The Employment Impact of Emerging Digital Technologies[*]

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

October 2024

Abstract

This paper measures the exposure of industries and occupations to a broad set of emerging digital technologies 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 international standard classifications, creating the open–access 'TechXposure' database. Through a shift–share design, we instrument regional exposure to estimate the effects of these technologies on employment across European regions. We find a net positive impact, with growth in low- and high-skilled employment at the expense of middle-skilled jobs, suggesting ongoing job polarization. At the technology level, we observe significant heterogeneity: robots and machine learning negatively impact employment (except for high-skilled workers), while workflow management and information processing systems have positive effects. Our results suggest that focusing narrowly on specific technologies like AI and robots may overlook broader positive employment impacts stemming from complementarities among diverse digital technologies.

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 Daisuke Adachi, Elliott Ash, Cecilia García-Peñalosa, Jeffrey Grogger, Florencia Jaccoud, Guy Michaels, Önder Nomaler, Ana Oliveira, Morgan Raux, Anna Salomons, 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, HKLM Seminar, UK DfE, CEP, University of Barcelona, University of Aarhus 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 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 international 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

[1] See, for instance, Autor et al. (1998), 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 Information and Communication Technologies (ICT); and Graetz and Michaels (2018), Acemoglu and Restrepo (2020), Vries et al. (2020), Aksoy et al. (2021), Dauth et al. (2021), Aghion et al. (2023), Adachi et al. (2024), Bonfiglioli et al. (2024) 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 technologies.

[3] A notable exception is Albanesi et al. (2023) who examine the relationship between labor markets and exposure to AI and software in 16 European countries, combining both Felten et al. (2018) and Webb (2019) US-based exposure metrics.

digital technologies that emerged over the past decade across several demographic groups. 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 and most likely to be used in the production of goods and services during 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* (Song et al. 2020).[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 correspondence 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 how *relevant* a specific technology is to an industry or occupation. For industries, relevance is determined by the integration of technology into the production process and/or its role in enhancing industry output. For occupations, relevance measures the importance of a technology in performing tasks and functions. These exposure scores serve as proxies for adoption, indicating the contextual relevance of each technology across industries and occupations. Yet, our exposure scores are neutral regarding the nature of the relationship between technology and workers in a given industry or occupation, meaning that they do not assume *ex-ante* whether technology and labor are complements or substitutes in production. In contrast, our estimates in the second part of the paper clarify this

[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 (Song et al. 2020).

relationship.

We estimate the causal effect of these digital technologies on European regional employment over the period 2012–2019. Our analysis proceeds in two steps. First, we assess the overall impact of emerging digital technologies on the regional employment-to-population ratio from 2012 to 2019 across several demographic groups. The sample covers 320 NUTS-2 regions in 32 European countries. Second, we conduct a detailed analysis to disentangle the effects of specific technologies, such as robotic and data-intensive technologies (including several AI applications), conditional on exposure to other emerging digital technologies.

To address endogeneity issues, we instrument the regional exposure to these technologies with a shift-share design, in which the industry exposure scores over the period are the *shocks* and the baseline employment shares of these industries are the *shares*. Our identification strategy relies on the quasi-random assignment of shocks, allowing employment shares to be endogenous. We argue that the development of emerging digital technologies is predominantly a global phenomenon, independent of 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. We also assume that regions more exposed to emerging digital technologies are not disproportionately affected by other labor market shocks or trends. Our approach leverages the equivalence proposed by Borusyak et al. (2021), and we apply the AKM0 inference method following Adão et al. (2019).

Our work reveals several new findings. First, we document which industries and occupations are the most exposed 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. However, the exposure of this latter group of high-paying occupations tends to be driven by the exposure of recurrent tasks rather than specialized tasks. 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 1.03 percentage point (pp.) change, corresponding to 2.05%, 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 0.72 pp. (+6.01%) and 0.74 pp. (+4.92%) in their employment-to-population ratios, while middle-skilled employment decreases by 0.41 pp. (−1.78%). 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 on the negative employment impact of certain automation technologies (e.g., industrial robots and AI), they also suggest that a narrow focus on these technologies may overlook 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 Acemoglu and Restrepo (2020), which focus on industrial robots, or Webb (2019), which focuses on AI and software—also find overall positive employment effects in US labor markets and occupation-industry cells. Our work underscores the crucial role of complementarities among these technologies in determining their employment effects.

Second, this paper uniquely addresses a gap in this literature regarding exposure metrics. While most existing metrics concentrate on US classifications and specific technologies,[6] our work is the first to provide detailed exposure scores based on international standard classifications, specifically NACE Rev. 2 and ISCO-08, with high granularity across a broad range of digital technologies. This contribution enhances the applicability of exposure metrics beyond the US context and prominent technologies, offering a broader basis for future research. Furthermore, our scores draw on global patent data, reflecting technological advances worldwide,

[6] See Jurkat et al. (2022) for the 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, including Large Language Models (LLM).

not limited to the US and Europe.

Third, we contribute methodologically by introducing a scalable approach that leverages advanced NLP techniques with sentence transformers to estimate exposure. Unlike traditional methods that rely on keywords (i.e., tokens) to match innovations with occupations and industries,[7] our methodology bypasses this requirement by leveraging semantic and contextual similarity, requiring only a relevant patent set. Moreover, our approach innovatively uses patents by clustering them based on semantic distance to define technology groups. This method identifies a broad range of digital technologies—not limited to AI or robotics—and allows for more precise and interpretable categorization.

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 define our set of emerging digital technologies, with each technology represented as a cluster of patents from the Derwent Innovation Index (DII) database.[8] For simplicity, we use the term 'patent' instead of 'patent family' to refer to a single invention across various patent offices. We first describe the components of patent texts and the characteristics of our sample, then explain our methodology for clustering patents based on semantic similarity to identify our emerging digital technologies.

We use a set $\mathcal{P}$ of 190,714 Derwent patents filed between 2012 and 2021.[9] This patent set, constructed by Chaturvedi et al. (2023), captures core emerging digital technologies and applications since 2011. Appendix A.1 provides further details on the patent corpus construction.

[7]For example, Kelly et al. (2021) and Kogan et al. (2021) use a token-based TF-IDF approach to estimate occupational exposure from breakthrough innovations, while Dechezleprêtre et al. (2023) measure automation innovation by analyzing keyword frequencies in patents, and Mann and Püttmann (2023) categorize patents as automation-related using tokens.

[8]DII covers over 120 million global patent publications from 59 worldwide patent-issuing authorities and assigns each invention to a unique patent family. These families, represented by standardized English titles and abstracts, are structured by experts into themed blocks (e.g., novelty, use, claims) to streamline searching. Alongside CPC and IPC classifications, DII employs Derwent Manual Codes, a custom hierarchical indexing system reflecting technical and application content for improved patent retrieval.

[9]Each patent document details the invention and its distinctions from prior inventions. Information includes a title, abstract, and metadata, such as applicants, inventors, filing year, authority, citations, and technical classifications (e.g., International Patent Classification or IPC). The abstract is segmented into labeled topical blocks like novelty, use, and claims.

We encode the semantic content of each patent title into a numerical representation, or *embedding*, using sentence transformers.[10] The Derwent Database provides titles (curated by experts) and abstracts as the main textual data for each patent. Patent titles are structured in two parts: the first part ($p_1 \in p$) provides a concise description of the technology, while the second part ($p_2 \in p$) explains *how the technology functions*. These two parts are separated by the first **comma–verb combination**.[11] This structure achieves a balanced representation of the invention, maintaining both generality and specificity. Incorporating additional abstract content, such as independent claims or novelty, would increase text length without enhancing precision, potentially reducing the signal-to-noise ratio.

For semantic matching, this concise representation of an invention—combining its essence and function—can be paralleled with industrial and occupational descriptions. Specifically, we represent an industry or occupation with sentences that follow the same structure: essence (from the industry/occupation title) combined with function (a task for an occupation or an activity/process for an industry). Section 3 details the treatment applied to these texts. Aligning the structure of patent titles with industrial and occupational texts enhances the matching between patents and these taxonomies. This is further facilitated by term standardization and text harmonization in the DII, which conveys technical details clearly and accessibly rather than with excessive jargon. Additionally, unlike abstracts, titles are consistently available 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 tran-*

[10] Previous studies primarily use a bag-of-words (BoW) approach, relying on token frequencies and weights (Kogan et al. 2019, Webb 2019, Arts et al. 2021, Dechezleprêtre et al. 2023, Mann and Püttmann 2023). Sentence transformers perform better than the BoW approach as they capture contextual relationships between words, allowing for a deeper understanding of semantic meaning. While BoW models rely solely on word frequency and lack context, sentence transformers generate embeddings that represent the overall meaning of a sentence, preserving word order and context. This results in more accurate and nuanced representations of patent content, particularly for identifying similarities and clustering technologies.

[11] Using Part-of-Speech (POS) tagging, we identify this pattern in 87.3% of our sample, commonly appearing as ', has', ', includes', ', involves', and ', comprises'. For the remaining patents, titles are split at the nearest midpoint.

scription system using word sequences and counter. (Patent ID 202048118D, 2020)

For each patent title in our sample $\mathcal{P}$, we obtain its *embedding* Emb_p, a 768-dimensional numerical representation of text, using the pre-trained sentence transformer model all-mpnet-base-v2 (Song et al. 2020). Sentence transformers encode word meaning in relation to surrounding context, offering an advantage over bag-of-words models that treat text as unordered words. The *all-mpnet-base-v2* model is particularly well-suited for sentence similarity and clustering tasks.[12]

We then cluster the embeddings using the k-means algorithm to obtain 40 clusters, which we designate as our set of emerging digital technologies $k \in \mathcal{K}$. Initially, we compute partitions ranging from 5 to 100 clusters and record each Davies-Bouldin Index (DBI) score (Davies and Bouldin 1979). The optimal range, based on the lowest DBI scores, lies between 30 and 45 clusters, indicating high within-cluster and low between-cluster similarity. We further analyze this range using the most representative phrases per cluster via c-TF-IDF.[13] We find that 40 clusters are optimal for our analysis, as they align well with commonly discussed technologies in digital and automation literature (Acemoglu and Restrepo 2019; Zolas et al. 2021; Martinelli et al. 2021; Acemoglu et al. 2022a).

Table 1 presents our set of emerging digital technologies grouped by technology families. Short descriptions of each technology are provided in Tables A.1 to A.3 in Appendix A.2. The grouping of these 40 technologies into 9 families is based on correlations in their co-occurrence within occupations (discussed further in the next section). Each family includes technologies with highly correlated occupational semantic links; see Appendix A.6 for a detailed discussion. Figure A.2 in the appendix shows the distribution of patents across the emerging digital technologies.

3 Semantic–based Exposure

In this section, we present the methodology for calculating the exposure scores of industries and occupations to emerging digital technologies. First, we calculate the cosine similarity scores between industries/occupations and patents using textual data, filtering for relevant

[12]We select *all-mpnet-base-v2* due to its high performance in Semantic Textual Similarity (STS) benchmarks (see https://huggingface.co/spaces/mteb/leaderboard), computational efficiency, open-source availability, and ease of use via the SentenceTransformers library. Its specialization in text similarity arises from the contrastive loss function used in training, which adjusts model weights based on sentence pairs or triplets, pulling embeddings closer for similar texts and pushing them apart for dissimilar ones. The model is trained on 1.17 billion sentence pairs from sources like WikiAnswers, Reddit, Stack Exchange, and Semantic Scholar.

[13]A modified term frequency-inverse document frequency (TF-IDF) measure, which identifies terms most relevant to clusters rather than individual documents. TF-IDF highlights terms that frequently appear and are unique to specific clusters within the corpus.

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.

pairs. We then aggregate these similarity scores at the technology level to derive semantic-based exposure scores.

Exposure scores indicate the *relevance* of each technology to a given industry or occupation, which we later use as a proxy for adoption. For industries, relevance depends on whether a technology is integrated into the production process or constitutes an improved industry output. For occupations, relevance reflects the importance of technology in performing tasks and functions specific to that occupation.

3.1 Industry Cosine Similarity Scores

Industry Descriptions. We select the 3-digit NACE Rev.2 classification as the most detailed level for the textual description of industries, based on two primary considerations. First, this approach enables us to incorporate titles and descriptions from the 4-digit level into the 3-digit descriptions, thereby expanding the text corpus for matching. Second, industry subsets within the same 3-digit category do not exhibit substantial differences in their connections to patents, allowing for consolidation without significant information loss.

In Section 2, we discussed the patent title structure and argued that mirroring this structure in industry (and occupation) texts facilitates matching. Therefore, for each industry $i \in \mathcal{I}$, we split the descriptions (both 3-digit and nested 4-digit) into individual sentences and concatenate each with its title. We represent these composite sentences as $s \in S_i \subset \mathcal{S}_{\mathcal{I}}$, where S_i is the set of sentences combining the title and a description for industry i. This yields 271 industries at the 3-digit level, each represented by an average of 11 composite sentences.

Embeddings. We generate embeddings for these composite sentences using the same pre-trained sentence transformer as in Section 2, namely all-mpnet-base-v2 (Song et al. 2020). 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 between all composite sentences $s \in \mathcal{S}_{\mathcal{I}}$ and both parts of the patent titles: p_1 (describing the invention) and p_2 (describing its function). Specifically, the cosine similarities are:

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

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

which quantify the semantic relationship between p_1, respectively p_2, and s. However, similarity may reflect different nuances of meaning, such as application, technical domain, or specific functions, whether central or ancillary. This is encapsulated into a scalar value, which *approximates* the degree of similarity between industry aspects (as described in the NACE 4-digit classification) and aspects of the invention (as described in the patent).

To reduce the noise and capture the most relevant similarity between an invention and an industry, we retain the composite sentence s with the highest cosine similarity for each (i, p_1) and (i, p_2) combination. Formally,

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

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

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

Redundancy. To filter out irrelevant pairs, we incorporate *redundancy* in calculating cosine similarity for industry–patent pairs (i, p). For each combination, we rank the sub-pairs (i, p_1) and (i, p_2) separately by their cosine similarity scores, $C_i^{p_1}$ and $C_i^{p_2}$. We then classify a pair (i, p) as relevant (denoted as $(i, p)^\star$) if *both* sub-pairs rank within the top 10 in their respective lists. This approach excludes pairs that do not achieve a top-10 rank for both components.[14] Thus, we retain only those inventions where both the description and function are relevant to the industry.

For each identified relevant pair, we calculate the harmonic mean of the cosine similarity scores for both the invention's description 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 given by Equations (3) and (4), respectively. Thus, Equation (5) establishes a connection between an invention, identified in a single patent $p \in \mathcal{P}$, and a set of relevant industries, where the innovation can enhance process, output, or organizational aspects.

[14] Additionally, we manually exclude three very specific connections to improve our exposure scores; see Appendix A.3 for details.

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

		Cosine Similarity		
Code	NACE Industry	$C_i^{p_1}$	$C_i^{p_2}$	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.

Table 2 illustrates the redundancy principle using the first patent example from Section 2, which details a targeted TV advertising method based on user profile information. For this patent, redundancy filters out industries irrelevant to the innovation. Redundancy filtering for the other two patent examples from Section 2 is shown in Tables A.4 and A.5 in the appendix.

3.2 Occupation Cosine Similarity Scores

Occupation Descriptions. We select the 4-digit ISCO-08 level as the most detailed for the textual description of occupations. Unlike industries, this level includes distinct occupations that provide valuable insights for our analysis. Each ISCO-08 occupation corresponds to a specific set of tasks, though some tasks may overlap across occupations.

For each occupation $o \in \mathcal{O}$, we consider two components of its description: the occupation title o_1 and the task description o_2. We split the task description into individual tasks $s \in S_o \subset \mathcal{S}_{\mathcal{O}}$, where S_o represents the set of tasks for occupation o. This process yields 433 occupations at the 4-digit level, each represented by a title and an average of 7.5 tasks.

Embeddings. As with industries, we generate embeddings using the same sentence transformer model. The embedding of the occupation title is denoted as Emb_{o_1}, and the embedding of each task s is represented as Emb_{s,o_2}.

Cosine Similarity. For each patent $p \in \mathcal{P}$, we compute the cosine similarity between the full patent title and both components describing occupations: the occupation title o_1 and each task $o_{s,2}$ separately. Specifically, the cosine similarities are:

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

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

which reflect the semantic connection between o_1 and p, as well as between s and p.

For each (o_2, p) combination, as with industries, we retain the task with the highest cosine similarity score. Formally,

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

where C^p_{s,o_2} is the cosine similarity between patent p and task s given by Equation (7). No aggregation is needed for o_1 as each occupation has only one title. These scalars summarize the quality of the semantic match between an occupation and a patent, either through the occupation's title or its associated tasks.

Redundancy. We apply the same methodology as for industries, designating occupation–patent pairs (o, p) as relevant (denoted $(o, p)^\star$) if *both* sub-pairs $(o, p)_1$ and $(o, p)_2$ rank within the top 10 of their respective lists. This way, we retain only inventions relevant to the occupation.[15]

For each relevant pair, we calculate the harmonic mean of both cosine similarity scores, yielding the composite cosine similarity score for occupation–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 given by Equations (6) and (8), respectively. Equation (9) thus establishes a connection between an invention, identified in a single patent $p \in \mathcal{P}$, and a set of relevant occupations, where the innovation can be used. Tables A.6 to A.8 in the appendix

[15] As with industries, we manually exclude three specific connections to improve our exposure scores; see Appendix A.3 for details.

show redundancy filtering of occupations for our patent examples from Section 2.

3.3 Aggregation by Technology

We aggregate cosine similarity scores C_i^p and C_o^p from Equations (5) and (9) to the technology level. To do this, we apply a weighting scheme based on the number of citations a patent receives, as a proxy for its relevance and likelihood of use across industries and occupations. Given the variation in patent impact, it is essential that their weights reflect this heterogeneity (Hall et al. 2005, OECD 2009).

We assign a weight to the cosine similarity score of each relevant patent–industry/occupation pair, proportional to the number of citations the patent has received relative to the total citations of all relevant patents associated with the same occupation/industry, technology, and year.[16] The weight for a relevant pair $(d,p)^\star$ is calculated 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 patent p has received, $\mathcal{P}_{dt}^k$ is the set of patents related to emerging digital technology k, filed in year t, and relevant to industry/occupation $d = \{i, o\}$.

We apply this weighting scheme to aggregate patent-level cosine similarity scores to the technology level. The cosine similarity of a technology k to an industry/occupation is then calculated 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 is the cosine similarity score of the pair (d,p) as defined in Equations (5) and (9), ω_d^p is the weight from Equation (10), and $|\mathcal{P}_{dt}^k|$ is the total number of patents associated with industry/occupation–technology pair (d,k) for $d = \{i,o\}$ in year t. This yields the cosine similarity score of industry/occupation d with technology k for year t.

Accounting for impact via patent citations does not significantly alter our exposure measures.[17] This is because our exposure measure is based on semantic similarity between patents and industries/occupations. The semantic content of patents within the same technology

[16] In our sample, 41% of patents have not received any citations. This includes 1,733 patents (0.91%) with an indeterminate citation count, treated as zero, and 77,307 patents (40.54%) with no citations. Figure A.4 in the appendix shows the distribution of these patents across technologies, and Figure A.3 shows the overall citation distribution.

[17] Aggregating without citation weighting yields yearly cosine similarity scores very similar to those obtained with weighting. Figure A.5 in the appendix shows the correlation between weighted and unweighted scores. Pearson correlations between the two methods are approximately 0.99 for both industries and occupations, and the Spearman rank correlation is around 0.89.

does not vary significantly between more and less impactful patents, leading to minimal differences between weighted and unweighted exposure measures.

We then aggregate cosine similarity scores across all years to obtain a cumulative measure for the period 2012–2021, as follows:

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

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

3.4 Exposure Scores

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

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

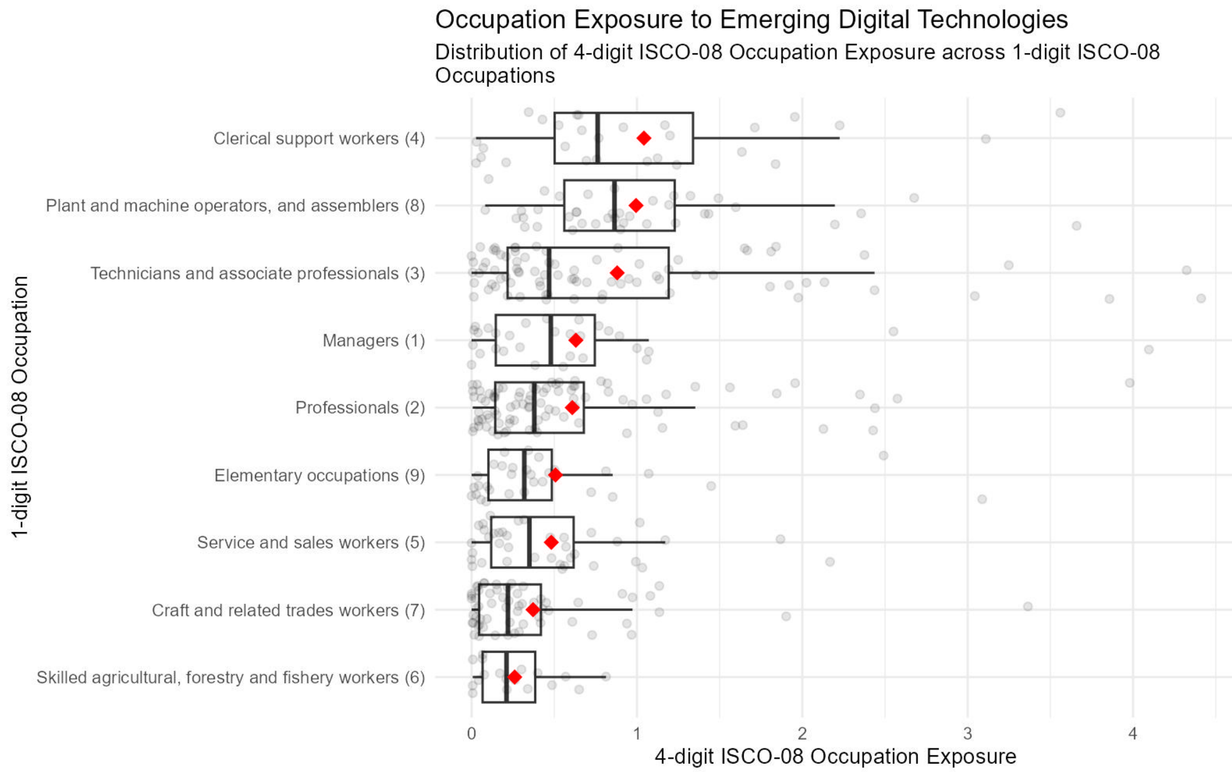
where C_d^k is the cumulative cosine similarity score for industry/occupation–technology pair (d, k) over 2012–2021 as defined in Equation (12).

While our exposure metric indicates the relevance of a specific technology to an industry or occupation, two clarifications are necessary. First, although exposure scores serve as a proxy for technology adoption across industries and occupations, they do not measure actual adoption. Second, our exposure scores are neutral regarding the relationship between technology and labor, meaning they do not assume *ex-ante* whether they are complements or substitutes in production. This neutrality is deliberate, allowing us to estimate the nature of this relationship later in Section 5.

We provide these data as an open–access resource, the 'TechXposure' database. The database also includes exposure measures at higher levels of aggregation, such as the 1-digit and 2-digit levels for industries, and the 1-digit to 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 but also capture additional dimensions of these technologies that previous studies have not addressed, either due to the nonexistence of these technological features or a narrower focus. For example, the AI exposure scores in Webb (2019) are limited to core aspects of AI, such as industrial automation, workflow management systems, cloud computing, and machine learning. In contrast, Felten et al. (2021) cover a broader scope but focus only on *intangible* AI applications, excluding AI embedded in *tangible* technologies like industrial and mobile robots, and IoT. For details on

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.

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 first examine the overall exposure of occupations, defined as the average exposure across all technologies: $X_o = \frac{1}{40}\sum_k X_o^k$, where X_o^k is defined by Equation (13). Figure 1 shows the distribution of exposure to emerging digital technologies across ISCO-08 occupations. In this figure, 4-digit occupations are grouped into their respective 1-digit categories, with their distribution presented as a boxplots. Occupation groups are ranked by their average 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 (Group 8) are the most exposed to emerging digital technologies. Occupations in these groups typically involve a high proportion of routine tasks related to information handling and production equipment supervision, respectively. Although these middle-paying jobs have already been significantly impacted by earlier ICT waves (Goos and Manning 2007, Goos et al. 2009, Goos et al. 2014), they remain strongly associated with newer ICT vintages, particularly emerging digital technologies that facilitate semi- or unsupervised information handling and equipment operation.

High-paying occupations, such as Managers (Group 1), Professionals (Group 2), and Technicians and Associate Professionals (Group 3), are the next most exposed to emerging digital technologies. These roles predominantly involve non-routine cognitive tasks that frequently require a variety of digital technologies. As technologies evolve and new vintages emerge, these occupations may experience shifts in task structure due to the introduction of new tasks.

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

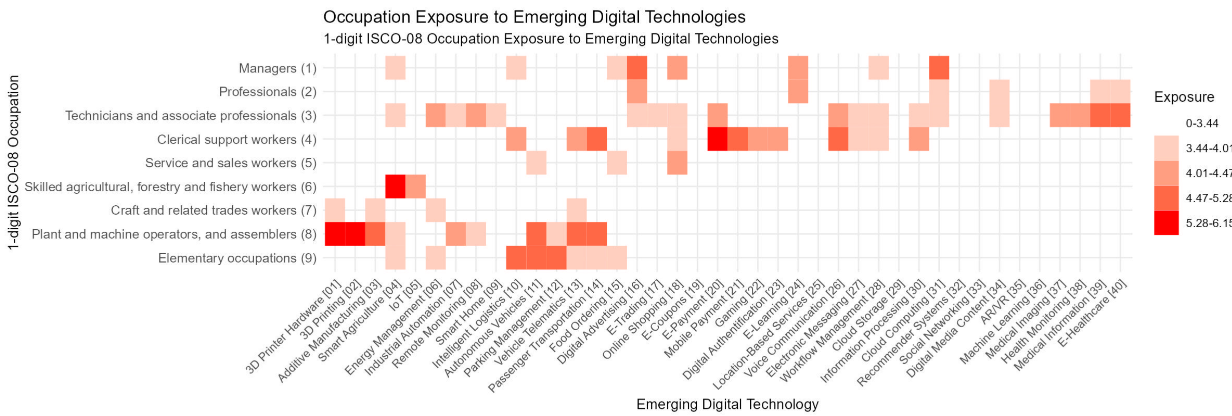
Lastly, we observe greater heterogeneity in exposure to emerging digital technologies within high-paying occupations (Groups 1, 2, and 3) compared to middling occupations (Groups 4 and 8). This suggests that only a subset of high-paying roles is closely associated with these technologies, whereas middling occupations display more generalized exposure.

We analyze the overall exposure of 1-digit ISCO Groups by examining their exposure to each of the 40 emerging digital technologies. Figure 2 displays this exposure as a heatmap, with exposure levels shown at the intersections of 1-digit occupations (rows) and emerging digital technologies (columns). This visualization reveals two distinct patterns.

First, we observe a clear distinction between *tangible* and *intangible* technologies in their relevance to different occupations. *Tangible* technology families, such as 3D Printing, Embedded Systems, and Smart Mobility, are more relevant to manual occupations in ISCO Groups 6 to 9. In contrast, *intangible* technology families, including E-Commerce, Payment Systems, Digital Services, Computer Vision, and HealthTech, are more pertinent to cognitive occupations, particularly within ISCO Groups 1 to 4.

Second, we observe that Technicians and Associate Professionals (Group 3) and Clerical Support Workers (Group 4) are exposed to a broad range of emerging digital technologies. In contrast, Managers (Group 1) and Professionals (Group 2) are associated with a narrower scope of relevant technologies, primarily within the domain of intangible technologies. Simi-

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.

larly, exposure within ISCO Groups 6 to 9 is exclusively focused on tangible technologies. Notably, this aggregated mapping conceals some variability in exposure within 1-digit ISCO-08 occupations due to aggregation; for a more detailed mapping at the 2-digit level, see Figure B.1 in the appendix.

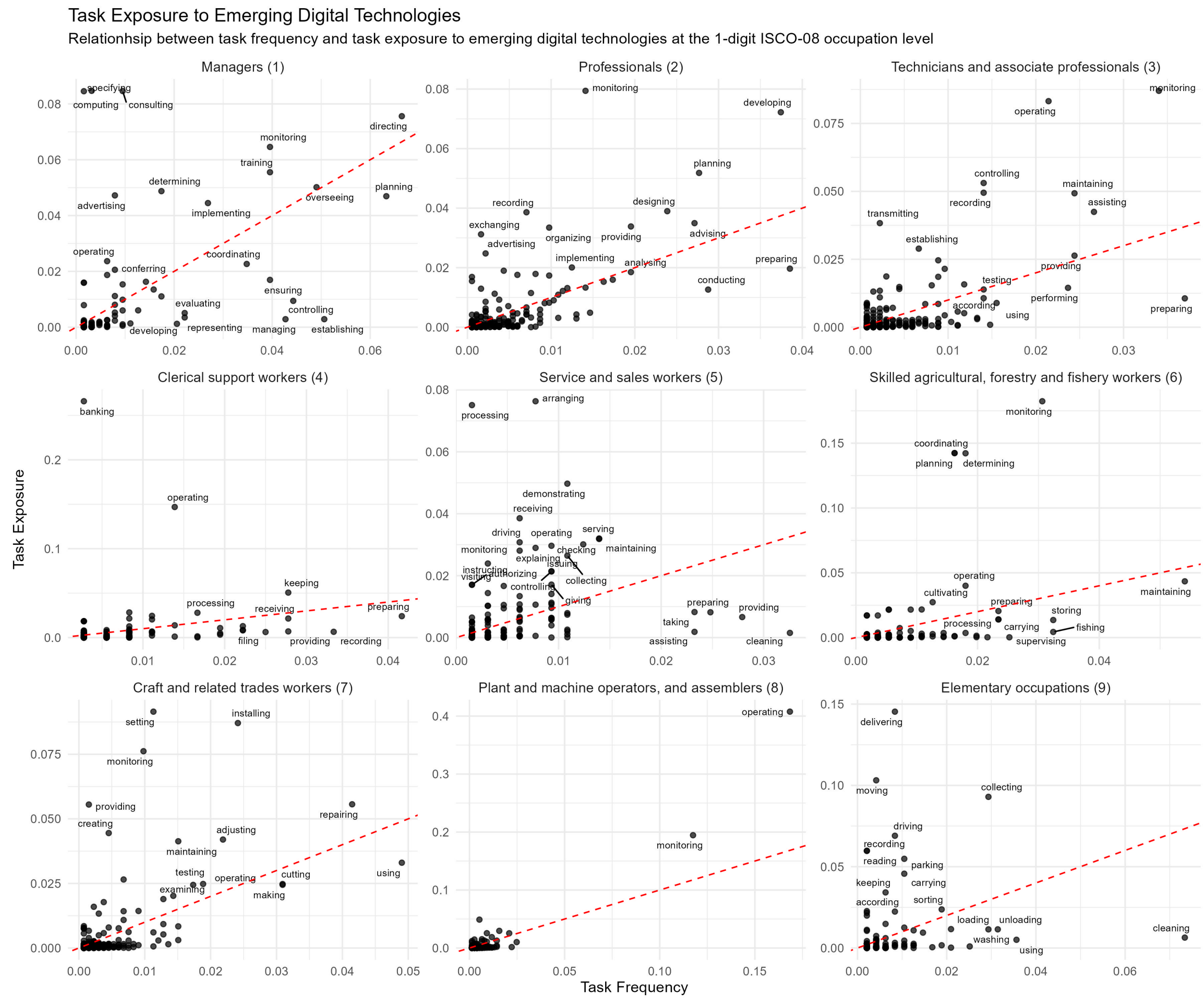
Leveraging the semantic structure of tasks in the 4-digit ISCO-08 taxonomy, we identify the actions within tasks most exposed to our emerging digital technologies. By construction of ISCO-08 taxonomy, task descriptions begin with *gerunds*, which define the primary action (e.g., planning, monitoring, developing, preparing, operating, cleaning). For each gerund, we calculate its *baseline frequency* or simply *task frequency*, representing its occurrence in a 1-digit group of ISCO-08 taxonomy, i.e., the baseline corpus. The baseline frequency reflects how often the taxonomy uses a gerund to describe tasks within a 1-digit ISCO group, with more frequent gerunds being core to that group. We also calculate the gerund's *target frequency* or *task exposure*, representing its occurrence among established task-technology pairs in a 1-digit group, i.e., the target corpus. The target frequency is high for tasks most exposed to emerging digital technologies. [18]

Figure 3 displays the most exposed tasks, expressed as gerunds, for 1-digit ISCO-08 groups. It highlights actions to which our emerging digital technologies are the most relevant.

We observe heterogeneity in task exposure both within and between 1-digit ISCO-08 groups. This suggests that the extent to which *recurrent* (rightmost) or *specialized* (leftmost) tasks are exposed to emerging digital technologies depends heavily on the 1-digit ISCO-08 group. However, all high-skilled occupation groups, including Managers (Group 1), Professionals (Group 2), and Technicians and Associate Professionals (Group 3), display a clear tendency for the

[18] All frequencies are relative.

Figure 3: Most Exposed Tasks by 1-digit ISCO-08 Occupation Groups

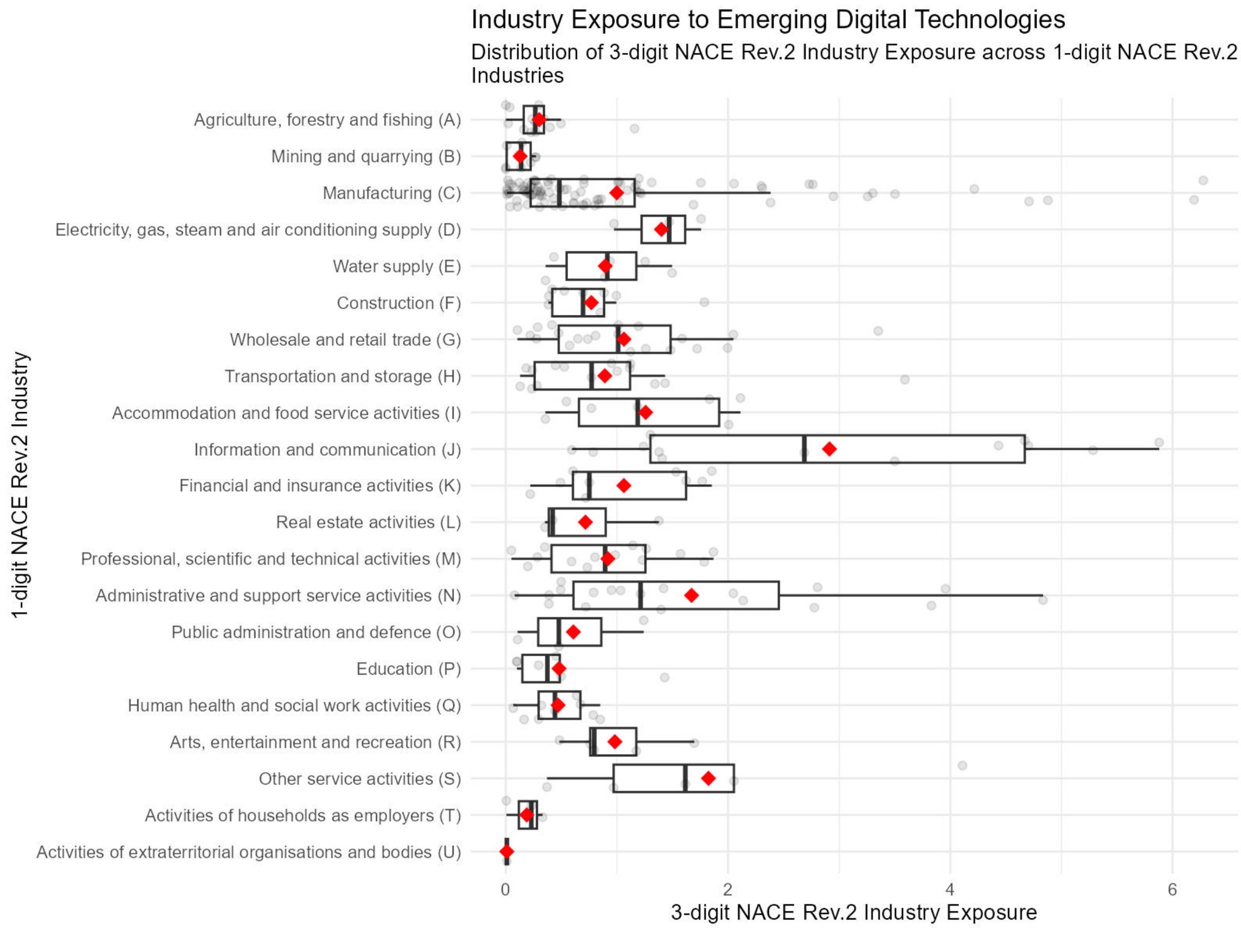


Notes: This figure displays the top exposed tasks, summarized with their gerunds, to all emerging digital technologies by 1-digit ISCO-08 group. The horizontal axis is the term's baseline frequency (i.e., ISCO-08 classification). The vertical axis is the term's target frequency. The probabilities in the target corpus are weighted by the cosine similarity between the task and the technology. The diagonal line indicates equality between the baseline and target frequencies.

majority of their *recurrent* tasks to be exposed (above the red line).

Overall, the set of most exposed tasks is fairly unique to each 1-digit ISCO-08 group, with the exception of Technicians and Associate Professionals (Group 3) and Plant and Machine Operators (Group 8), who both prominently feature operating and monitoring tasks among their top exposures. This finding aligns with previous results, as these two groups have been jointly identified as highly exposed to tangible emerging digital technologies, whereas Managers (Group 1) and Professionals (Group 2) are associated with a smaller, more specific subset

Figure 4: 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.

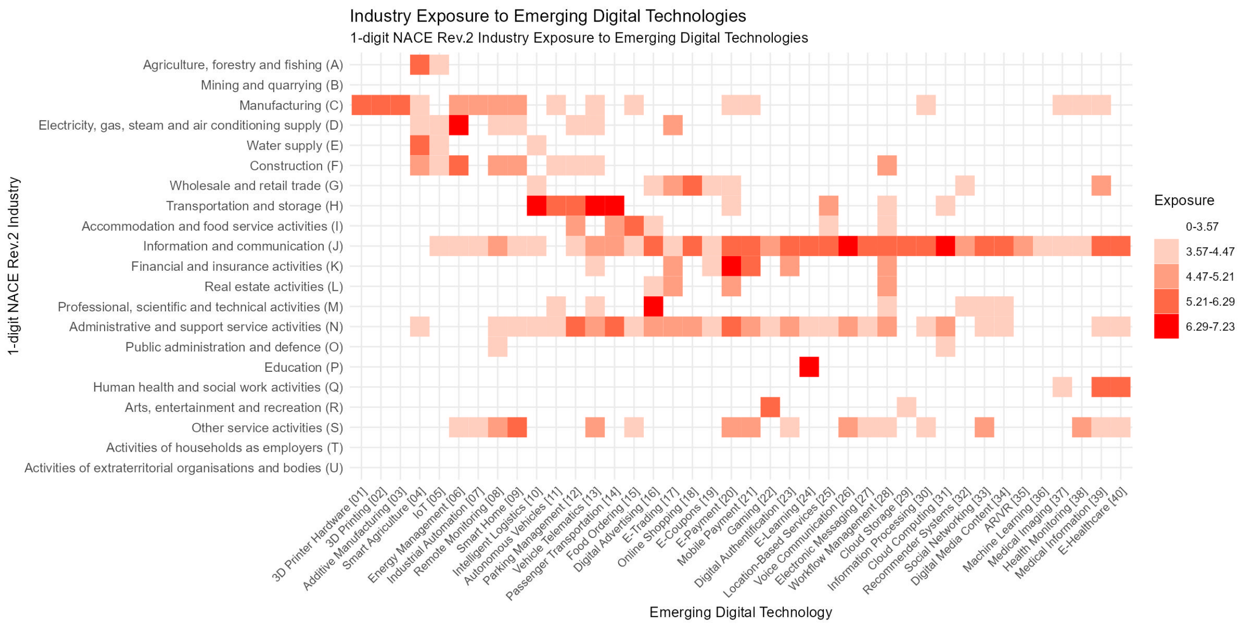
of technologies (see Figure 2).[19]

4.2 Industry Exposure to Emerging Digital Technologies

For industries, we examine overall exposure as the average exposure across all technologies: $X_i = \frac{1}{40}\sum_k X_i^k$, where X_i^k is given by Equation (13). Figure 4 shows 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, with distributions presented as a

[19]Further insights can be gained from examining the tasks most exposed to individual emerging digital technologies. We briefly discuss a few technologies frequently addressed in automation and labor market literature. Monitoring and oversight tasks are particularly prominent for Internet of Things technology across all occupational groups. Industrial Automation and Robot Control technology primarily relates to operation control tasks, while Machine Learning and Neural Networks technology generally involves various processing tasks. These figures are available in the online appendix, and the complete set of figures is available upon request.

Figure 5: 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.

boxplot.

We observe that the Information and Communication (J) and Manufacturing (C) sectors contain the most exposed 3-digit industries. This finding is notable given the substantial heterogeneity in exposure within these 1-digit sectors. Such differences in exposure may reflect whether industries act as producers or intensive users, rather than light users, of emerging digital technologies. Specifically, industries within the Information and Communication (J) sector are likely to produce intangible technologies, while certain industries within the Manufacturing (C) sector likely 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).

We analyze the overall exposure of 1-digit NACE sectors by examining their exposure to each of the 40 emerging digital technologies. Figure 5 shows the exposure heatmap for 1-digit sectors; see Figure B.2 in the appendix for a more detailed mapping at the 2-digit level.

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

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.[20] Our sample includes 320 NUTS-2 regions in 32 European countries.[21]

Our outcome variable is the change in the regional employment-to-population ratio between 2012 and 2019. This ratio is defined as the number of employees within the group of interest (e.g., the total population or the youth population) divided by the total number of individuals aged 15 or older.

Our analysis uses a long-difference approach for the period between 2012 and 2019. We begin in 2012, which is the starting year of our patent sample and therefore serves as the baseline for measuring exposure to emerging digital technologies. We conclude in 2019 to avoid potential confounding factors associated with employment and population changes due to the COVID-19 pandemic.[22]

The EU-LFS also provides data on the number of employees across 1-digit NACE indus-

[20] 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).

[21] The countries in the sample are (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.

[22] Although our exposure metrics in Section 3 span 2012–2021, we recalculate them for the subperiod 2012–2019 to ensure consistency with the timeframe in this analysis.

tries, categorized into 10 distinct sectors.[23]

Estimating the causal impact of technology on employment involves two main challenges: reverse causality and omitted variable bias. Reverse causality implies that technological advancements could be driven by labor shortages or rising labor costs. Furthermore, unobserved factors—such as shifts in industry organization or infrastructure investments—might simultaneously influence both technological change and employment levels.

To address these concerns, we adopt a shift-share strategy, leveraging recent advancements in this methodology (Adão et al. 2019; Goldsmith-Pinkham et al. 2020; Borusyak et al. 2021). Specifically, we use the Bartik instrument to measure the region's exposure 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.[24] The term X_j denotes 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 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, rely on the semantic similarity between patents and industry descriptions. Notably, only 7.1% of the patents in our sample originate from Europe, indicating that the advancement of these technologies is largely a global phenomenon. Consequently, global technological trends are unlikely to be driven solely by regional labor markets in Europe. To reinforce this point, we recalculated our exposure measure after excluding European patents.[25]

Since our shocks are assumed to be exogenous to local employment changes in European

[23]These sectors include 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).

[24]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 Market Services (average employment share of 23.8%), the Public Sector (23.7%), and Industry (17.9%). The Information and Communication sector, which is highly exposed to emerging digital technologies, accounts for only 2.3% of employment on average.

[25]In the Online Appendix, we compare the 1-digit industry exposure scores with and without European patents (i.e., patents filed with the European Patent Office). The correlation remains approximately 0.99 across all 40 emerging digital technologies, underscoring the global nature of these technological advancements.

labor markets, we apply the equivalence proposed by Borusyak et al. (2021) and can thus consider our shift–share as a valid instrument.[26] In addition to the quasi-random assignment of shocks, our second identifying assumption is 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.[27]

Figure 6 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 exposure levels are predominantly located in Western Europe, specifically in countries such as Germany, Italy, Spain, Switzerland, and the UK.

Figure 7 shows a positive relationship between the change in the employment-to-population ratio from 2012 to 2019 and the regional exposure to emerging digital technologies.[28] However, although this 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 change in the regional employment-to-population ratio 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 characteristics,[29] $\phi_{c(r)}$ represents country fixed effects, and u_r is the error term.

Table 3 presents estimates of the effect of regional exposure to emerging digital technologies on the change in the employment-to-population ratio from 2012 to 2019. Since exposure is standardized across regions, the estimated coefficient $\hat{\beta}$ can be interpreted as the effect of a one-standard-deviation increase in regional exposure on the employment-to-population ra-

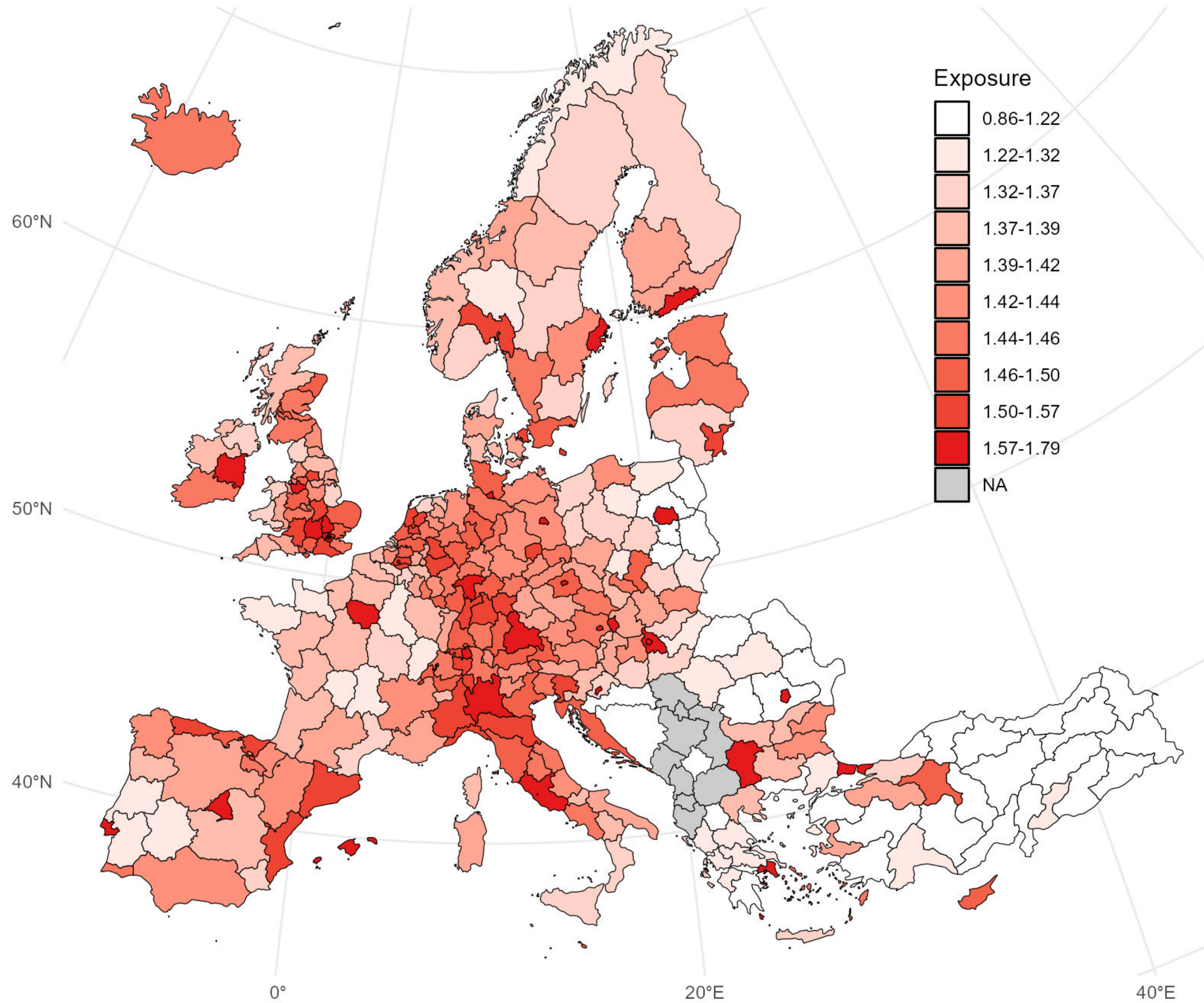
[26]Figure C.1 in the appendix illustrates the positive correlation between employment changes and shocks at the sector level in Europe.

[27]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 shown in Table C.1. This HHI can be considered relatively small, as the minimum index under a uniform distribution would be $1/|J| = 0.1$. Thus, the latter part of our assumption is realistic. The effective sample size, corresponding to the inverse of the HHI, is 5.95.

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

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

Figure 6: Geographic Distribution of Regional Exposure to Emerging Digital Technologies across Europe from 2012 to 2019

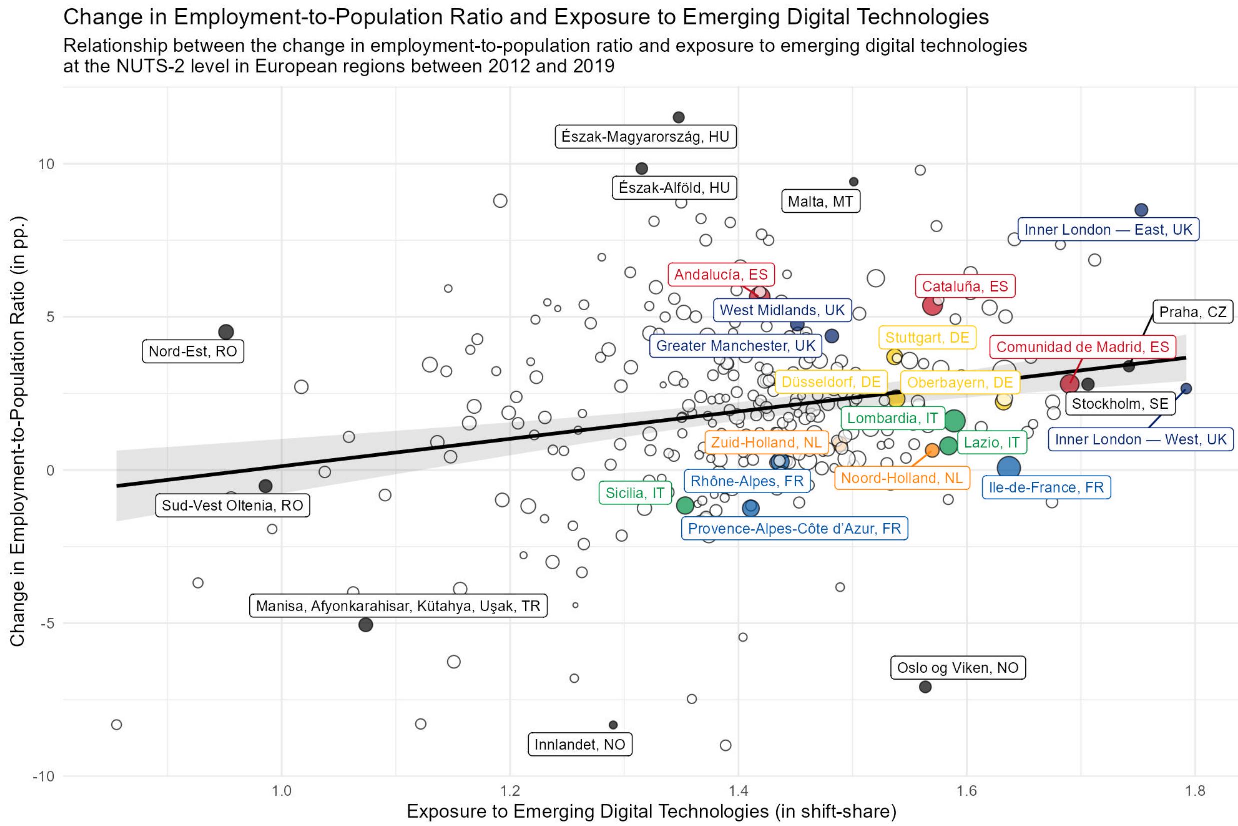


Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 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.

tio, measured in percentage points (pp.). Following recent literature on shift-share designs, we control for the sum of exposure shares (Borusyak et al., 2021) and report 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 7 remains robust when fixed effects and various covariates, such as regional demographic characteristics and industry share, are included. In the specification with all covariates (shown in the last column), a one-standard-deviation

Figure 7: 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.

increase in regional exposure corresponds to a 1.029 pp. change, or 2.05%, in the employment-to-population ratio from 2012 to 2019.

The latter estimation suggests that the overall impact of emerging digital technologies on employment is positive at the regional level. However, it remains to be seen whether this positive relationship holds uniformly across all demographic groups. Table 4 provides estimates of the same empirical specification, including the full set of control variables, for various 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 leads to a 0.673 pp. change (equivalent to 3.03%) in the employment-to-population ratio for women and a 0.355 pp. change (1.27%) for men. Although the impact is twice as large for women, there is greater regional heterogeneity in this effect, as indicated by the larger standard errors

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).

compared to those for men.

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

Emerging digital technologies positively impact employment only at the extremes of the skill distribution, specifically among low- and high-skilled workers, with respective changes of 0.715 pp. (+6.01%) and 0.738 percentage points (+4.92%) in their employment-to-population ratios following a one-standard-deviation increase in regional exposure. Conversely, a similar increase in regional exposure results in a decline of 0.412 percentage points (−1.78%) in the employment-to-population ratio for middle-skilled workers. This differentiated effect indicates that job polarization continues to be driven by emerging digital technologies.

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).

As a robustness check, we conduct a placebo test by estimating the effect of regional exposure to emerging digital technologies from 2012 to 2019 on the change in the employment-to-population ratio during the pre-period, specifically from 2002 to 2009. These estimates, presented in Table C.2 in the appendix, show null effects for all demographic groups in the pre-period, reinforcing 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 with our expectations, as regions more exposed to emerging digital technologies are likely those where the share of high-skilled workers has increased the most, given that these technologies are primarily developed and produced by workers in this demographic group.

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 apply the same shift-share strategy independently to each technology.[30] The regional exposure to a specific technology k is given 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 in 2010, and X_j^k is the exposure of sector j to technology k.[31]

Estimating the individual effect of a single technology on labor is challenging because technologies can be complementary and are often implemented together. For example, recent literature on the employment impact of robots, a specific technology, typically 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, specific emerging digital technologies, such as Cloud Storage, may complement other digital technologies. Additionally, the degree of complementarity may vary within the same technology family or with other emerging 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. Therefore, we propose an empirical approach that accounts for these complementarities to mitigate bias in estimating the individual impact

[30]We also estimate the employment impacts at the emerging digital technology family level; see Appendix C.4 for further details.

[31]Figures C.4 to C.8 in the appendix report the geographic distributions of exposure to individual emerging digital technologies.

of a specific technology on employment.

We estimate the impact of regional exposure to each emerging digital technology on the regional employment-to-population ratio using 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 (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 technologies, and Z includes the same set of covariates as in Equation (15). Both $X_r^{K\setminus\{k\}}$ and X_r^{-K} are calculated as shift-share variables.

The estimated coefficient of interest, $\widehat{\beta}_k$, represents the employment effect, measured as a percentage point change, resulting from a one-standard-deviation increase in regional exposure to a specific emerging digital technology k. This estimate is conditional on regional exposure to both its technology family and all other emerging technologies. Controlling for these additional exposures is essential for isolating the causal effect of regional exposure to a specific technology, independent of the effects of other emerging digital technologies or their combinations.

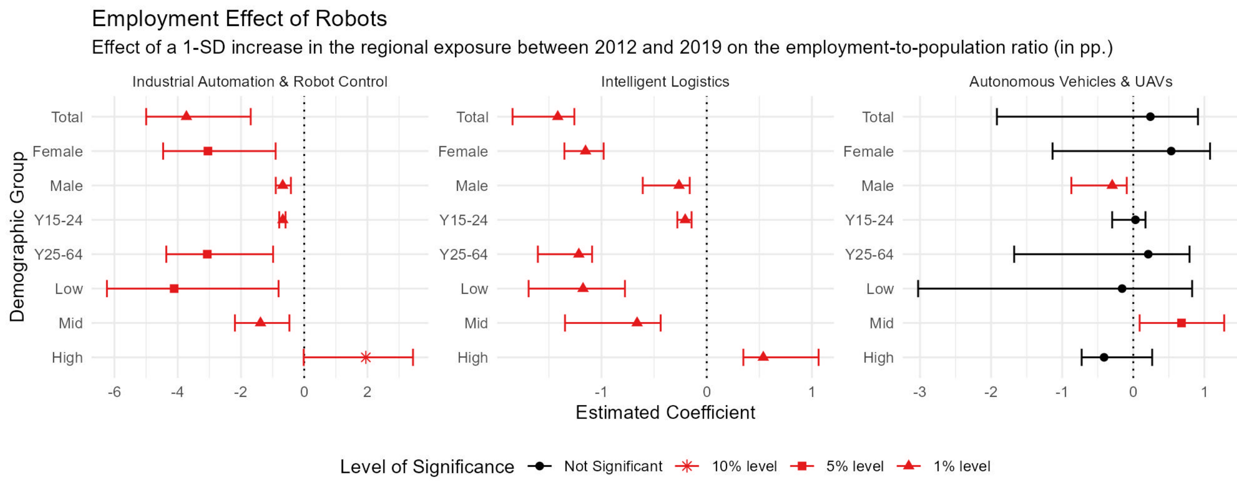
We present results at the individual technology level for two groups of technologies that have received significant attention in the literature and exhibit noteworthy patterns: robots and data-intensive technologies. Estimates for all individual technologies are provided in the appendix (see Figures C.9 to C.13).

Robots. Figure 8 displays the estimated coefficients and their corresponding 95% AKM0 confidence intervals for the employment effects of three technologies that encompass three different types of robots. The figure is interpreted as follows: each panel represents a different technology, with demographic groups listed on the vertical axis and the estimated coefficients shown on the horizontal axis.

Both industrial automation (which includes industrial robots) and intelligent logistics (which encompasses mobile robots) have negative impacts on employment, particularly for female and mature workers, with the impact of industrial automation being twice as large as that of intelligent logistics. Increased regional exposure to these robot-inclusive technologies raises employment among high-skilled workers while reducing it for low- and middle-skilled workers.

We do not find any significant effect of autonomous vehicles on total employment, except for a small decrease in male employment and a slight increase in middle-skilled employment.

Figure 8: 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.

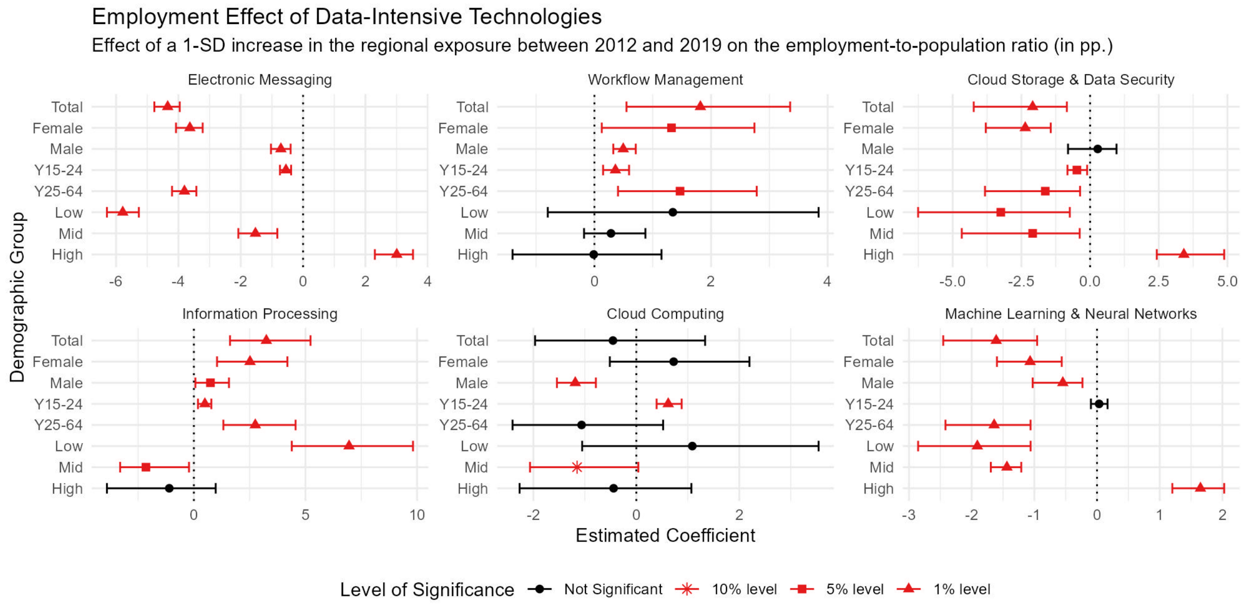
This may reflect the limited diffusion of this technology across European regions, beyond a few highly advanced areas.

Data-Intensive Technologies. Figure 9 shows the estimated coefficients for the employment effects of data-intensive technologies. Among these, Electronic Messaging, Cloud Storage & Data Security, and Machine Learning & Neural Networks have significant negative impacts on the overall employment-to-population ratio. Similar to robots, these technologies tend to displace female and mature workers rather than male and young workers. Additionally, as with the impact of robots, greater regional exposure to these technologies increases employment among high-skilled workers, while employment declines for low- and middle-skilled workers.

While Cloud Storage has a sizeable impact on employment, we do not detect any impact from Cloud Computing, as indicated by the bottom middle panel. This suggests that Cloud Computing, on its own, neither creates employment opportunities nor displaces workers, conditional on the presence of other emerging digital technologies. However, it may act as a complementary or *enabling* technology—one that amplifies the employment effects of other technologies when used in combination.

Lastly, Information Processing and Workflow Management exhibit positive impacts on employment. A one-standard-deviation increase in exposure to these technologies raises the

Figure 9: 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.

employment-to-population ratio by 3.25 pp. and 1.81 pp., respectively. For Information Processing, the employment gains are concentrated among low-skilled workers. Although the coefficients for middle- and high-skilled employment are negative but insignificant, this may suggest that Information Processing enables low-skilled workers to engage in more complex and abstract tasks, thereby increasing labor demand at the lower end of the skill distribution. Workflow Management also positively impacts employment, benefiting all demographic groups. When examining skill levels, we find no evidence of skill-biased technological change, as all coefficients are positive, though not statistically significant. This suggests that Workflow Management has a uniformly positive impact across the skill distribution.

6 Conclusion

Recent advancements in digital technologies, particularly in AI, have raised significant public and academic interest in understanding the impact of these emerging technologies on future employment. Determining whether these technologies will generate more jobs than they displace is a crucial issue for both individuals and policymakers. However, previous research has primarily concentrated on examining either specific technologies, such as industrial robots or certain AI applications, or a broad range of digital technologies commonly referred to as "automation technologies".

In this study, we identify a broader array of emerging digital technologies than previous research and analyze their impact on employment. Our empirical results show that these technologies generally have a positive effect on the employment-to-population ratio, thereby creating jobs rather than eliminating them. However, we observe considerable heterogeneity in these impacts: while some technologies, such as robots, negatively affect employment—especially among low- and middle-skilled workers—others, like information processing and workflow management systems, contribute to employment growth. Furthermore, our findings suggest that focusing solely on specific technologies, such as AI and robotics, may overlook the broader positive employment effects arising from complementarities among diverse digital technologies.

A key component of this work is a new measure of the exposure of industries and occupations to 40 digital technologies that have emerged over the past decade. By leveraging state-of-the-art NLP tools, such as sentence transformers, we introduce a novel methodology to assess exposure at a granular level. Our pioneering dataset is available as an open–access resource, called the 'TechXposure' database.

We outline the advantages and limitations of the exposure scores in the 'TechXposure' database. First, because our scores are based on text data from standard international classifications, they are universally applicable and not specific to any European country. Second, our method does not rely on keywords or tokens and only requires relevant patents, making it adaptable to other contexts, such as green technologies or future ISCO/NACE classifications.

However, our exposure scores do not account for the augmentation or automation effects on occupations and industries; they solely reflect the relevance of technologies to a given industry or occupation. This limitation allows us to make fewer assumptions during data construction, acknowledging that some technologies may positively impact employment in one context but negatively in another.

Additionally, our set of emerging digital technologies excludes recent developments in Large Language Models (LLMs), like ChatGPT, as our analysis covers technologies up to 2021. Nonetheless, it includes several other AI applications, particularly in Machine Learning (for computer vision), Information Processing, and Workflow Management. Finally, while our exposure metrics serve as proxies for the adoption of these digital technologies across industries and occupations, they do not measure actual adoption—a topic we plan to explore in future research.

We consider our paper a foundational contribution to future research on technological change and labor markets. By constructing this open–access database, we anticipate that future studies will greatly benefit from its unprecedented detail in analyzing the exposure of occupations and industries to a wide range of emerging digital technologies. These include not only widely discussed technologies, such as robots and AI, but also lesser-studied ones like social networks, cloud computing, and health technologies. Since our database is based on international classifications of occupations and industries, it offers valuable potential for research beyond the US context. Such research could yield insights into the economic impact of emerging technologies, especially given Europe's institutional diversity, which may significantly shape technology adoption and labor market effects. We believe our database is accessible and user-friendly 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. Working Paper 2049427, National Bureau of Economic Research.

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.

Adachi, D., Kawaguchi, D., and Saito, Y. U. (2024). Robots and Employment: Evidence from Japan, 1978–2017. *Journal of Labor Economics*, 42(2):591–634.

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.

Aksoy, C. G., Özcan, B., and Philipp, J. (2021). Robots and the Gender Pay Gap in Europe. *European Economic Review*, 134:103693.

Albanesi, S., Silva, A. D. D., Jimeno, J. F., Lamo, A., and Wabitsch, A. (2023). New Technologies and Jobs in Europe. Working paper, 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*, 139(3):1399–1465.

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., Katz, L. F., and Krueger, A. B. (1998). Computing Inequality: Have Computers Changed the Labor Market? *The Quarterly Journal of Economics*, 113(4):1169–1213.

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.

Bonfiglioli, A., Crinò, R., Fadinger, H., and Gancia, G. (2024). Robot Imports and Firm-Level Outcomes. *The Economic Journal*, page ueae055.

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.

Cockburn, I. M., Henderson, R., and Stern, S. (2019). *4. The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis.* University of Chicago Press.

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. (2023). Induced Automation: Evidence from Firm-level Patent Data. *University of Zurich, Department of Economics, Working Paper.*

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? *arXiv preprint arXiv:2303.01157*.

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. Working paper.

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.

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. *France: OECD Publishing*.

Savona, M., Ciarli, T., Steinmueller, E., and Vannuccini, S. (2022). The Design of Digital Automation Technologies: Implications for the Future of Work. *EconPol Forum*, 23(5):4–10.

Song, K., Tan, X., Qin, T., Lu, J., and Liu, T.-Y. (2020). Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33:16857–16867.

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. (2021). Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey. Working paper, National Bureau of Economic Research.

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 patent corpus in Chaturvedi et al. (2023) is constructed by querying the Derwent Innovation Index (DII) database. The query has two components, using patent codes (Derwent Manual Codes and International Patent Classification codes) and keywords from previous studies on digital automation technologies and Industry 4.0 (Cockburn et al. 2019; Webb 2019; Martinelli et al. 2021). The first component retrieves digital automation inventions related to *i)* process and machine control in physical production sectors like manufacturing, agriculture, mining, and construction, and *ii)* process and workflow control in services. The second component narrows the sample to large technology families, such as AI, computing, networking, data management, and user interfaces, based on prior research on emerging digital technologies (Savona et al. 2022). The final sample includes 1,143,033 patent families from 2000 to 2021. Figure **??** illustrates the SQL-style structure of the query, with the full details available in the Online Appendix.

Figure A.1: SQL-stylized Structure of the Patent Query

```
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 the structure of the patent query used to construct the total sample in Chaturvedi et al. (2023). The list of CPC codes in A, B, and C is available in the Online Appendix.

Patent Embeddings. To analyze emerging digital technologies, Chaturvedi et al. (2023) concatenate patent titles and abstracts to create embeddings. Using the pre-trained sentence transformer model *all-mpnet-base-v2* (Song et al. 2020), each patent text is mapped into a 768-dimensional space, converting text into semantic vectors. This transformation enables large-scale analysis and comparison of document meanings using other ML and NLP methods.

Core Digital Patents. To identify the backbone of the corpus of digital automation inventions, the 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 core among patents. Thus, it measures the local density of a focal document compared to the local density of its k-nearest neighbors in the semantic space. The locality (i.e., the size of the neighborhood) is set by the parameter k. A document with a notably dense neighborhood is considered part of the backbone. For identification of the backbone among digital innovations, larger values of k are more suitable as they allow for larger neighborhoods and hence, a wider reference group of patents to compute LOF measure. Chaturvedi et al. (2023) use $k = 1000$ and the LOF measure is computed for each patent in year t using the 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 identify established/core digital patents in the most recent decade of the patent sample, i.e. 2012–2021. They begin with a base sample of 258,344 patents from 2001-–2011 and calculate the LOF measure for each year from 2012 to 2021, updating the base sample iteratively. For example, to compute the LOF measure for patents filed in 2014, the base sample includes patents from 2001–2013.

Lastly, *core patents* are defined as those in the bottom 10% of the LOF measure for each year over the 2012–2021 period. These patents form the backbone of the patent corpus, being the most representative of digital automation technologies. A low LOF measure indicates a dense semantic neighborhood, meaning these patents are highly central within their local semantic spaces.

Offshoots. To track the development of these core technological innovations throughout the 2012–2021 period, Chaturvedi et al. (2023) identify their offshoots (i.e., subsequent inventions that build on and are semantically similar to the core ones). For each core patent, 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.5 presents the correlation between citation-weighted and unweighted yearly cosine similarity scores for both industries and occupations.

A.6 Technology Co-Occurrence

Using cosine similarity scores, we analyze 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.6 shows technology groupings based on cosine semantic scores, revealing distinct segments 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 aggregation levels within the ISCO and NACE classifications, we apply the inverse hyperbolic sine transformation to the average cosine similarity score, aggregated from the most granular classification level up to the desired level of interest.

For example, consider the calculation of the exposure score for a 1-digit NACE industry $I \subset \mathcal{I}$ to an emerging digital technology k. We start by aggregating the cosine similarity score to the higher classification level 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 approach is similarly used to calculate exposure scores for 2-digit industries and for occupation exposures at higher aggregation levels.

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. Challenges in comparison are the different occupational classifications and variations in the definitions of technologies among the studies. To address classification differences, we use crosswalks between systems, aggregating exposure scores within a 4-digit ISCO-08 occupation by averaging exposures across all matched occupations. Table A.9 summarizes key aspects of these studies, including descriptions of the technologies represented, data and methodology used for exposure score construction, and score interpretations, facilitating comparison and aiding result interpretation.

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 standardized 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 standardized exposure scores. We also standardize 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.7. 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 occupation exposure 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 as they do not account for embedded AI (see Table A.9), reflecting that their exposure scores are more oriented 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

Code	NACE Industry	Cosine Similarity		
		$C_i^{p_1}$	$C_i^{p_2}$	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

		Cosine Similarity		
Code	ISCO Occupation	$C^p_{o_1}$	$C^p_{o_2}$	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

Code	ISCO Occupation	Cosine Similarity		
		$C^p_{o_1}$	$C^p_{o_2}$	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	

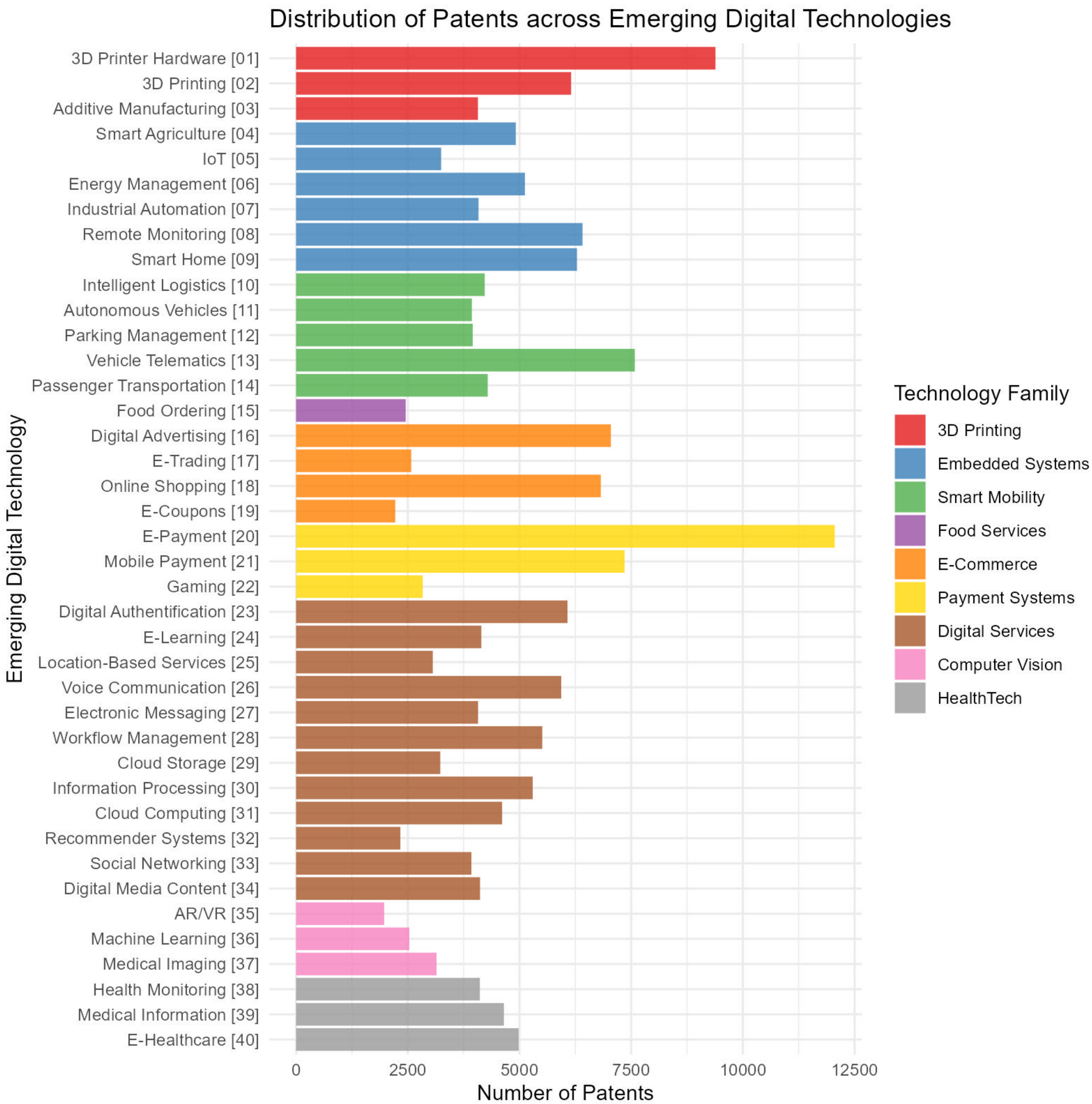
Notes: This table presents the redundancy filtering of occupations for the Patent ID 201713859U (i.e., "Vehicle intelligent logistics control device, has GPS locating module for obtaining position information of transport vehicle through main control chip, RFID reader for reading RFID tag information, 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	

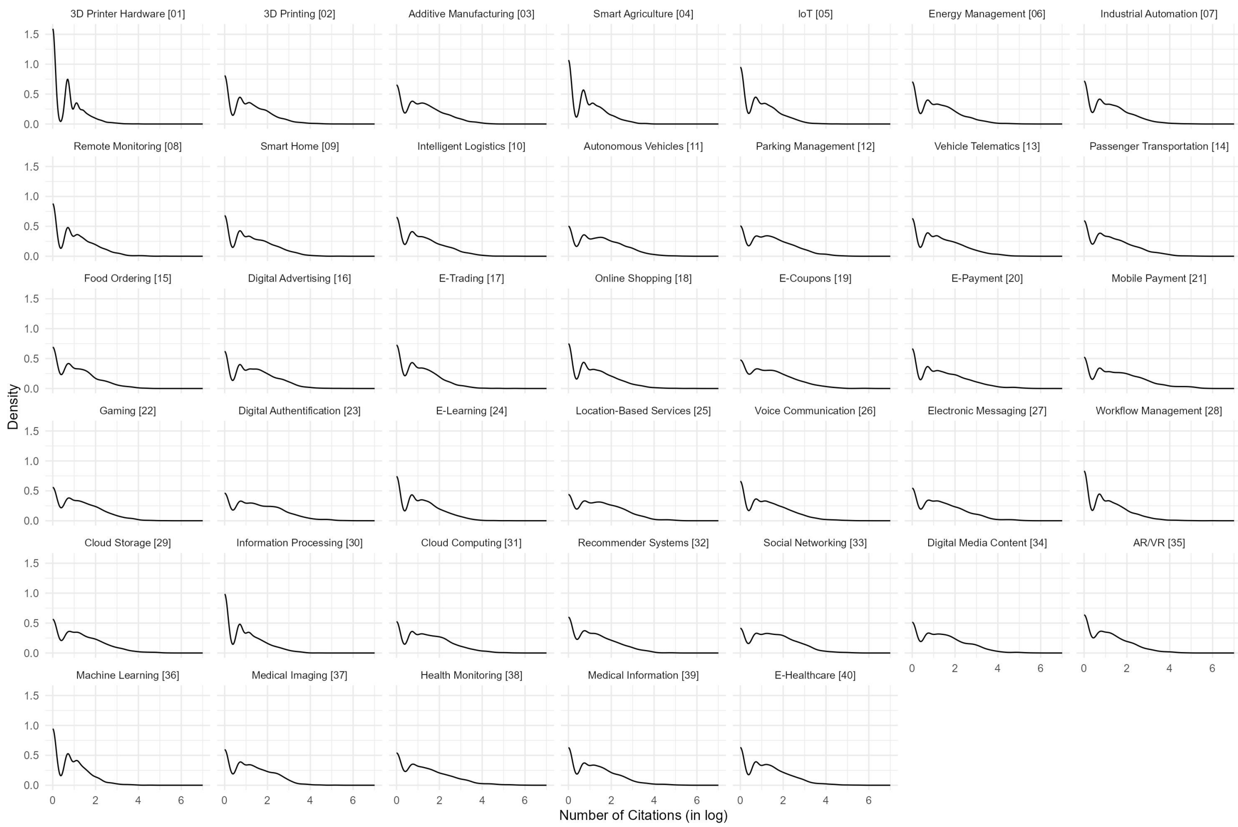
Notes: This table presents the redundancy filtering of occupations for the Patent ID 202048118D (i.e., "System for recognizing training speech, has process or which is configured to increment counter associated with word sequences, and train language model of automatic transcription system using word 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



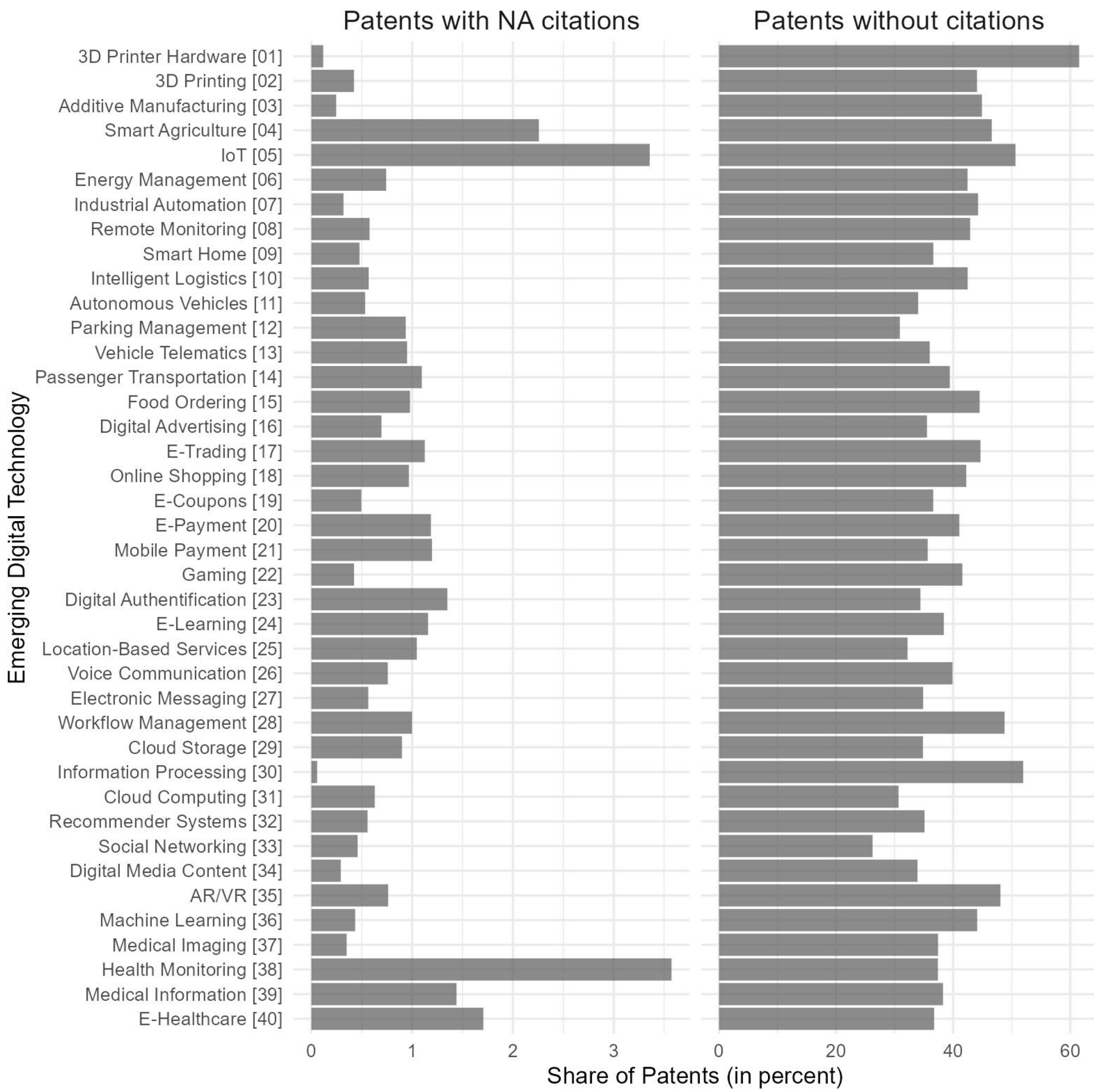
Notes: This figure presents the distribution of patents across emerging digital technologies. The set of patents includes 190,714 Derwent patents, filed between 2012 and 2021. This patent set constructed by Chaturvedi et al. (2023) comprises the core patents related to digital innovations, together with 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



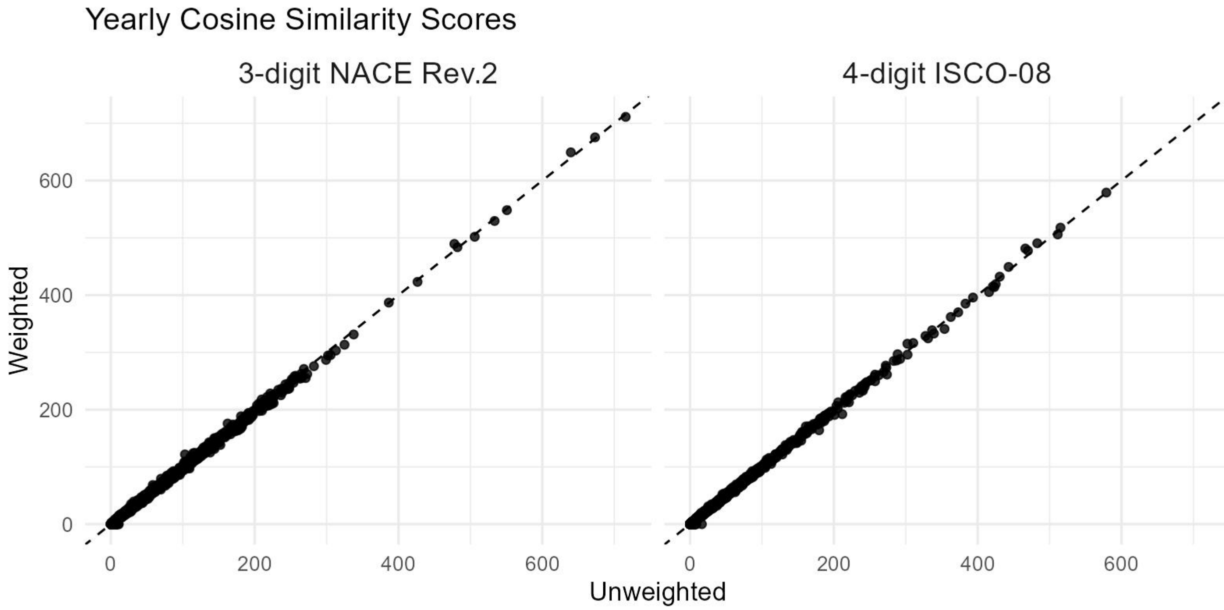
Notes: This figure presents the log distribution of patent citations across emerging digital technologies.

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



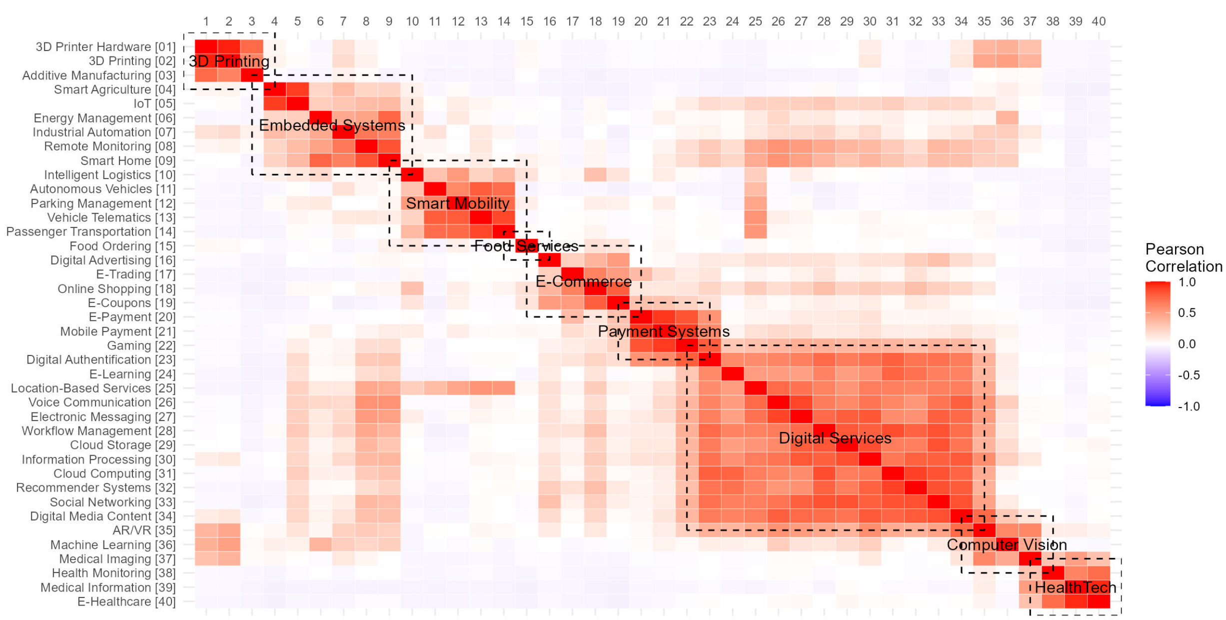
Notes: This figure presents the distribution of non-cited and undetermined-count patents across emerging digital technologies.

Figure A.5: 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.6: Semantic Co-Occurrence of Technologies in 3–digit ISCO-08 Occupations



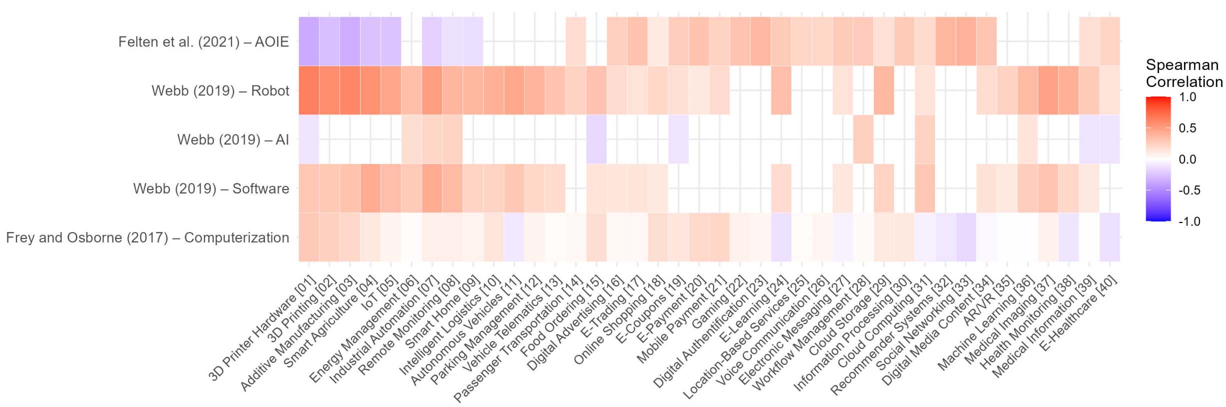
Notes: This figure shows all pairwise semantic-based technology co-occurrences as a correlation matrix, which is symmetric with diagonal values of 1. The matrix categorizes technologies into blocks, grouping them according to their semantic associations with occupations.

Table A.9: Comparison of the Exposure Studies

	Technology	Methodology	Exposure Scores
Webb (2019)	**Software** refers to computer programs that execute manually specified "if-then" rules. A program is considered software (as opposed to AI) if every action it performs is predetermined by a human. **Robots** are defined as "automatically controlled, reprogrammable, multipurpose manipulators with three or more programmable axes, which may be fixed or mobile for industrial automation". **AI** encompasses machine learning algorithms, specifically supervised and reinforcement learning algorithms.	**Data: Patents.** The author uses a set of keywords for each technology to retrieve a patent sample. He applies dependency parsing on patent titles and O*NET task descriptions to extract verb-noun pairs that represent actions performed by technologies and occupations, respectively. To improve matching, nouns are aggregated into broader categories using the WordNet lexical database. The matching is then established between the aggregated verb-noun pairs of patents (technologies) and O*NET tasks.	An occupation's exposure score for a technology reflects the intensity of patenting activity in this technology related to the tasks within that occupation. This measure is agnostic to substitution or augmentation effects.
Felten et al. (2021)	**AI** (10 applications as defined by EFF): abstract strategy games, real-time video games, image recognition, visual question answering, image generation, reading comprehension, language modeling, translation, speech recognition, and instrumental track recognition.	**Data: Survey.** To estimate exposure scores, the authors created a crowd-sourced dataset using survey responses from gig workers on Amazon's Mechanical Turk (mTurk). Participants evaluated the relevance of each AI application to specific workplace abilities as defined by the Occupational Information Network (O*NET). These responses generated a relatedness score between 0 and 1 for each AI application and occupational ability. The exposure score is then adjusted to account for the breadth of abilities required in an occupation, ensuring that occupations with a broader range of abilities are not over-weighted in exposure.	An occupation's exposure score indicates how closely AI applications relate to the abilities required for that occupation. This measure is agnostic to substitution or augmentation effects.
Frey and Osborne (2017)	**Computerization** is driven by **Machine Learning** (ML: including Data Mining, Machine Vision, Computational Statistics, and other subfields of AI) and Mobile Robotics (MR). These definitions are based on prior literature, discussions, and expert insights.	**Data: Expert Knowledge.** Susceptibility to computerization is manually assessed for 70 occupations based on expert knowledge, which serves as training data. For the remaining 632 occupations, the probability of computerization is estimated using a Gaussian Process Classifier, considering factors such as creativity, dexterity, and social intelligence from O*NET.	The probability of computerization indicates the likelihood that a given occupation could be automated based on current technological capabilities. This probability depends on the tasks involved and the presence of engineering bottlenecks in automating those tasks, such as those requiring creativity, social intelligence, or complex perception and manipulation.

Notes: This table provides summary of the three exposure studies with regard to technologies, methodology, and interpretation of the constructed exposure scores.

Figure A.7: 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

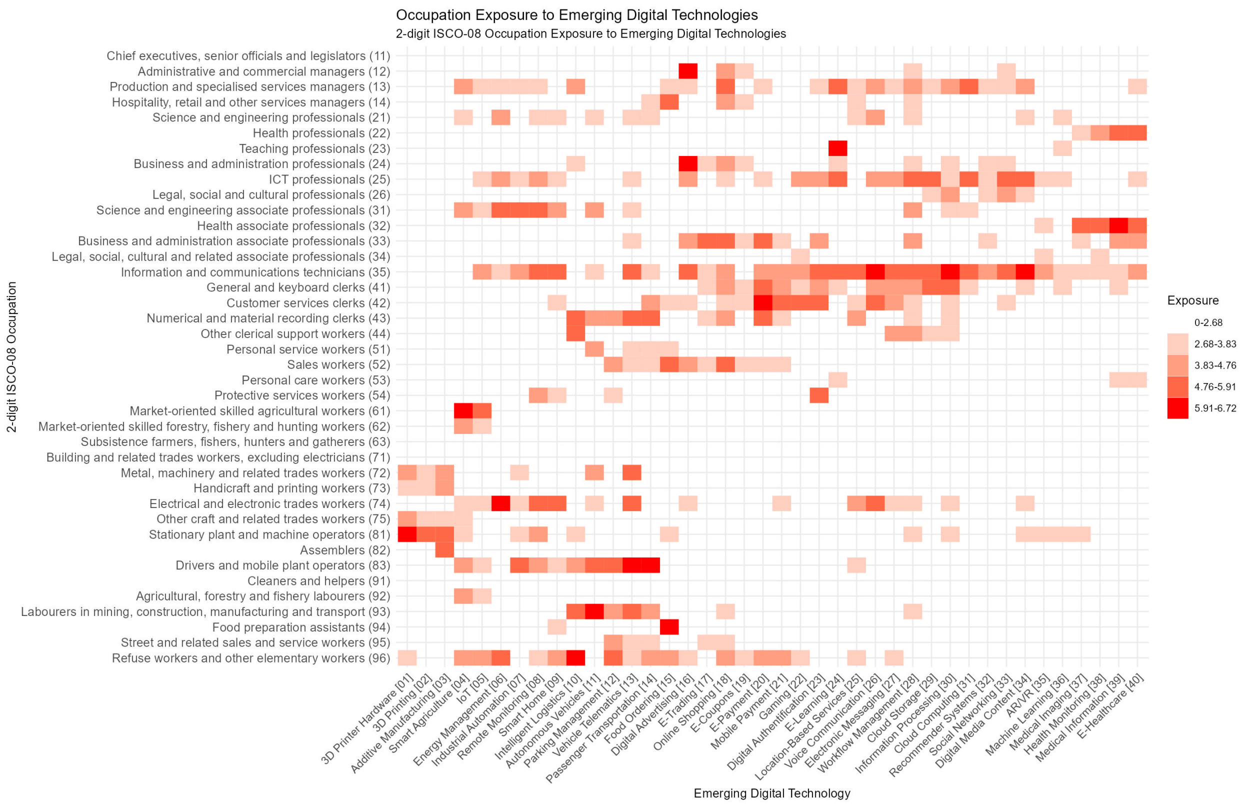
Notes: This table presents the top 30 4-digit ISCO-08 occupations ranked by exposure to all emerging digital technologies. Columns (from left to right) correspond to occupation code, occupation title, average exposure to emerging digital technologies, top-3 concentration ratio which represents the sum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

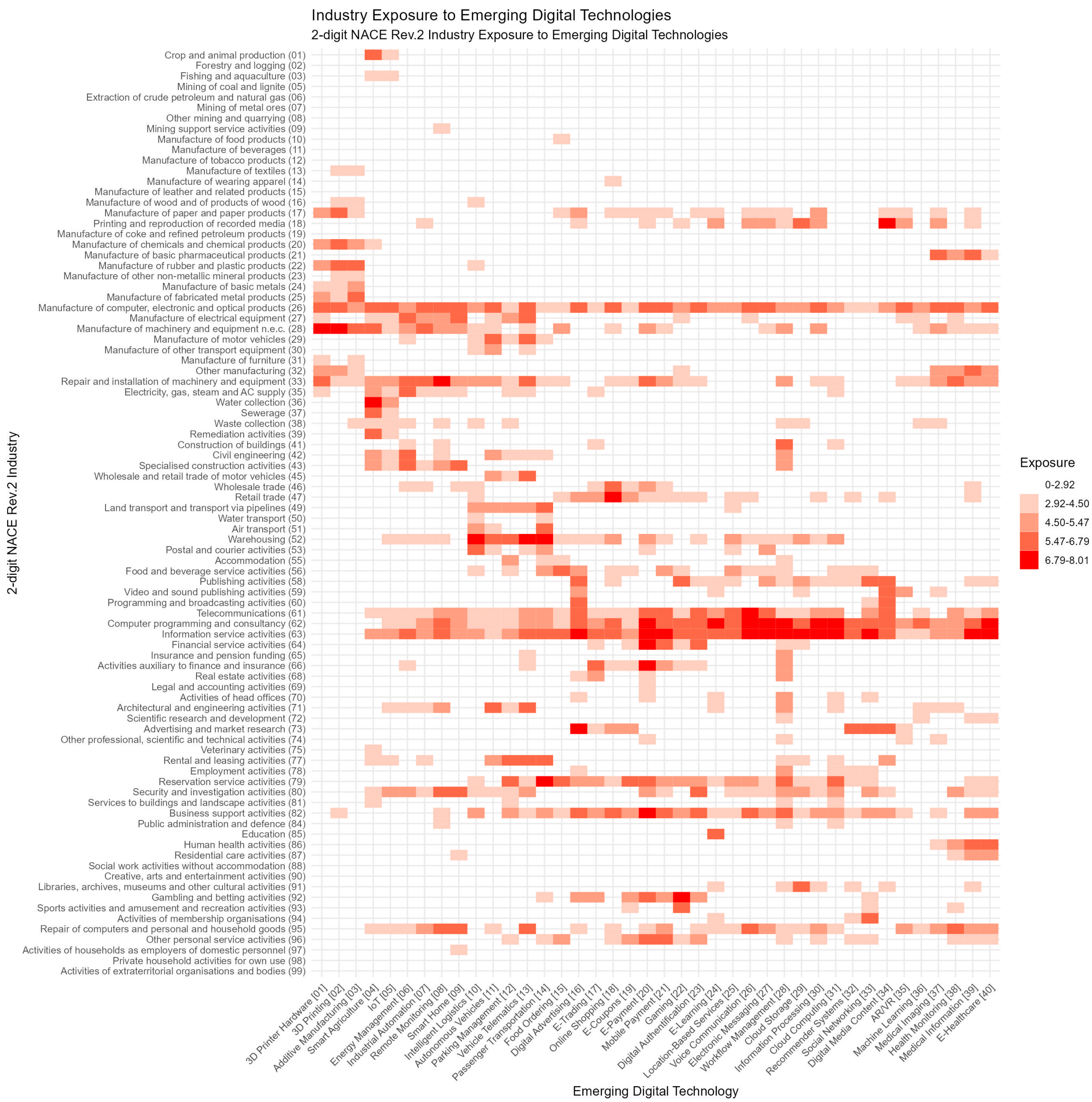
Notes: This table presents the top 30 3-digit NACE Rev.2 industries ranked by exposure to all emerging digital technologies. Columns (from left to right) correspond to industry code, industry title, average exposure to emerging digital technologies, top-3 concentration ratio which represents the sum 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 shows the employment shares of our 10 sectors across European regions in 2010. The three largest sectors are Market Services (G-I), with an average of 23.8% of employment; the Public Sector (O-Q), at 23.7%; and Industry (B-E), averaging 17.9%. Following these, Agriculture (A), Construction (F), and Professional, Scientific, Technical, Administration, and Support Services (M-N) each contribute between 6% and 9% to employment. The remaining four sectors collectively account for 11.4% of employment. Notably, the Information and Communication sector (J), crucial for digital technologies, represents only 2.6% of average regional employment—similar to the Financial and Insurance 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.

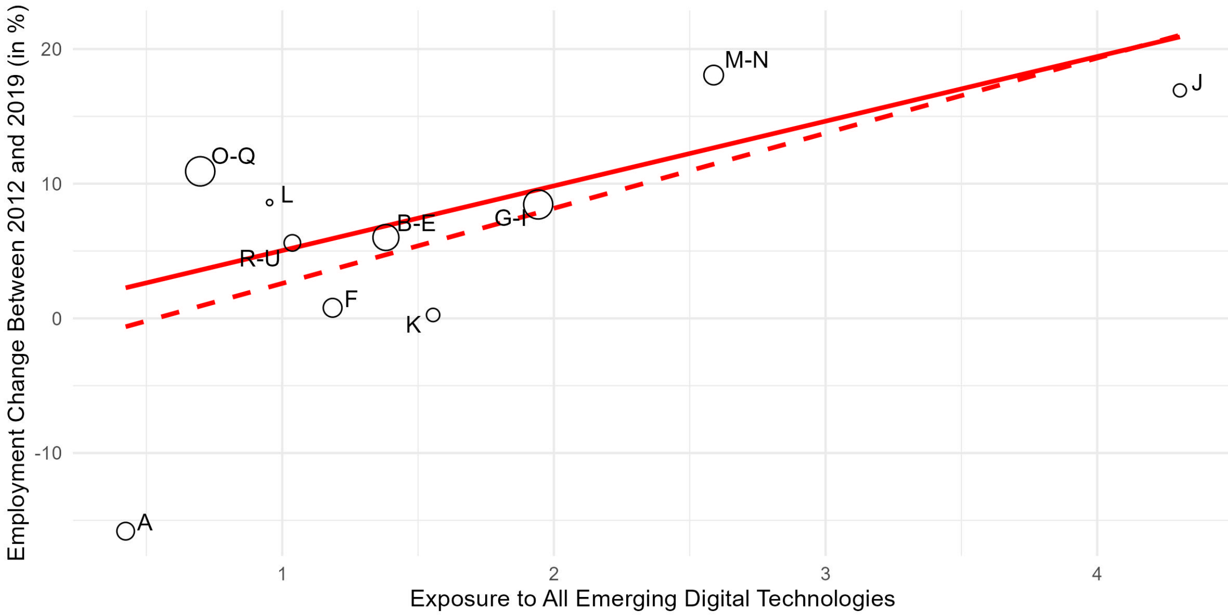
C.2 Employment Change and Exposure

Figure C.1 shows the positive correlation between the change in employment and the exposure to all emerging digital technology across 1-digit grouped industries in Europe.

C.3 Placebo Estimates

To further validate the shift-share approach, we conduct a placebo analysis by estimating the effect of regional exposure to emerging digital technologies on the change in the employment-to-population ratio during the pre-period (2002–2009). The results are presented in Table C.2.

Figure C.1: Employment Change and Exposure to All Emerging Digital Technologies in 1-digit NACE (Grouped) Industries in Europe between 2012 and 2019



Notes: This figure shows the positive relationship between employment change (2012–2019, in percentage) and exposure to emerging digital technologies across 1-digit (grouped) industries at the European level. The dashed line represents the unweighted regression, while the solid line shows the regression weighted by sector employment size in 2010.

We find no effect of regional exposure in the pre-period for any demographic group, except for high-skilled workers. This significant result is important, as high-skilled workers are those who develop and produce emerging digital technologies. The positive estimate suggests reverse causality, indicating that employment growth among high-skilled workers—who are pivotal to these technologies—is higher in regions with greater exposure.

The placebo analysis is based on the pre-period (2002–2009), but employment data for 2002 are unavailable in 62 regions, requiring us to restrict the sample to 258 regions. Table C.3 presents baseline specification estimates for this restricted sample, as in Table 4. The estimates are similar to those in Table 4, suggesting that the placebo analysis results are not influenced 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.4 Assessing the Impact of Emerging Digital Technology Families

We conduct the analysis at the technology family level, using a shift-share design to calculate 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 family: $X_j^K = \frac{1}{|K|} \sum_{k \in K} X_j^k$.

Figure C.2 presents 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.

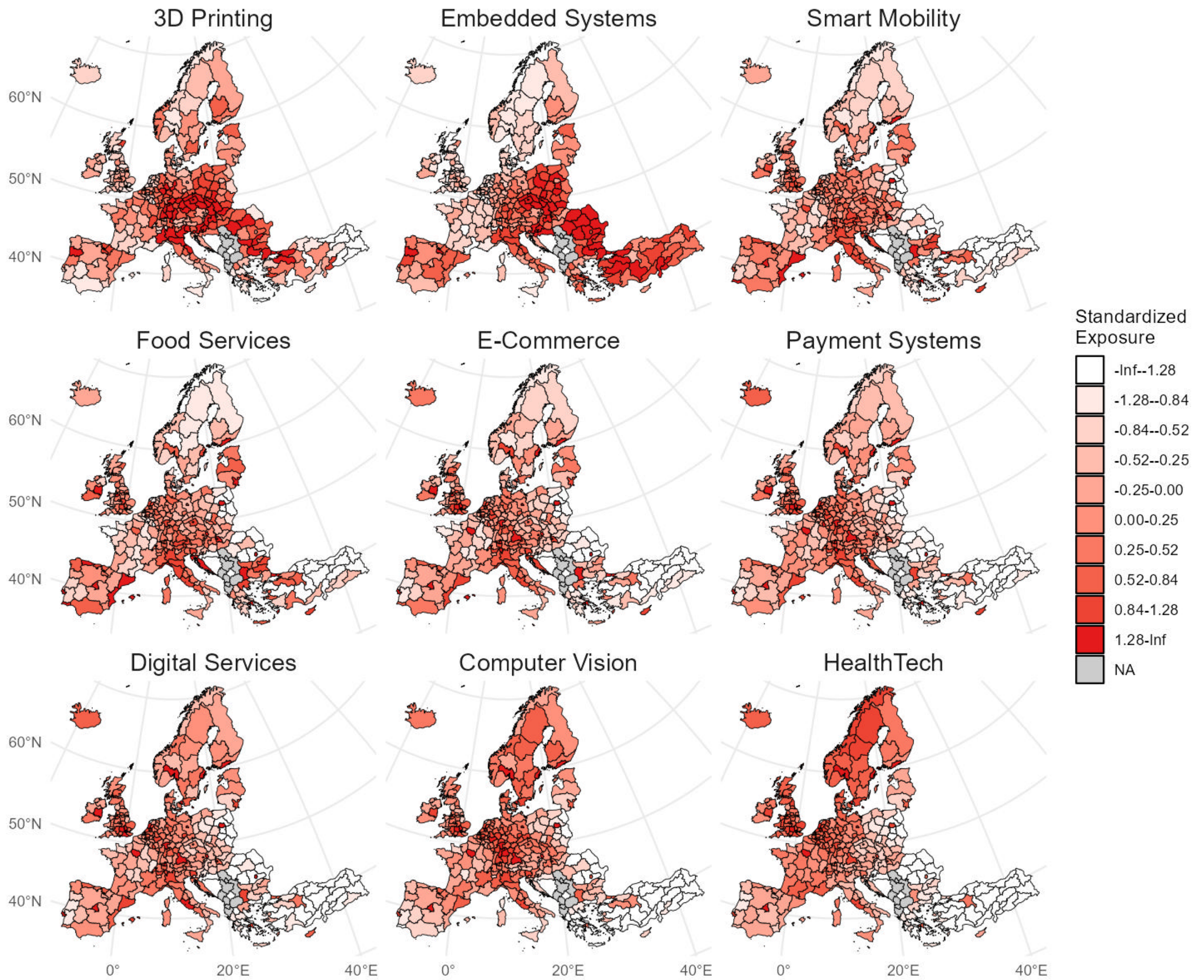
Furthermore, spatial differences in exposure are also evident within countries, characterized by disparities between rural and urban areas. Exposure to E-Commerce, Payment Systems, and Digital Services is predominantly concentrated in capital cities and financial hubs. In contrast, exposure to Smart 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.

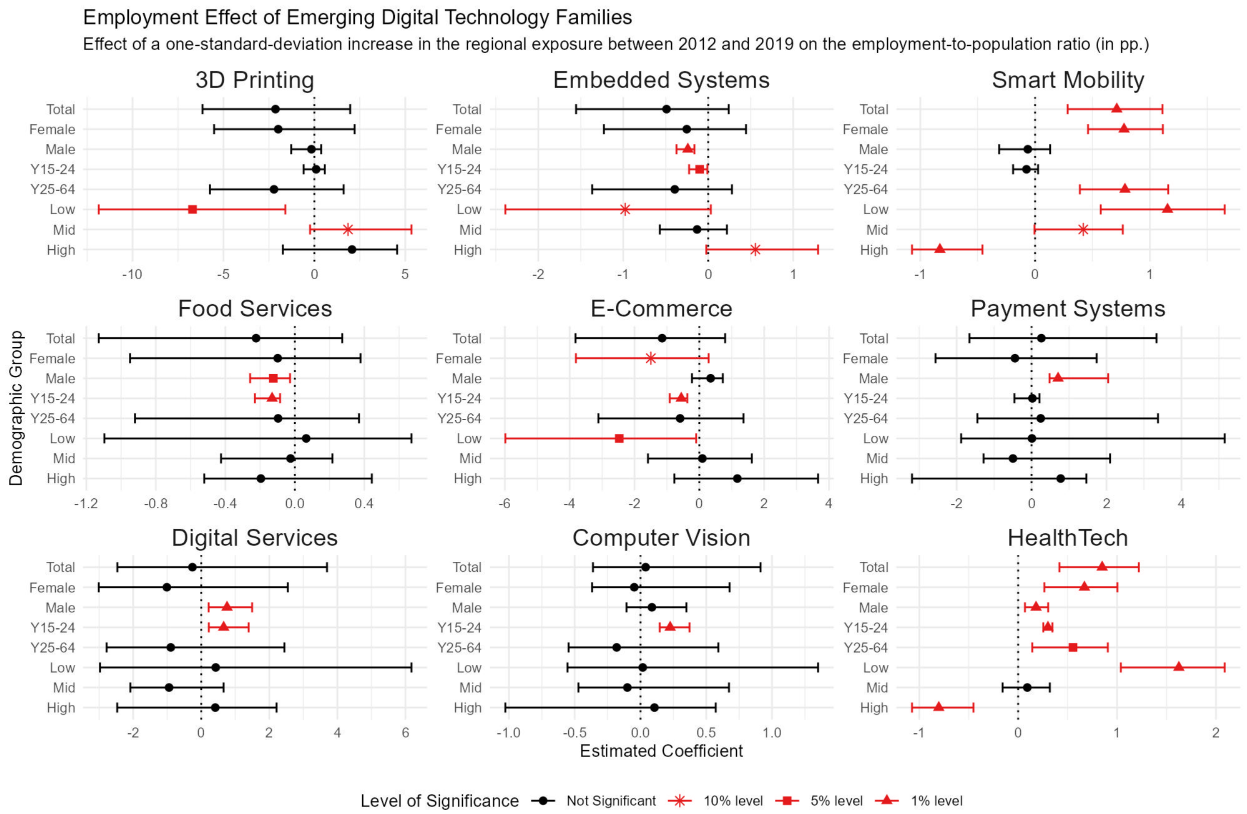
Figure C.2: Geographic Distribution of Regional Exposure to Families of Emerging Digital Technologies across Europe from 2012 to 2019



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* 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.3: 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.3 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.

While we do not find any effect of Embedded Systems on the overall employment-to-population ratio, we find that regional exposure to them reduces employment among young, male, and low-skilled workers. The opposite signs for low- and high-skilled workers (significant at the 10% level) suggest that Embedded Systems are skilled-biased. The 10% significance level reflects the heterogeneity in impacts across individual technologies.

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.

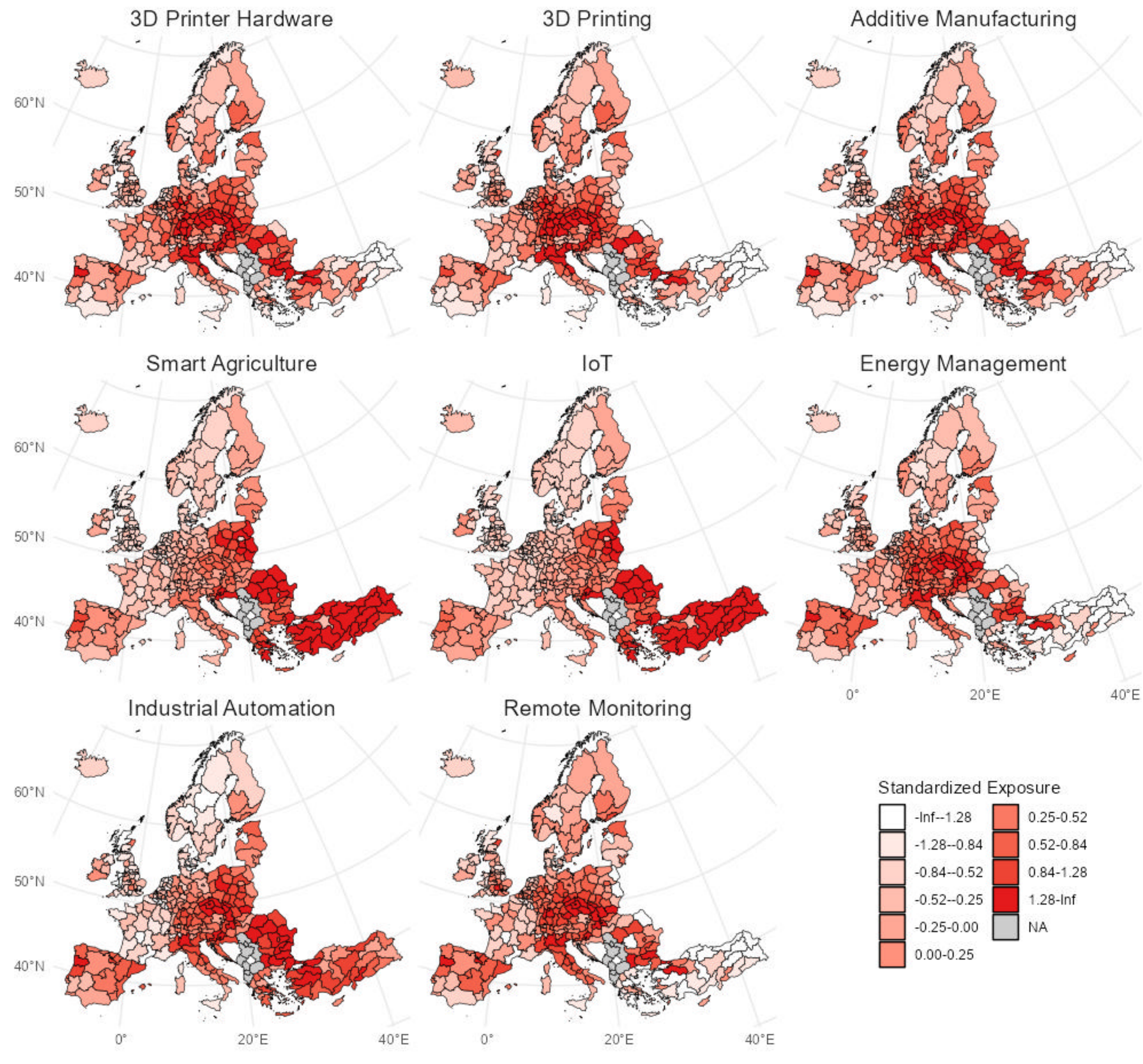
C.5 Geographic Distribution of Regional Exposure to Individual Emerging Digital Technologies

Figures C.4 to C.8 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.6 Individual Effects of Emerging Digital Technologies

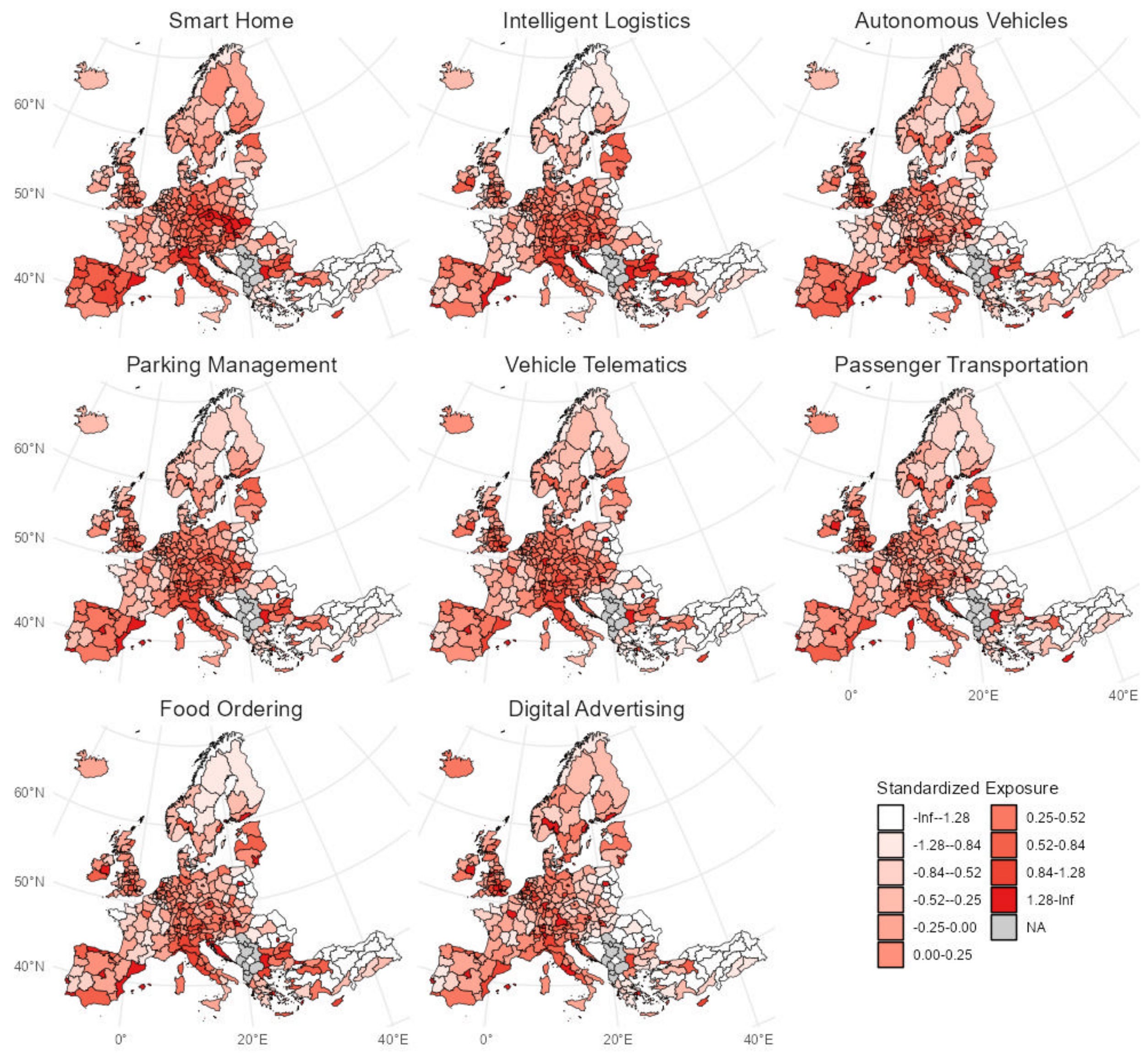
Figures C.9 to C.13 present the effect of individual emerging digital technologies.

Figure C.4: Geographic Distribution of Exposure to Emerging Digital Technologies (1/5)



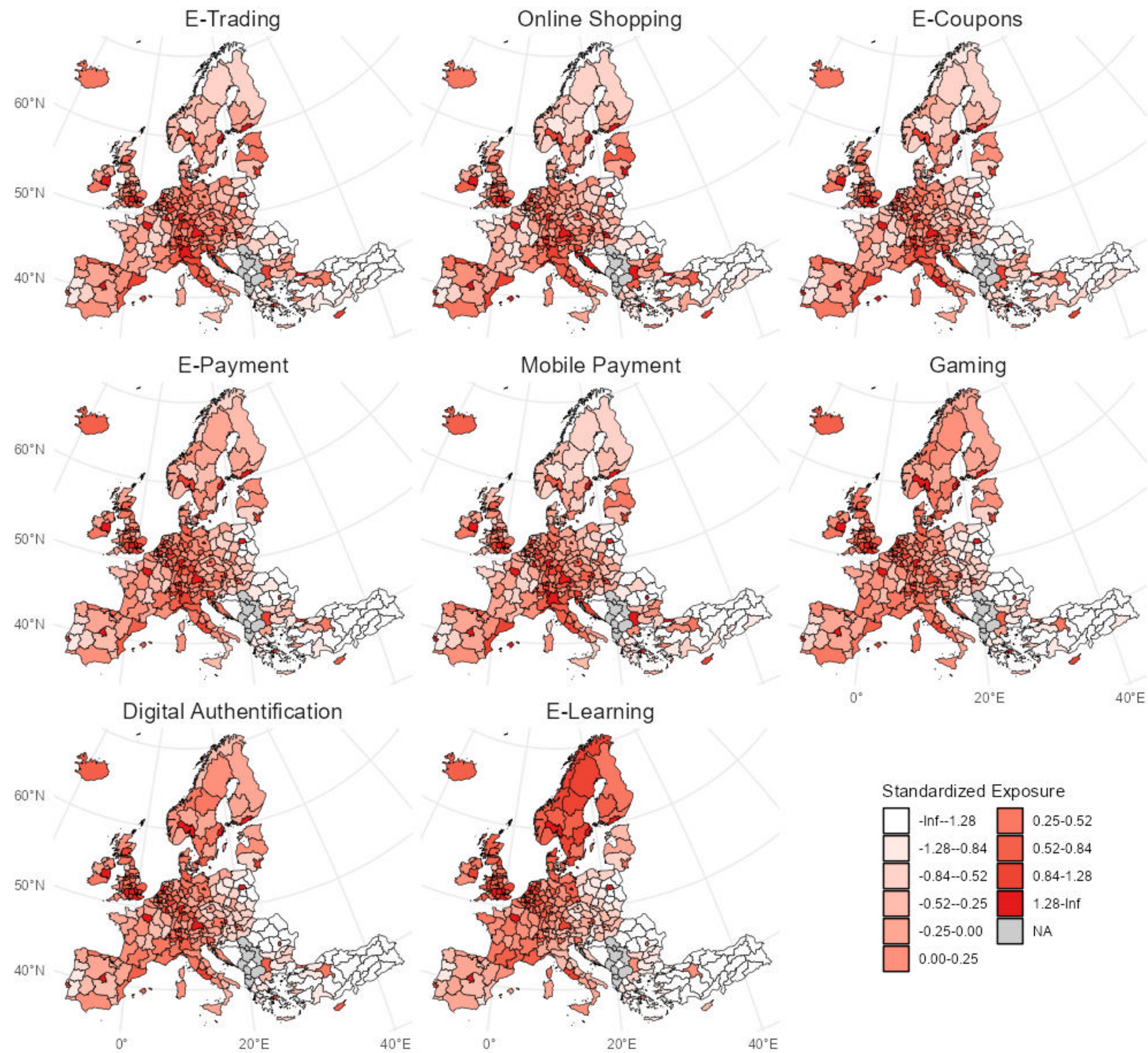
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 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.5: Geographic Distribution of Exposure to Emerging Digital Technologies (2/5)



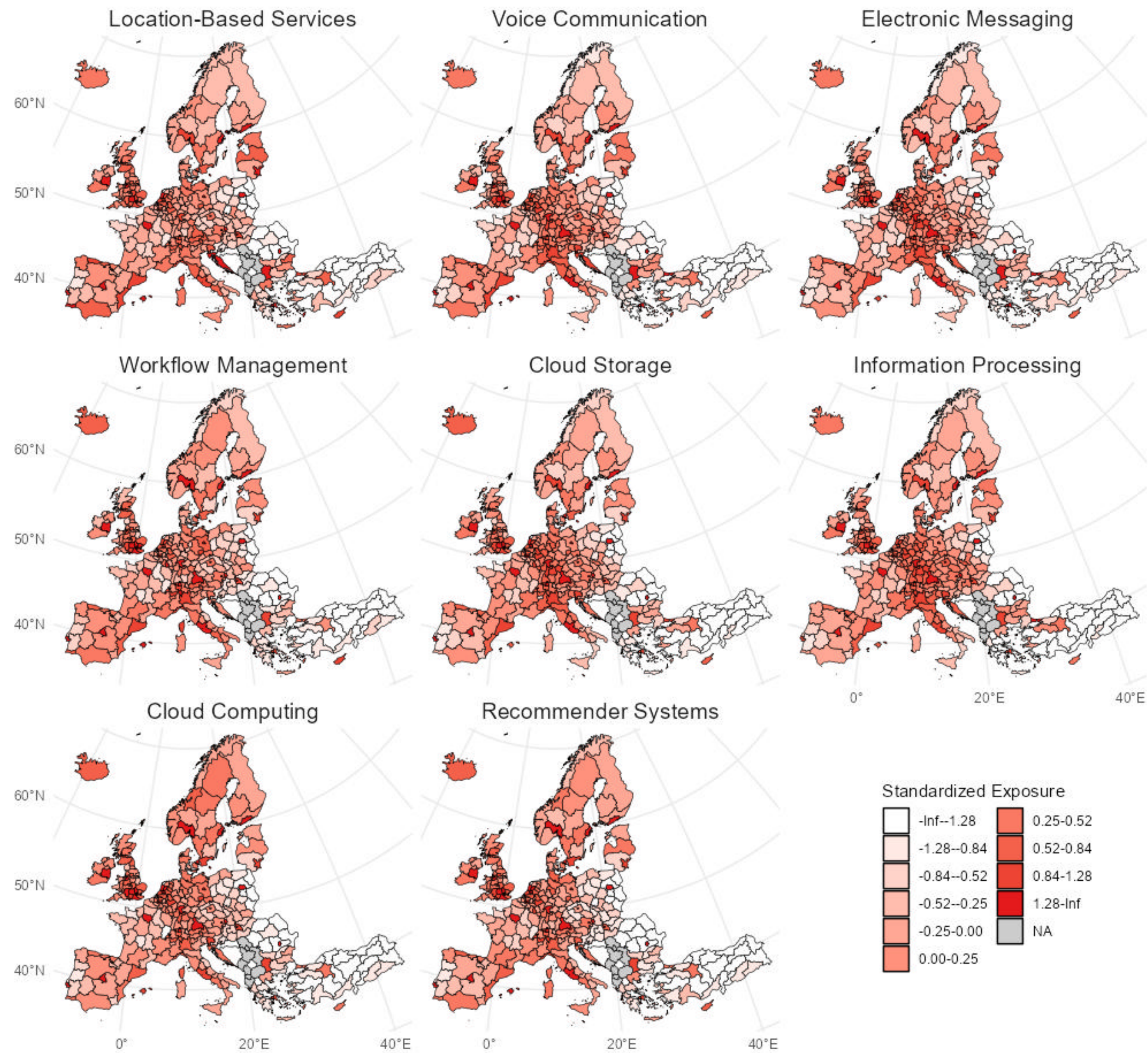
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 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.6: Geographic Distribution of Exposure to Emerging Digital Technologies (3/5)



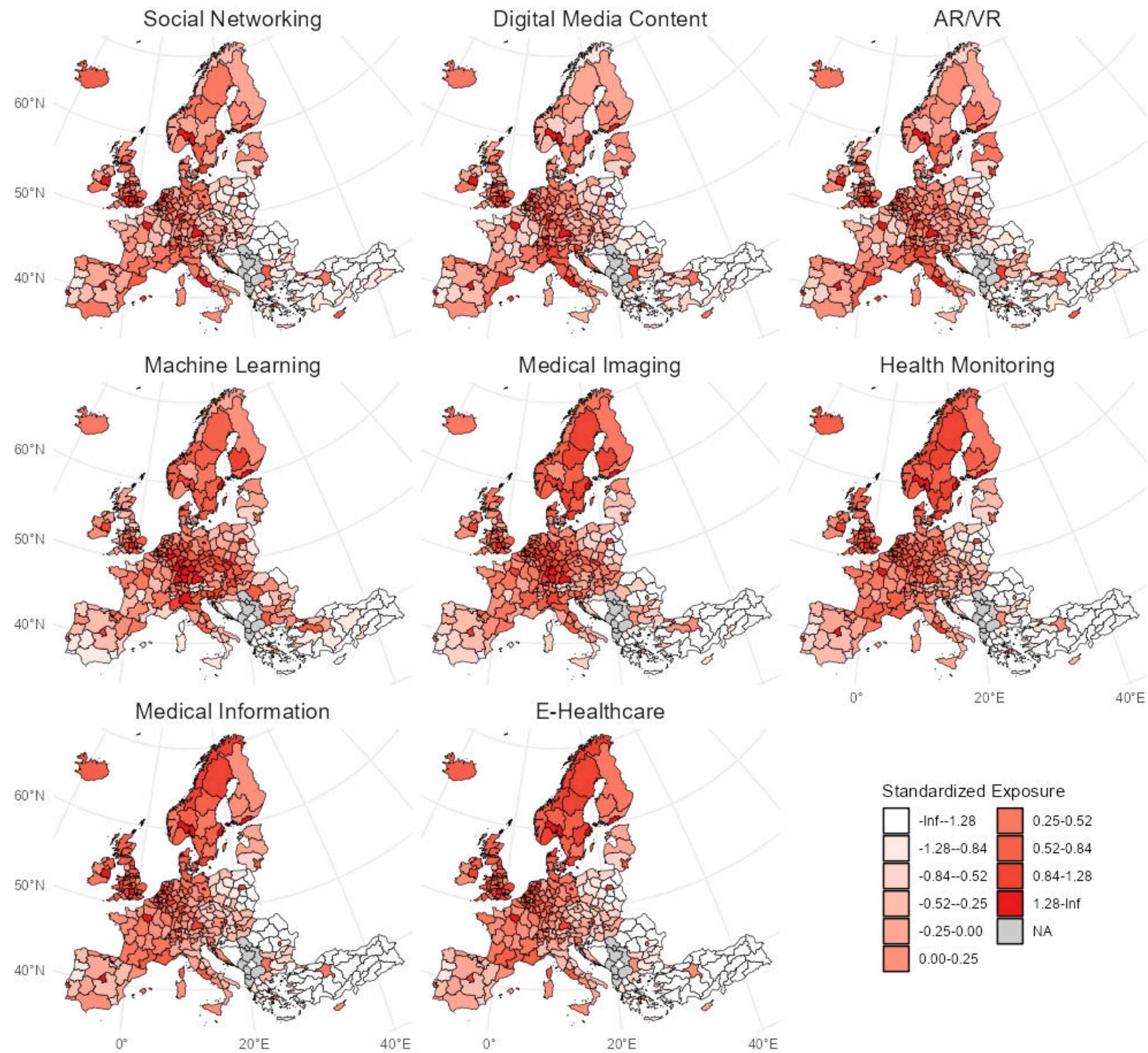
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 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.7: Geographic Distribution of Exposure to Emerging Digital Technologies (4/5)



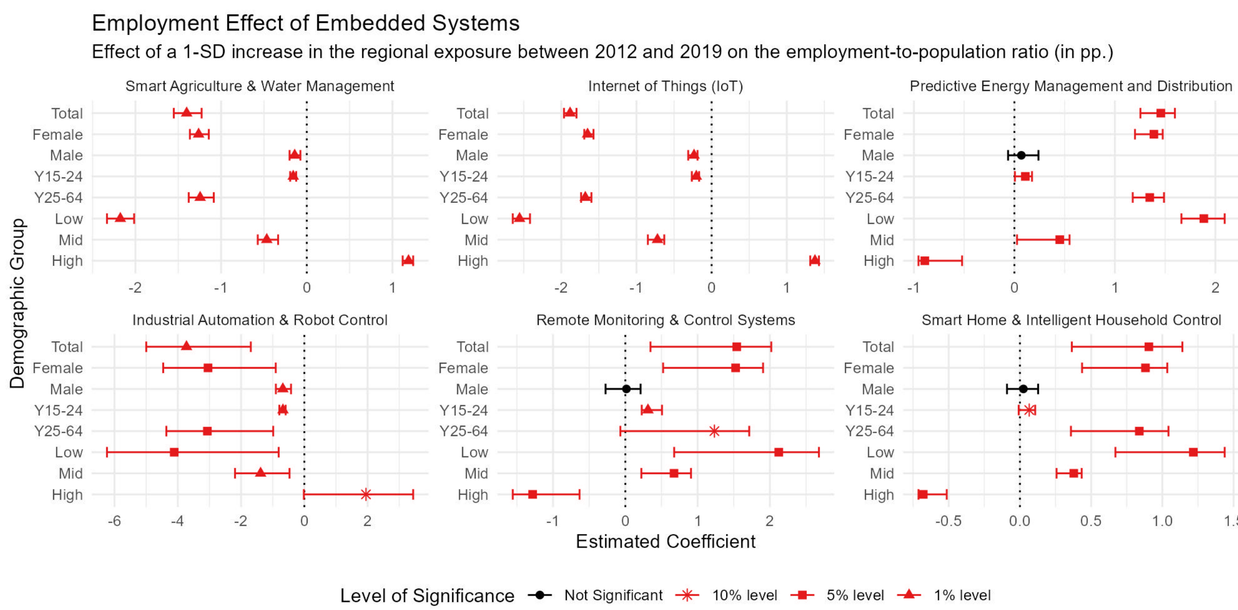
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 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.8: Geographic Distribution of Exposure to Emerging Digital Technologies (5/5)



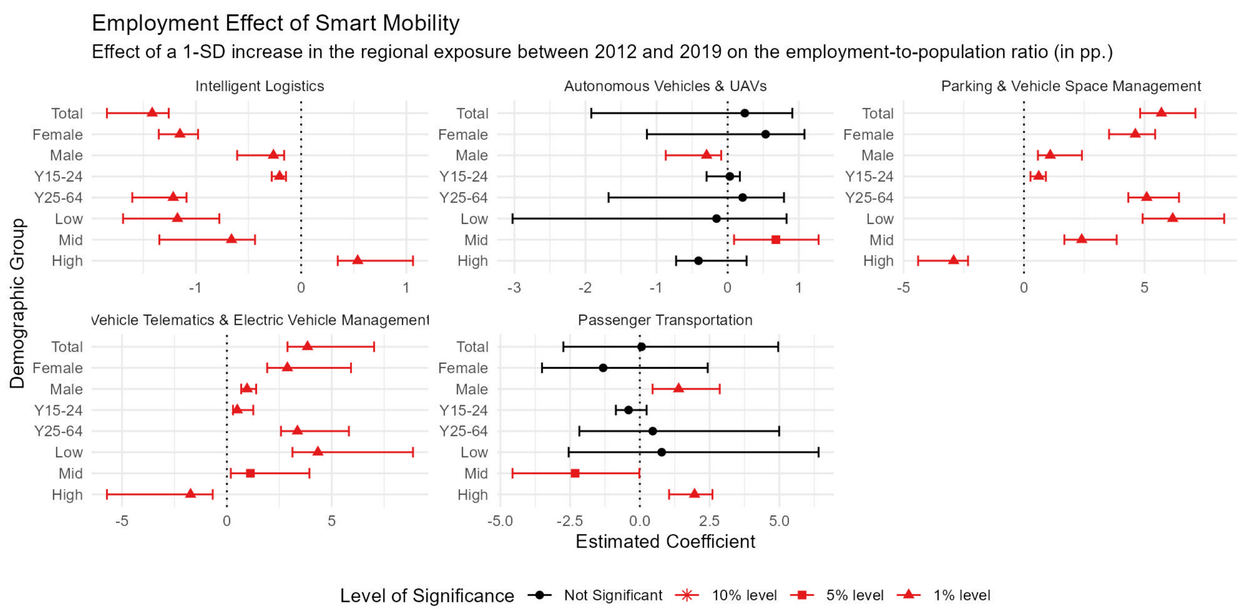
Notes: This figure illustrates the geographic distribution of exposure to emerging digital technologies for NUTS-2 regions. Regional exposure is constructed as a shift-share variable by interacting the sectoral employment shares in the baseline year 2010 and sectoral exposure to these technologies from 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.9: Employment Effect of Embedded Systems



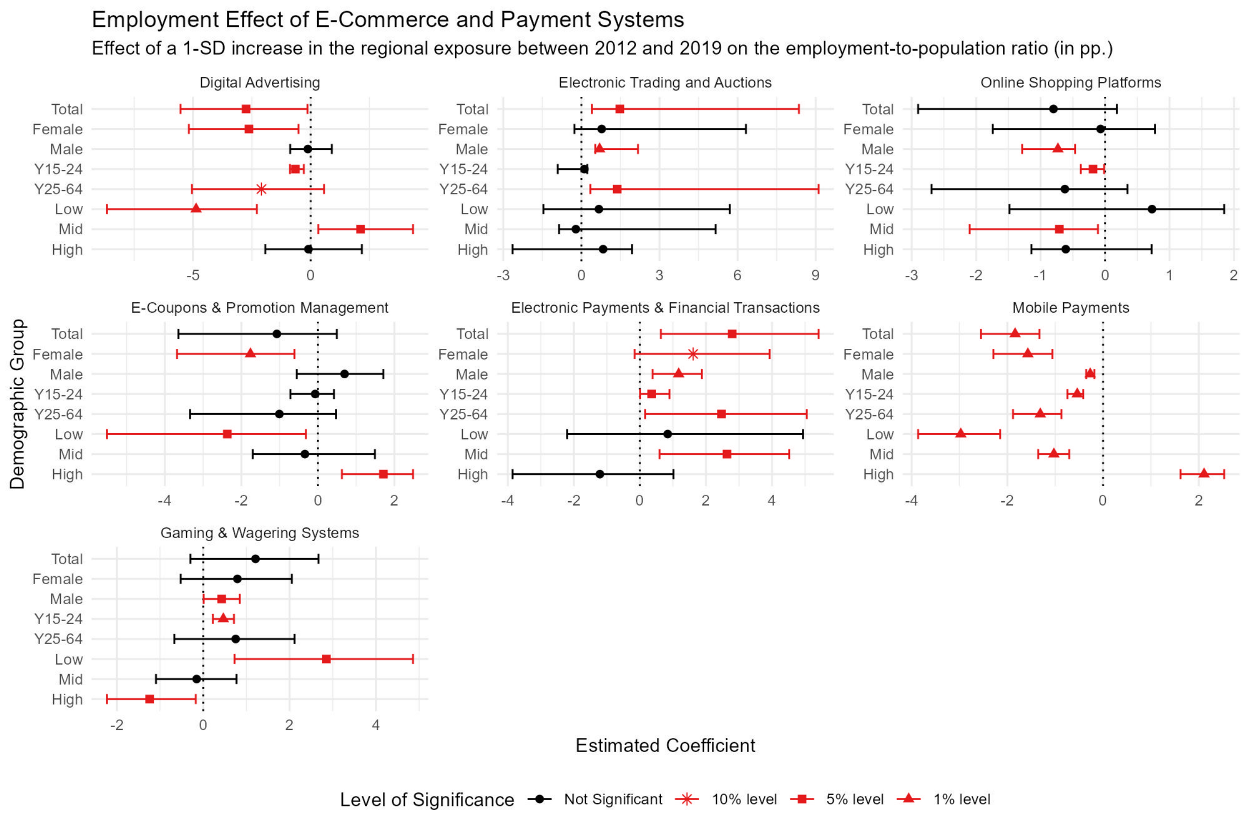
Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all 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.10: Employment Effect of Smart Mobility



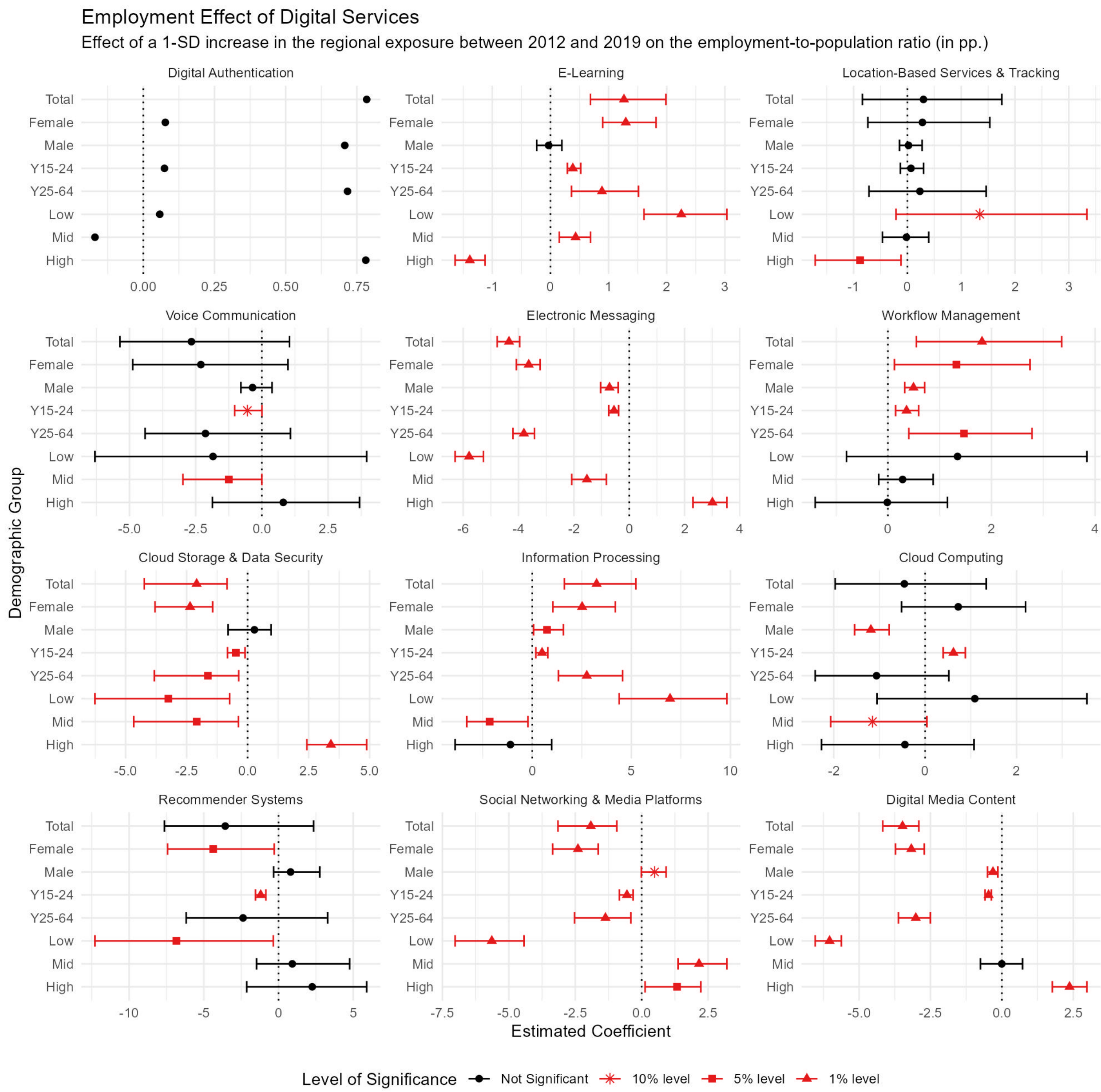
Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all 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.11: Employment Effect of E-Commerce and Payment Systems



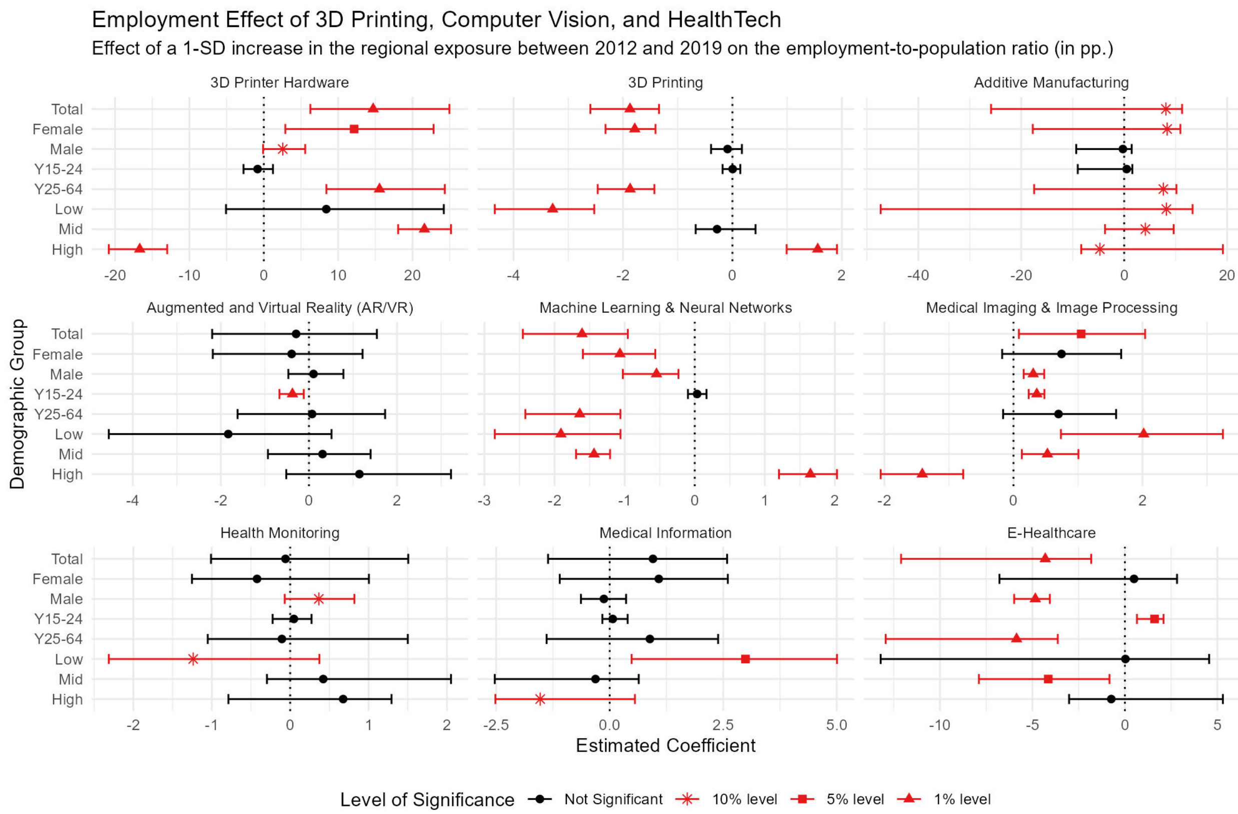
Notes: This figure presents the coefficients measuring the effect of regional exposure to emerging digital technology, constructed as shift-shares and standardized, on changes in the employment-to-population ratio, expressed in percentage points (pp.), between 2012 and 2019 in European regions, for all 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.12: 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 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. Confidence intervals for Digital Authentification are not displayed since the standard errors cannot be computed under the AKM0 inference procedure.

Figure C.13: 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.