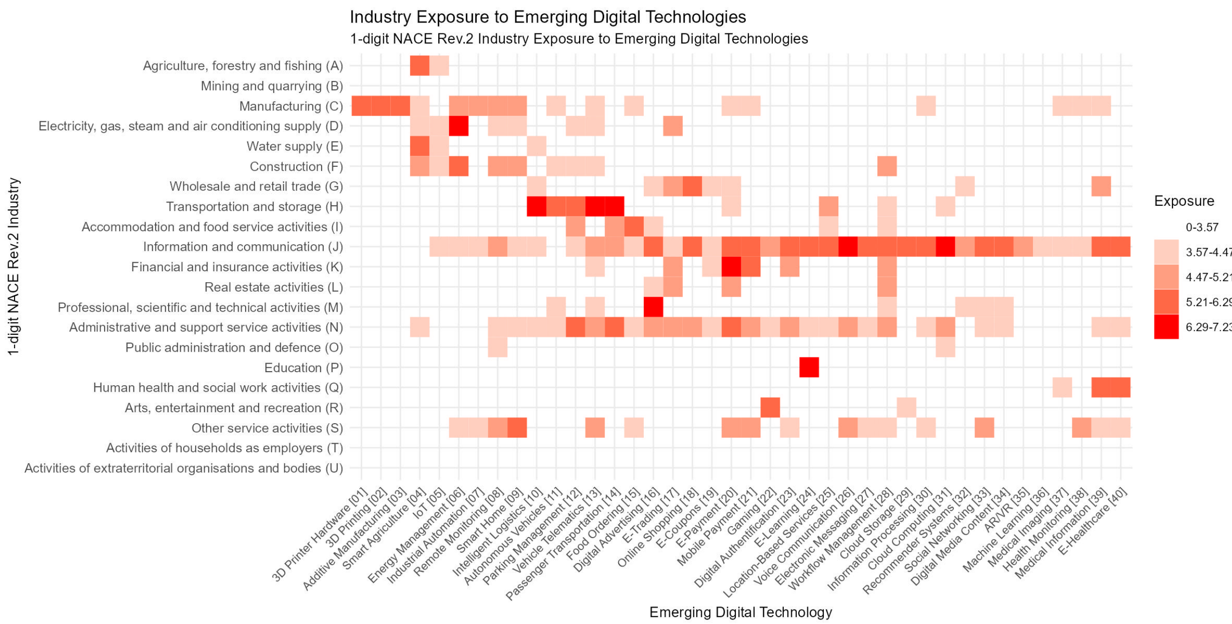


Notes: This figure presents the distribution of exposure to emerging digital technologies across 3-digit NACE Rev.2 industries, with each 1-digit industry displayed separately in boxplots. Vertical bars indicate the median exposure for all 3-digit industries within the same 1-digit industry, and diamond points represent the average exposure for these 4-digit industries.

We disaggregate the overall exposure of 1-digit NACE sectors into their exposure to each of the 40 emerging digital technologies. Figure 4 replicates the exposure heatmap for 1-digit sectors; see Figure B.2 in the appendix for a more detailed mapping at the 2-digit level.

Similar to occupations, we observe a divide between tangible and intangible emerging digital technologies. In the figure, exposure cells follow a top-left to bottom-right diagonal pattern, thereby associating tangible technologies with sectors like Agriculture (A), Mining and Quarrying (B), and Manufacturing (C), and aligning intangible technologies with service sectors from Financial and Insurance Activities (K) to Other Service Activities (S). In between these extremes, we find sectors ranging from Electricity, Gas and Air Conditioning Supply (D) to Information and Communication (J) operate physical infrastructures and are thus exposed to more tangible but distributed technology families, such as Embedded Systems and Smart Mobility.

Figure 4: Industry Exposure by Emerging Digital Technologies (1-digit NACE Rev.2)



Notes: Each cell shows the exposure of a 1-digit NACE Rev.2 industry (row) to a given emerging digital technology (column). Exposure scores below the 80th percentile (0-3.57) are transparent, whereas the four other groups represent respectively the 80th (3.57-4.47), 90th (4.47-5.21), 95th (5.21-6.29), and 99th (6.29-7.23) percentile of the distribution. Figure B.2, in the appendix, presents the same figure at the 2-digit level.

5 Impact on Employment

In this section, we estimate the causal effect of emerging digital technologies on regional employment using an instrumental variable (IV) shift-share approach.

5.1 Overall Impact of Emerging Digital Technologies

We use employment data from the Regional European Labour Force Survey (EU-LFS), which provides information on the number of employees and population across several demographic groups.[19] Our sample comprises 320 NUTS-2 regions in 32 European countries.[20]

We consider the change in the regional employment-to-population ratio between 2012 and 2019 as our outcome variable. The regional employment-to-population ratio is defined by the number of employees in the group of interest (e.g., the total population of young age) as the numerator and the total number of individuals aged 15 or older as the denominator.

Our analysis uses a long-difference approach between 2012 and 2019. We begin in 2012,

[19]These demographic groups include male, female, young (aged 15 to 24 years), mature (aged 25 to 64 years), and low-, middle-, and high-skilled workers, defined by educational level (i.e., primary, secondary, and tertiary).

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

which marks the starting year for our patent sample and thus serves as our measure of exposure to emerging digital technologies. We conclude our analysis in 2019 to avoid confounding factors related to employment and population fluctuations caused by the COVID-19 pandemic.[21]

Additionally, the EU-LFS provides information on the number of employees across 1-digit NACE industries, which are grouped into 10 distinct sectors.[22]

Estimating the causal impact of technology on employment presents two main challenges: reverse causality and omitted variable bias. Reverse causality suggests that technological advancements may also result from labor shortages or rising labor costs. Additionally, unobserved factors, such as changes in the organization of industries or investments in infrastructures, could simultaneously affect both technological change and employment levels.

To address these concerns, we adopt a shift-share strategy, leveraging recent advancements in this literature (Adão et al. 2019; Goldsmith-Pinkham et al. 2020; Borusyak et al. 2021). Specifically, we employ the Bartik instrument to measure the exposure of a region X_r as follows:

$$X_r = \sum_j l_{rj} X_j, \tag{14}$$

where l_{rj} is the employment share of sector j in region r in the baseline year 2010,[23] and X_j is defined as the average exposure of sector j to emerging digital technologies from 2012 to 2019, calculated as

$$X_j \equiv \frac{1}{40} \times \sum_{k \in \mathcal{K}} X_j^k,$$

where X_j^k represents the average exposure of sector j to each technology k across all 1-digit NACE industries $i \in j$ during this period.

We argue that the sectoral exposure to emerging digital technologies, X_j, which represents the *shock* in our shift-share design, is quasi-exogenous to changes in regional employment within Europe. Our metrics for industrial exposure, as derived in Section 3, are based on the semantic similarity between patents and industry descriptions. However, only 7.1% of the patents in our sample originate from Europe, suggesting that the advancement of these tech-

[21] Although our exposure metrics from Section 3 cover the period 2012–2021, we recompute them for the subperiod 2012–2019 to ensure consistency with our analysis timeframe in this section.

[22] These sectors are Agriculture (A); Industry (B-E); Construction (F); Market Services (G-I); Information and Communication (J); Financial and Insurance Activities (K); Real Estate Activities (L); Professional, Scientific, Technical, Administration, and Support Service Activities (M-N); Public Administration, Defence, Education, Human Health, and Social Work Activities (O-Q); and Other Services (R-U).

[23] Table C.1 in the Appendix provides details on the average employment share by economic sector across European regions in 2010. The three largest sectors are the Public Sector (with an average employment share of 25.7%), Market Services (24%), and Industry (17.1%). The Information and Communication sector, which is highly exposed to emerging digital technologies, accounts for only 2.3% of employment on average.

nologies is predominantly a global phenomenon. Consequently, global technological trends are unlikely to be influenced solely by regional labor markets in Europe. To reinforce this point, we recompute our exposure measure excluding European patents.[24]

Since our shocks are (assumed to be) exogenous to local employment change in European labor markets, we use the equivalence proposed by Borusyak et al. (2021) and can therefore consider our shift–share as a valid instrument. In addition to the quasi-random assignment of shocks, our second identifying assumption holds that regions more exposed to emerging digital technologies are not disproportionately affected by other labor market shocks or trends, and that the number of observed shocks is sufficiently large.[25]

Figure 5 shows the geographic distribution of exposure across European regions. Emerging digital technologies are more prevalent in industries concentrated in European capital cities, which typically have larger service sectors compared to more peripheral regions. Beyond capital cities, regions with the highest levels of exposure levels are predominantly found in Western Europe, specifically in countries such as Germany, Italy, Spain, Switzerland, and the UK.

Figure 6 depicts a positive relationship between the change in the employment-to-population ratio from 2012 to 2019 and the regional exposure to emerging digital technologies.[26] Although the observed correlation is statistically significant, it is not adjusted for country fixed effects and regional demographic characteristics.

We estimate the impact of regional exposure to emerging digital technologies on the regional employment-to-population ratio change using the following empirical specification:

$$\Delta Y_r = \alpha + \beta X_r + Z\delta + \phi_{c(r)} + u_r, \tag{15}$$

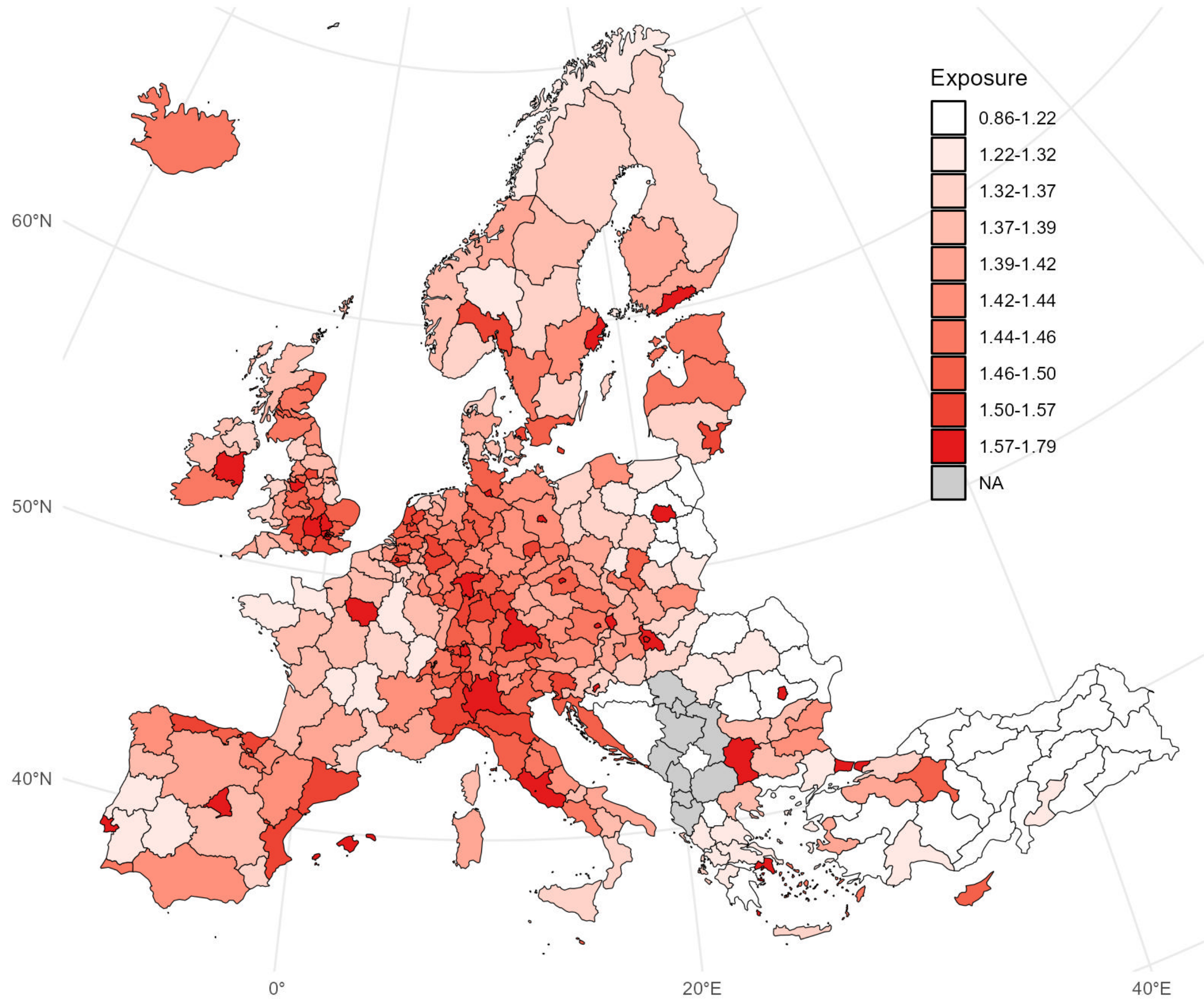
where ΔY_r represents the change in the employment-to-population ratio (in pp.) for region r between 2012 and 2019, X_r denotes the regional exposure to emerging digital technologies as defined in Equation (14) and standardized, Z is a set of covariates which capture regional

[24]In the Online Appendix, we compare the 1-digit industry exposure scores with and without European patents (i.e., patents filed in the European Patent Office). The correlation is approximately 0.99 for all 40 emerging digital technologies, suggesting that these technologies are predominantly a global phenomenon.

[25]The Herfindahl index (HHI) of average shock exposure is calculated as $\sum_j l_j^2 = 0.168$, where l_j represents the average employment share in sector j in 2010 across all regions, as detailed in Table C.1. This HHI can be considered relatively small since the lowest index we could obtain with a uniform distribution is $1/|J| = 0.1$, suggesting that the latter part of the assumption is realistic. The effective sample size, which corresponds to the inverse of the HHI, is 5.95.

[26]In the Online Appendix, we show that this positive association persists even after excluding regions with exceptionally low exposure levels — specifically, those with exposure below -2 standard deviations (i.e. below 1.149), which typically includes rural areas in Romania, Turkey, and overseas French territories.

Figure 5: Geographic Distribution of Regional Exposure to Emerging Digital Technologies across Europe from 2012 to 2019



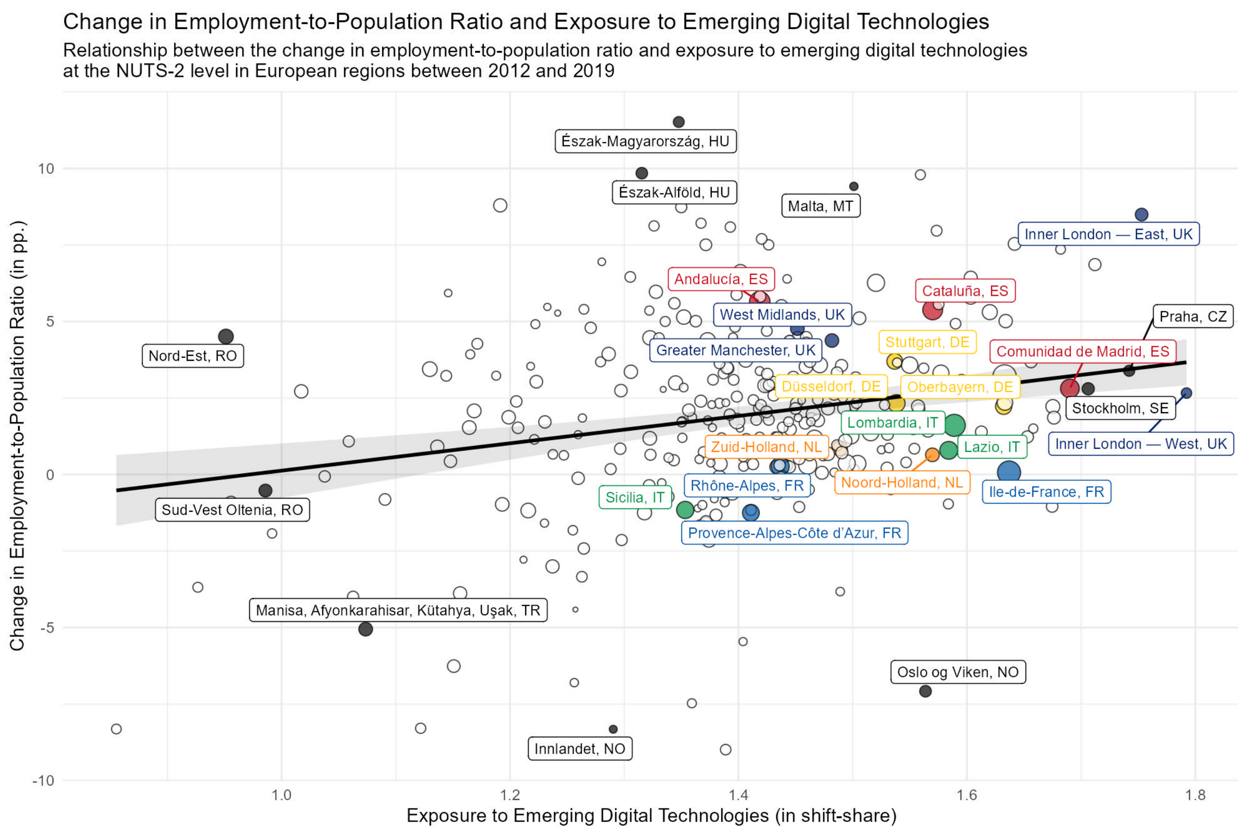
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

characteristics,[27] $\phi_{c(r)}$ are country fixed effects, and u_r is the error term.

Table 3 presents the estimates of the effect of regional exposure to emerging digital technologies on the change in the employment-to-population ratio (2012–2019). As the exposure is standardized across regions, the estimated coefficient of interest $\hat{\beta}$ can be interpreted as

[27]Our set of control variables, fixed at their 2010 values to avoid endogeneity, includes the log of population (in thousands), the proportion of females, the proportion of the population aged over 65, the proportion of the population with secondary and tertiary education levels, and the proportion of employment in the industry sector.

Figure 6: Change in Employment-to-Population Ratio and Exposure to Emerging Digital Technologies



Notes: This figure shows the relationship between the change in the employment-to-population ratio and the exposure to emerging digital technologies in European NUTS-2 regions between 2012 and 2019. Each point represents a region. The size of the point is proportional to the population in 2010. The horizontal axis measures the exposure to emerging technologies calculated by the shift-share method, while the vertical axis represents the change in the employment-to-population ratio in percentage points (pp.). The solid line indicates a positive correlation between regional exposure to emerging technologies and employment growth. The grey shaded area indicates the 95% confidence interval.

the effect of a one-standard-deviation increase in regional exposure on the employment-to-population ratio, expressed in percentage points. Following the recent literature on shift–share designs, we control for the sum of exposure shares (Borusyak et al. 2021), and report the AKM0 shift–share standard errors which account for arbitrary cross-regional correlation in the regression residuals (Adão et al. 2019).

The positive relationship observed in Figure 6 remains robust upon including fixed effects, and various covariates, such as demographic characteristics of the region and the industry share. In the specification encompassing all covariates, in the last column, a one-standard-deviation increase in regional exposure implies a 1.029 pp. change, equivalent to 2.05%, in the employment-to-population ratio from 2012 to 2019.

The latter estimation indicates that the overall impact of emerging digital technologies on employment is positive at the regional level. However, it remains to be determined whether

Table 3: Effect of Emerging Digital Technologies on Regional Employment

	Δ Emp-to-pop. ratio (2012-2019) $\times$ 100		
	(1)	(2)	(3)
Exposure to Emerging Technologies	0.640^{**}	0.926^{***}	1.029^{***}
	(0.239)	(0.139)	(0.120)
Country FE	✓	✓	✓
Demographics		✓	✓
Industry share			✓
R^2	0.668	0.696	0.698
Adj. R^2	0.629	0.655	0.656
Num. obs.	320	320	320

Notes: This table presents the estimates of exposure to emerging digital technologies on regional employment. It presents the coefficients measuring the effect of regional exposure to emerging technologies, constructed as shift-shares and standardized, on changes in the employment-to-population ratio between 2012 and 2019 in European regions, expressed in percentage points. Regressions are weighted by population in 2010. Column (1) includes country fixed effects; Column (2) adds demographics controls in 2010, including the logarithm of population, the proportion of females, the proportion of the population aged over 65, and the proportions of the population with secondary and tertiary education levels; Column (3) adds the share of employment in the industry sector. All columns control for the sum of exposure shares. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses are derived following the AKM0 inference procedure from Adão et al. (2019).

this positive relationship between emerging digital technologies and employment is uniform across all demographic groups. Table 4 presents estimates of the same empirical specification with the full set of control variables for different demographic groups.

Emerging digital technologies have an overall positive impact on both female and male employment. A one-standard-deviation increase in regional exposure over the period generates a 0.673 pp. (equivalent to 3.03%) change in the employment-to-population ratio for the former group and a 0.355 pp. (1.27%) change for the latter group. Although the impact is twice as large for women, there is more heterogeneity across regions regarding this effect, as indicated by the much larger standard errors than those for men.

Both young (aged 15 to 24) and mature workers (aged 25 to 64) experience a positive impact from emerging digital technologies. The former demographic group experiences a 0.181 pp. change in the employment-to-population ratio, representing a 3.8% increase, whereas the latter group experiences a 0.849 pp. change, representing only a 1.87% increase. This is consistent with Adão et al. (2024) who show that labor market adjustments to technological innovations (i.e., technological transitions) tend to be driven by the gradual entry of younger generations.

Emerging digital technologies have a positive impact on employment only at the extremes of the skill distribution, specifically for low- and high-skilled workers, with respective changes of 0.715 pp. (6.01%) and 0.738 pp. (4.92%) in their employment-to-population ratios due to a one-standard-deviation increase in regional exposure. Conversely, an increase of similar magnitude in regional exposure results in a decline in the employment-to-population ratio for middle-skilled workers by 0.412 pp. (-1.78%). This differentiated effect indicates a pattern of job polarization still evident due to emerging digital technologies.

As a robustness check, we conduct a placebo test where we estimate the effect of regional exposure to emerging digital technologies from 2012 to 2019 on the change in the employment-to-population ratio in the pre-period, specifically, from 2002 to 2009. These estimates are presented in Table C.2 in the appendix. As expected, we observe null effects on the pre-period for all demographic groups, which reinforces the validity of our shift-share approach. The only notable exception is a positive and significant effect on the employment of high-skilled workers. We interpret this result as consistent, considering that regions more exposed to emerging digital technologies are likely those where the share of high-skilled workers has increased the most since these technologies are developed and produced by workers in this demographic group.

Table 4: Effect of Emerging Digital Technologies on Regional Employment by Demographic Groups

	Δ Emp-to-pop. ratio (2012-2019) × 100							
	All	Gender		Age		Skill		
	Total	Female	Male	Y15-24	Y25-64	Low	Mid	High
Exposure to Emerging Technologies	1.029***	0.673***	0.355***	0.181***	0.849***	0.715**	−0.412***	0.738***
	(0.120)	(0.105)	(0.027)	(0.026)	(0.103)	(0.215)	(0.043)	(0.124)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓	✓	✓
Industry share	✓	✓	✓	✓	✓	✓	✓	✓
Emp-to-pop. ratio in 2012	50.14	22.22	27.92	4.76	45.38	11.89	23.11	15.00
Change (in %)	2.05	3.03	1.27	3.80	1.87	6.01	−1.78	4.92
R^2	0.698	0.558	0.726	0.333	0.722	0.630	0.754	0.647
Adj. R^2	0.656	0.497	0.688	0.240	0.684	0.579	0.720	0.598
Num. obs.	320	320	320	320	320	320	320	320

Notes: This table presents the estimates of exposure to emerging digital technologies on regional employment by demographic groups. It presents the coefficients measuring the effect of regional exposure to emerging technologies, constructed as shift-shares and standardized, on changes in the employment-to-population ratio between 2012 and 2019 in European regions, expressed in percentage points, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Regressions are weighted by population in 2010. All columns include a control for the sum of exposure shares; country fixed effects; demographics controls in 2010, including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels; and the share of employment in the industry sector. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses are derived following the AKM0 inference procedure from Adão et al. (2019).

5.2 Disentangling the Individual Effects of Emerging Digital Technologies

To estimate the individual effects of regional exposure to each emerging digital technology on employment, we use the same shift-share strategy previously described, applying it independently to each technology (see Table 1 for the full list).[28] The regional exposure to a specific technology k is represented by:

$$X_r^k = \sum_j l_{rj} X_j^k,$$

where l_{rj} denotes the employment share of the sector j in region r, and X_j^k is the exposure of sector j to technology k.[29]

Estimating the individual effect of a single technology on labor is challenging because technologies can be complementary. When technologies are complementary, they tend to be implemented together. For example, recent literature on the employment impact of robots, a specific technology, always controls for the use of ICT to account for complementarities between the two technologies (Acemoglu and Restrepo 2020, Dauth et al. 2021, among others). Similarly, one could argue that a specific emerging digital technology may be complementary to other emerging digital technologies (e.g., Cloud Storage).

Moreover, the degree of complementarity may vary within the same technology family and with other emerging digital technologies. For instance, Cloud Storage is likely more complementary with technologies within Digital Services, such as Cloud Computing, rather than with those from other families, like 3D Printing or Payment Systems. Consequently, we propose an empirical approach that accounts for these complementarities to avoid bias in estimating the individual impact of a specific technology on employment.

We proceed to estimate separately the impact of regional exposure to each emerging digital technology on the regional employment-to-population ratio with the following empirical specification:

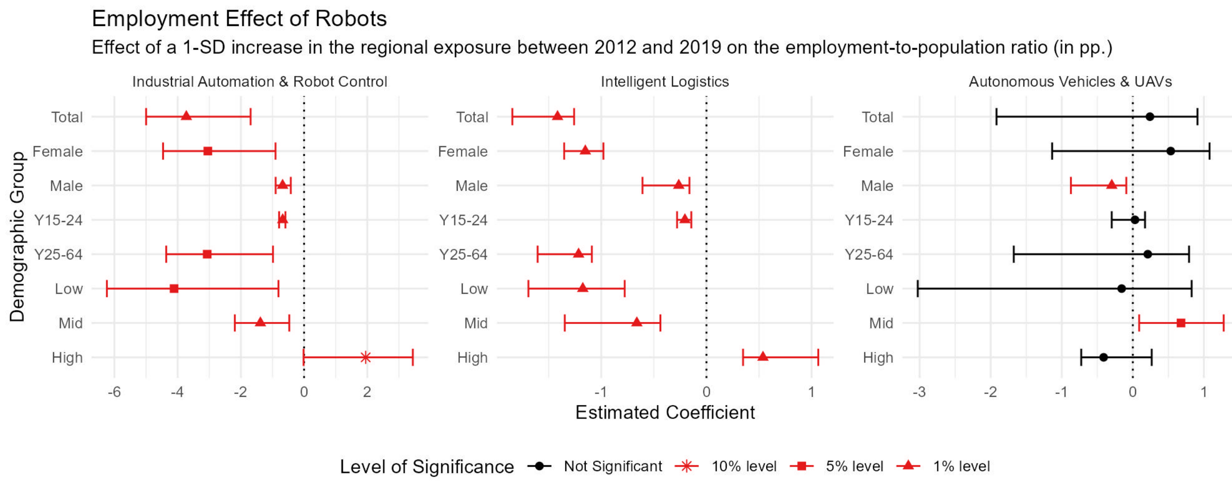
$$\Delta Y_r = \alpha + \beta_k X_r^k + \gamma_{1k} X_r^{K \setminus \{k\}} + \gamma_{2k} X_r^{-K} + Z\delta + \phi_{c(r)} + u_r, \tag{16}$$

where X_r^k is the regional exposure to technology k (i.e., our variable of interest), $X_r^{K \setminus \{k\}}$ represents the regional exposure to all other technologies within the same family (excluding the one of interest), X_r^{-K} indicates the regional exposure to all remaining emerging digital tech-

[28]We also estimate the employment impacts at the emerging digital technology family level; see Appendix C.3 for more details.

[29]Figures C.3 to C.7, in the appendix, report the geographic distributions of exposure to individual emerging digital technologies.

Figure 7: Employment Effect of Robots



Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares.

nologies, and Z includes the same set of covariates as in Equation (15). Both $X_r^{K\setminus\{k\}}$ and X_r^{-K} are also calculated as shift-share variables.

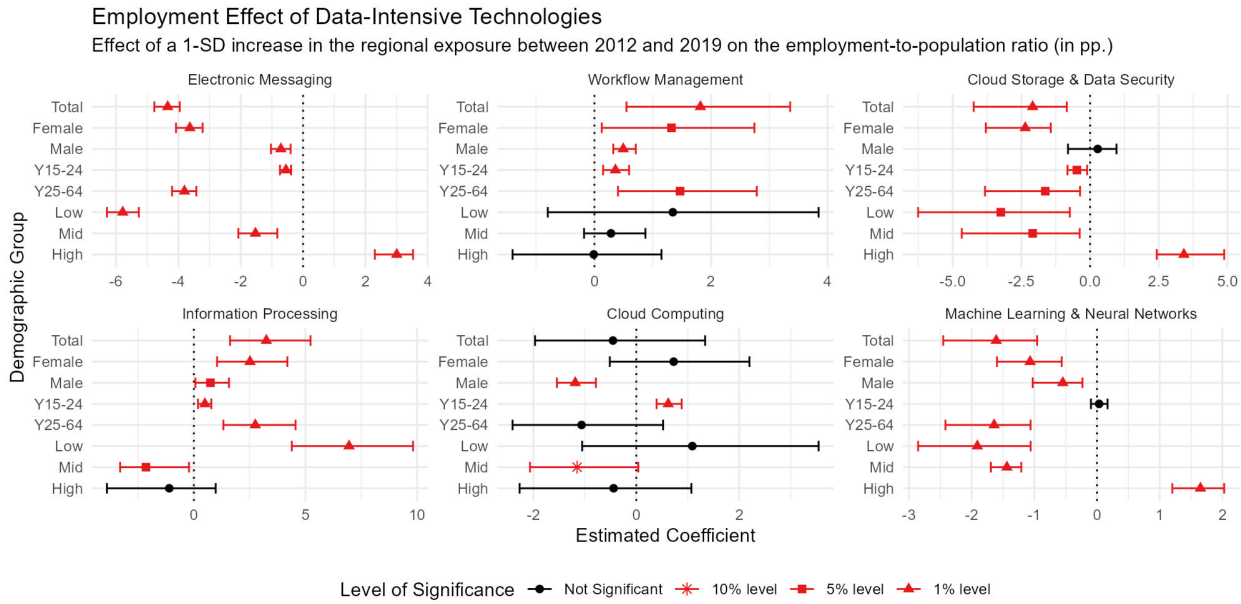
Our estimated coefficient of interest, denoted as $\hat{\beta}_k$, represents the employment effect, measured in pp. change, of a one-standard-deviation increase in the regional exposure to a specific emerging digital technology k. This is conditional on the regional exposure to both its technology family and all other emerging technologies. Accounting for these latter is key to obtaining the causal effect of regional exposure to a specific technology, independent of the effects of other emerging digital technologies or combinations thereof.

We report the results at the individual technology level for two sets of technologies that have received considerable attention in the literature and reveal interesting patterns, namely robots and data-intensive technologies. We include the estimates for all individual technologies in the appendix; see Figures C.8 to C.12.

Robots. Figure 7 displays the estimated coefficients, along with their corresponding 95% AKM0 confidence intervals, for the employment effects of three types of robotic technologies. The figure is interpreted as follows. Each panel corresponds to a technology. The vertical axis lists the demographic groups, while the horizontal axis depicts the estimated coefficients.

Both industrial and mobile robots have negative impacts on employment. The first panel,

Figure 8: Employment Effect of Data-Intensive Technologies



Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares.

labeled "Industrial Automation & Robot Control", corresponds to industrial robots, while the second panel, labeled "Intelligent Logistics", refers to mobile robots. Both types of robots negatively impact employment, particularly for female and mature workers, with the magnitude of the impact being twice as large for industrial robots as for mobile robots. Greater regional exposure to these robots increases the employment of high-skilled workers while decreasing it for both low- and middle-skilled workers.

We do not find any effect of Automation Vehicles & Unmanned Autonomous Vehicles (UVAs) on total employment—except for a small decrease in male employment and a small increase in middle-skilled employment.

Data-Intensive Technologies. Figure 8 presents the employment impact of data-intensive technologies. Among these, Electronic Messaging, Cloud Storage & Data Security, and Machine Learning & Neural Networks have significant negative impacts on the total employment-to-population ratio. Similarly to robots, these technologies tend to displace female and mature workers rather than male and young workers. Also akin to the impact of robots, we observe that the employment of high-skilled workers increases with greater regional exposure to these technologies, while employment decreases for low- and middle-skilled workers.

Although we observe a sizeable employment impact from Cloud Storage, we do not detect any impact of Cloud Computing on employment, as indicated by the bottom middle panel. This suggests that this latter technology neither creates employment opportunities nor displaces workers per se, that is, conditional on other emerging digital technologies. However, we cannot rule out that this technology might be complementary to other technologies and thus be an "enabling" technology—i.e., a technology that amplifies the employment impact of other technologies when combined with them.

Lastly, Information Processing and Workflow Management show positive impacts on employment. We find that a one-standard-deviation increase in exposure to these technologies increases the employment-to-population ratio by 3.25 pp. and 1.81 pp., respectively. For Information Processing, the increase in employment is concentrated among low-skilled workers. While the coefficients are negative but insignificant for middle- and high-skilled employment, this may suggest that Information Processing enables low-skilled workers to perform more complex and abstract tasks, thereby increasing the labor demand for workers at the bottom of the skill distribution.

Workflow Management also has a positive impact on total employment, which is shared across all demographic groups. When examining skill levels, we find no evidence of skilled-biased technological change, as all the coefficients are positive, though insignificant. This suggests that this technology has a uniformly positive impact across the skill distribution.

6 Conclusion

Recent developments in digital technologies, notably AI, have raised public and academic interest in the impact of emerging digital technologies on future employment. Determining whether these technologies will create more jobs than they eliminate is a crucial issue for both individuals and policymakers. However, prior research has largely focused on analyzing either very specific technologies, such as industrial robots or certain applications of AI, or a diverse array of digital technologies commonly labeled as "automation technologies".

In this paper, we measure the exposure of industries and occupations to 40 digital technologies that have emerged over the past decade and investigate their effects on European employment. Using state-of-the-art NLP tools, such as sentence transformers, we introduce a novel methodology to measure the exposure of industries and occupations at a granular level. We have made our pioneering data available as an open–access resource, named the 'TechXposure' database. Using this new data source, we estimate the employment impact of these emerging digital technologies. Our main findings reveal that emerging digital technologies have an overall positive impact on the employment-to-population ratio, thereby creating employment opportunities rather than destroying jobs. However, when examining the specific effects of these technologies, we find considerable heterogeneity in the employment impact of these technologies, with, for instance, robots having a negative impact on employment (except for high-skilled workers). Yet, our paper does not address the question of the quality of these employment opportunities, which is a research question we intend to investigate in the future.

We highlight the advantages and limitations of our exposure scores present in the 'TechXposure' database. First, since our exposure scores are based on text data from standard European classifications, they are universal and not influenced by any specific European country. Second, our method does not rely on keywords (or tokens) and therefore only requires a set of relevant patents, making it replicable in other contexts, such as for green technologies or using future ISCO/NACE classifications. However, our exposure scores do not account for the augmentation or automation effect on occupations and industries; they solely reflect the relevance of technologies to a given industry or occupation. This limitation in capturing their employment effects allows us to make fewer assumptions in data construction, leaving the question open as some technologies may have positive effects on employment in one context and negative ones in another. Additionally, our set of technologies does not include recent developments in Large Language Models (LLM), such as ChatGPT, as our analysis period focuses on technologies that emerged until 2021. However, our set does include several other applications of AI, specifically in areas such as Machine Learning (for computer vision), Information Processing, and Workflow Management. Lastly, our exposure metrics do not measure the adoption of these emerging digital technologies, which is a topic we intend to address in future research.

We regard our paper as a foundational contribution to new avenues for future research on technological change and labor markets. By constructing this open–access database, we anticipate that future studies will greatly benefit from its use. It offers an unprecedented level of detail in analyzing the exposure of occupations and industries to emerging digital technologies, encompassing not only those frequently discussed in economic literature, such as robots and AI, but also less-studied technologies like social networks, cloud technologies, and health technologies. Given that our database is based on European classifications of occupations and industries, it presents a valuable opportunity for research focused on Europe. This research could provide deeper insights into the impact of emerging digital technologies on the economy, particularly considering Europe's rich diversity in institutional contexts that may significantly influence technology development, adoption, and labor market effects. We believe our database is user-friendly and accessible for both researchers and policymakers.

References

Acemoglu, D., Anderson, G., Beede, D., Buffington, C., Childress, E., Dinlersoz, E., Foster, L., Goldschlag, N., Haltiwanger, J., Kroff, Z., Restrepo, P., and Zolas, N. (2022a). Automation and the Workforce : A Firm-Level View from the 2019 Annual Business Survey.

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022b). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acemoglu, D. and Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33:3–30.

Acemoglu, D. and Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128:2188–2244.

Adão, R., Beraja, M., and Pandalai-Nayar, N. (2024). Fast and Slow Technological Transitions. *Journal of Political Economy Macroeconomics*.

Adão, R., Kolesár, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Aghion, P., Antonin, C., Bunel, S., and Jaravel, X. (2023). Modern Manufacturing Capital, Labor Demand, and Product Market Dynamics: Evidence from France. Working paper.

Akerman, A., Gaarder, I., and Mogstad, M. (2015). The Skill Complementarity of Broadband Internet. *The Quarterly Journal of Economics*, 130(4):1781–1824.

Albanesi, S., Silva, A. D. D., Jimeno, J. F., Lamo, A., and Wabitsch, A. (2023). New Technologies and Jobs in Europe. Technical report, National Bureau of Economic Research.

Alekseeva, L., Azar, J., Giné, M., Samila, S., and Taska, B. (2021). The Demand for AI Skills in the Labor Market. *Labour Economics*, 71:102002.

Arts, S., Hou, J., and Gomez, J. C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, 50(2):104144.

Autor and David, H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3):3–30.

Autor, D., Chin, C., Salomons, A., and Seegmiller, B. (2024). New Frontiers: The Origins and Content of New Work, 1940–2018. *The Quarterly Journal of Economics*, Forthcoming.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2006). The Polarization of the U.S. Labor Market. *American Economic Review*, 96(2):189–194.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118:1279–1333.

Borusyak, K., Hull, P., and Jaravel, X. (2021). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies*, 89(1):181–213.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104.

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

Dauth, W., Findeisen, S., Suedekum, J., and Woessner, N. (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association*, 19:3104–3153.

Davies, D. L. and Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.

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

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, 132:1593–1640.

Dorn, D. (2009). *Essays on inequality, spatial interaction, and the demand for skills*. PhD thesis, University of St. Gallen.

Felten, E., Raj, M., and Seamans, R. (2021). Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and its Potential Uses. *Strategic Management Journal*, 42(12):2195–2217.

Felten, E. W., Raj, M., and Seamans, R. (2018). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings*, 108:54–57.

Felten, E. W., Raj, M., and Seamans, R. (2023). How will Language Modelers like ChatGPT Affect Occupations and Industries? *SSRN Electronic Journal*.

Foster-McGregor, N., Nomaler, Ö., and Verspagen, B. (2019). Measuring the creation and adoption of new technologies using trade and patent data.

Frey, C. B. and Osborne, M. A. (2017). The Future of Employment: How Susceptible are Jobs to Computerisation? *Technological Forecasting and Social Change*, 114:254–280.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, 57(3):535–74.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Goos, M. and Manning, A. (2007). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *Review of Economics and Statistics*, 89:118–133.

Goos, M., Manning, A., and Salomons, A. (2009). Job Polarization in Europe. *American Economic Review*, 99:58–63.

Goos, M., Manning, A., and Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, 104:2509–2526.

Graetz, G. and Michaels, G. (2018). Robots at work. *Review of Economics and Statistics*, 100:753–768.

Hall, B., Jaffe, A., and Trajtenberg, M. (2005). Market Value and Patent Citations: A First Look. *The RAND Journal of Economics*, 36(1):16–38.

Jurkat, A., Klump, R., and Schneider, F. (2022). Tracking the Rise of Robots: The IFR Database. *Jahrbücher für Nationalökonomie und Statistik*.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, 3(3):303–320.

Kogan, L., Papanikolaou, D., Schmidt, L., and Seegmiller, B. (2019). Technological Change and Occupations over the Long Run. Working paper.

Kogan, L., Papanikolaou, D., Schmidt, L., and Seegmiller, B. (2021). Technology-Skill Complementarity and Labor Displacement: Evidence from Linking Two Centuries of Patents with Occupations. Technical report.

Mann, K. and Püttmann, L. (2023). Benign Effects of Automation: New Evidence from Patent Texts. *Review of Economics and Statistics*, 105(3):562–579.

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

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., and Clark, J. (2024). The AI Index 2024 Annual Report. Technical report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA.

Ménière, Y., Rudyk, I., and Valdes, J. (2017). Patents and the Fourth Industrial Revolution. Technical Report December, EPO, Munich.

Michaels, G., Natraj, A., and Reenen, J. V. V. (2014). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years. *Review of Economics and Statistics*, 96:60–77.

OECD (2009). OECD Patent Statistics Manual.

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

Vries, G. J. d., Gentile, E., Miroudot, S., and Wacker, K. M. (2020). The Rise of Robots and the Fall of Routine Jobs. *Labour Economics*, 66:101885.

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

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D., Buffington, C., Goldschlag, N., Foster, L., and Dinlersoz, E. (2020). Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey. Technical report, National Bureau of Economic Research, Cambridge, MA.

Appendices

A Data Appendix

We provide additional information on the set of emerging digital technologies and the derivation of exposure scores.

A.1 Patent Corpus Construction

Query and Patent Corpus. The Chaturvedi et al. (2023) patent corpus is constructed by querying Derwent Innovation Index (DII) database. The query consists of two components, and comprises patent codes (Derwent Manual Codes, and International Patent Classification (IPC) codes) and keywords collected from earlier studies on digital automation technologies and Industry 4.0 (CITATIONS). The first component retrieves digital automation inventions in (a) process and machine control implemented in physical production such as manufacturing, agriculture, mining, construction, and (b) process and workflow control in services. The second component narrows the sample to a set of large technology families studied in the previous literature on emerging technologies (CITATIONS) such as AI, computing, networking, data acquisition and management, user interfaces. The total sample comprises 1,143,033 patent families between 2000 and 2021. Figure A.1 presents the SQL-stylized structure of the patent query.

Figure A.1

```
SELECT Process and machine control in production AS A,
           Process and workflow control in services    AS B
FROM Derwent Innovation Index
WHERE technology IN (Networking,
                          Data acquisition,
                          Data management,
                          AI and Intelligent Systems,
                          User Interfaces,
                          Computing)
    AND year BETWEEN 2000 AND 2021

UNION

SELECT Additive manufacturing AS C
FROM Derwent Innovation Index
WHERE year BETWEEN 2000 AND 2021
```

Notes: This figure presents .

Patent Embeddings. To leverage patent texts for the analysis of emerging digital technologies, Chaturvedi et al. (2023) concatenate patent titles and abstracts and produce their embeddings. Using the pre-trained sentence transformer model all-mpnet-base-v2 (Reimers and Gurevych, 2019), each patent text is mapped onto a 768-dimensional space, transforming text into a semantic vector called embedding. This allows for analysis and comparison of documents' meaning at scale using other ML and NLP methods.

Novel patents. To identify novel patents among the corpus of digital automation inventions, Local Outlier Factor (LOF) algorithm is employed. Proposed by Breunig et al. (2000), LOF is an anomaly detection algorithm applied by Chaturvedi et al. (2023) to search for semantic outliers among patents. Thus, it measures local density of a focal document compared to local density of its k-nearest neighbors in the semantic space. The locality or the size of neighborhood is set by the parameter k. A document is considered anomalous (i.e. local outlier) if it has a substantially lower density than its neighbors.

In application to radical innovation search, larger values of k are more suitable as it allows for larger neighborhoods and hence wider reference group of patents to compute LOF measure. Chaturvedi et al. (2023) use $k = 1000$.

The LOF measure is computed for each patent in year t using cumulative set of patents up to year $(t-1)$.

Since Chaturvedi et al. (2023) are interested in emerging digital automation technologies whose impact on labor markets is unfolding, they search for novel patents in the latest decade of the patent sample, i.e. 2012-2021. Thus, they start with the base sample from 2001–2011 period that comprises 258,344 patents to compute LOF measure in each year from 2012-2021 period, updating the base sample at every iteration. For example, the base sample includes patents from 2001-2013 period to compute LOF measure of patents filed in 2014.

Lastly, the novel patents are defined as those in the bottom 10% of the LOF measure within each year, resulting in 88,413 novel patents in the 2012-2021 period. This set of novel patents $\mathcal{P}^n$ received on average 20% more citations. This property is robust to the variation in identification of the $\mathcal{P}^n$: (a) varying locality value k, (b) using continuous measure of novelty instead of the threshold one, (c) year fixed effects.

Offshoots. To track the development of technological innovations in $\mathcal{P}^n$ throughout the 2012-2021 period, Chaturvedi et al. (2023) identify offshoots of $\mathcal{P}^n$, i.e. subsequent inventions that build on and are semantically similar to $\mathcal{P}^n$. For each novel patent in $\mathcal{P}^n$, the authors compute cosine similarity to all patents in each subsequent year, and define as offshoots patents in the top 10% of cosine similarity within each year.

The final patent corpus $\mathcal{P}$ comprises 190,714 core digital automation patents and their offshoots.

A.2 Description of Emerging Digital Technologies

Tables A.1 to A.3 present the 40 emerging digital technologies from the TechXposure database as well as their descriptions.

A.3 Manual Exclusions

Industry. For industries, we make the following manual adjustments:

- We exclude the exposure scores that relate to 'Printing and service activities related to printing' (18.1) due to the persistent conflation of its intended meaning (i.e. printing products with text, symbols (e.g. musical notation), and imagery (e.g. maps, engraving, etc.)) with emerging digital technologies.
- We exclude the sentence "*manufacture of computer printout paper ready for use*" (Sentence ID 17.2_11) from the industry description text of 'Manufacture of articles of paper and paperboard' (17.2) when combining tasks with patents belonging to the technologies within the 3D Printing family.
- We exclude the sentence "*units giving this type of instructions might be named "schools", "studios", "classes" etc.*" (Sentence ID 85.5_17) from the industry description text of 'Other education' (85.5) when combining tasks with patents belonging to the technology Machine Learning.

Occupation. For occupations, we make the following manual adjustments:

- Analogously with industry 18.1, we exclude the exposure scores that relate to 'Printing trades workers' (732) and its nested occupations (7321, 7322, 7323) due to the persistent conflation of its intended meaning with emerging digital technologies.
- We exclude the task "*creating the blueprint or pattern pieces for a particular apparel design with the aid of a computer;*" (Task ID 7532_2) from the occupation description text of 'Printers' (7532) when combining tasks with patents belonging to the technology Machine Learning.
- We exclude the task "*preparing and developing instructional training material and aids such as handbooks, visual aids, online tutorials, demonstration models and supporting training reference documentation;*" (Task ID 2424_3) from the occupation description text of 'Training and staff development professionals' (2424) when combining tasks with patents belonging to the technology Machine Learning.

A.4 Redundancy Filtering Examples

Tables A.4 and A.5 present additional examples of redundancy filtering for industries. Tables A.6 to A.8 present examples of redundancy filtering for occupations.

A.5 Distribution of Patents and Citation-based Weighting Scheme

Figure A.2 presents the distribution of patents across emerging digital technologies. Figure A.4 presents the distribution of non-cited and undetermined-count patents across emerging digital technologies. Figure A.3 presents the log distribution of patent citations across emerging digital technologies. Figure A.6 presents the correlation between citation-weighted and unweighted yearly cosine similarity scores for both industries and occupations.

A.6 Technology Co-Occurrence

Using our cosine similarity scores, we examine the semantic co-occurrence of emerging digital technologies across occupations. Let $C^k_{\mathcal{O}} = (C^k_1, \ldots, C^k_o, \ldots, C^k_O)$ represent the vector of cosine similarity scores for all occupations related to technology k. We define the pairwise semantic-based technology co-occurrence as the correlation between $C^k_{\mathcal{O}}$ and $C^{k'}_{\mathcal{O}}$ for each pair of technologies (k, k'). These pairwise correlations are computed for all technologies using semantic similarity scores at the 3-digit occupational level.

Figure A.7 presents the result of technology grouping based on cosine semantic scores. We observe a distinct segmentation within the figure, categorized as 'technology families'. Starting from the top-left corner and moving along the diagonal, the first group encountered includes technologies related to 3D Printing. Subsequent to this, the range from Smart Agriculture to Smart Home falls within the Embedded Systems family. A significant block then emerges, spanning from Intelligent Logistics to Passenger Transportation, and encompasses Smart Mobility technologies. Following this, a standalone block dedicated to Food Ordering appears. The next two blocks represent E-Commerce and Payment Systems, respectively. This sequence is succeeded by the most extensive block, which includes 12 technologies and relates to Digital Services. Afterward, AR/VR, Machine Learning, and Medical Imaging are grouped under Computer Vision technologies. Finally, the figure concludes with HealthTech technologies.

A.7 Exposure Scores at Higher Levels of Aggregation

To calculate exposure scores at higher levels of aggregation within the ISCO and NACE classifications, we apply the inverse hyperbolic sine transformation to the average cosine similarity score aggregated across all industries/occupations from the most granular classification level up to the level of interest.

For example, consider the derivation of the exposure score for a 1-digit NACE industry $I \subset \mathcal{I}$ to an emerging digital technology k. We begin with the cosine similarity score, aggregating it to a higher level of classification as follows:

$$C_I^k = \frac{1}{|I|} \sum_{i \in I} C_i^k,$$

where C_i^k is cosine similarity score between a 3–digit industry i (belonging to the 1-digit industry I) and technology k, as obtained in Equation (12). We then apply the inverse hyperbolic sine transformation to obtain the exposure score, namely, $X_I^k = \sinh^{-1}(C_I^k)$. This methodology is similarly employed to derive exposure scores for 2-digit industries, as well as for occupation exposures at higher levels of aggregation.

A.8 Comparing Exposure Scores with Other Metrics

We compare our occupational exposure scores with metrics from Frey and Osborne (2017), Webb (2019), and Felten et al. (2021). These studies provide exposure scores for specific digital technologies that are subsets of our list. A challenge in comparison is the different occupational classifications. To address this, we use crosswalks between classification systems. Then, we aggregate exposure scores within a 4-digit ISCO-08 occupation by averaging the exposures across all matched occupations.

Webb (2019). Exposure scores in Webb (2019) cover three broad technologies: robots, AI, and software. These scores are expressed in percentiles from 0 to 100, with 100 representing the highest exposure. Occupations are classified using the "occ1990dd" system developed by Dorn (2009) and extended by Deming (2017). We link these occupations to the 2010 Census Occupational Classification using the crosswalk from Autor and David (2015). From there, we derive the 2010 SOC and then the ISCO-08 occupations through two crosswalks provided by the Bureau of Labor Statistics (BLS). Once ISCO-08 occupations are linked to the initial "occ1990dd" occupations, we aggregate the exposure scores for each 4-digit ISCO-08 occupation by averaging them separately for the three technologies. Finally, we recompute the exposure scores as percentiles and transform the TechXposure scores into percentiles for comparison.

Felten et al. (2021). Exposure scores in Felten et al. (2021) cover ten AI applications and are normalized with a zero mean and a standard deviation of one. Occupations are classified using the 2010 SOC. Using the BLS crosswalk, we convert 2010 SOC occupations to 4-digit ISCO-08. We aggregate by taking the average, before recomputing the normalized exposure scores. We also normalize the TechXposure scores for comparison.

Frey and Osborne (2017). Exposure scores in Frey and Osborne (2017) measure the risk of computerization of occupations, expressed as probabilities between 0 and 1. Occupations are classified using the 2010 SOC. We apply the same procedure as for Felten et al. (2021), normalizing the exposure scores.

We compute the correlation between our exposure scores for each technology and those obtained with these metrics at the 4-digit ISCO-08 level and report the correlations as a heatmap in Figure A.8. The figure reveals several insights. First, our exposure metrics correlate overall with those in the literature. The robot and software exposure scores in Webb (2019) align with our metrics across a range of emerging digital technologies. Specifically, Webb's robot exposure scores are highly correlated with our *tangible* emerging digital technologies and capture the occupational exposure to Smart Mobility technologies, which include robotization in transportation and mobility. This indicates that Webb's robot exposure scores are not limited to industrial robots.

Conversely, we find that AI exposure scores in Webb (2019) are confined to core AI applications, such as some embedded technologies (i.e., energy management, industrial automation, and remote monitoring) and data-intensive technologies (i.e., machine learning, workflow management systems, and cloud computing), thus missing broader AI applications like medical imaging or information processing.

Exposure scores in Felten et al. (2021) correlate with a broader set of our technologies, indicating they cover a wider spectrum of AI applications as compared to Webb (2019). However, they are negatively correlated with embedded systems, suggesting they do not account for embedded AI, potentially biasing their exposure scores toward high-skilled jobs.

Lastly, software exposure in Webb (2019) and computerization exposure in Frey and Osborne (2017) correlate with a large segment of our emerging digital technologies. However, the magnitudes of these correlations are smaller, as both computerization and software are inherent to emerging digital technologies.

Table A.1: Description of the Emerging Digital Technologies (1/3)

	Technology	**Description**
1	3D Printer Hardware	Three-dimensional printers and their components, such as printing heads, pens, nozzles, platforms, and devices for printing, extruding, cleaning, recycling, heating, and cooling.
2	3D Printing	Printing systems for creating three-dimensional objects using a variety of materials and techniques, like photocuring and powder spreading.
3	Additive Manufacturing	Technologies and processes for additive manufacturing, with applications such as prostheses and building materials.
4	Smart Agriculture & Water Management	Various Internet of Things (IoT) technologies for intelligent and remote management in agriculture, and water and sewage systems.
5	Internet of Things (IoT)	Systems and devices interconnected via IoT for data collection, remote control, and real-time monitoring in diverse applications, including agriculture, home automation, and environmental monitoring.
6	Predictive Energy Management and Distribution	A combination of network, data management, and AI technologies for monitoring, distribution, and efficient use of electrical power and energy, including renewable energy sources, and for consumption prediction in intelligent power management.
7	Industrial Automation & Robot Control	Industrial process automation, including robots, programmable logic controllers, and related control apparatuses such as remote control and fault diagnosis.
8	Remote Monitoring & Control Systems	Real-time remote monitoring and management technologies for factories, building management, warehouses, intelligent homes, disaster management, and network security.
9	Smart Home & Intelligent Household Control	Various IoT technologies for the intelligent control of homes and buildings, including household appliances, home environments, and smart home integrations, often utilizing wireless communication and monitoring.
10	Intelligent Logistics	A combination of monitoring, remote control technologies, data acquisition, and mobile robot technologies for logistics and delivery applications, including supply chain management, warehouse operations, package tracking, and courier services.
11	Autonomous Vehicles & UAVs	Developments in unmanned aerial vehicles (UAVs), drones, and autonomous driving technologies, with an emphasis on vehicle control, navigation, and sensor integration.
12	Parking & Vehicle Space Management	Networking technologies for parking management, including systems for monitoring available spaces and intelligent parking solutions.
13	Vehicle Telematics & Electric Vehicle Management	Technologies for intra-vehicle information management, especially in electric vehicles, including aspects of real-time monitoring, traffic information, and vehicle diagnostics.
14	Passenger Transportation	Technologies for ride-sharing, taxi hailing, and public transportation reservations using real-time information, electronic ticketing, and route optimization.

Notes: This table provides descriptions of emerging digital technologies ranging from 1 to 14.

Table A.2: Description of the Emerging Digital Technologies (2/3)

	Technology	Description
15	Food Ordering & Vending Systems	Wireless infrastructures, encryption, monitoring, and remote control technologies for food order management, such as automatic vending, self-service ordering, meal preparation, and delivery.
16	Digital Advertising	Automated tracing and tagging, and AI technologies for digital advertisements, including targeted delivery on mobile devices.
17	Electronic Trading and Auctions	Online trading platforms, financial instrument exchanges, and auction mechanisms, focusing on real-time bidding, trading, and market data.
18	Online Shopping Platforms	Wireless technologies (e.g., RFID and mobile terminals), encryption (e.g., blockchain), and AI technologies for e-commerce transactions, and digital tools related to the purchase, sale, and display of product information, including recommendation systems.
19	E-Coupons & Promotion Management	Data management platforms for electronic coupon distribution, management, redemption, and associated loyalty programs.
20	Electronic Payments & Financial Transactions	A combination of wireless (e.g., mobile) and encryption (e.g., blockchain) technologies for processing electronic payments (e.g., credit card transactions) and interfacing with financial institutions.
21	Mobile Payments	A combination of mobile technologies for processing electronic payments.
22	Gaming & Wagering Systems	A combination of user interface and data management technologies for gaming, both online and physical, including gambling and gaming machines.
23	Digital Authentication	Encryption and robotic processing technologies for verifying user identities, securing transactions, and safeguarding data through various authentication mechanisms, such as biometrics and cryptographic methods.
24	E-Learning	A combination of AI and data management technologies for digital platforms and systems in education, including teaching, learning, and classroom management.
25	Location-Based Services & Tracking	Technologies that provide location-based content and services, often relying on global positioning and navigation systems and related communication technology.
26	Voice Communication	Technologies focusing on voice communication, including communication protocols and user interfaces.
27	Electronic Messaging	Digital communication methods, infrastructure, and user interfaces for services such as email and conferences.
28	Workflow Management	A combination of AI and network technologies for management applications, including workflow automation, recruitment, event scheduling, and building and property management.

Notes: This table provides descriptions of emerging digital technologies ranging from 15 to 28.

Table A.3: Description of the Emerging Digital Technologies (3/3)

	Technology	**Description**
29	Cloud Storage & Data Security	Cloud-based data storage, distributed data management, encryption, and backup, often integrated with blockchain technology.
30	Information Processing	Systems for managing, processing, and delivering data and information across various domains, potentially including content generation, transmission, and verification.
31	Cloud Computing	Cloud computing and virtual machines, focusing on cloud platforms and resource allocation in cloud environments.
32	Recommender Systems	Algorithms and systems for providing recommendations and personalized content delivery based on user behavior, search queries, and similarity metrics.
33	Social Networking & Media Platforms	User interfaces for online social networking services, content sharing, and recommendation systems.
34	Digital Media Content	Tools and platforms for digital media content creation, management, distribution, and access.
35	Augmented and Virtual Reality (AR/VR)	Augmented reality (AR) and virtual reality (VR) models, devices, interfaces, and experiences, including head-mounted displays and interactions in virtual environments.
36	Machine Learning & Neural Networks	Machine learning training techniques, model architectures, and data processing for computer vision applications.
37	Medical Imaging & Image Processing	Diverse applications for acquiring and analyzing medical images from various modalities, such as computed tomography (CT), ultrasound, magnetic resonance imaging (MRI), and virtual reality (VR), for purposes including diagnosis, surgical planning, and the design of prostheses.
38	Health Monitoring	Wearable and implantable devices and systems for real-time health monitoring that track vital signs such as blood pressure, heart rate, and temperature, coupled with comprehensive medical data management.
39	Medical Information	A combination of data sharing, encryption, and Natural Language Processing (NLP) technologies for the storage, retrieval, and management of medical and patient information, encompassing electronic medical records, prescription management, and remote healthcare services.
40	E-Healthcare	An integration of data sharing, wireless communication, monitoring, and user interface technologies for healthcare and health management systems, including those used in hospitals and cloud-based platforms.

Notes: This table provides descriptions of emerging digital technologies ranging from 29 to 40.

Table A.4: Example of Redundancy Filtering of Industries for Intelligent Vehicular Control Device

		Cosine Similarity		
Code	NACE Industry	$C_i^{p_1}$	$C_i^{p_2}$	C_i^{p}
52.2	Support activities for transportation	0.531	0.454	0.489
49.4	Freight transport by road and removal services	0.371	0.418	0.393
29.1	Manufacture of motor vehicles	0.409	0.371	0.389
27.9	Manufacture of other electrical equipment	0.358	0.375	0.366
30.9	Manufacture of transport equipment n.e.c.	0.452		
29.2	Manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers	0.389		
33.1	Repair of fabricated metal products, machinery and equipment	0.379		
45.3	Sale of motor vehicle parts and accessories	0.377		
49.1	Passenger rail transport, interurban	0.371		
47.3	Retail sale of automotive fuel in specialised stores	0.362		
26.5	Manufacture of instruments and appliances for measuring, testing and navigation; watches and clocks		0.472	
26.3	Manufacture of communication equipment		0.434	
26.2	Manufacture of computers and peripheral equipment		0.410	
56.1	Restaurants and mobile food service activities		0.392	
61.2	Wireless telecommunications activities		0.378	
49.3	Other passenger land transport		0.369	

Notes: This table presents the redundancy filtering of industries for the Patent ID 201713859U. It displays the cosine similarity of distinct 3-digit NACE Rev.2 industry descriptions with the patent description "Vehicle intelligent logistics control device" (Column 3) and the function principle "GPS locating module for obtaining position information of transport vehicle through main control chip, RFID reader for reading RFID tag information, and 4G module connected with server" (Column 4). Industries are ranked according to Column 3 in decreasing order. Cosine similarity scores in Columns 3 and 4 are displayed only for sub-pairs belonging to their respective top 10. Column 5 shows the composite patent-industry cosine similarity score, which corresponds to the harmonic mean of Columns 3 and 4. Cosine similarity scores in Column 5 are displayed only for pairs that rank simultaneously in both top 10.

Table A.5: Example of Redundancy Filtering of Industries for Speech Recognition System

		Cosine Similarity		
Code	NACE Industry	$C_i^{p_1}$	$C_i^{p_2}$	C_i^{p}
26.3	Manufacture of communication equipment	0.256	0.333	0.289
28.2	Manufacture of other general-purpose machinery	0.246	0.344	0.286
82.9	Business support service activities n.e.c.	0.279	0.285	0.282
26.4	Manufacture of consumer electronics	0.250	0.295	0.271
63.9	Other information service activities	0.245	0.269	0.257
62.0	Computer programming, consultancy and related activities	0.276		
85.5	Other education	0.250		
61.9	Other telecommunications activities	0.225		
58.1	Publishing of books, periodicals and other publishing activities	0.224		
26.5	Manufacture of instruments and appliances for measuring, testing and navigation; watches and clocks		0.303	
28.9	Manufacture of other special-purpose machinery		0.294	
72.1	Research and experimental development on natural sciences and engineering		0.276	
18.2	Reproduction of recorded media		0.265	

Notes: This table presents the redundancy filtering of industries for the Patent ID 202048118D. It displays the cosine similarity of distinct 3-digit NACE Rev.2 industry descriptions with the patent description "System for recognizing training speech" (Column 3) and the function principle "process or which is configured to increment counter associated with word sequences, and train language model of automatic transcription system using word sequences and counter" (Column 4). Industries are ranked according to Column 3 in decreasing order. Cosine similarity scores in Columns 3 and 4 are displayed only for sub-pairs belonging to their respective top 10. Column 5 shows the composite patent-industry cosine similarity score, which corresponds to the harmonic mean of Columns 3 and 4. Cosine similarity scores in Column 5 are displayed only for pairs that rank simultaneously in both top 10.

Table A.6: Example of Redundancy Filtering of Occupations for Targeted TV Advertising

Code	ISCO Occupation	Cosine Similarity		
		$C^p_{o_1}$	$C^p_{o_2}$	C^p_o
2431	Advertising and marketing professionals	0.413	0.502	0.453
1222	Advertising and public relations managers	0.308	0.420	0.356
3521	Broadcasting and audio-visual technicians	0.274	0.380	0.318
3322	Commercial sales representatives	0.250	0.394	0.306
2434	ICT sales professionals	0.297		
7422	ICT installers and servicers	0.282		
4227	Survey and market research interviewers	0.279		
2656	Announcers on radio, television and other media	0.278		
1330	ICT service managers	0.262		
3512	ICT user support technicians	0.252		
5242	Sales demonstrators		0.396	
1420	Retail and wholesale trade managers		0.393	
3432	Interior designers and decorators		0.388	
2153	Telecommunications engineers		0.374	
3323	Buyers		0.358	
9520	Street vendors (excluding food)		0.357	

Notes: This table presents the redundancy filtering of occupations for the Patent ID 2013B87254 (i.e., "Method for targeting television advertisement based on profile linked to online device, involves selecting television advertisement to be directed to set-top box based on profile information pertaining to user or online activity"). It displays the cosine similarity of the patent title with the 4-digit ISCO-08 title (Column 3) and the task with the highest cosine similarity (Column 4). Occupations are ranked according to Column 3 in decreasing order. Cosine similarity scores in Columns 3 and 4 are displayed only for sub-pairs belonging to their respective top 10. Column 5 shows the composite patent-occupation cosine similarity score, which corresponds to the harmonic mean of Columns 3 and 4. Cosine similarity scores in Column 5 are displayed only for pairs that rank simultaneously in both top 10.

Table A.7: Example of Redundancy Filtering of Occupations for Intelligent Vehicular Control Device

		Cosine Similarity		
Code	ISCO Occupation	$C^p_{o_1}$	$C^p_{o_2}$	C^p_o
8322	Car, taxi and van drivers	0.354	0.525	0.423
4323	Transport clerks	0.333	0.440	0.379
9333	Freight handlers	0.333	0.420	0.371
9621	Messengers, package deliverers and luggage porters	0.308	0.412	0.353
8332	Heavy truck and lorry drivers	0.301	0.405	0.345
7422	ICT installers and servicers	0.371		
8341	Mobile farm and forestry plant operators	0.332		
1330	ICT service managers	0.314		
1324	Supply, distribution and related managers	0.298		
8160	Food and related products machine operators	0.273		
8344	Lifting truck operators		0.496	
9329	Manufacturing labourers not elsewhere classified		0.481	
4321	Stock clerks		0.420	
9520	Street vendors (excluding food)		0.409	
8331	Bus and tram drivers		0.405	

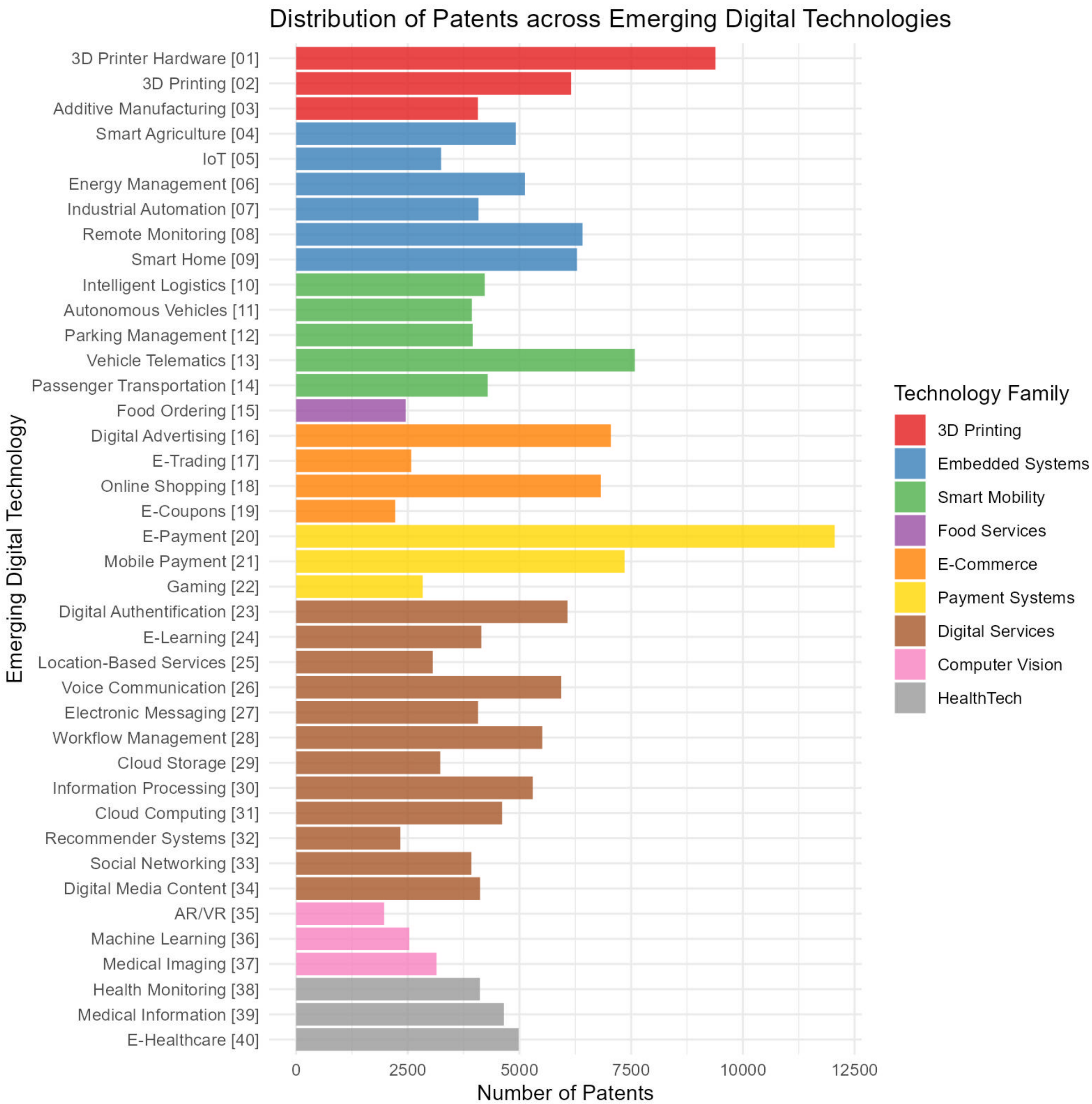
Notes: This table presents the redundancy filtering of occupations for the Patent ID 201713859U (i.e., "Vehicle intelligent logistics control device, has GPS locating module for obtaining position information of transport vehicle through main control chip, RFID reader for reading RFID tag information, and 4G module connected with server"). It displays the cosine similarity of the patent title with the 4-digit ISCO-08 title (Column 3) and the task with the highest cosine similarity (Column 4). Occupations are ranked according to Column 3 in decreasing order. Cosine similarity scores in Columns 3 and 4 are displayed only for sub-pairs belonging to their respective top 10. Column 5 shows the composite patent-occupation cosine similarity score, which corresponds to the harmonic mean of Columns 3 and 4. Cosine similarity scores in Column 5 are displayed only for pairs that rank simultaneously in both top 10.

Table A.8: Example of Redundancy Filtering of Occupations for Speech Recognition System

Code	ISCO Occupation	Cosine Similarity		
		$C^p_{o_1}$	$C^p_{o_2}$	C^p_o
4131	Typists and word processing operators	0.309	0.452	0.367
2643	Translators, interpreters and other linguists	0.245	0.379	0.298
4413	Coding, proofreading and related clerks	0.232	0.343	0.277
2266	Audiologists and speech therapists	0.218	0.363	0.273
8153	Sewing machine operators	0.214		
7532	Garment and related patternmakers and cutters	0.209		
4223	Telephone switchboard operators	0.198		
8143	Paper products machine operators	0.197		
8131	Chemical products plant and machine operators	0.193		
7422	ICT installers and servicers	0.193		
4110	General office clerks		0.396	
3252	Medical records and health information technicians		0.339	
4120	Secretaries (general)		0.329	
4132	Data entry clerks		0.324	
4311	Accounting and bookkeeping clerks		0.304	
2152	Electronics engineers		0.302	

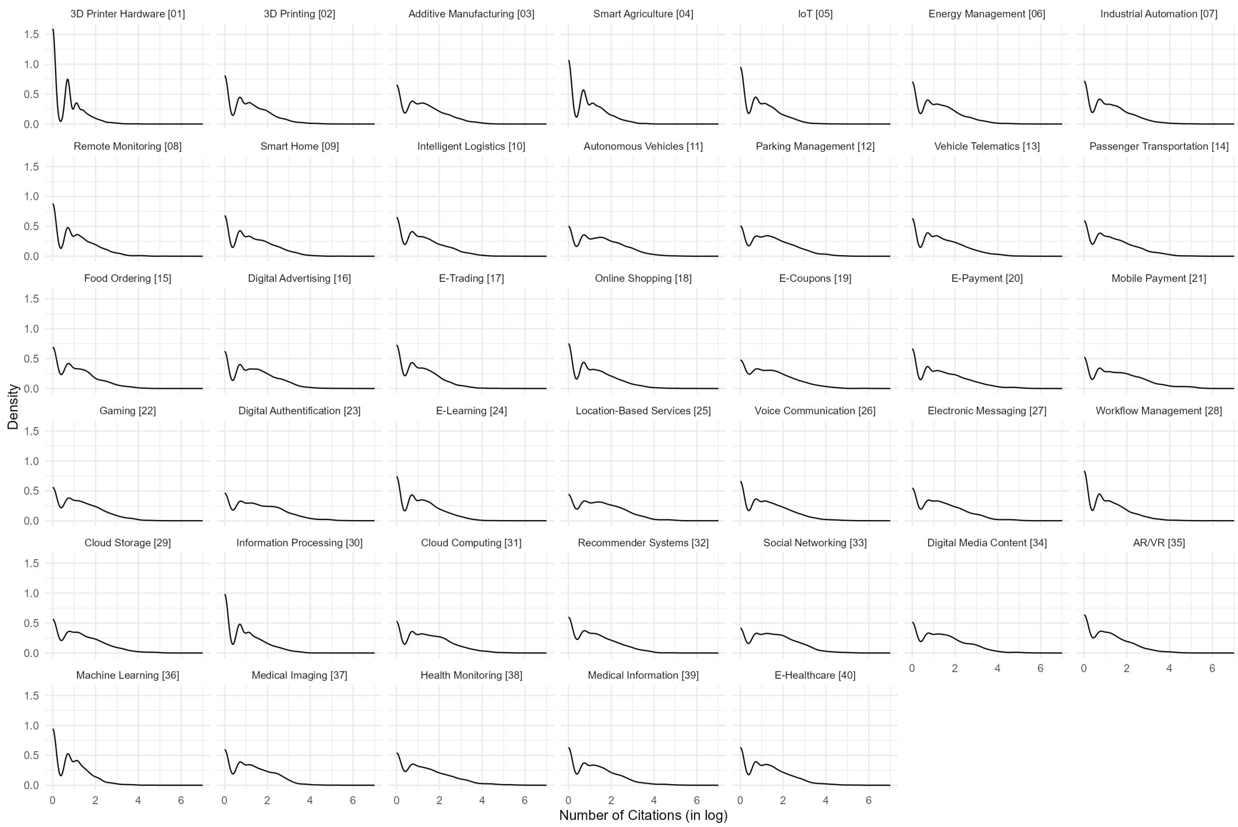
Notes: This table presents the redundancy filtering of occupations for the Patent ID 202048118D (i.e., "System for recognizing training speech, has process or which is configured to increment counter associated with word sequences, and train language model of automatic transcription system using word sequences and counter"). It displays the cosine similarity of the patent title with the 4-digit ISCO-08 title (Column 3) and the task with the highest cosine similarity (Column 4). Occupations are ranked according to Column 3 in decreasing order. Cosine similarity scores in Columns 3 and 4 are displayed only for sub-pairs belonging to their respective top 10. Column 5 shows the composite patent-occupation cosine similarity score, which corresponds to the harmonic mean of Columns 3 and 4. Cosine similarity scores in Column 5 are displayed only for pairs that rank simultaneously in both top 10.

Figure A.2: Distribution of Patents across Emerging Digital Technologies



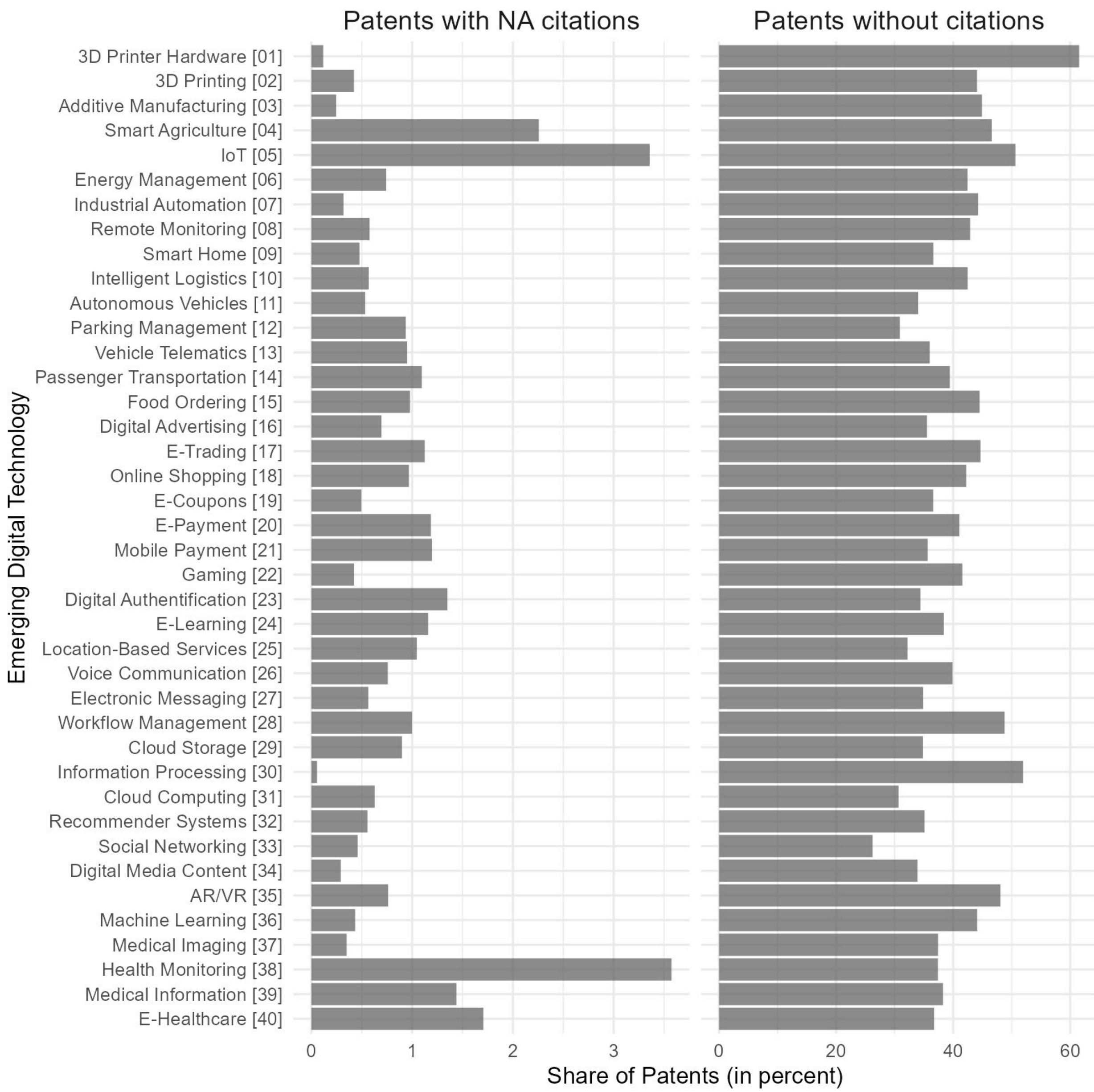
Notes: This figure presents the distribution of patents across emerging digital technologies. The set of patents includes 190,714 Derwent patents, filed between 2012 and 2021. This patent set constructed by Chaturvedi et al. (2023) comprises the core patents related to digital innovations, together with the patents that follow their semantic trajectory, that is, the most similar patents filed in subsequent years.

Figure A.3: Log Distribution of Patent Citations across Emerging Digital Technologies



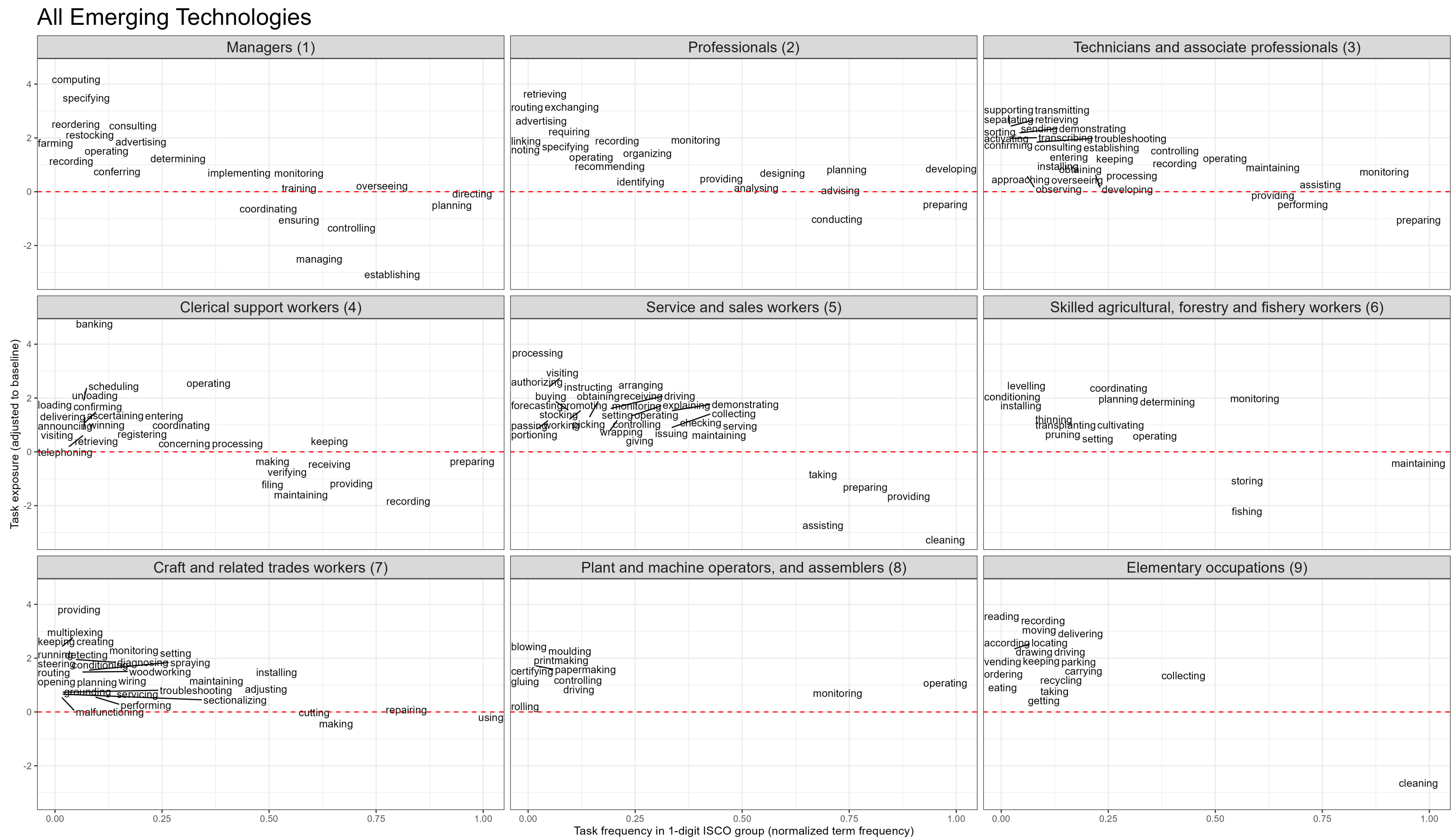
Notes: This figure presents the log distribution of patent citations across emerging digital technologies.

Figure A.4: Distribution of Non-Cited and Undetermined-Count Patents across Emerging Digital Technologies



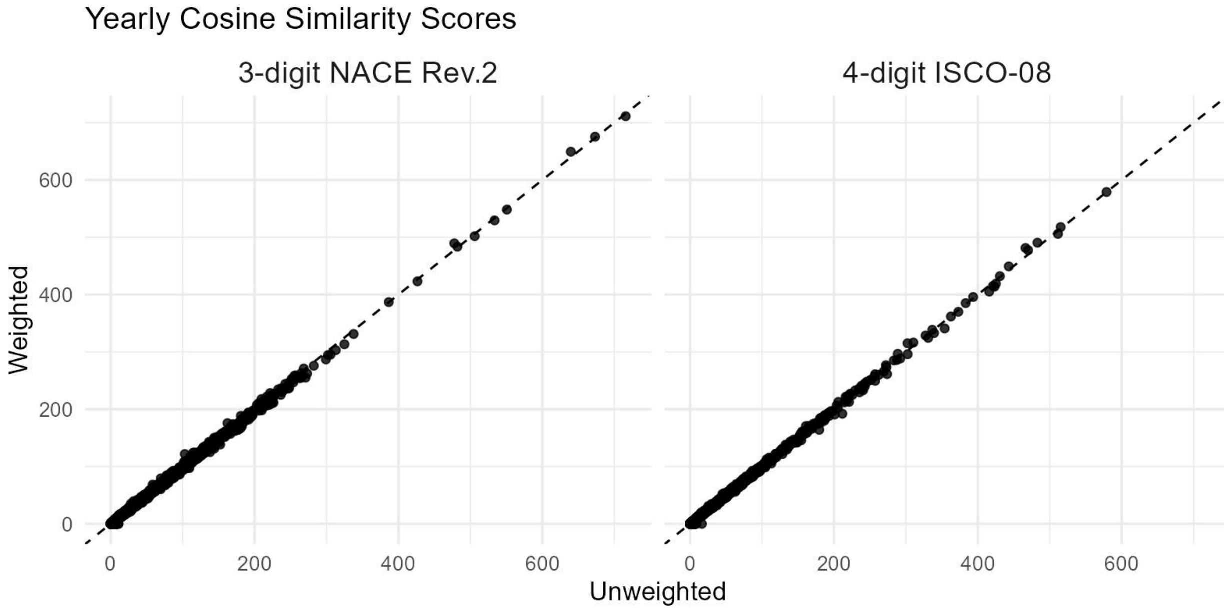
Notes: This figure presents the distribution of non-cited and undetermined-count patents across emerging digital technologies.

Figure A.5: The Most Exposed Tasks (1-digit ISCO-08)



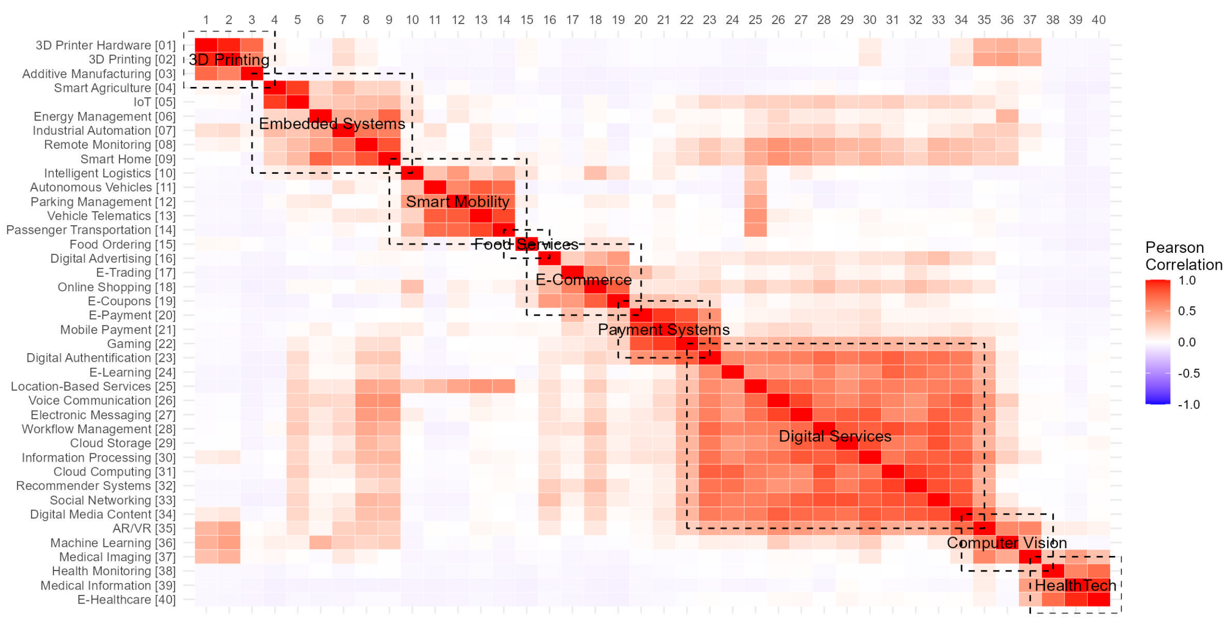
Notes: This figure displays the top exposed tasks to all emerging digital technologies by 1-digit ISCO-08 group. The baseline corpus is ISCO-08 classification while the target corpus is the set of occupational tasks exposed to the emerging digital technologies. Therefore, the target corpus is essentially a sample with replacement where the sampling is the exposure identification procedure outlined in Section 3.2. The horizontal axis is the term frequency in the baseline corpus normalized to the frequency maximum within the 1-digit group. The vertical axis is the logarithm of the prevalence ratio, i.e. ratio of token's probabilities to occur in the target corpus over the baseline corpus: $log(\frac{p^{target}}{p_{base}})$. The probabilities in the target corpus are weighted by the cosine similarity between the task and the technology.

Figure A.6: Weighted versus Unweighted Yearly Cosine Similarity Scores



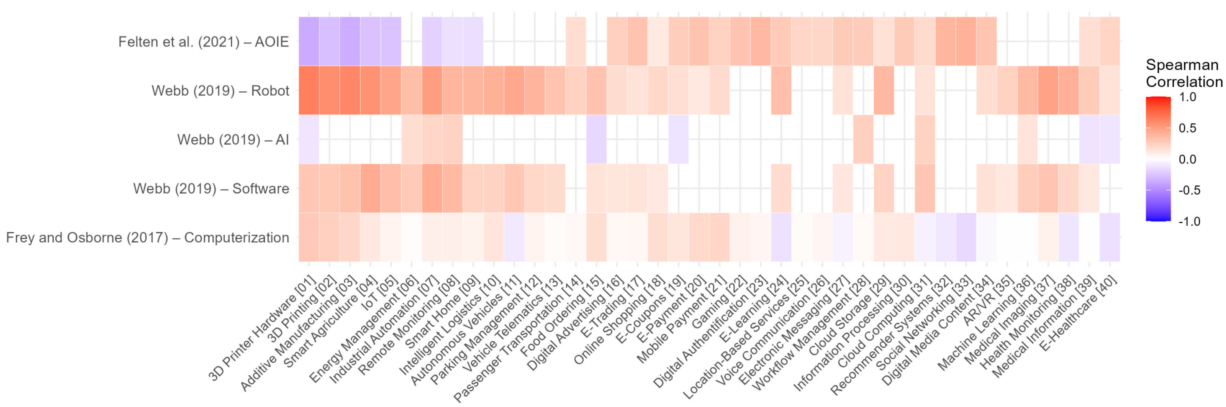
Notes: This figure presents the correlation between citation-weighted and unweighted yearly cosine similarity scores for both industries and occupations. The dashed line is the 45-degree line.

Figure A.7: Semantic Co-Occurrence of Technologies in 3–digit ISCO-08 Occupations



Notes: This figure shows all pairwise semantic-based technology co-occurrences as a correlation matrix, which is symmetric with diagonal values of 1. The matrix categorizes technologies into blocks, grouping them according to their semantic associations with occupations.

Figure A.8: Correlation of Occupation Exposure with Other Metrics in the Literature



Notes: This figure presents the correlation between occupational exposure scores to emerging digital technologies (column) and other occupational exposure metrics available in the literature (rows), both measured at the 4-digit ISCO-08 level. Each cell shows the Spearman correlation ranging from -1 to 1. Correlations with a p-value above 0.05 are transparent. Exposure scores in the literature are from Felten et al. (2021), Webb (2019), and Frey and Osborne (2017) and are converted into 4-digit ISCO-08 exposure scores using several crosswalks.

B Descriptive Statistics Appendix

In this Appendix, we provide additional descriptive statistics on the exposure of industries and occupations to emerging digital technologies.

B.1 Top-30 Most Exposed

Tables B.1 and B.2 display the top 30 exposed 4-digit ISCO-08 occupations and 3-digit NACE Rev.2 industries, respectively, according to their average exposure to all emerging digital technologies, denoted as

$$X_o = \frac{1}{40} \sum_k X_o^k,$$

where X_o^k is the exposure of occupation o to technology k given by Equation (13) and

$$X_i = \frac{1}{40} \sum_k X_i^k,$$

where X_i^k is the exposure of industry i to technology k also given by Equation (13). Tables also include their top-3 concentration ratio (CR3) expressed in percent.

B.2 Exposure Scores at the 2-Digit Level

Figures B.1 and B.2 present the exposure of 2-digit ISCO-08 occupations and 2-digit NACE Rev.2 industries, respectively, to the 40 emerging digital technologies.

Table B.1: Top 30 Exposed 4-digit ISCO-08 Occupations

Code	ISCO Occupation	X_o	$CR3_o$
3513	Computer network and systems technicians	4.41	11.7
3511	ICT operations technicians	4.32	12.4
1330	ICT service managers	4.10	13.1
2523	Computer network professionals	3.98	12.7
3512	ICT user support technicians	3.86	12.4
8132	Photographic products machine operators	3.66	15.9
4223	Telephone switchboard operators	3.56	14.6
7422	ICT installers and servicers	3.36	14.3
3514	Web technicians	3.25	13.3
4132	Data entry clerks	3.11	15.6
9623	Meter readers and vending-machine collectors	3.09	16.9
3133	Chemical processing plant controllers	3.04	18.0
8322	Car, taxi and van drivers	2.68	21.8
2153	Telecommunications engineers	2.57	17.1
1324	Supply, distribution and related managers	2.55	19.8
9621	Messengers, package deliverers and luggage porters	2.49	19.7
2513	Web and multimedia developers	2.44	19.5
3311	Securities and finance dealers and brokers	2.44	22.7
2521	Database designers and administrators	2.43	17.8
3252	Medical records and health information technicians	2.38	25.3
8183	Packing, bottling and labelling machine operators	2.36	18.0
2622	Librarians and related information professionals	2.35	20.9
4323	Transport clerks	2.23	24.5
8312	Railway brake, signal and switch operators	2.20	21.0
5244	Contact centre salespersons	2.17	20.7
3522	Telecommunications engineering technicians	2.13	19.4
2529	Database and network professionals n.e.c.	2.13	20.7
3135	Metal production process controllers	2.03	20.2
3114	Electronics engineering technicians	1.98	19.7
2522	Systems administrators	1.96	17.6

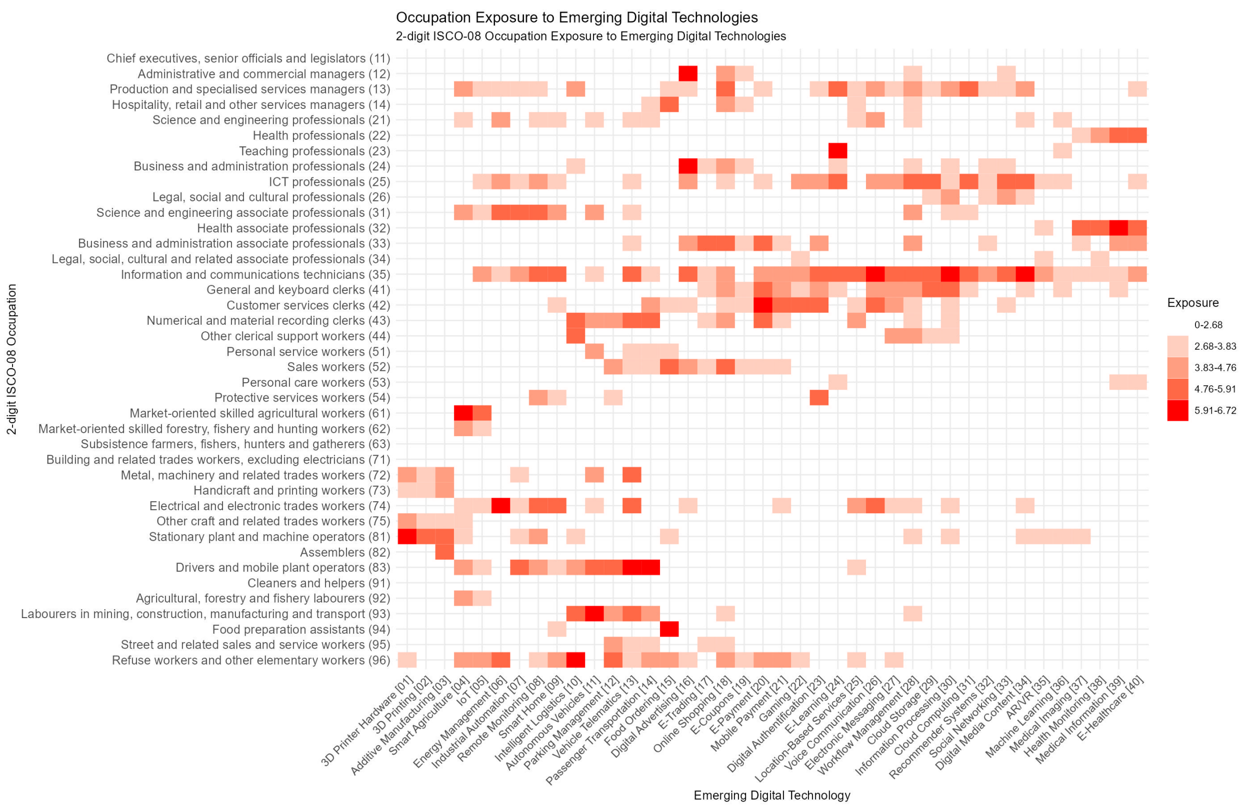
Notes: This table presents the top 30 4-digit ISCO-08 occupations ranked by exposure to all emerging digital technologies. Columns (from left to right) correspond to occupation code, occupation title, average exposure to emerging digital technologies, top-3 concentration ratio which represents the sum of top-3 technology exposure shares (in percent).

Table B.2: Top 30 Exposed 3-digit NACE Rev.2 Industries

Code	NACE Industry	X_i	CR3$_i$
26.3	Manufacture of communication equipment	6.28	9.7
26.2	Manufacture of computers and peripheral equipment	6.19	9.5
63.1	Data processing, hosting and related activities	5.88	10.0
62.0	Computer programming, consultancy and related activities	5.28	10.6
26.5	Manufacture of instruments and appliances for measuring	4.88	11.8
82.9	Business support service activities n.e.c.	4.83	11.5
28.2	Manufacture of other general-purpose machinery	4.71	12.7
63.9	Other information service activities	4.70	11.8
61.2	Wireless telecommunications activities	4.67	11.7
61.9	Other telecommunications activities	4.43	12.2
33.1	Repair of fabricated metal products, machinery and equipment	4.22	12.1
95.1	Repair of computers and communication equipment	4.11	12.2
79.9	Other reservation service and related activities	3.96	13.4
80.2	Security systems service activities	3.83	14.2
52.2	Support activities for transportation	3.59	16.0
27.9	Manufacture of other electrical equipment	3.50	15.1
61.1	Wired telecommunications activities	3.50	13.8
47.4	Retail sale of information and communication equipment	3.35	15.2
26.4	Manufacture of consumer electronics	3.30	13.1
28.9	Manufacture of other special-purpose machinery	3.26	18.3
27.1	Manufacture of electric motors, generators and transformers	2.95	18.3
82.2	Activities of call centres	2.81	16.2
80.1	Private security activities	2.78	16.6
26.1	Manufacture of electronic components and boards	2.76	16.8
17.2	Manufacture of articles of paper and paperboard	2.73	14.8
58.1	Publishing of books, periodicals and other publishing activities	2.69	17.4
27.3	Manufacture of wiring and wiring devices	2.38	17.7
18.2	Reproduction of recorded media	2.31	20.0
33.2	Installation of industrial machinery and equipment	2.30	20.6
82.1	Office administrative and support activities	2.14	18.8

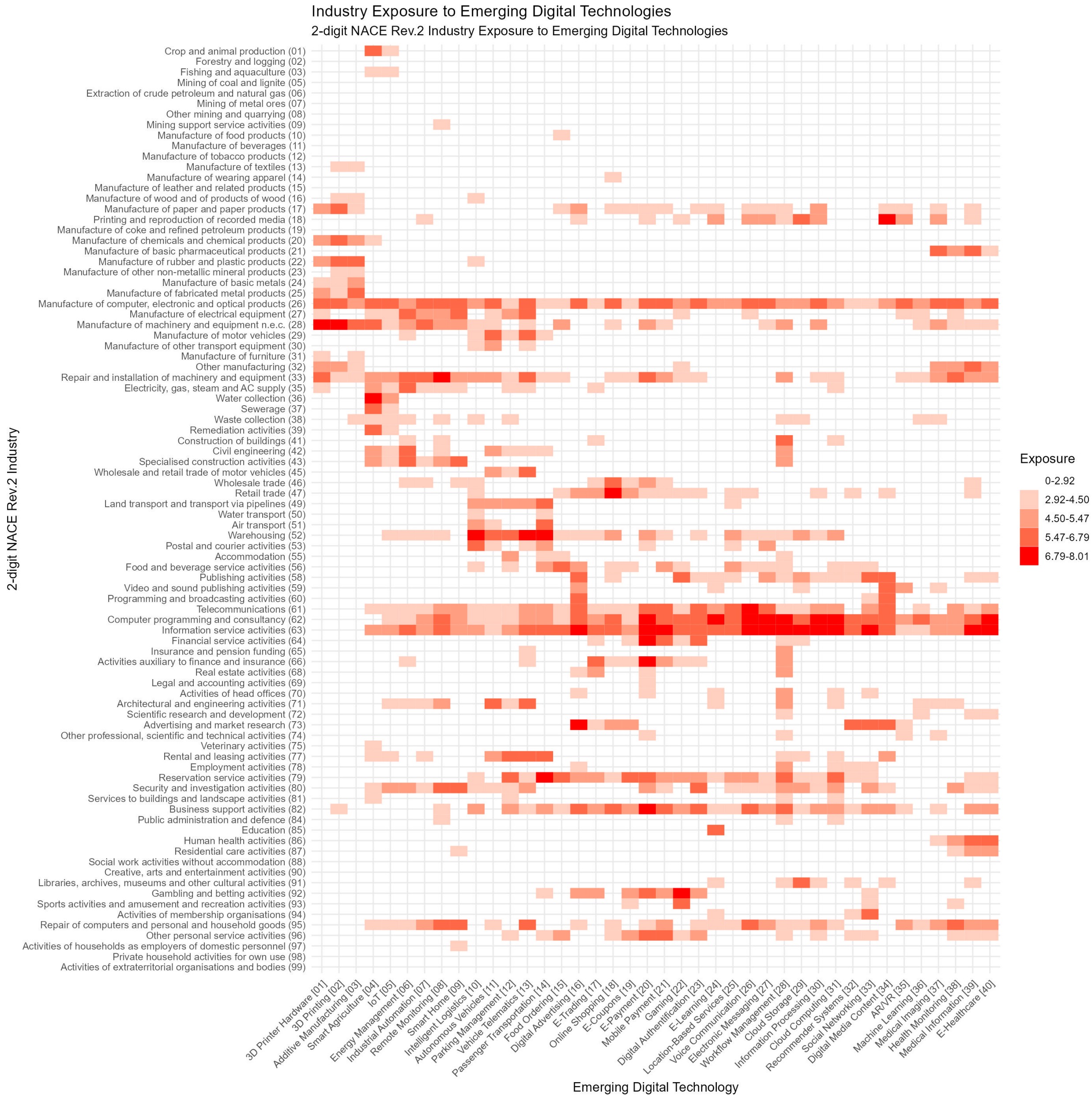
Notes: This table presents the top 30 3-digit NACE Rev.2 industries ranked by exposure to all emerging digital technologies. Columns (from left to right) correspond to industry code, industry title, average exposure to emerging digital technologies, top-3 concentration ratio which represents the sum of top-3 technology exposure shares (in percent).

Figure B.1: Occupation Exposure by Emerging Digital Technologies (2-digit ISCO-08)



Notes: Each cell shows the exposure of a 2-digit ISCO-08 occupation (row) to a given emerging digital technology (column). Exposure scores below the 80th percentile (0-2.68) are transparent, whereas the four other groups represent respectively the 80th (2.68-3.83), 90th (3.83-4.76), 95th (4.76-5.91), and 99th (5.91-6.72) percentile of the distribution.

Figure B.2: Industry Exposure by Emerging Digital Technologies (2-digit NACE Rev.2)



Notes: Each cell shows the exposure of a 2-digit NACE Rev.2 industry (row) to a given emerging digital technology (column). Exposure scores below the 80th percentile (0-2.92) are transparent, whereas the four other groups represent respectively the 80th (2.92-4.50), 90th (4.50-5.47), 95th (5.47-6.79), and 99th (6.79-8.01) percentile of the distribution.

C Employment Impact Appendix

In this Appendix, we provide additional information on the regional employment analysis in Section 5.

C.1 Employment Shares

Table C.1 presents the employment shares of our 10 sectors of activities averaged across all the European regions in 2010. The three largest sectors in Europe are the Public Sector (O-Q), accounting for an average of 23.9% of employment, Market Services (G-I), with an average share of 23.9%, and Industry (B-E), representing 18% on average. Subsequently, there is a group of sectors each contributing between 6% and 9% on average to employment, comprising Agriculture (A), Construction (F), and Professional, Scientific, Technical, Administration, and Support Service Activities (M-N). The remaining four sectors collectively account for 11.3% of employment. Notably, the Information and Communication sector (J), pivotal to emerging digital technologies, comprises only 2.6% of average regional employment in Europe. This figure is comparable to the Financial and Insurance Activities sector (K), which averages 2.8%.

Table C.1: Average Employment Share by Sector of Activities in 2010

	NACE Sector	Mean	SD
A	Agriculture	0.068	0.010
B-E	Industry, excluding Construction	0.179	0.006
F	Construction	0.076	0.000
G-I	Market Services, excluding Information and Communication	0.238	0.001
J	Information and Communication	0.026	0.000
K	Financial and Insurance Activities	0.028	0.000
L	Real Estate Activities	0.007	0.000
M-N	Professional, Scientific, Technical, Administration and Support Service Activities	0.083	0.001
O-Q	Public Administration, Defence, Education, Human Health and Social Work Activities	0.237	0.004
R-U	Other Services	0.053	0.000

Notes: This table presents the employment share by sector of activities averaged across all the European regions in 2010. Regions are weighted by population in 2010. The first column indicates the 1-digit NACE codes, the second column is the name of the NACE sector, the third column is the average employment share in 2010, and the fourth column gives the standard errors.

C.2 Placebo Estimates

To provide further evidence supporting the validity of the shift-share approach, we conduct a placebo analysis. We estimate the effect of regional exposure to emerging digital technologies on the employment-to-population ratio change during the pre-period, specifically between 2002 and 2009. Table C.2 presents the results.

Consistently, we observe no effect of regional exposure on the pre-period for all demographic groups, with the exception of high-skilled workers. The presence of a significant result for this group is crucial, as these are the individuals who produce and develop emerging digital technologies. Thus, the significant and positive estimate suggests the reverse causality where employment growth among high-skilled workers, who are central to the development of these technologies, tends to be higher in regions that are more exposed to them.

Since the placebo analysis is conducted on the pre-period (2002-2009), employment data are not available for the year 2002 in 62 regions. This necessitates restricting our sample to 258 regions to conduct the placebo analysis. In Table C.3, we present the estimates of the baseline specification, as obtained in Table 4, for the restricted sample. These estimates are close to those obtained in Table 4. This suggests that our results in the placebo analysis are not driven by the exclusion of specific regions.

Table C.2: Placebo Estimates of the Effect of Emerging Digital Technologies on Regional Employment by Demographic Groups

	Δ Emp-to-pop. ratio (2002-2009) $\times$ 100							
	All	Gender		Age		Skill		
	Total	Female	Male	Y15-24	Y25-64	Low	Mid	High
Exposure to Emerging Technologies	-0.080	-0.012	-0.068	0.001	-0.090	0.078	-0.058	0.257^{***}
	(0.179)	(0.105)	(0.078)	(0.039)	(0.144)	(0.104)	(0.046)	(0.071)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓	✓	✓
Industry share	✓	✓	✓	✓	✓	✓	✓	✓
Emp-to-pop. ratio in 2012	50.93	22.25	28.69	5.70	45.23	14.22	24.52	11.39
Change (in %)	-0.16	-0.05	-0.24	0.02	-0.20	0.55	-0.23	2.26
R^2	0.717	0.713	0.768	0.676	0.696	0.756	0.768	0.630
Adj. R^2	0.672	0.668	0.732	0.625	0.649	0.717	0.731	0.571
Num. obs.	258	258	258	258	258	258	258	258

Notes: This table presents the placebo estimates of exposure to emerging digital technologies on regional employment by demographic groups. It presents the coefficients measuring the effect of regional exposure to emerging technologies, constructed as shift-shares and standardized, on changes in the employment-to-population ratio between 2002 and 2009 in European regions, expressed in percentage points, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Regressions are weighted by population in 2010. All columns include country fixed effects; demographics controls in 2010, including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels; and the share of employment in the industry sector. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses are derived following the AKM0 inference procedure from Adão et al. (2019).

Table C.3: Effect of Emerging Digital Technologies on Regional Employment by Demographic Groups (Placebo Sample)

	Δ Emp-to-pop. ratio (2012-2019) $\times$ 100							
	All	Gender		Age		Skill		
	Total	Female	Male	Y15-24	Y25-64	Low	Mid	High
Exposure to Emerging Technologies	0.931***	0.547***	0.384***	0.055**	0.878***	0.041	−0.478***	1.374***
	(0.085)	(0.071)	(0.041)	(0.033)	(0.101)	(0.090)	(0.053)	(0.027)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓	✓	✓
Industry share	✓	✓	✓	✓	✓	✓	✓	✓
Emp-to-pop. ratio in 2012	50.93	22.25	28.69	5.70	45.23	14.22	24.52	11.39
Change (in %)	1.83	2.46	1.34	0.97	1.94	0.29	−1.95	12.06
R^2	0.795	0.717	0.757	0.428	0.778	0.793	0.788	0.734
Adj. R^2	0.762	0.673	0.718	0.338	0.743	0.761	0.754	0.692
Num. obs.	258	258	258	258	258	258	258	258

Notes: This table presents the estimates of exposure to emerging digital technologies on regional employment by demographic groups. It presents the coefficients measuring the effect of regional exposure to emerging technologies, constructed as shift-shares and standardized, on changes in the employment-to-population ratio between 2012 and 2019 in European regions, expressed in percentage points, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Regressions are weighted by population in 2010. All columns include country fixed effects; demographics controls in 2010, including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels; and the share of employment in the industry sector. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors between parentheses are derived following the AKM0 inference procedure from Adão et al. (2019).

C.3 Assessing the Impact of Emerging Digital Technology Families

We conduct the analysis at the level of technology families. We use the same shift-share design to calculate the regional exposure to technology family X_r^K, defined as

$$X_r^K = \sum_j l_{rj} X_j^K,$$

where l_{rj} is the employment share of sector j in region r, and X_j^K is the exposure of sector j to technology family K, which is computed as the average sectoral exposure across technologies within the same family (i.e., $X_j^K = \frac{1}{|K|}\sum_{k\in K} X_j^k$).

Figure C.1 illustrates the geographic distribution of exposure to the 9 families of emerging digital technology. Regional exposure is standardized at the family level to facilitate comparisons and account for variations in exposure magnitudes across different technology families.

Exposure to emerging digital technologies exhibits significant variation across European regions and between technology families. For instance, regions with the highest exposure to tangible technologies, such as 3D Printing and Embedded Systems, are predominantly located in Central and Eastern European countries, as well as in certain areas of Southern Europe, including Northern Portugal and Turkey. These are the regions with the highest manufacturing shares. Conversely, Western and Northern European countries show greater exposure to Computer Vision and HealthTech, which correlates with their more service-oriented economies and digitized healthcare systems.

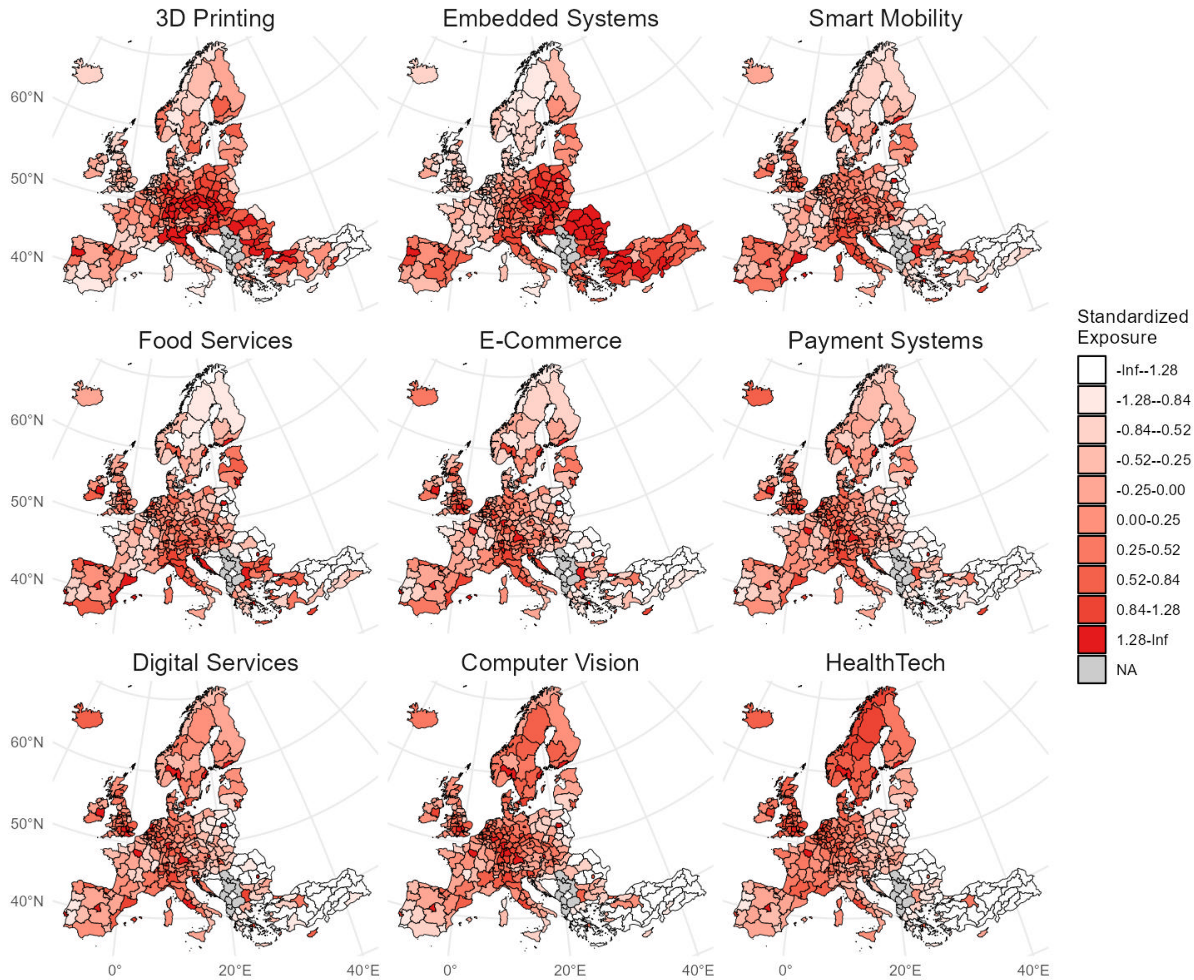
Furthermore, spatial differences in exposure are also evident within countries, characterized by disparities between rural and urban areas. Exposure to E-Commerce, Payment Systems, and Digital Services is predominantly concentrated in capital cities and financial hubs. In contrast, exposure to Smart Mobility and Food Services is relatively more pronounced in the rural regions of Western countries, such as France, Italy, Spain, and the United Kingdom.

We estimate the impact of the regional exposure to a specific emerging technology family on the employment-to-population ratio using an empirical specification analogous to that of Equation (15). However, instead of using the exposure to all technologies X_r, we focus on the regional exposure to a particular family X_r^K. More specifically, the empirical specification is:

$$\Delta Y_r = \alpha + \beta_K X_r^K + \gamma_K X_r^{-K} + Z\delta + \phi_{c(r)} + u_r, \tag{17}$$

where X_r^{-K} is regional exposure to all *other* emerging digital technologies. This latter variable is constructed as a shift–share variable, similar to that of Equation (14), but specifically excluding the exposure from the technology family of interest K. For interpretability, we standardize our variable of interest X_r^K.

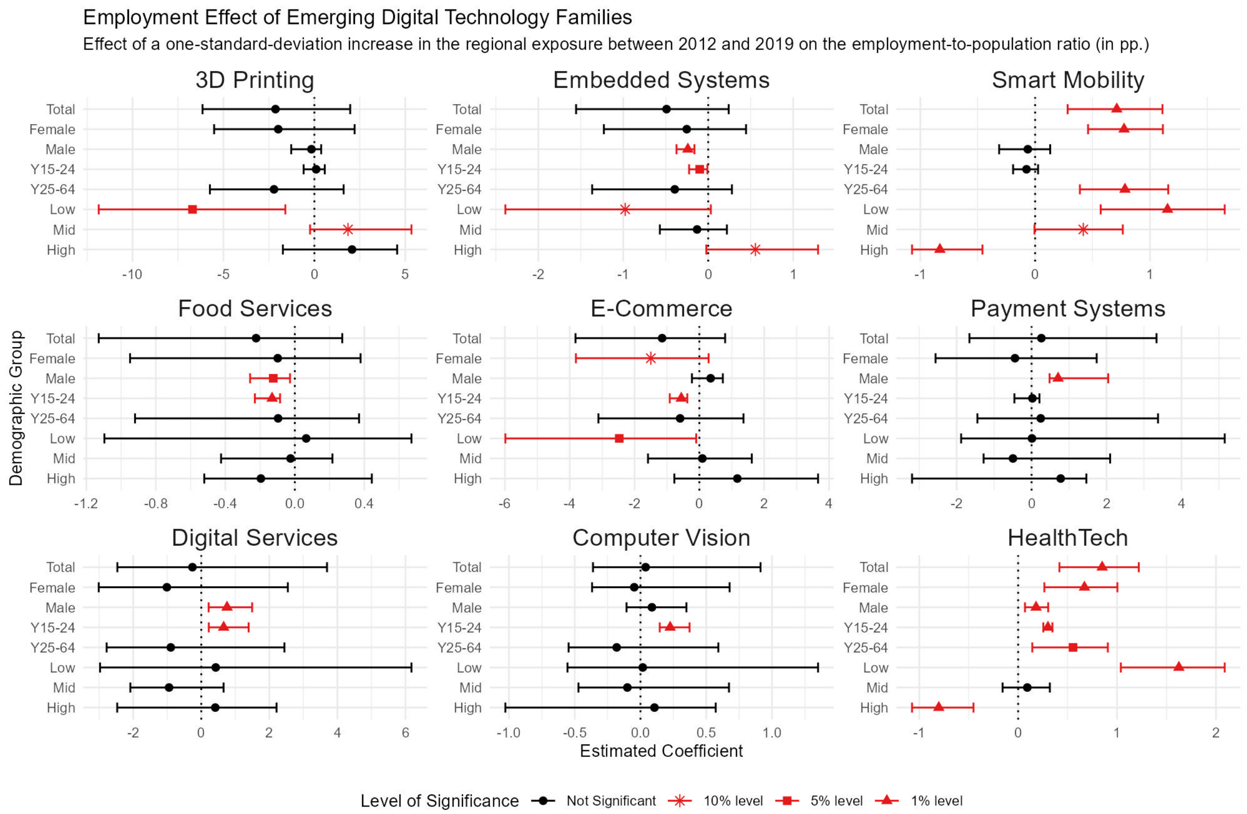
Figure C.1: Geographic Distribution of Regional Exposure to Families of Emerging Digital Technologies across Europe from 2012 to 2019



Notes: This figure illustrates the geographic distribution of exposure to families of emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technology families from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

Our estimated coefficient of interest, denoted as $\hat{\beta}_K$, represents the employment effect, measured in pp. change, of a one-standard-deviation increase in the regional exposure to a specific emerging technology family K, conditional on the regional exposure to all *other* families of emerging digital technologies. This empirical approach allows us to identify the causal effect of technology family K on employment at the regional level, net of the overall effect of emerging digital technologies. Our approach is consistent with methodologies applied in the recent literature, which assess the impact of a particular technology, such as robots, on employment, while also accounting for exposure to contemporaneous technologies, such as ICT (see, for example, Acemoglu and Restrepo 2020; Dauth et al. 2021).

Figure C.2: Employment Effect of Emerging Digital Technology Families



Notes: This figure the coefficients measuring the effect of regional exposure to emerging digital technology families, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies, also constructed as a shift-share.

Figure C.2 displays the estimated coefficients, along with their corresponding 95% AKM0 confidence intervals, for the employment effects of emerging digital technology families for the different demographic groups. The figure is interpreted as follows. Each panel corresponds to a technology family. The vertical axis lists the demographic groups, while the horizontal axis depicts the estimated coefficients.

Smart Mobility has a positive and significant impact on total employment. This positive impact is driven by the increase in employment of low- and middle-skilled workers as well as female and mature workers. We find no effect on young workers and male workers. However, we find a negative impact of Smart Mobility on high-skilled workers.

HealthTech also has a positive and significant impact on total employment. Similar to the former technology, the employment of low-skilled, female, and mature workers increases with exposure to HealthTech. Additionally, male and young workers are also positively impacted.

While we find no effect on middle-skilled workers, we find a negative impact on high-skilled workers.

Although we do not find any effect of Embedded Systems on the total employment-to-population ratio, we find that regional exposure to them reduces the employment of young, male, and low-skilled workers. The opposite signs for the estimates of low- and high-skilled workers (although significant at the 10% level) suggest that Embedded Systems are skilled-biased.

We find no effect on the total employment-to-population ratio for the other technology families. However, we observe positive employment effects on specific demographic groups. For male workers, we find a positive effect of Payment Systems and Digital Services. For young workers, we find positive effects of Computer Vision and Digital Services.

Conversely, some technology families have a negative impact on certain demographic groups. Food Services harm the employment of male and young workers, whereas E-Commerce reduces the employment of young and low-skilled workers, as well as female workers (although significant at the 10% level). Lastly, we find a negative and significant effect of 3D Printing on low-skilled workers.

C.4 Geographic Distribution of Regional Exposure to Individual Emerging Digital Technologies

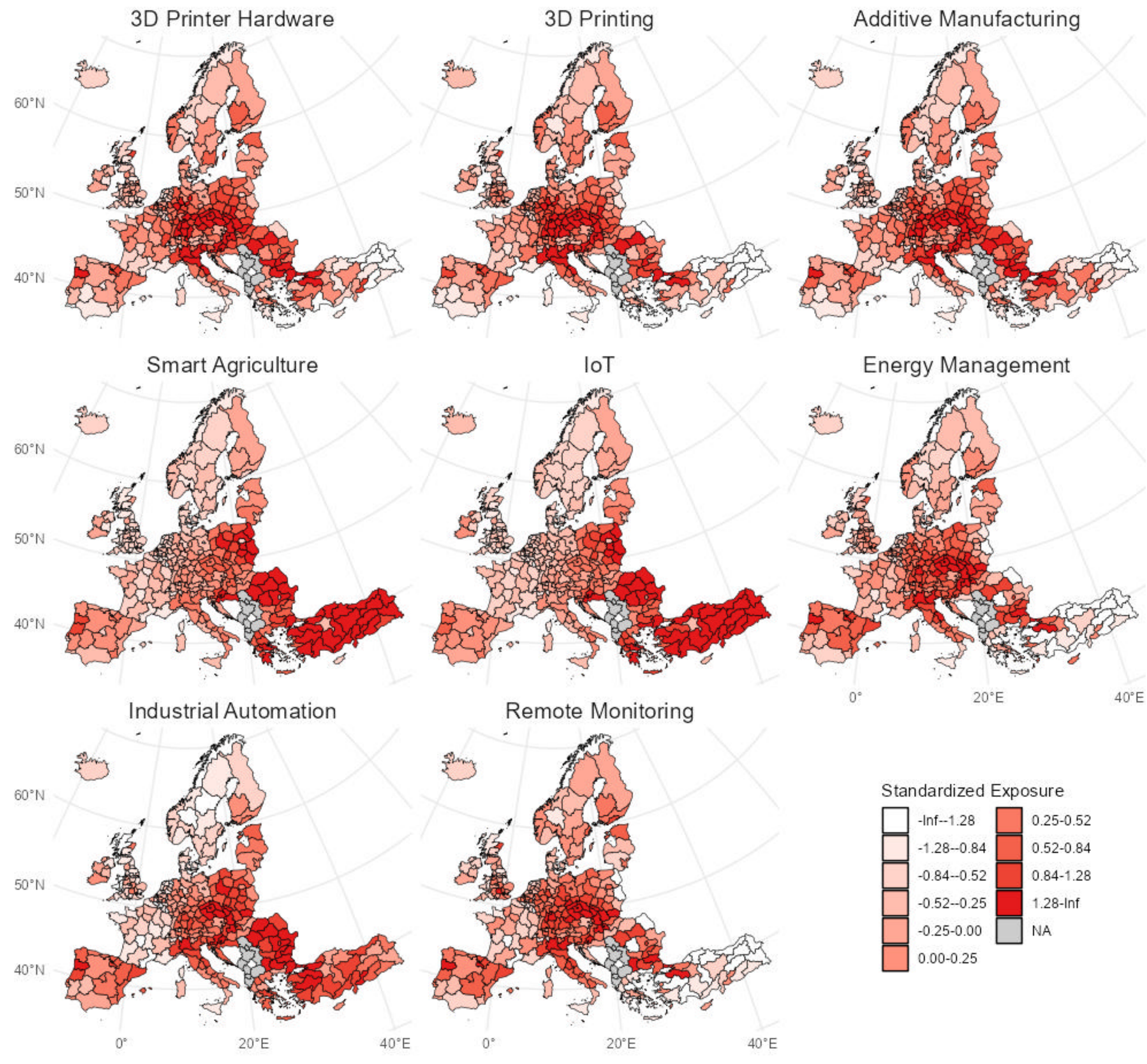
Figures C.3 to C.7 present the geographic distribution of regional exposure to individual emerging digital technologies, constructed as a shift-share. Regional exposure scores are standardized to allow comparability between technologies.

C.5 Individual Effects of Emerging Digital Technologies

??

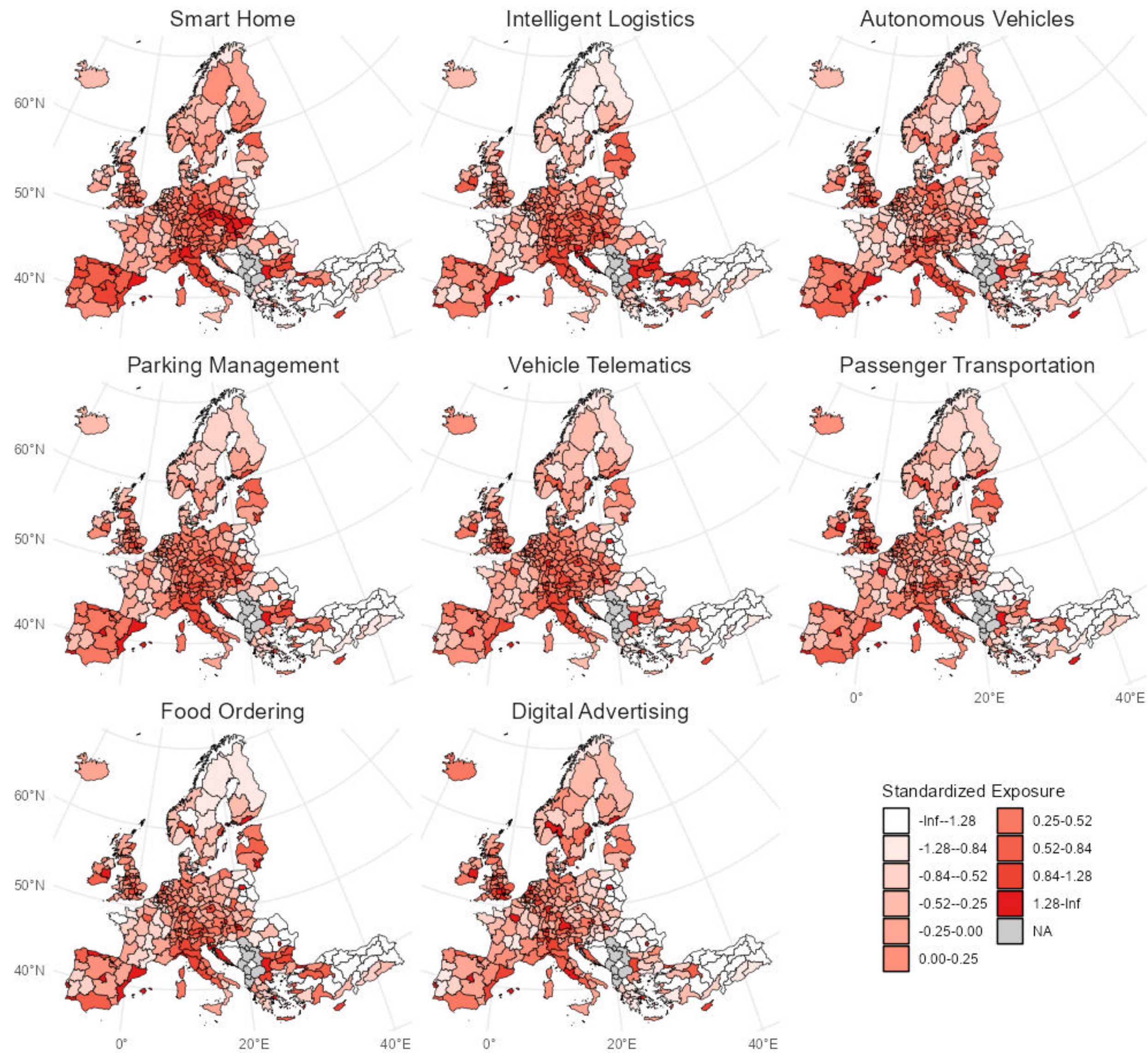
Figures C.8 to C.12 present the effect of individual emerging digital technologies.

Figure C.3: Geographic Distribution of Exposure to Emerging Digital Technologies (1/5)



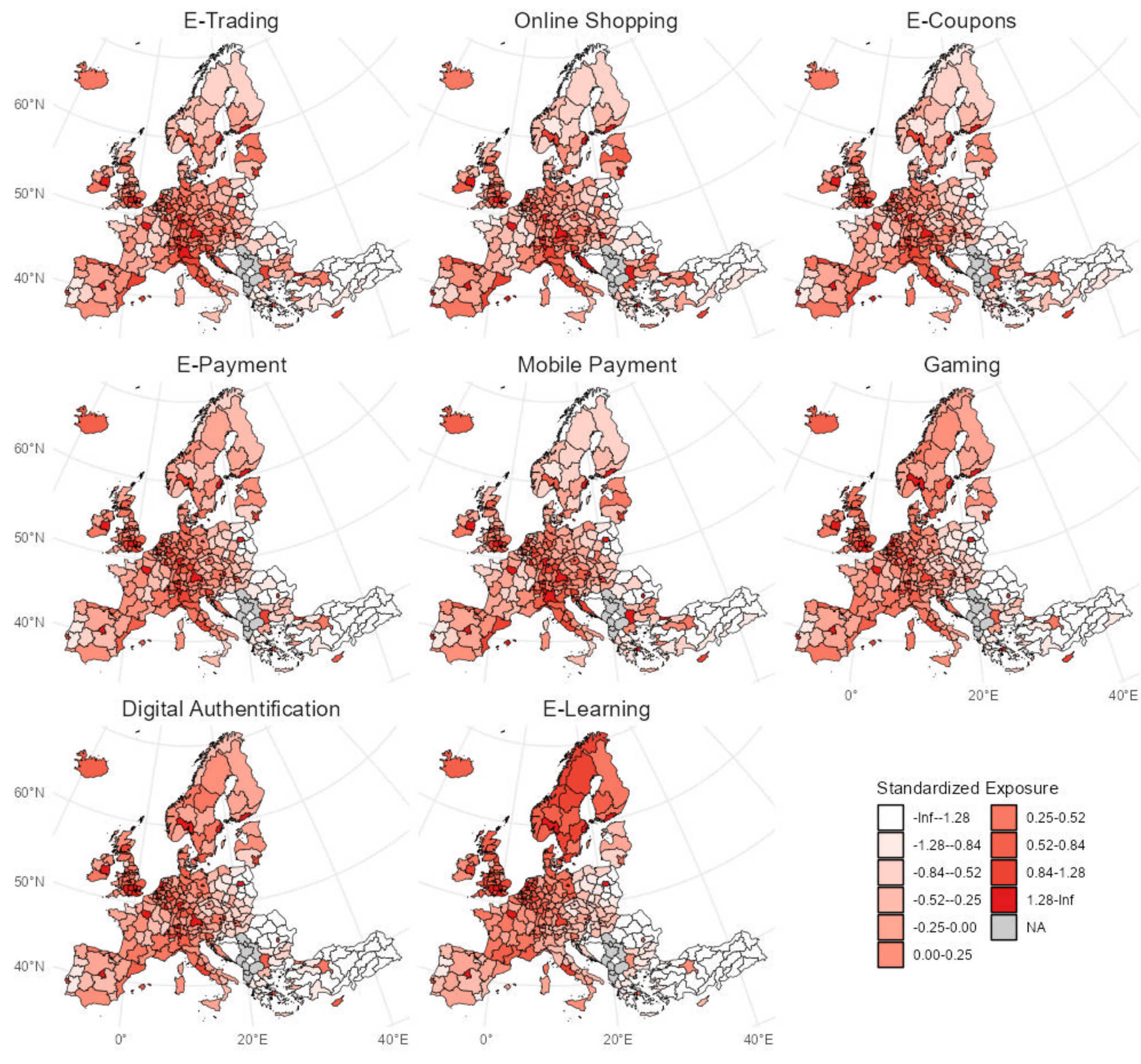
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

Figure C.4: Geographic Distribution of Exposure to Emerging Digital Technologies (2/5)



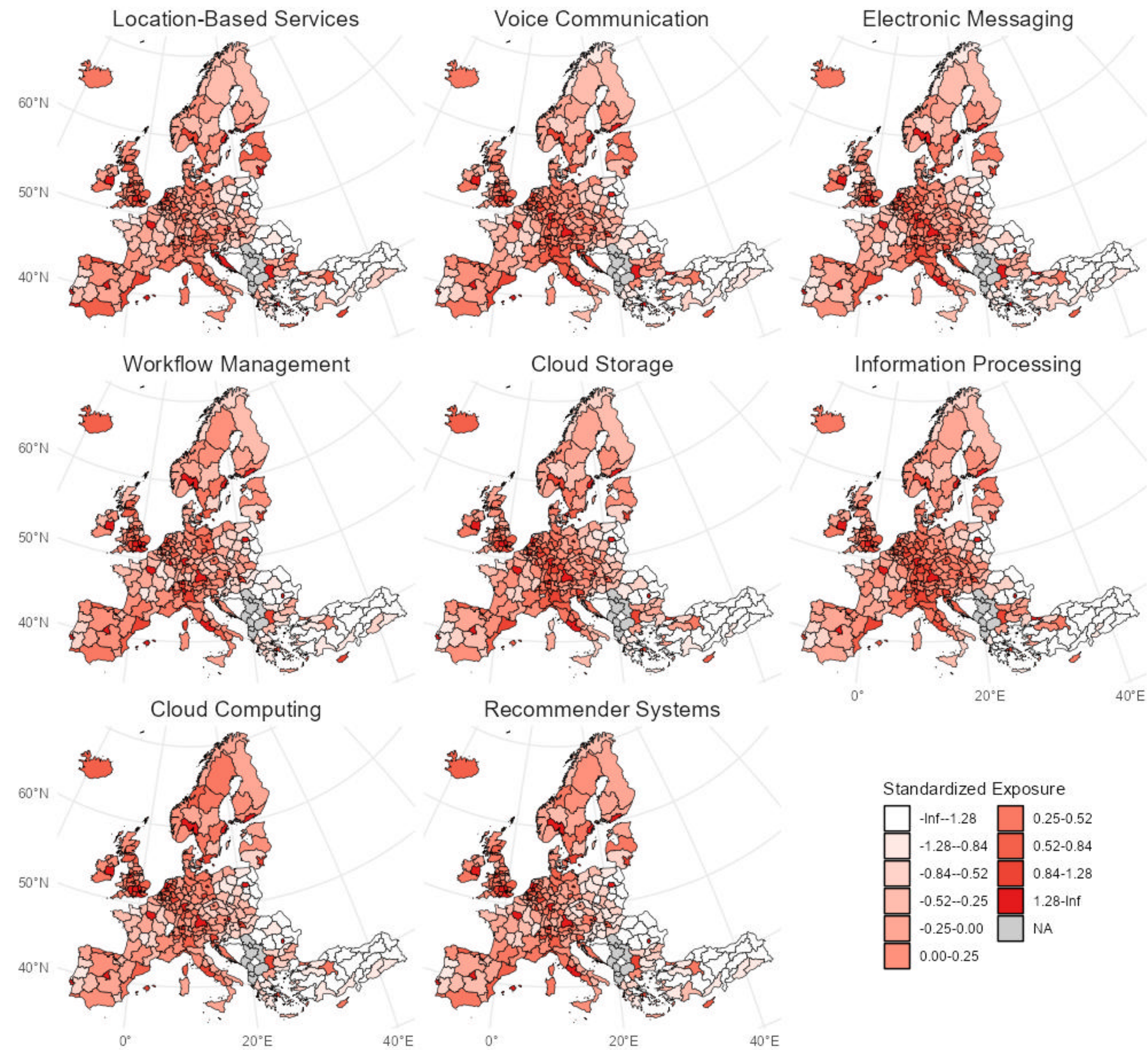
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

Figure C.5: Geographic Distribution of Exposure to Emerging Digital Technologies (3/5)



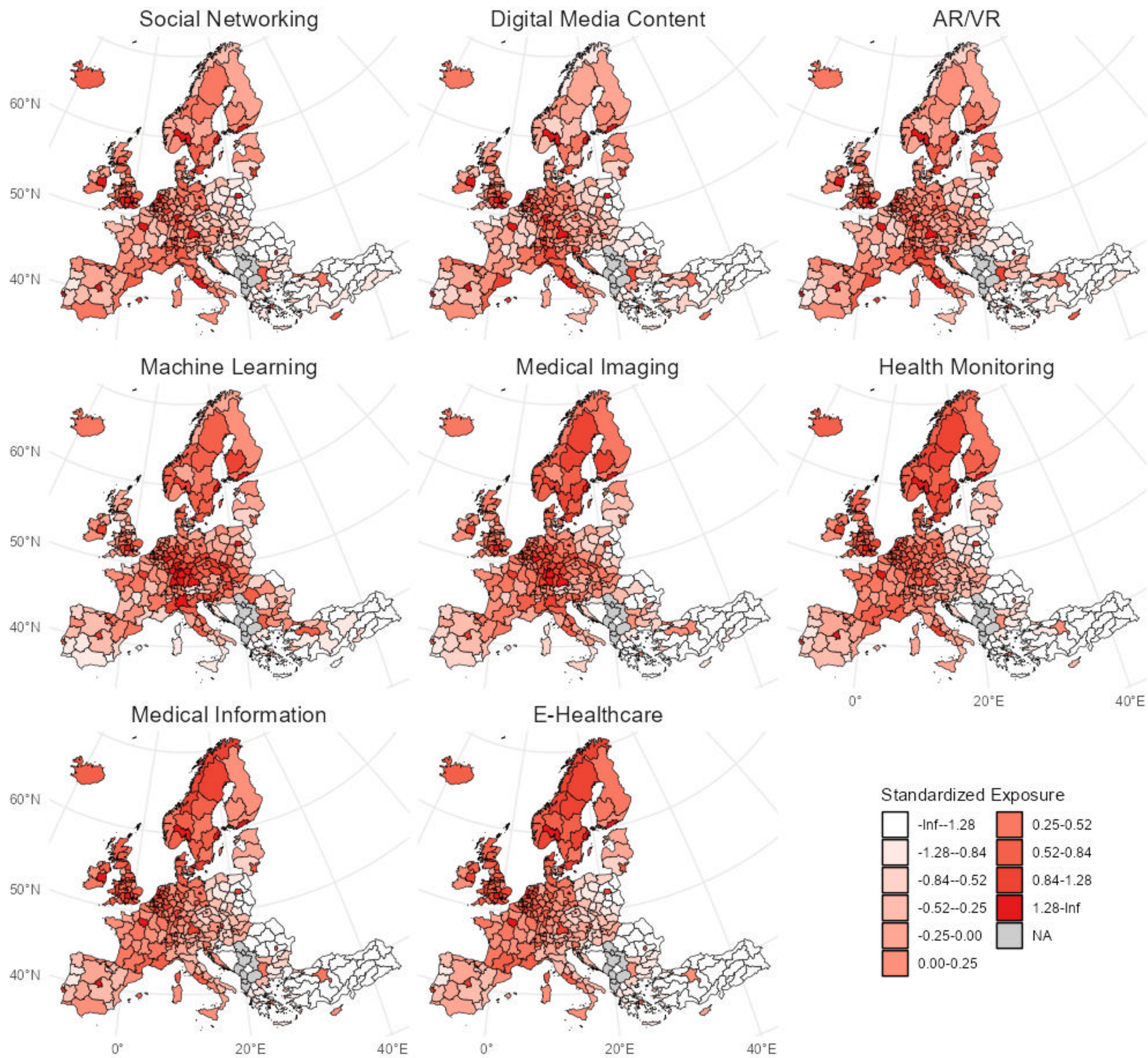
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

Figure C.6: Geographic Distribution of Exposure to Emerging Digital Technologies (4/5)



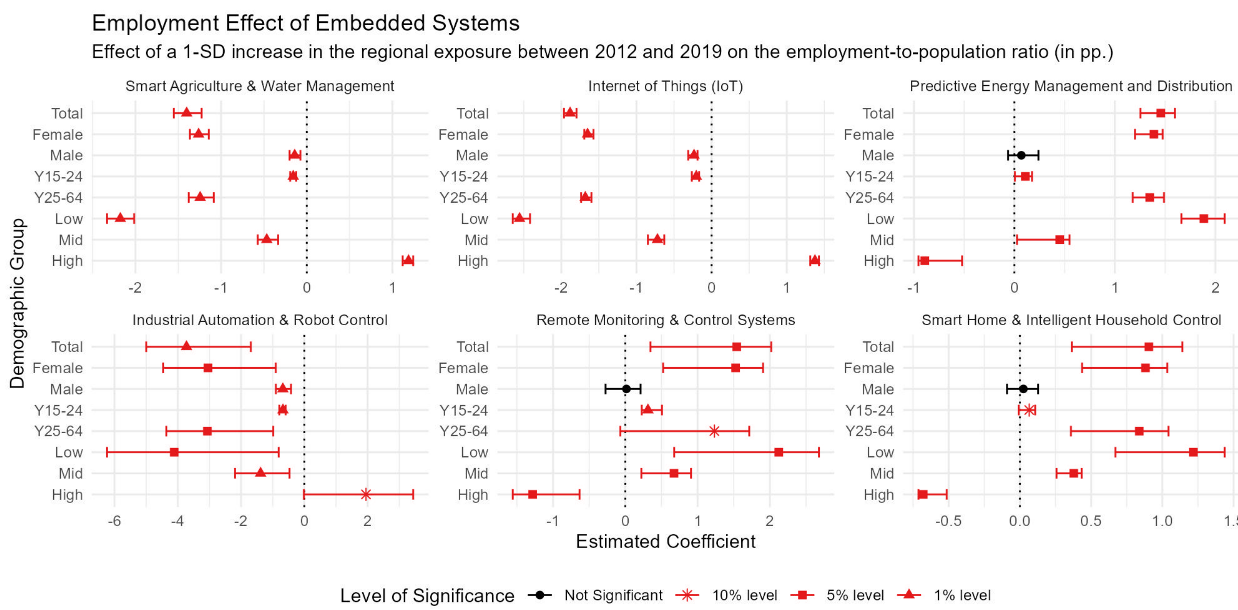
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

Figure C.7: Geographic Distribution of Exposure to Emerging Digital Technologies (5/5)



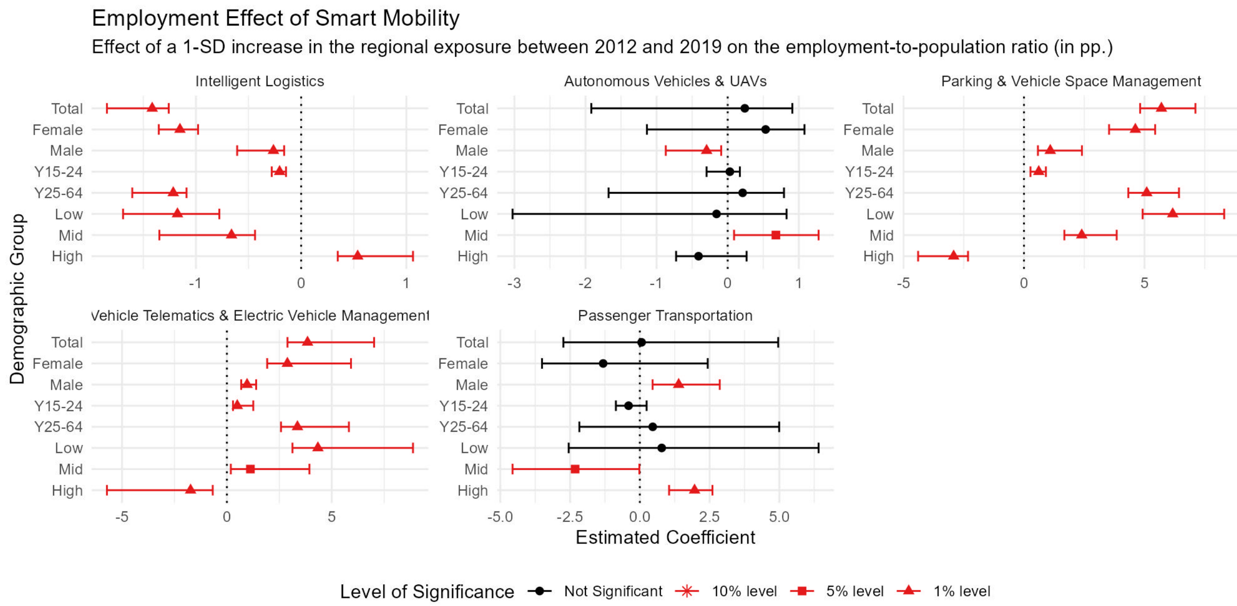
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 2012 to 2019. Regions are categorized into deciles. Regions are shaded according to their exposure level, with the legend indicating the range of exposure. Areas not applicable (NA) are marked in grey.

Figure C.8: Employment Effect of Embedded Systems



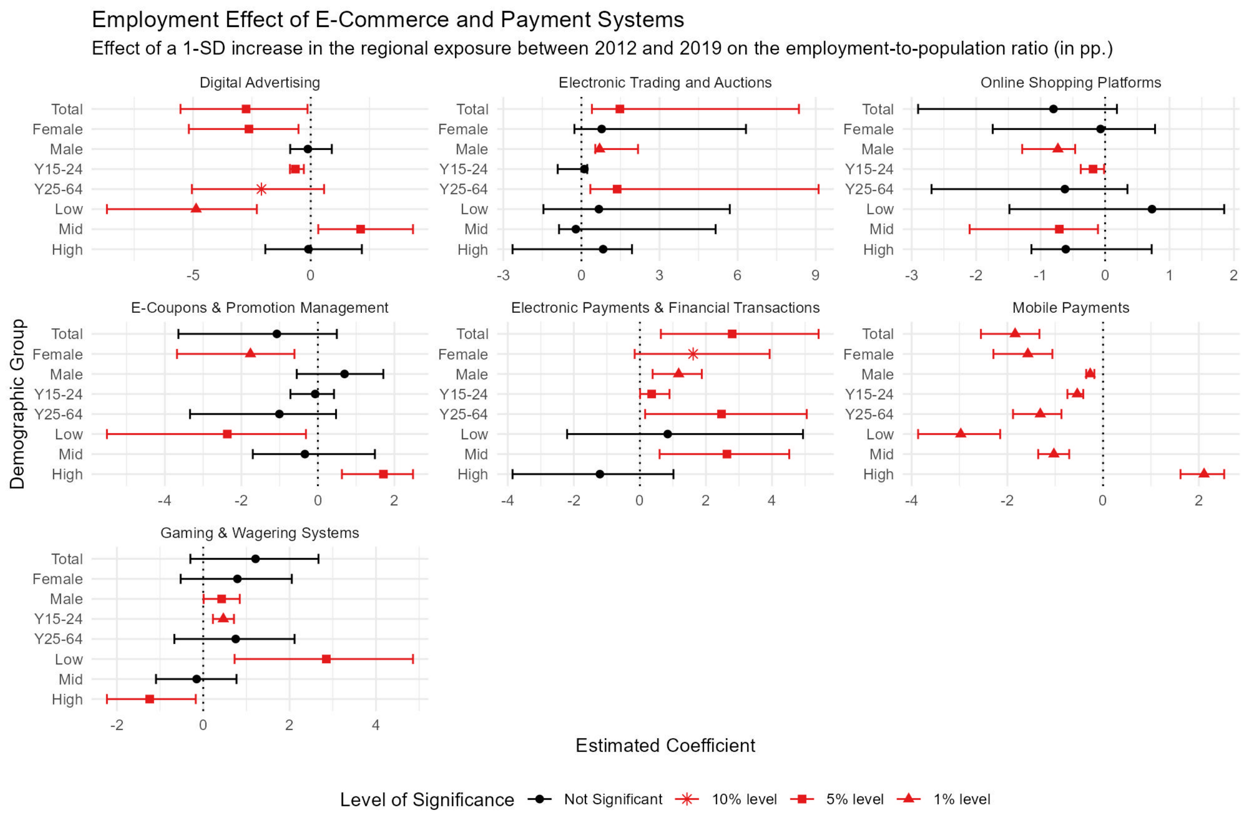
Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares.

Figure C.9: Employment Effect of Smart Mobility



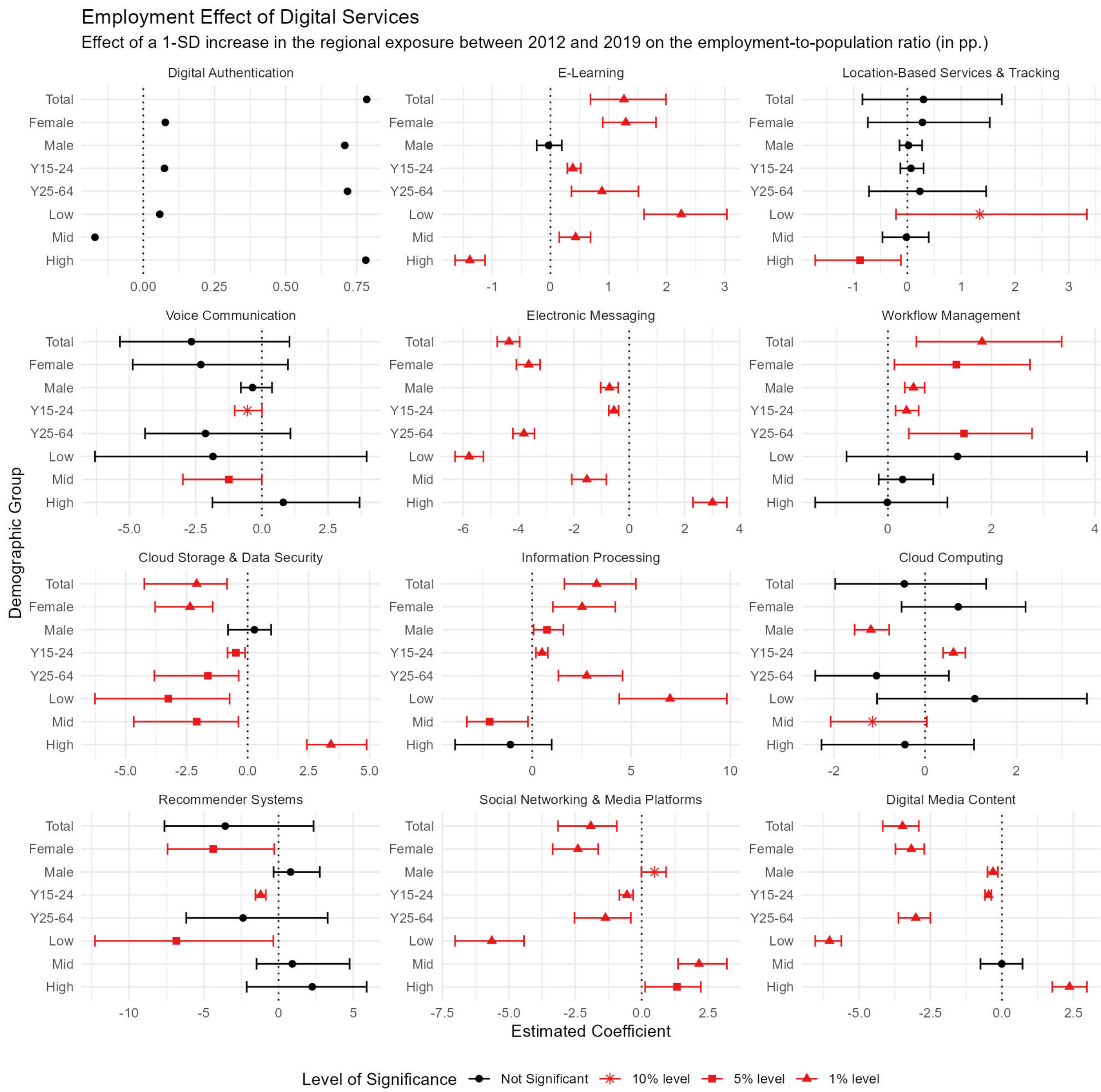
Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares.

Figure C.10: Employment Effect of E-Commerce and Payment Systems



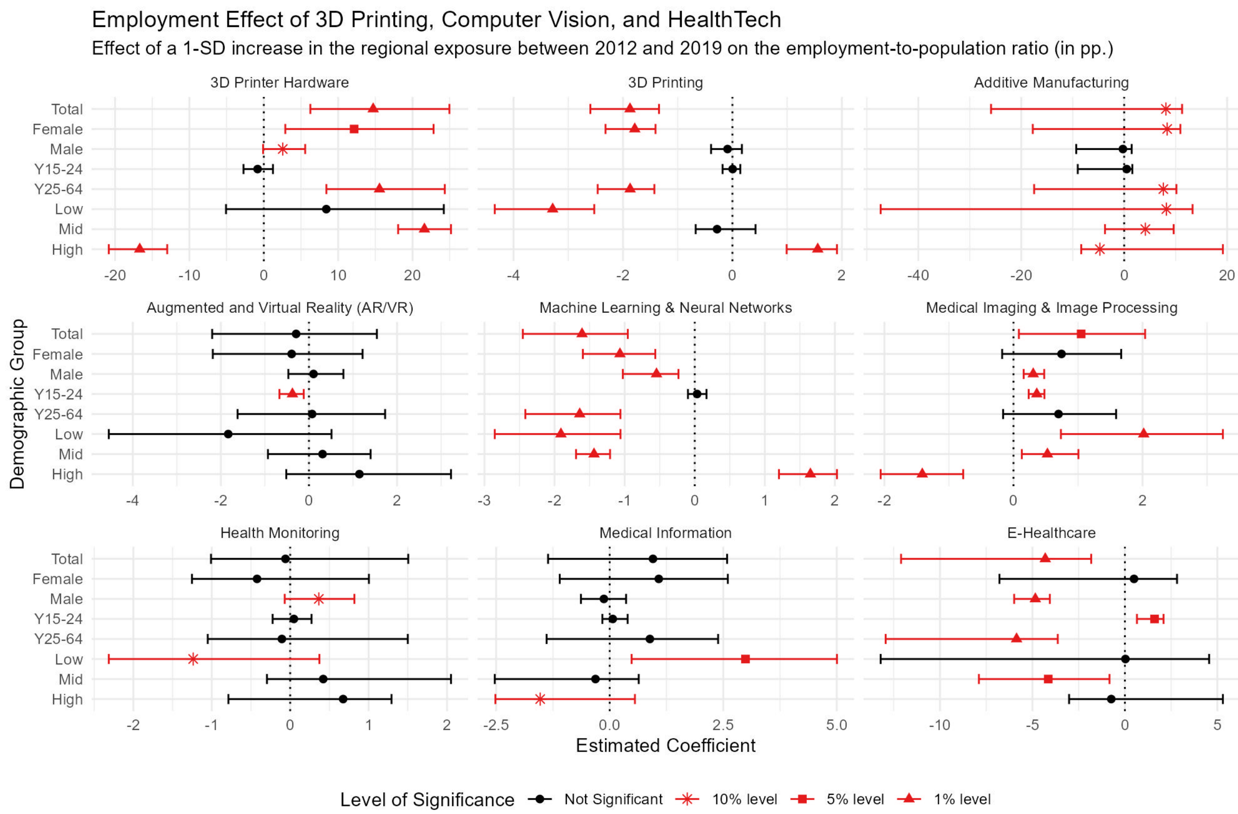
Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares.

Figure C.11: Employment Effect of Digital Services



Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares. Confidence intervals for Digital Authentification are not displayed since the standard errors cannot be computed under the AKM0 inference procedure.

Figure C.12: Employment Effect of 3D Printing, Computer Vision, and HealthTech



Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all workers, female and male workers, young (aged 15-24) and mature (aged 25-64) workers, and low-, middle-, and high-skilled workers. Each panel represents a technology. The confidence intervals are reported at the 5% significance level using the AKM0 inference procedure from Adão et al. (2019). Regressions are weighted by population in 2010 and the set of control variables include country fixed effects, the sum of exposure shares as a control, demographics controls in 2010 (including the logarithm of population, the proportion of females, the proportion of the population aged over 65, the proportions of the population with secondary and tertiary education levels), the share of employment in the industry sector, and the regional exposure to all other emerging digital technologies within the same technology family and outside, both also constructed as shift-shares.

Online Appendix

The Employment Impact of Emerging Digital Technologies

Ekaterina Prytkova, Fabien Petit, Deyu Li, Sugat Chaturvedi, Tommaso Ciarli

D Additional Figures and Tables

Tables OA.1 and OA.2 present the structured queries discussed in Appendix A.1 that retrieve the patent sample.

Table OA.1: The structured patent queries in Derwent Innovation Index database (1/2)

1. **Process and machine control in production.** MAN=(T06* NOT P36*) NOT DC=(X26 OR X27 OR W07) OR MAN=(T01-J07 OR T01-J07A* OR T01-J07B OR T01-J07B1 NOT P36*) NOT DC=(X26 OR X27 OR W07)
2. **Process and workflow control in services.** IP=(G06Q*) OR MAN=(T01-n01a* OR T01-N01B3* OR T01-N01D3A OR T01-n01e* OR T01-J05A* OR T01-J06A* OR S05-D06A*)
3. **Additive manufacturing.** MAN=(T01-J07B3* OR X25-A08*) OR IP=(B33Y*)
Technology Constraints.
Networking MAN=(W01-A* OR T01-N* OR W05-D06E* OR W05-D06F* OR W05-D07* OR T06-A11* NOT W05-D07A* NOT W05-D07C*) OR IP=(H04L* OR H04W*) OR TS=(MWSN OR WSN OR (SENSOR NEAR/1 (CLUSTER$ OR NETWORK$ OR NODE$)) OR (NETWORK NEAR/3 (TRANSDUCER$ OR PROBE$)) OR (DETECTOR$ NEAR/0 NETWORK$) OR (METER NEAR/1 NETWORK$)) OR IP=(H04B-001/00 OR H04B-001/02 OR H04B-001/03 OR H04B-001/034 OR H04B-001/036 OR H04B-001/04 OR H04B-001/06 OR H04B-001/08 OR H04B-001/10 OR H04B-001/12 OR H04B-001/14 OR H04B-001/16 OR H04B-001/18 OR H04B-001/20 OR H04B-001/22 OR H04B-001/24 OR H04B-001/26 OR H04B-001/28 OR H04B-001/30 OR H04B-001/38 OR H04B-001/3805 OR H04B-001/3816 OR H04B-001/3818 OR H04B-001/3822 OR H04B-001/3827 OR H04B-001/3877 OR H04B-001/3883 OR H04B-001/3888 OR H04B-001/40 OR H04B-001/401 OR H04B-001/403 OR H04B-001/405 OR H04B-001/408 OR H04B-001/44 OR H04B-001/46 OR H04B-001/48 OR H04B-001/50 OR H04B-001/52 OR H04B-001/525 OR H04B-001/54 OR H04B-001/56 OR H04B-001/58 OR H04B-001/59 OR H04B-001/60 OR H04B-001/62 OR H04B-001/64 OR H04B-001/66 OR H04B-001/68 OR H04B-001/72 OR H04B-001/74 OR H04B-001/76 OR H04B-001/59 OR G01S-013/74 OR G01S-013/75 OR G01S-013/76 OR G01S-013/78 OR G01S-013/79 OR G01S-013/82 OR G01S-013/84 OR G01V-015*) OR TS=(D2D OR IOT OR M2M OR M2MI OR MTC OR MTM OR INTER-VEHIC* OR DEVICE-2-DEVICE OR DEVICE-TO-DEVICE OR MACHINE-TO-MACHINE OR MACHINE-2-MACHINE OR MACHINE-TYPE-COMMUNICATION$ OR PEER-TO-PEER OR P2P OR (VEHICLE-TO- NEAR/0 (ANYTHING OR SERVER OR SOMETHING)) OR (INTER NEAR/0 (VEHICLE OR CAR)) OR (INTERNET NEAR/1 (THINGS OR EVERYTHING)) OR (WEB NEAR/1 THING$) OR (UBIQUITOUS NEAR/0 COMPUT*) OR (AMBIENT NEAR/1 INTELLIGENCE) OR (INTER-VEHIC* NEAR/0 COMMUNIC*))
Data acquisition TS=(MWSN OR WSN OR (SENSOR NEAR/1 (CLUSTER$ OR NETWORK$ OR NODE$)) OR (NETWORK NEAR/3 (TRANSDUCER$ OR PROBE$)) OR (DETECTOR$ NEAR/0 NETWORK$) OR (METER NEAR/1 NETWORK$)) OR IP=(G01S-013/74 OR G01S-013/75 OR G01S-013/76 OR G01S-013/78 OR G01S-013/79 OR G01S-013/82 OR G01S-013/84 OR G06T-011* OR G06T-013* OR G01V-015*)
Data management IP=(H04L-009* OR H04W-012*)

Notes:

Figures ref display the most exposed tasks to each technology by 1-digit ISCO-08 group.

Table OA.2: The structured patent queries in Derwent Innovation Index database (2/2)

AI and intelligent systems
MAN=(T01-J16* OR T06-A05* OR T06-A07*) OR IP=(G06K-009/00 OR G06K-009/03 OR G06K-009/18 OR G06K-009/46 OR G06K-009/48 OR G06K-009/50 OR G06K-009/52 OR G06K-009/54 OR G06K-009/56 OR G06K-009/58 OR G06K-009/60 OR G06K-009/62 OR G06K-009/64 OR G06K-009/66 OR G06K-009/68 OR G06K-009/70 OR G06K-009/72 OR G06K-009/74 OR G06K-009/76 OR G06K-009/78 OR G06K-009/80 OR G06K-009/82 OR G06T-007/30 OR G06T-007/32 OR G06T-007/33 OR G06T-007/35 OR G06T-007/37 OR G06T-007/38 OR G06T-007/40 OR G06T-007/41 OR G06T-007/42 OR G06T-007/44 OR G06T-007/45 OR G06T-007/46 OR G06T-007/48 OR G06T-007/49 OR G06T-007/50 OR G06T-007/507 OR G06T-007/514 OR G06T-007/521 OR G06T-007/529 OR G06T-007/536 OR G06T-007/543 OR G06T-007/55 OR G06T-007/557 OR G06T-007/564 OR G06T-007/571 OR G06T-007/579 OR G06T-007/586 OR G06T-007/593 OR G06T-007/60 OR G06T-007/62 OR G06T-007/64 OR G06T-007/66 OR G06T-007/68 OR G08G* OR G06N-003/00 OR G06N-003/02 OR G06N-003/04 OR G06N-003/06 OR G06N-003/063 OR G06N-003/067 OR G06N-003/08 OR G06N-003/10 OR G06N-003/12 OR G10L-025/63 OR G10L-025/66 OR G06N5* OR G10L15* OR G10L17*) OR TS=(((ARTIFIC* OR COMPUTATION*) NEAR/1 INTELLIGEN*) OR (NEURAL NEAR/1 NETWORK*) OR (BAYES* NEAR/1 NETWORK$) OR (CHATBOT$) OR (DATA NEAR/1 MINING) OR (DECISION NEAR/1 MODEL*) OR (DEEP NEAR/1 LEARNING*) OR (GENETIC NEAR/1 ALGORITHM$) OR ((INDUCTIVE NEAR/1 LOGIC) NEAR/0 PROGRAM*) OR (MACHINE NEAR/1 LEARNING) OR ((NATURAL NEAR/1 LANGUAGE) NEAR/1 (GENERATION OR PROCESSING)) OR (REINFORCEMENT NEAR/1 LEARNING) OR ((SUPERVISED OR UNSUPERVISED) NEAR/1 (LEARNING OR TRAINING)) OR (SWARM NEAR/1 INTELLIGEN*) OR ((SEMI-SUPERVISED OR SEMISUPERVISED) NEAR/1 (LEARNING OR TRAINING)) OR CONNECTIONIS* OR (EXPERT NEAR/1 SYSTEM$) OR (FUZZY NEAR/1 LOGIC) OR (TRANSFER NEAR/1 LEARNING) OR (LEARNING NEAR/2 ALGORITHM$) OR (LEARNING NEAR/1 MODEL*) OR (SUPPORT VECTOR MACHINE$) OR (RANDOM FOREST$) OR (DECISION TREE$) OR (GRADIENT TREE BOOSTING) OR (XGBOOST) OR ADABOOST OR RANKBOOST OR (LOGISTIC REGRESSION) OR (STOCHASTIC GRADIENT DESCENT) OR (MULTILAYER PERCEPTRON$) OR (LATENT SEMANTIC ANALYSIS) OR (LATENT DIRICHLET ALLOCATION) OR (MULTIAGENT SYSTEM$) OR (HIDDEN MARKOV MODEL$)) OR TS=((ARTIFI* NEAR/1 INTELLI*) OR (AUTO* NEAR/1 LEARNING*) OR BAYESIAN OR (DATA NEAR/1 MINING) OR (DEEP NEAR/1 LEARNING) OR (MACHINE NEAR/1 LEARNING) OR (ARTIFICIAL* NEAR/1 LOGIC) OR (INTELLIG* NEAR/1 NEURONAL*), (NEURAL* NEAR/1 REASON*) OR (FUZZY NEAR/2 NETWORK*) OR (DATA NEAR/1 MINING) OR (ARTIFICIAL NEAR/2 INTELLIGENCE) OR (INDUCTIVE NEAR/2 LOGIC) OR (DEEP NEAR/1 LEARNING) OR (GENETIC NEAR/1 ALGORITHM*) OR (SUPPORT VECTOR MACHINE$) OR (NEURONAL NEAR/1 NETWORK$) OR (FUZZY NEAR/1 LOGIC)) OR IP=(G06F-015/18 OR G10L-013/027 OR G06N-020* OR G06F-017/00 OR G06F-017/10 OR G06F-017/11 OR G06F-017/12 OR G06F-017/13 OR G06F-017/14 OR G06F-017/15 OR G06F-017/16 OR G06F-017/17 OR G06F-017/18)
User interfaces
MAN=(T01-J10C4A OR T01-J40 OR T01-J40C OR T04-F02B7) OR TS=((AUGM* NEAR/1 REALITY) OR (DATA NEAR/0 EYEGLASS$) OR (DATA NEAR/0 SPECTACLE$) OR (GOOGLE NEAR/0 GLASS$) OR (HEAD NEAR/0 MOUNT* NEAR/0 DISPLAY$) OR (HEAD$UP NEAR/0 DISPLAY$) OR HMD OR HUD OR (HEAD NEAR/0 DISPLAY$) OR (WEARABLE NEAR/1 DISPLAY$) OR (ENVIRONMENT$ NEAR/3 VIRTUAL) OR (DISPLAY NEAR/2 HELMET) OR (MIXED NEAR/1 REALITY) OR (VIRTUAL NEAR/1 REALITY) OR (ENHANCED NEAR/1 REALITY) OR (AUGMENTED NEAR/1 ENVIRONMENT$) OR (MEDIATED NEAR/1 REALITY) OR (MIXED NEAR/1 ENVIRONMENT) OR (VIRTUAL NEAR/1 WORLD)) OR IP=(G06K-011* OR G06T-011* OR G06T-013*)
Computing
MAN=(T01-M06C OR T01-M06Q) OR TS= (IAAS OR PAAS OR SAAS OR ((SOFTWARE OR PLATFORM OR INFRASTRUCTURE) NEAR/2 SERVICE$) OR (SERVER$ NEAR/1 CLUSTER) OR (CLOUD NEAR/1 (COMPUT* OR DATA OR DISTRIBUT* OR GRID* OR POINT OR SERVER OR SERVICE$ OR STOR*)) OR (COMPUT* NEAR/1 GRID$) OR (DATA NEAR/0 CENTER$) OR (SERVER NEAR/0 FARM$)) OR IP=(G06E* OR G06J* OR G06N*)

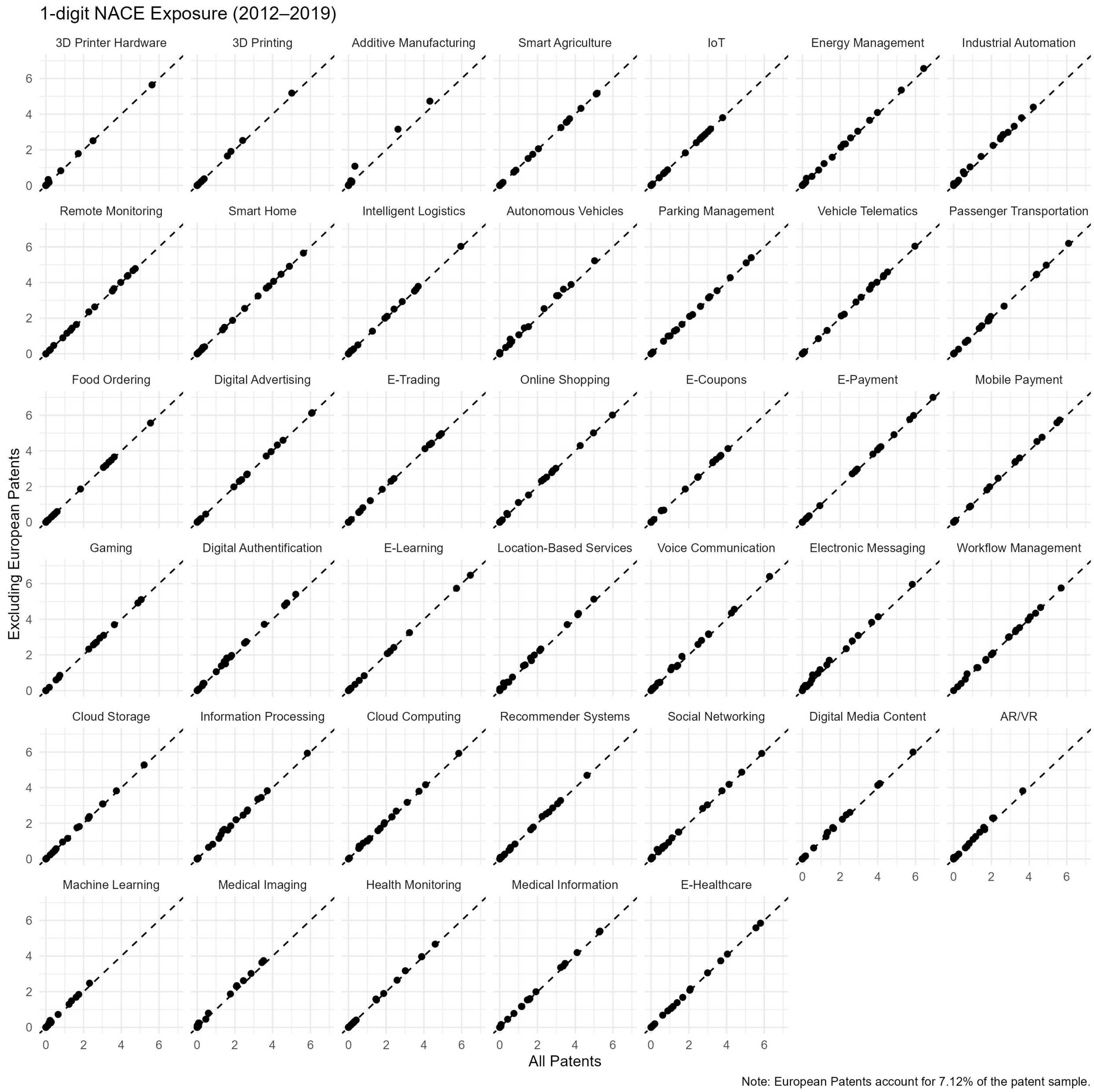
Notes:

Figure OA.1 presents the correlation between the 1-digit exposure scores with all patents and those while excluding the European patents.

Tables OA.3 to OA.16 presents the top three most-cited patents by emerging digital technologies.

Figure OA.2 depicts a positive relationship between the change in the employment-to-population ratio from 2012 to 2019 and the regional exposure to emerging digital technologies, after excluding regions with exceptionally low exposure levels—specifically, those with an ex-

Figure OA.1: 1-digit Industry Exposure with and without European Patents (2012–2019)



Notes: Each panel depicts the correlation between the 1-digit NACE exposure scores over the period 2012–2019 for each emerging digital technology with (x-axis) and without (y-axis) European patents. European patents are identified as those filed in the European Patent Office (EPO).

posure index below -2 standard deviations (i.e. below 0.929), which typically includes rural areas in Romania, Turkey, and overseas French territories.

Table OA.3: Most cited patents by technology (1/14)

Patent ID	Patent title	Year	Cited
[01] 3D Printer Hardware			
201736370E	Three-dimensional object printer, has actuator for moving outlet into alignment with two channels in set of channels at different times to supply extrusion material from extrusion material supply to channels in set of channels	2017	54
2017363641	Multi-nozzle extrusion printhead for use in three-dimensional object printer, has electromechanical actuator for moving unit to position to enable flow of extrusion material through corresponding fluid outlet in fluid outlets	2017	53
201641448Q	Three-dimensional printer, has set of status pin connections for transferring data comprising identity of each cartridge, properties of associated build material dispenser, and properties of build material	2016	44
[02] 3D Printing			
201757292P	Method for generating three-dimensional object, involves accessing alteration of characteristic of three-dimensional printing based on measurement during three-dimensional printing and generating three-dimensional object	2017	216
201742766M	Method for printing three-dimensional object involves using energy beam to transform at least portion of exposed surface to transformed material, in which transformed material is portion of three-dimensional object	2017	137
201800449Q	Apparatus for printing e.g. three-dimensional objects in three-dimensional printer system, has load-lock for defining volume, and energy source for generating energy beam that irradiates to facilitate printing of three-dimensional object	2018	115
[03] Additive Manufacturing			
201730027V	Method for forming three-dimensional object, involves altering three-dimensional model of requested three-dimensional object to form altered model, and transforming portion of material bed with energy beam according to altered model	2017	85
201723486N	Formation of three-dimensional object, such as medical devices e.g. stents, involves providing carrier and optically transparent component defining a build region in between and irradiating build region with light through the component	2017	80
201835956B	Forming three-dimensional object for e.g. medical devices involves filling build region of transparent component with polymerizable liquid comprising polymerizable component, upconverting particles and photoinitiator and irradiating region	2018	79

Notes: This table presents the three most cited patents by technology.

Table OA.4: Most cited patents by technology (2/14)

Patent ID	Patent title	Year	Cited
[04] Smart Agriculture & Water Management			
2015385330	User display graphical configuration system for use in plant monitoring system for monitoring e.g. chemical process, has editor for presenting interface, and configuration form application using information to create graphical element usage	2015	41
201616464K	Computer-implemented method for managing agricultural activities, involves determining multiple field condition data based on subset of input data and providing multiple field condition data to user device	2016	41
2013H90027	System for controlling direct-drinking water equipment, has remote monitoring platform module to receive data obtained by monitoring control module so as to remotely monitor operation status of direct-drinking water device	2013	39
[05] Internet of Things (IoT)			
2017509033	ZigBee and cloud computing based intelligent household control system, has zigBee three-level tree wireless sensing network provided with network framework for shortening communication distance and reducing electromagnetic pollution	2017	89
2020669092	IoT system configured for monitoring and creating a digital twin of an industrial setting, comprises multiple sensors that capture sensor data and transmit the sensor data via a self-configuring sensor kit network	2020	54
2016224608	Internet-of-things based method for controlling and monitoring group of internet-of-things devices or mash-up service through computer, involves providing messaging service participated in internet-of-things devices through group chat room	2016	52
[06] Predictive Energy Management and Distribution			
2013V51221	System for electric power grid element and network registration and management of grid elements, has active grid element that is constructed within housing, and update message transforms function by updating attribute of grid element	2013	243
2013F50499	Method for interactively and graphically displaying e.g. energy consumption of Heating, Ventilation and Air conditioning system to user, in home, involves gathering information relating to system usage using thermostat	2013	173
201570114J	System for analyzing building energy consumption information, has real-time energy efficiency plan providing device providing energy efficiency improvement plan for building in real-time based on energy consumption analysis result	2015	138

Notes: This table presents the three most cited patents by technology.

Table OA.5: Most cited patents by technology (3/14)

Patent ID	Patent title	Year	Cited
[07] Industrial Automation & Robot Control			
2014R39999	System for operating process plant e.g. chemical process plant, has one user interface (UI) device that is operated, so that status information indicating one multiple routines on one UI device is passed to another UI device	2014	190
2012M05358	Control system for controlling operation of pallet truck in e.g. shipping dock, has input device responding to operational data by producing commands that are transmitted for remotely controlling operation of industrial vehicle	2012	154
2018A3032C	Method for performing device operations management in schedule of tasks involving devices, involves unlocking machine for use by operator by using computer at scheduled time by sending unlock message to machine at updated start time	2018	132
[08] Remote Monitoring & Control Systems			
2017783476	Data collection system for industrial environment, comprises platform that is provided with a computing environment, where computing environment of the platform compares the relative phases of the first and second sensor signals	2017	523
201913893E	Data collection system for use in industrial production environment, has analysis response circuit which is structured to adjust sensor scaling value or sensor sampling frequency value, in response to sensor performance value	2019	309
2013N11698	Method of operating heating, ventilation or air conditioning (HVAC) monitoring system installed in e.g. residential building, involves analyzing stored data to selectively identify problems and predict faults of HVAC system	2013	189
[09] Smart Home & Intelligent Household Control			
2014T60296	Method for controlling home automation system by vehicle control system, involves monitoring status of person in home and determining whether status of person is changed or not, by microprocessor executable home automation system	2014	257
201532495K	Method for controlling smart-home environment of smart devices in e.g. resource-consuming physical systems, involves automatically adjusting functionality of one of smart devices using computing system based on analyzing	2015	230
2013K37085	Intelligent control method for intelligent home system and household appliance, involves controlling gateway by intelligent control unit to control and enable household appliances for executing operations according to habit order table	2013	195

Notes: This table presents the three most cited patents by technology.

Table OA.6: Most cited patents by technology (4/14)

Patent ID	Patent title	Year	Cited
[10] Intelligent Logistics			
201542986L	Shelving system for package-delivery vehicle e.g. delivery truck, has central processing unit that is configured to identify package-location information, in response to vehicle-location information	2015	280
2013Q07655	Method for delivering items stored in e.g. bin of pickup location to customers for electronic-commerce and mail-order companies, involves providing instructions at location to place item into storage compartments of identified location	2013	222
201537957C	Computerized electronic locker system for parcel delivery and pick-up, has computer and software with internet connection, which are configured to transmit parcel related data and notifications to parcel recipient	2015	148
[11] Autonomous Vehicles & UAVs			
2013J15200	Vehicle control method of vehicle system, involves automatically determining person within vehicle, automatically identifying person, determining if setting to be stored for person, and storing setting	2013	1101
201749377L	Navigation system for host vehicle, has processing device determining actual navigational action having modification of desired navigational action, and causing adjustment of navigational actuator in response to actual navigational action	2017	259
201648714V	Autonomous guidance system for operating a vehicle in an autonomous mode, comprises a controller to determine an object-location of object on a map of area based on a vehicle-location of vehicle on map, image signal, and reflection signal	2016	231
[12] Parking & Vehicle Space Management			
2012E53618	Computer-implemented system for managing motor vehicle e.g. car parking reservations, has availability module to indicate availability of parking space through nearest parking availability indicator	2012	200
2014P33734	Parking meter for monitoring and managing vehicle parking, has processor which detects vehicle presence in parking space, captures identification (ID) of vehicle, times initial grace period, and receives payment for parking time period	2014	152
2012J60343	Auto-valet parking (AVP) server device for AVP system, has parking map management unit that receives information about final slot selected by user from among customized parking slots, and provides route to final slot	2012	134

Notes: This table presents the three most cited patents by technology.

Table OA.7: Most cited patents by technology (5/14)

Patent ID	Patent title	Year	Cited
[13] Vehicle Telematics & Electric Vehicle Management			
2012D36559	Risk management system for monitoring and facilitating review of data collected from vehicle, has server processes selected vehicle data and generates rating factor based on selected vehicle data stored in database	2012	342
2013K60165	Apparatus for detecting usage of mobile phone during driving of car, has detection system including processor and set of sensors that are operative to be used by processor, where detection system is operated to communicate to remote server	2013	236
2012K82179	System for monitoring vehicle data that is used to determine e.g. level of risk, in operating vehicle e.g. automobile, has wireless transceiver to encrypt and encode relationship data and vehicle data and transmit encoded data	2012	208
[14] Passenger Transportation			
2012C13650	Method for coordinating transportation service e.g. taxi service, involves determining suitable transportation vehicle for trip, and delaying the dispatching if request for trip specifies delayed pick-up	2012	266
2015381635	Method for automobile sharing server, involves crediting financial account of owner with payment of renter in threshold radial distance from driverless vehicle when predicted at non-transitory location for available period of time	2015	200
2014M09285	Method for operating dispatch server used in system for facilitating short-term automobile rentals, involves dispatching private vehicle of set of private vehicles in geo-spatial vicinity of geo-spatial location	2014	189
[15] Food Ordering & Vending Systems			
2014E23249	Method for tracking of delivery of menu item from restaurant to dwelling of customer, involves sending displacement notification to customer interface including indication of geographical position of available delivery vehicle	2014	77
2013F52957	Method for offering and managing reservations for restaurant on e.g. smartphone, involves receiving approval from user to have reservation to be transferred to another user, and arranging incentive to be provided to user	2013	76
2015641449	System for presenting smart recurrent orders for purchasing e.g. baby wipe by consumer, has analysis module executed by processor for determining adjustment to order schedule, and order module facilitating adjustment to schedule for item	2015	70

Notes: This table presents the three most cited patents by technology.

Table OA.8: Most cited patents by technology (6/14)

Patent ID	Patent title	Year	Cited
[16] Digital Advertising			
2013B87254	Method for targeting television advertisement based on profile linked to online device, involves selecting television advertisement to be directed to set-top box based on profile information pertaining to user or online activity	2013	382
201522675P	Method for targeting advertising content, involves determining individual sets of advertising content from individual sets of advertising content, and transmitting individual sets of advertising content to fictitious user name	2015	329
2013T30678	Method for delivering marketing information to customers, involves receiving sighting message and transmitting message to computing device including identified marketing information relevant to wireless identity transmitter	2013	244
[17] Electronic Trading and Auctions			
2014M20580	Data processing system e.g. desktop computer for managing electronic offer, has logic module that automatically suggest offer to offer provider based on subset of historical transaction records and subset of offer data	2014	252
201619462W	Computer-implemented method for trading assets using decentralized escrow service, involves receiving notification from central processing server of trade order match module using order matching module	2016	114
2013M28862	Exchange data processing system for e.g. trading valuation of consumer-directed trading financial products, has trading software for executing on computer for trading financial product of system using monetary value and transparency index	2013	109
[18] Online Shopping Platforms			
2014A42720	Method for purchasing product or service in e.g. social network, involves electronically providing mobile and internet posting of location based-customized, promotion or offer comprising website for products or services to user	2014	373
2013M09177	Method for processing e.g. point-of-sale transaction for purchase of grocery item in supermarket, involves processing transaction based on data regarding information regarding transaction and physical location of electronic device	2013	320
2014T48720	System for fulfilling sale request for item in e.g. internet based online store, has processing element that determines whether item is suitable for use in fulfilling request based on fulfillment confidence score	2014	255

Notes: This table presents the three most cited patents by technology.

Table OA.9: Most cited patents by technology (7/14)

Patent ID	Patent title	Year	Cited
[19] E-Coupons & Promotion Management			
2013E25218	Information storage and display device for managing and redeeming bar-coded coupons displayed from light emitting display surfaces of information display devices, comprise computing platform for running computer applications	2013	373
2013E12801	Method for generating wireless and internet posted-location based customized promotion and offer e.g. coupon for product and service, involves assigning unique identifier to user of client mobile device that receives request from user	2013	327
2014V55925	Device e.g. smart-phone for displaying digital images of bar-coded store coupon for coupon redemption operation, has computing platform that is configured to convert barcode symbol to pulse code modulation formatted barcode symbol	2014	263
[20] Electronic Payments & Financial Transactions			
201604016J	Method for sending funds or credits relating to good or service to e.g. location of participating entity, involves delivering electronic communication to electronic address, where communication comprises data pertaining to instruction	2016	694
2013X40687	Bill payment apparatus for e.g. facilitating transactions relating to effectuating payments of bills to consumers, has processor to expire temporary postponement payment account if balance value reaches zero, by deleting account	2013	334
2014E57471	Cloud-based virtual wallet secure transaction processor-implemented method for processing electronic purchase transaction in e.g. online shopping, involves providing transaction bounding token to transaction security server	2014	295
[21] Mobile Payments			
2012D97057	Mobile payment account activation system has account activation unit that automatically authenticates user associated with inactive mobile payment account by transmitting validation data to portable electronic device	2012	324
2015135153	Method for processing transaction in e.g. portable electronic device, involves processing transaction based on data regarding information regarding transaction and physical location when obtaining information regarding time of day	2015	311
2012J91915	Mobile device e.g. mobile phone, for authorizing payment for transaction in e.g. restaurants by e.g. point of sale terminal, has processor using communication device to transmit predetermined payment information and authorize payment	2012	298

Notes: This table presents the three most cited patents by technology.

Table OA.10: Most cited patents by technology (8/14)

Patent ID	Patent title	Year	Cited
[22] Gaming & Wagering Systems			
2013B47685	Game networking system for combining games based on levels of interactivity of e.g. gambling games has integrating module that provides player with option to participate in secondary game while having idle time with respect to primary game	2013	183
2013M91402	Method for performing casino game using personal digital assistant, involves allowing cellular telephone to access gaming services based on location of cellular telephone being in approved location by computing device	2013	107
2016106891	Wireless communication system for lottery ticket selling with a single platform, has computer system with a workflow server, and workflow module with sets of workflow instructions for processing different types of lottery game packets	2016	106
[23] Digital Authentication			
2014R64759	Apparatus for performing advanced authentication techniques for banking applications, has authentication logic attesting to model and/or integrity of component to another components prior to allowing components to form authenticator	2014	647
2014B84307	Proxy wallet transaction authentication method for finding, storing and applying discounts in transaction, involves receiving transaction authentication request associated with proxy payment identifier and authenticating transaction	2014	358
2012R18811	System for acquiring digital credential data to perform securing authorization of card present financial transaction with card issuing bank in financial transaction industry, has communication device receiving repository response with data	2012	311
[24] E-Learning			
2014K73633	Learning management system for managing multiple e.g. tablet computers simultaneously in smart classroom, has server system for interpreting and handling communications between controllable devices and controller device	2014	181
201519486C	Interactive learning platform, has teaching resource management module connected with online classroom management module that is connected with collecting unit, and data package storing and uploading process performed	2015	101
2013V12318	System for interactive class support and education management using e.g. portable terminal, has specific device that performs correction guidance for assignment and test and stores teacher-correction-guided content in server	2013	98

Notes: This table presents the three most cited patents by technology.

Table OA.11: Most cited patents by technology (9/14)

Patent ID	Patent title	Year	Cited
[25] Location-Based Services & Tracking			
2015159503	Method for distributing micro-location-based notification to computing device, involves resolving unique identifier that is collected from wireless beacon by computing device, into identity of wireless beacon using rule	2015	162
2013E12714	Method for identifying e.g. airport visited by user of e.g. pager, involves receiving indicated named location from device corresponding to business location visited by user of device, and transmitting reference number to device	2013	150
201563193M	Method for power management of mobile clients using location-based services in social networking environment, by sending location history to location server of online social network based on current status of mobile-client system	2015	149
[26] Voice Communication			
2013L87369	Personal safety notification system i.e. mobile alert system, for alerting e.g. students, during emergency situation, has server providing alert/event details to mobile device in active sub-network in response to occurrence of alert/event	2013	336
2014E49356	Communication device e.g. cell phone has mixed array that is provided with different types of array units and provided to occupy area coinciding with in plan view as viewed perpendicular to major surface, and entire major surface	2014	205
2014G07765	Wireless communication device e.g. smart phone, for use in communication system, has processor for establishing connection between device and end-point device and applying control to traffic, and memory coupled to processor	2014	195
[27] Electronic Messaging			
2016569813	Messaging system used for instructing staff with enterprise related matters, has distributed network gateway server that validates client device with message management policy before authorizing transmission of impermanent text message	2016	221
2013C74472	Method for e.g. integrating conversation view with voice over-internet protocol calls and communication, involves selecting contact from list of contacts, and displaying graphical user interface including interaction events	2013	188
2013B03892	On-line system for providing group interaction e.g. group chats, around common online content in computing device, has user devices accessing and loading content via computer network in response to receiving reference to content	2013	184

Notes: This table presents the three most cited patents by technology.

Table OA.12: Most cited patents by technology (10/14)

Patent ID	Patent title	Year	Cited
[28] Workflow Management			
2016535312	System for providing enhanced security for enterprise computing environment has modules which provide services to computing environment and interfaces, which provide access to collected data and services	2016	391
2016112028	Inventory management system for managing tasks in facilities, has computer system for generating connectivity improvement information and providing connectivity improvement information to one of access points or mobile drive unit	2016	174
2014R46796	Computer implemented method for interacting with records from user interface, involves receiving second information associated with first record or second record from publisher, and updating database system based on second information	2014	143
[29] Cloud Storage & Data Security			
2013M79136	Method for enabling provision of keys between users to enable data security, involves providing specific key encrypted with specific further key to predetermined user in manner independent of data	2013	208
2014P51999	Method for executing application program in public cloud network without moving private dataset of application program from data storage, involves executing application program with data processor to access data blocks of private dataset	2014	138
2014W31419	Method for providing cloud storage service in communication system in hybrid cloud computing environment, involves deploying cloud storage gateway in cloud, where gateway facilitates secure migration of data associated with virtual machine	2014	136
[30] Information Processing			
201865517X	Readable and rewriteable card blank for use with hand-held electronic device, has personal electronic data sets that are read into card to form factor collectively for electronic transacting or fulfillment of electronic identification query	2018	81
2014Q18150	System for processing data in connection with insurance information submissions to generate insurance policy, has storage devices for storing data relating to accessing of entity data where insurance form is outputted for display on device	2014	45
2019427506	Method for processing information of terminal device, involves obtaining first information, where first information is processed by terminal device, and transmitting operation instruction to computing device to obtain second information	2019	43

Notes: This table presents the three most cited patents by technology.

Table OA.13: Most cited patents by technology (11/14)

Patent ID	Patent title	Year	Cited
[31] Cloud Computing			
2014D78354	Multi-tenant cloud computing system, has cloud controllers for managing cloud infrastructures utilizing virtual resources to operate other resources that provide access to physical resource pool through controllers	2014	297
2014A40902	Computer-implemented method for implementing hybrid-cloud computing network infrastructure, involves installing user application in provisioned computing resource in accordance with application blueprint	2014	221
201516824J	Method for providing cloud service brokering service by computer-implemented cloud service brokering system, involves allocating selected cloud computing resource for use by customer by computer-implemented cloud service brokering system	2015	187
[32] Recommender Systems			
201541187F	Method for generating client-side structured search queries involves generating structured queries by matching unstructured text query to accessed nodes and grammar templates having non-terminal tokens by mobile client system	2015	131
2012P90150	Computer-implemented method for searching e.g. application object of website, involves ordering objects of combined result set, and providing portion of combined result set to client device in response to query	2012	119
201764769V	System for providing personalized content recommendation, has memory that executable by processors generates user interface in which individual cards of group of cards are corresponds to individual containers in group of containers	2017	108
[33] Social Networking & Media Platforms			
2015197305	Method for presenting real-time interface for sending invitation for adding users to contacts and groups or social networks, involves enabling requestor users and users of wireless networks for participating with activities of each other	2015	240
2013L99116	Method for detecting social graph elements for queries to perform search for e.g. text, within e.g. internet, involves identifying nodes including score greater than node-threshold score and generating query including references to nodes	2013	195
2013H44326	Method for generating structured queries based on social-graph information, involves identifying identified edges corresponding to grams, and generating structured queries with references to identified nodes and identified edges	2013	192

Notes: This table presents the three most cited patents by technology.

Table OA.14: Most cited patents by technology (12/14)

Patent ID	Patent title	Year	Cited
[34] Digital Media Content			
201673838R	Non-transitory computer-readable storage medium for proactively identifying and surfacing relevant content on electronic device, comprises multiple executable instructions that are executed for executing an application on electronic device	2016	332
201726949F	Playback device e.g. mobile device for accessing content using decentralized blockchain right ledger, has ledger modification application to decrypt content from digital media work using decrypted content key and play back decrypted content	2017	239
201547609G	System for sharing digital user content through distinct network-accessible sharing platforms, has external content exposure tracker that is configured for tracking integration of external content with user content for distinct users	2015	228
[35] Augmented and Virtual Reality (AR/VR)			
201553908H	Wall-mounted interactive sensing and audio-visual node device for networked e.g. living space, has faceplate mounted to wall such that outer surface of faceplate is placed away from wall, where wall power input is connected to power line	2015	115
2015364438	Method for controlling operation of display screens of vehicle, involves processing instructions to automatically move item of graphical content rendered in display screen to being rendered in dashboard display screen	2015	88
2014T69316	Multimedia data aesthetic and synchronous display method for graphical user interfaces of smart TV set of social network user, involves displaying arranged content of video channel by graphical user interface according to display design	2014	67
[36] Machine Learning & Neural Networks			
202050356J	Method for image recognition in image or video recognition platform, involves obtaining match for image of search engine for images based on contradiction, uncertainty analysis and datas, and outputting match for image of engine for images	2020	94
201937401H	Method for generating augmented training dataset for training convolutional neural network model to recognize target object, involves training convolutional neural network model to recognize target object based on training dataset	2019	49
201923013S	Method for training neural network, involves obtaining trained neural network, and continuing input of first training input image and second training input image to repeat training process when loss value does not satisfy preset condition	2019	48

Notes: This table presents the three most cited patents by technology.

Table OA.15: Most cited patents by technology (13/14)

Patent ID	Patent title	Year	Cited
[37] Medical Imaging & Image Processing			
2014V59740	System for displaying three-dimensional point cloud image of biopsy needle and ultrasonic image of patient's anatomy on movable display, has processor repositioning image of patient's anatomy on display screen in real-time	2014	177
2019567486	Surgical image acquisition system for use in operating theater of hospital, has computing system for determining depth location of structure within tissue sample and calculating visualization data regarding structure and depth location	2019	174
2017165997	Computer-implemented method of providing ensuring medical device position and functionality, involves confirming position of medical device within patient using imaging device	2017	91
[38] Health Monitoring			
2013V85481	System for electronic patient care in hospital, has medical device that operatively receives and communicates measured physiological parameter from medical sensor, and server receives and stores measured physiological parameter	2013	407
201549924Y	Electronic device for use in network environment, has main body for obtaining bio-information of user by using communication module of bio-signal detection sensor, and processor for providing service of bio-information	2015	333
2013G61970	Wearable device e.g. radio, to provide physical or physiological characteristics associated with e.g. speed data, has instructions displaying metric of one mode in response to determining that time-out period has expired	2013	268
[39] Medical Information			
201832322L	System for facilitating synthetic interaction between patient, and computer-implemented program, has processor for identifying and executing action including instructing interactive device to present subsequent script to user	2018	227
2019575586	Method for collecting data within healthcare facility i.e. hospital, involves determining trends associated with surgical procedures performed in facility by computer system according to perioperative data and procedural context data	2019	226
2013U13719	Method for processing medical documentation about patient in healthcare industry, involves receiving structured data set including information relating to medical facts, from medical documentation system	2013	179

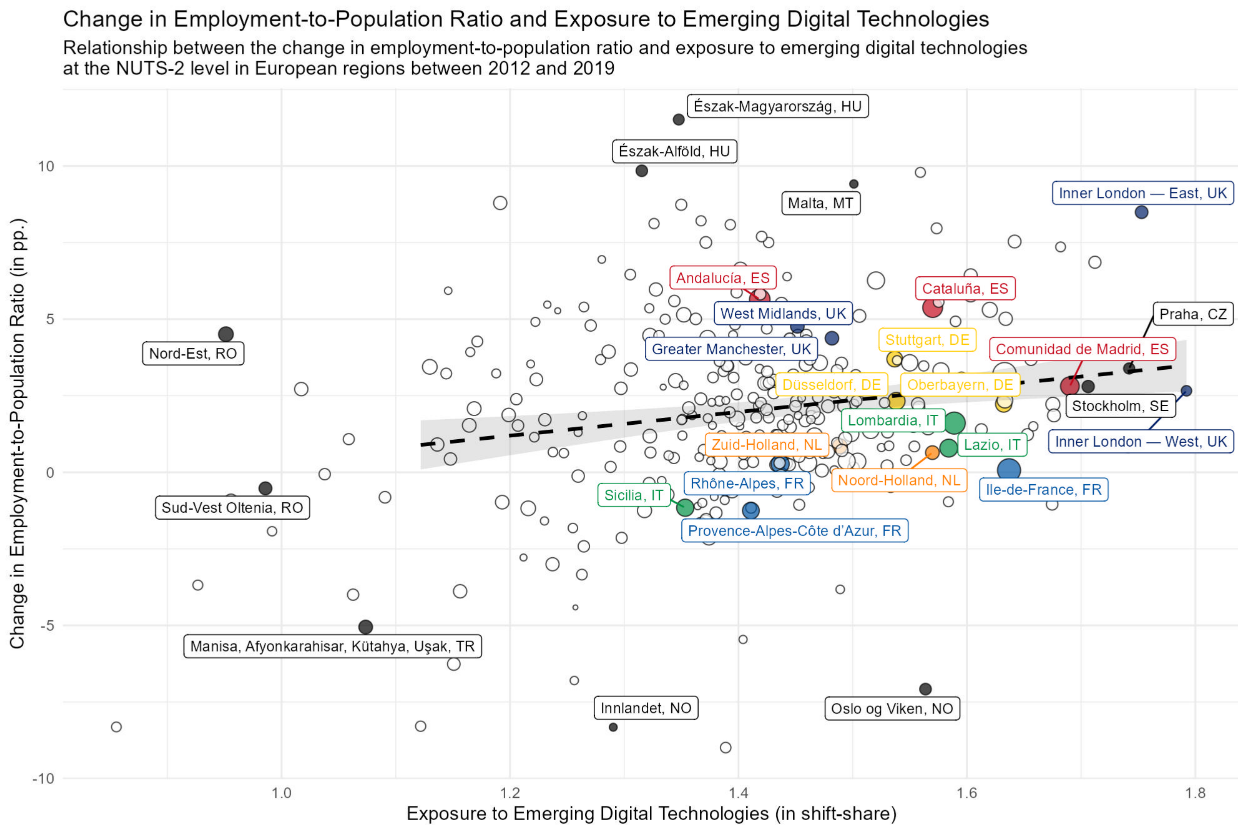
Notes: This table presents the three most cited patents by technology.

Table OA.16: Most cited patents by technology (14/14)

Patent ID	Patent title	Year	Cited
[40] E-Healthcare			
201521223E	Method for using mobile information gateway for home healthcare for treating patient by e.g. nurse, involves retrieving information using captured information, and presenting retrieved information using human interface module	2015	255
2013C62758	Medical image exchange system for exchanging medical image data in medical institution, has medical image transmission system, which is provided with terminal in hospital, relay server, medical image display terminal and information system	2013	252
2013F70757	System for providing automatic messaging to patient on behalf of e.g. doctors from hospital regarding health instructions to continuously monitor patient, has processor selecting agents to send instruction promoting healthy client behavior	2013	242

Notes: This table presents the three most cited patents by technology.

Figure OA.2: Change in Employment-to-Population Ratio and Exposure to Emerging Digital Technologies (excluding regions below -2 standard deviations)



Notes: This figure shows the relationship between the change in the employment-to-population ratio and the exposure to emerging technologies in European NUTS-2 regions between 2012 and 2019. Each point represents a region, with select regions labeled for emphasis. The size of the point is proportional to the population in 2010. The horizontal axis measures the exposure to emerging technologies calculated by the shift-share method, while the vertical axis represents the change in the employment-to-population ratio. The solid line indicates a positive correlation between increased regional exposure to emerging technologies and employment growth. Regressions lines are weighted by population in 2010. Data points are color-coded by country. Outliers are highlighted and labeled for clarity.