UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Contributions towards Navigating the Fourth Industrial Revolution
Technology trends, research directions, workforce implications

Benjamin Meindl

Supervisor: Doctor Joana Serra da Luz Mendonça
Co-Supervisor: Doctor Daron Acemoglu

Thesis approved in public session to obtain the PhD Degree in Leaders for Technical Industries

Jury final classification: Pass with Distinction

2022

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Contributions towards Navigating the Fourth Industrial Revolution
Technology trends, research directions, workforce implications

Benjamin Meindl

Supervisor: Doctor Joana Serra da Luz Mendonça
Co-Supervisor: Doctor Daron Acemoglu

Thesis approved in public session to obtain the PhD Degree in Leaders for Technical Industries

Jury final classification Pass with Distinction

Jury

Chairperson

Doctor Hélder Carriço Rodrigues, Instituto Super Técnico, Universidade de Lisboa

Members of the Committee

Doctor Melanie Arntz, Alfred-Weber-Institute for Economics, University of Heidelberg, Germany

Doctor Joana Serra da Luz Mendonça, Instituto Superior Técnico, Universidade de Lisboa

Doctor Ana Isabel Cerqueira de Sousa Gouveia Carvalho, Instituto Superior Técnico, Universidade de Lisboa

Doctor Hugo Miguel Fragoso de Castro Silva, Instituto Superior Técnico, Universidade de Lisboa

Doctor Alejandro Germán Frank, Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Brasil

Funding Institutions
Studienstiftung des Deutschen Volkes

2022

Acknowledgments

I could not have conducted this research without my supervisor, Joana Mendonça. She helped me find a research topic that, four years later, still excites me almost every day. I would like to thank her for her guidance, flexibility during the past years, and caring supervision.

I would also like to thank my co-supervisor, Daron Acemoglu, who particularly inspired my chapter exploring the impact of technologies on jobs and challenged the approaches I tried. I am also grateful to him for facilitating my unique experience staying at MIT in fall 2018.

Many colleagues supported me throughout this dissertation—in particular:

- Alejandro Frank, who has been a warm-hearted collaborator. I value his expertise and focus.
- Morgan Frank, whose creative and pragmatic attitude inspired me.
- Ulrich Zierahn and Ingrid Ott, who inspired me through various discussions regarding how patents can serve as indicators of technological change.
- Elsa Henriques, who has supported me since our first call during the application phase of the MIT Portugal Program in 2016.
- My CAT committee members—Ana Carvalho, Melanie Arntz, and Michael Breunig—for their comments and feedback.

I would like to thank the colleagues, staff, and friends I met at IN+ and in the MIT Portugal Program—in particular, my close friend Antonio, who introduced me to Lisbon. Together with Carlos, he and I had many wonderful discussions during our lunch breaks at Técnico.

I am also grateful to McKinsey & Company, who gave me the opportunity to conduct my PhD as part of the Fellow program, and the Studienstiftung des Deutschen Volkes, who supported me with a scholarship. I enjoyed interacting with the friendly staff at Studienstiftung and meeting fellow scholars during seminars.

I am grateful for my parents and family, who have always been there for me and always welcomed me for short “holidays” at home.

Finally, I would like to thank Alena, who brought a great deal of fun to my life and has been a wonderful partner, particularly during the lockdown.

Abstract

Technologies of the Fourth Industrial Revolution (4IR) offer many opportunities, such as reducing production costs and addressing skill shortages. However, many companies have faced difficulties implementing these technologies and risk falling behind. One challenge in this area is understanding 4IR trends, concepts, and research opportunities. Such an understanding would help researchers to direct their attention, practitioners to focus their implementation efforts, and policymakers to create appropriate measures. With the implementation of more 4IR technologies, it is increasingly important to understand the impact of technological change on the workforce in order to prepare for a successful and sustainable transformation. This will enable workers to plan their careers and will support policymakers and companies in navigating a successful and sustainable labor market transition. This dissertation addresses both challenges by exploring the evolution of technological trends and concepts, identifying future research directions, and analyzing the exposure of various occupations to 4IR technologies.

To address the first challenge, I researched two key topics: technology trends and research directions. For both, I reviewed scientific articles in the field of Industry 4.0 using natural language processing (NLP), network analysis, and machine learning. First, I analyzed 4IR technology trends and found that, based on the initial definitions, the industrial Internet of Things is the current core of the 4IR technology landscape, but strong growth in the field of artificial intelligence (AI) suggests that AI may evolve as the new core technology. Second, I examined research directions using the "four smarts" framework. Smart Working is the least explored dimension, with many opportunities for future research. Further opportunities lie in the intersections of the smart dimensions, particularly between Smart Working and Smart Supply Chain. To address the second challenge, I analyzed the third key topic: workforce implications. I used NLP to mapping of patents and occupations in order to calculate scores of occupations' exposures to 4IR technologies. My analysis shows that exposure to 4IR technologies differs from traditional technology exposure. Occupations that involve many manual tasks, such as construction and production, have been exposed mainly to traditional (i.e., non-4IR) technologies and have low exposure to 4IR technologies.

Overall, the dissertation depicts the evolution of 4IR technologies and concepts; identifies research directions to assure an integrated, holistic evolution of the 4IR; and analyzes the exposure of occupations to 4IR technologies. Thus, the dissertation supports workers, practitioners, and policymakers in navigating the 4IR.

Keywords: fourth industrial revolution; patent occupation mapping; technology exposure; technology trends; Industry 4.0 research directions

Resumo

As tecnologias da quarta revolução industrial (4IR) trazem muitas oportunidades, tais como a redução dos custos de produção e o combate à escassez de competências. No entanto, muitas empresas enfrentam dificuldades na implementação destas tecnologias e correm o risco de ficar para trás. Um dos desafios é compreender as tendências, conceitos, e oportunidades de investigação da 4IR. Esta compreensão ajuda os investigadores a dirigir a sua atenção, os profissionais a concentrar os esforços de implementação, e os decisores políticos a criar medidas em conformidade. Com a implementação de mais tecnologias da 4IR, torna-se cada vez mais importante compreender o impacto da mudança tecnológica para a preparação da força de trabalho para uma transformação bem sucedida e sustentável. Este entendimento permitirá aos trabalhadores planear carreiras e apoiar os decisores políticos e as empresas a navegar numa transição bem sucedida e sustentável do mercado de trabalho. A dissertação aborda ambos os desafios explorando a evolução das tendências e conceitos tecnológicos, identificando direcções de investigação, e analisando a exposição das profissões às tecnologias da 4IR.

Para enfrentar o primeiro desafio, pesquiso dois tópicos-chave, tendências tecnológicas e orientações de investigação. Para ambos revejo artigos científicos no campo da Indústria 4.0 utilizando processamento de linguagem natural (NLP), análise de rede, e aprendizagem de máquina. Primeiro, analiso as tendências das tecnologias da 4IR e descubro que, seguindo as definições iniciais, a internet industrial das coisas é actualmente o núcleo do panorama das tecnologias da 4IR. O forte crescimento no campo da inteligência artificial (AI) sugere que a AI pode evoluir como o novo núcleo da tecnologia. Em segundo lugar, sugiro orientações de investigação utilizando o panorama dos quatro "smarts". O trabalho inteligente é a dimensão menos explorada, com muitas oportunidades para investigação futura. Outras oportunidades existem nas interfaces entre as diferentes dimensões inteligentes, particularmente entre o Trabalho Inteligente e a Cadeia de Fornecimento Inteligente. Para enfrentar o segundo desafio, analiso o terceiro tópico chave, implicações de mão-de-obra. Utilizo o NLP para criar um mapeamento de patentes e profissões. Este mapeamento serve como base para calcular as pontuações da exposição das profissões às tecnologias da 4IR. A minha análise mostra que a exposição às tecnologias da 4IR difere da exposição às tecnologias tradicionais. As profissões com muitas tarefas manuais, tais como construção e produção, são expostas principalmente a tecnologias tradicionais (não 4IR) e têm uma baixa exposição a tecnologias da 4IR.

Globalmente, a dissertação descreve a evolução das tecnologias e conceitos relacionados à 4IR; identifica direções de investigação para assegurar uma evolução integrada e holística da 4IR; e analisa a exposição das profissões às tecnologias da 4IR, ajudando trabalhadores, profissionais, e decisores políticos a navegar com sucesso pela 4IR.

Keywords: quarta revolução industrial; mapeamento de patentes para ocupações; exposição tecnológica; tendências tecnológicas; direcções de investigação da indústria 4.0

Contents

List of Tables

List of Figures

Acronyms

4IR Fourth Industrial Revolution. ii, 1–7, 9, 61, 62, 64, 68, 69, 71–91, 98–101, 129, 135, 136, 138

ACATECH German Academy of Science and Engineering. 38, 56

AGV automated guided vehicle. 25, 31, 42

AI artificial intelligence. ii, 1, 2, 5, 9, 12, 13, 16, 25, 27, 32–36, 39–42, 54, 57, 58, 61, 73–75, 78, 79, 82, 84–87, 89, 90, 98–100, 135

AIOE AI occupational exposure. 78, 79

amod adjectival modifiers. 20

API application programming interface. 6, 17, 43, 47, 96

BernoulliNB Bernoulli naive Bayes. 95–97

BERT Bidirectional Encoder Representations from Transformers. 19, 95

BLS U.S. Bureau of Labor Statistics. 87, 88

CAD computer-aided design. 25, 30, 33, 34, 85, 91

CAM Computer-aided manufacturing. 22, 25–28, 30, 33, 34

CMS condition monitoring systems. 25, 27, 30, 31

CNC computerized numerical control. 25, 30

CNN convolutional neural network. 9, 10, 47, 95–97

CP computerization probability. 76, 81–84, 137

CPC Cooperative Patent Classification. 94

CPS cyber-physical systems. 1, 4, 11, 13, 25, 30, 31, 35–37, 53

DWA detailed work activity. 66, 67, 69–72, 84

EFF Electronic Frontier Foundation. 78

EPO European Patent Office. 64, 68, 126

ERP enterprise resource planning. 14, 30, 31

EV eigenvector. 21, 22

IC2S2 5th International Conference on Computational Social Science. 8, 9

ICT information and communication technologies. 5

IIoT Industrial Internet of Things. 22, 25–35, 98, 100

IoT Internet of Things. 5, 7, 11, 14, 16, 21, 25, 30, 31, 33–37, 39, 40, 42, 43, 53, 68

IR industrial revolution. 1

IT information technology. 1, 15, 19, 22, 34, 129, 130

k-NN k-nearest neighbor. 9, 95–97

LMAO Language Modelling with Approximate Outputs. 95–97

MES manufacturing execution system. 25, 30, 31

MGI McKinsey Global Institute. 62, 84, 88, 138

MNB multinominal naive Bayes. 9, 95

NER named entity recognition. 6, 19, 20

NIST National Institute of Standards and Technology. 13

NLP natural language processing. ii, 5–7, 9–13, 15, 19, 34, 35, 47, 60, 64, 97, 99, 100

nmod nominal modifiers. 20

npadvmod noun phrases as adverbial modifiers. 20

OPC UA Open Platform Communications Unified Architecture. 25, 27, 31

PLM product lifecycle management. 27, 30

rca relative comparative advantage. 26, 30, 49, 50, 55

RQ research question. 3, 4, 6–8, 98, 99

RRTC routine replacing technological change. 86, 87

SBTC skill-biased technological change. 86, 87

SDN software-defined network. 27

SLR systematic literature review. 13, 37, 38, 40, 43, 100

SML suitability for machine learning. 69, 76, 79, 81, 82, 84, 138

SVM support vector machines. 9, 95–97

TF-IDF term frequency–inverse document frequency. 19, 67, 95, 96

TMP Technology, Management, and Policy Consortium. 9

ULMFiT Universal Language Model FIne-Tuning. 95

USPC United States Patent Classification. 94

USPTO United States Patent and Trademark Office. 64, 126

WIPO World Intellectual Property Organization. 64, 126

Chapter 1

Introduction

1.1 Context

We are currently seeing the implementation of a wave of digital technologies that are becoming part of our lives and economies. Just as artificial intelligence (AI) selects what news we see on social media, it can also be used to make managerial decisions and optimize manufacturing processes. Entire sectors might change across the globe through these new technologies—for example, when companies backshore manufacturing to developed economies driven by new automation technologies (Mendonça & Heitor, 2016). Recently developed technologies have also been of use during the COVID-19 pandemic, where they have allowed many people to work remotely (Narayanamurthy & Tortorella, 2021) and have even enabled contactless surgeries using surgery robots (Wang & Wang, 2021). This wave of technological change is often referred to as the Fourth Industrial Revolution (4IR). Like the previous industrial revolutions (IRs), the 4IR is expected to significantly impact daily life and the economy. The first IR followed the introduction of water- and steam-powered mechanical manufacturing facilities and enabled new opportunities through the use of non-human energy in the production process. The second IR was defined by the introduction of electric-powered mass production and the division of labor (i.e., through assembly lines). The third IR achieved further automation of manufacturing through electronics and information technology (IT). The fourth IR is commonly defined as following the introduction of cyber-physical systems (CPS) (Kagermann et al., 2013). This dissertation uses primarily the term *4IR* as a general framing of the field. Similarly, I treat *Industry 4.0* a general conceptual term, albeit more narrowly focused on the field of operations and manufacturing research. For Chapters 2 and 3 (and other associated paragraphs, such as those in the introduction), I mostly use the term "Industry 4.0," as this phrase is more widely used among the intended audiences of these chapters.

Some key Industry 4.0 technologies include cloud computing, cybersecurity solutions, AI, 3D printing, and advanced robotics (Culot et al., 2020; Xu et al., 2018; Zheng et al., 2020). These technologies are not all new; rather, they have been in development for years. The 4IR is thus not only characterized by these technologies but also by their combination and diffusion into industry (Alcácer & Cruz-Machado, 2019; Bai et al., 2020; Culot et al., 2020). For many companies, implementing these technologies is a top priority (Behrendt et al., 2017), and technological innovation is an important driver of firm success. Companies who are not able to implement 4IR technologies risk falling behind (Brynjolfsson et al., 2019), while investment in new technologies and the development of a strong digital strategy helps maintain competitiveness over the long term (Llopis-Albert et al., 2021). There will be winners and losers in this regard (Brynjolfsson et al., 2019). In particular, early adopters who have already built up some capabilities have an excellent opportunity to benefit most from the diffusion of AI technologies (Bughin et al., 2017, 2019). However, many companies face challenges in implementing 4IR technologies (Behrendt et al., 2018). Policymakers thus need to create measures to support innovation and diffusion in order for their economies to flourish. One challenge for companies and policymakers is therefore to understand current trends and gaps as well as future potentials so that they can adjust innovation and investment policies accordingly. Technological considerations inherently involve economic implications. Understanding the trends in this area should guide policies for investment, production, and hiring, among other considerations.

The recent technological progress has raised many questions about the impact of these new technologies. On the one hand, 4IR technologies can help solve current challenges. Intelligent systems can, for example, address climate issues and increase energy efficiency (Plitsos et al., 2017). In addition, sectors with a lack of labor supply might benefit from new technologies. For example, AI can be used in programming contexts to allow developers to focus on highly creative and innovative ideas, rather than spending most of their time on testing and bug fixes (Frey & Osborne, 2017). On the other hand, 4IR technologies might automate some activities previously performed by humans and make some jobs obsolete. These technologies may impact not only repetitive office activities or blue-collar jobs but also more complex tasks that require a skilled workforce. McAfee et al. (2012) note that intelligent systems will take over highly complex tasks. Moreover, occupations in the service industry, which have recently been prone to automation, might soon be further disrupted; for example, chat-bots might substitute for call center agents, and robots that can directly interact with patients might replace healthcare personnel (Hengstler et al., 2016). Even some managerial activities may come to be performed by machines (Leyer & Schneider, 2021). Therefore, current technological trends have caused fears related to their broad potential impact on the workforce and the rapidity of technology adoption. *The Guardian* (Seager, 2016), for example, asked, "Will Jobs Exist in 2050?", and Frey & Osborne (2017) found that 47% of U.S. jobs are at high risk of automation. These fears, particularly the fear of mass unemployment, have been present for over a century. Already in 1900, people feared mass unemployment due to automated weaving in factories, yet employment in the textile industry continued to grow for decades (Bessen, 2019). Keynes (1930) expressed the fear that improvements in technical efficiency were "too quick," stating that the changes are "faster than we can deal with the problem of labor absorption."

Researchers have explored whether the fears of mass unemployment are valid in this scenario (Mokyr et al., 2015) and, if so, which occupations will be most impacted by new technologies. There are various labor market mechanisms, and 4IR technologies are expected to automate some occupations and create demand for new labor in other sectors (Acemoglu & Restrepo, 2019a; Acemoglu et al., 2020). The size of these effects will determine whether there is an overall increase or decline in the demand for labor. In any case, these changes are expected to have an enormous impact on the workforce and may force many workers to reskill or transition to other sectors (Brynjolfsson & Mitchell, 2017; Bessen, 2019). However, there is currently no clear understanding of their impact, and "policymakers are flying blind into [...] the fourth industrial revolution" (Mitchell & Brynjolfsson, 2017). More research on the 4IR's impact on the labor market is crucial in navigating this transition. Therefore, another challenge is to identify which occupations are most impacted by technological change. This can help policymakers adjust the education system and shape policies to prepare for transitions through the 4IR. These insights can help companies build up their capacities accordingly—for example, through training or hiring skilled employees, which is a key success factor in implementing these technologies and possibly creating a competitive advantage (Bughin et al., 2019; Teece, 1998). In addition, workers can benefit from having a better understanding of changing skill requirements so that they can retrain accordingly and avoid fears of an uncertain future. This area is particularly relevant in the context of the insights in Chapters 2 and 3, which suggest a growing importance of human-robot systems and highlight the high research potential in the area of Smart Working.

Overall, there is a twofold challenge.

- First, researchers and practitioners need to understand 4IR trends and research opportunities so that they can direct their research and implementation efforts.
- Second, with the implementation of additional technologies, it is becoming increasingly important to understand the impact of technological change on the workforce in order to prepare for a successful and sustainable transformation.

This dissertation addresses both challenges, exploring technology trends, research directions, and workforce implications with the overarching goal of *contributing to navigating the fourth industrial revolution.*

1.2 Research questions and significance

This dissertation addresses the overarching goal through three research questions (RQs), which are addressed in three main chapters. Chapter 2 describes the evolution of 4IR concepts, evaluates the technology landscape, and highlights technological trends. This analysis of 4IR technologies can help researchers, practitioners, and policymakers better navigate the 4IR and reap its benefits. Chapter 3 explores the research on Industry 4.0, highlights the integration of various sub-fields of Industry 4.0 research, and suggests further research avenues to navigate towards a holistic development of the field. Shifting the focus from generally understanding 4IR trends, concepts, and research opportunities, Chapter 4 investigates the impact of the 4IR on occupations. The chapter shows which occupations are exposed to which technologies and therefore contributes to navigating the workforce transition toward the 4IR. Each chapter includes a separate introduction and conclusion section, addressing one RQ. The current section describes the relevance of the three RQs:

RQ 1	What are 4IR technologies, and how did 4IR concepts evolve over time?
RQ 2	Which topics need to be addressed to support an integrated holistic evolution of the 4IR?
RQ 3	Which occupations are particularly exposed to 4IR technologies?

RQ 1 is relevant because advancing technological progress requires a clear understanding of Industry 4.0 technologies, trends, and concepts (Osterrieder et al., 2020; Culot et al., 2020; Thoben et al., 2017). As there are many articles, frameworks, and concepts in this area, understanding the different visions of Industry 4.0 can be challenging (Zheng et al., 2020; Thoben et al., 2017). A variety of conceptual terms have been used to label these visions, including Industry 4.0 (Kagermann et al., 2013), Advanced Manufacturing (Shah, 1983), and Smart Manufacturing (Kang et al., 2016). However, the definition of Industry 4.0 varies within and across these concepts. Kang et al. (2016) provide a summary of government-driven Industry 4.0 frameworks, showing that frameworks from Germany, the United States, and South Korea have some overlaps but do not use identical definitions. Further, the German Industrie 4.0 taskforce introduced CPS as the core of the Industry 4.0 concept, while, for example, Rüßmann et al. (2015) identified nine high-level technology drivers (core technologies), such as autonomous robots and cybersecurity. Moreover, some authors have described Smart Manufacturing as a sub-concept of Industry 4.0 (Frank et al., 2019a; Lasi et al., 2014), while others see the two terms as identical (Thoben et al., 2017; Mittal et al., 2018). Chapter 2 addresses these issues. The chapter maps the technologies frequently associated with Industry 4.0 and identifies clusters and relations between technologies. It also shows how the technology landscape has evolved and highlights recent trends as well as areas of potential future growth. Finally, the chapter identifies the differences between 4IR concepts such as smart manufacturing and digital manufacturing.

RQ 2 is relevant because the success of the 4IR depends on an integrated evolution across research areas (Frank et al., 2019a). Research related to Industry 4.0 has significantly increased in the last decade and encompasses various fields, including operations management, technology management, and information systems (Liao et al., 2017). Frank et al. (2019a) provide a holistic framework that accounts for different fields in this area and includes four dimensions: Smart Manufacturing, Smart Working, Smart Supply Chain, and Smart Products and Services. Smart Products and Services, for example, are products that are connected to the internet, may collect and control user data (Hofmeister Kahle et al., 2020), and allow the producer to provide additional services (Frank et al., 2019b). Companies and customers can benefit most if different perspectives are integrated. For example, product data from smart products can help manufacturers enhance production planning and improvement activities. Therefore, research should not only focus on specific areas but should also provide a holistic view of Industry 4.0 and conduct research at the intersection of different fields. Chapter 3 addresses this issue and evaluates the Industry 4.0 research landscape with a particular focus on the evolution of different research fields and their integration.

RQ 3 is important because data on various occupations' exposure to 4IR technologies is crucial for modeling the impact of such technologies on the workforce. This modeling can assist policymakers in navigating the transition through the 4IR and can support companies in building up their capabilities accordingly. Previous research has taken various approaches to estimating the impact of technologies on occupations. Some articles have directly evaluated the share of tasks or occupations that could be automated by future technologies. Frey & Osborne (2017) estimated, for example, that 47% of U.S. jobs are at risk of being computerized, though Arntz et al. (2016) modified this approach and identified only 9% of jobs in OECD countries as at high risk of being computerized. Beyond these theoretical potentials for automation, additional labor market mechanisms appear when new technologies are introduced. In summarizing those effects, Acemoglu & Restrepo (2022) (for example) described new technologies as having an *automation effect* wherein machines learn to perform new tasks. Next, there is a *deepening effect* that occurs when machines become better at performing these tasks, which may create additional demand for labor to conduct existing tasks. Finally, a *reinstatement effect* occurs when humans take over new tasks created through these technologies. In addition, the *diffusion* of new technologies creates a demand for workers who can produce or maintain those technologies (capital accumulation effect). Researchers can build on these ideas to model the impact of technologies on the labor market. Accurate modeling requires indicators reflecting the adoption of new technologies. These indicators may include investments in robots (Dauth et al., 2017; Acemoglu & Restrepo, 2019b; Dauth et al., 2018), AI maturity indices (Acemoglu et al., 2020), information and communication technologies (ICT) intensity (Gregory et al., 2016), and patent data (Acemoglu et al., 2020; Mann & Püttmann, 2017; Webb, 2019). Even though some technology indicators exist, the lack of high-quality data remains a barrier to better understanding the impact of 4IR technologies (Mitchell & Brynjolfsson, 2017; Frank et al., 2019c). For example, robot investment data is only available at an industry level, and the AI maturity index by Felten et al. (2021) describes future automation potential but not current technology diffusion. Chapter 4 addresses this issue and provides an indicator of 4IR patent exposure for occupations and tasks. The approach builds on patent data and refines existing methods to increase the granularity and quality of the mapping of patents to occupations. This indicator itself provides information about the potential impact of technologies on occupations and can be a valuable input source for modeling labor markets.

1.3 Methodological relevance

This dissertation relies on textual datasets to better understand 4IR technology trends, research directions, and the relation of technologies to occupations. Further, I demonstrate how existing tools can help analyze and visualize data. Therefore, the work contributes to research on both the 4IR and research methodologies. The methodological contributions include using natural language processing (NLP) and machine learning to obtain insights from textual data and demonstrating ways to provide interactive visualizations of the findings.

The rate of data generation has exploded in recent years. The Internet of Things (IoT) helps companies collect data from sensors and machines along the supply chain; social media collects user-generated data; and public institutions build up datasets of economic, environmental, and societal indicators. Some of these datasets can be easily accessed. For example, there are freely accessible databases with millions of social media entries. Banda et al. (2021), for example, collected more than 100 million tweets related to COVID-19 that are available for research use. Further, using application programming interfaces (APIs) allows easy access to data from millions of research articles on Scopus or Web of Science. These datasets offer significant opportunities for research. For example, Tumasjan et al. (2010) drew implications about political sentiments using Twitter data, Prabhakaran et al. (2015) identified paradigm shifts in the field of information technology using data from more than 11,000 scientific articles, and Yun et al. (2016) built on patent data to learn about the impact of firms' open innovation research networks on performance.

For my goal (i.e., to contribute to navigating the 4IR), insights from patents and scientific articles offer high value. Scientific articles reflect research progress and therefore enable learning about research gaps and opportunities, differences across countries, and the shift of research focus on technologies and concepts over time. Patents contain detailed descriptions of technological inventions, therefore providing an indication of state-of-the-art technologies and technology trends. Additionally, patent descriptions offer insights into the content of the inventions—such as what the technology can do—while patent metadata contains information on the innovators.

These large textual datasets offer rich opportunities to create insights. However, benefiting from these datasets requires efficient and automated methods. For example, many literature reviews are still conducted manually. Analysis steps include article screening and information extraction, which are time consuming for large datasets. In addition, textual data are not always straightforward to analyze; for example, different words can have similar meanings, and similar words can have different meanings. In recent years, several NLP algorithms have been introduced for conducting textual analysis. Thanks to increasing computational capacities and improved algorithms, these methods can be used without the need for extensive resources. In addition, libraries such as spaCy[1] and Gensim[2] have high usability and can thus be used by a broad range of people without requiring extensive knowledge in the field of NLP.

The chapters of this dissertation rely on various NLP libraries for different tasks. Chapter 2 builds on named entity recognition (NER) using spaCy and prodigy[3]. NER helped extract technology terms ("named entities") from research articles. Chapter 3 used text categorization to screen the articles for relevance to our literature review and Chapter 5 used the same libraries for patent classification. Chapter 4 relies on Gensim to calculate text similarity scores for mapping patent descriptions to occupational task descriptions. My work builds on existing tools and uses them in new ways or to answer questions different from those addressed in previous work. I hope that this work shows researchers the potential of using NLP, serves as an example of how to do so, and provides a motivation to use these methods in future work.

RQ 1 shows the relations between technologies and identifies trends and clusters, while RQ 2 shows the integration of research fields. These complex relations can be best analyzed and visualized through network analysis, a technique that allows insights to be made from large amounts of data. Network analysis has grown in popularity in recent years because of the increased availability of data, though it has been widely used in the social sciences for more than a decade (Borgatti et al., 2009; Broniatowski et al., 2014). Other fields have also realized the

[1] `https://spacy.io/`
[2] `https://radimrehurek.com/gensim/`
[3] `https://prodi.gy/`

potential of these methods. Kim et al. (2017), for example, used patent network data to research the influence of standards on the convergence of the IoT technology. Network visualizations offer an excellent way to communicate results. However, with large datasets, a static view can reflect only a fraction of the insights. I used Gephi, a tool for network analysis, to export data for visualization on a web page[4]. This provides a simple way for readers to explore the work and interact with the graph—for example, by zooming in on the graph to view relations between nodes.

RQ 3 shows the exposure of occupations to 4IR technologies. Chapter 4 addresses this question using data for more than 900 occupations and differentiating among more than 300 technologies and technology groups. My objective was to allow the reader to interact with the data to obtain a deep understanding of the findings. Therefore, I created visualizations in Tableau Public[5]. This platform enables data visualization and allows for intuitive interaction with a dataset—for example, by applying filters and reviewing various slices of the data. These visualizations make the insights accessible not only to researchers but also to other actors, such as companies, employees, and policymakers. As making data and findings publicly accessible becomes more common, I hope this work also inspires researchers to provide their findings on platforms that allow for an intuitive exploration of the insights.

1.4 Structure of the dissertation

The dissertation is structured around the research articles I wrote during my PhD program. Each chapter can be read stand-alone, including an introduction and conclusion. The chapters relate to work that have been (or will be) published with co-authors; therefore, the chapters use the first person plural rather than the singular. Chapters 2, 3, and 4 answer RQs 1–3. In addition, Chapter 5 includes contents from a short article presented at a conference and offers a methodological contribution. Figure 1.1 illustrates the structure of the dissertation. Below, I briefly describe the chapters.

1.4.1 Mapping Industry 4.0 technologies

Chapter 2 addresses RQ 1 and researches how 4IR technologies and concepts have evolved over time. Due to the growing research and fast evolution in the field of 4IR, there are as yet no clear definitions of concepts like Industry 4.0 and digital manufacturing. This work provides a clear description of technological trends and concepts to better align research efforts and foster communication among researchers, industry, and policymakers. I introduce a novel method to create a map of Industry 4.0 technologies: namely, using NLP to extract technology terms from 14,667 research articles. Network analysis identified eight clusters of Industry 4.0 technologies. I show the relations between these clusters and provide insights into technological trends that serve as a basis for the analysis. The chapter introduces a novel approach for literature reviews, and the results will enable researchers, industry, and policymakers to better navigate the large corpus of work; reveal the differences between concepts such as advanced and

[4] `https://bmeindl.github.io/technology_network/`
[5] `https://public.tableau.com/app/profile/benjamin.meindl/viz/4IR_tech/Landing`

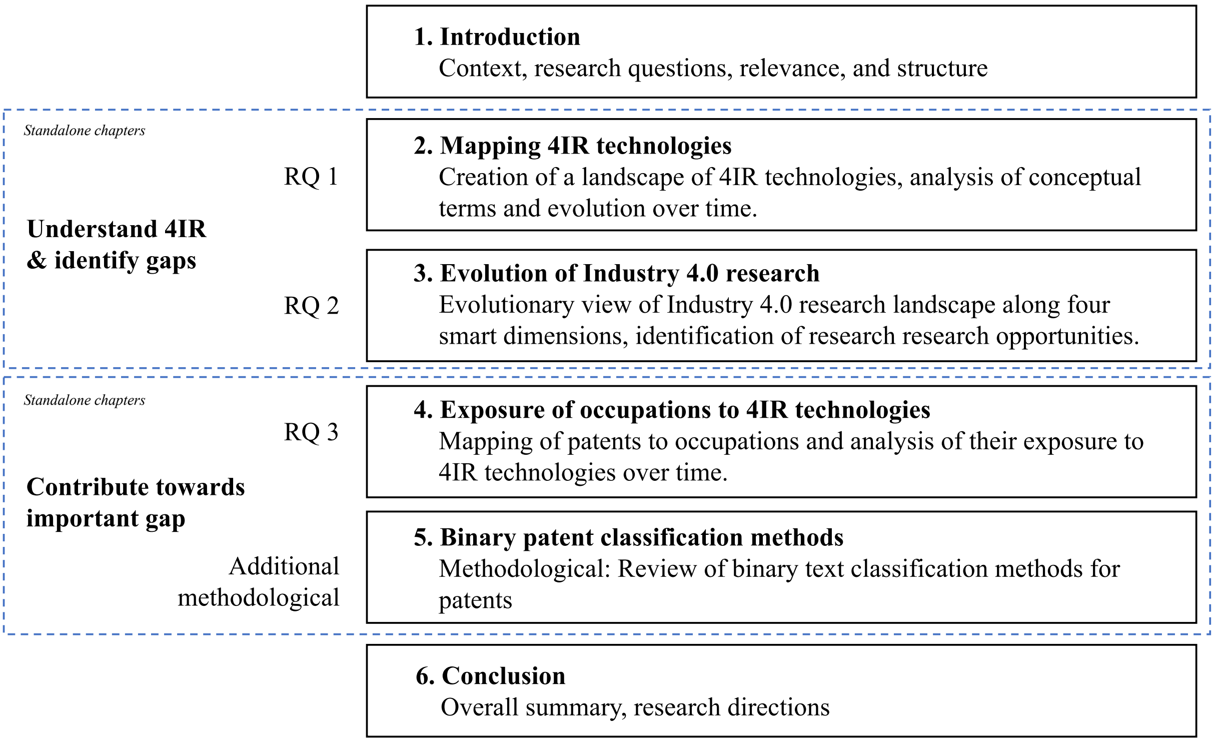


Figure 1.1: Structure of the dissertation. Each box describes a Chapter; Chapters 2–4 address the research questions, while Chapter 5 describes an additional methodological contribution.

intelligent manufacturing; and highlight trends and research gaps to help these actors reap the benefits of digital transformations.

The content of this chapter overlaps with an article uploaded on arXiv (Meindl & Mendonça, 2021). I presented an earlier version of this work at the *Innovation, Entrepreneurship and Knowledge Academy* (INEKA) 2019 Conference in Verona and as a poster presentation at the *5th International Conference on Computational Social Science (IC2S2)* 2019 in Amsterdam. I received additional feedback presenting the work at *Building the future @ CEiiA* and *LEIS meetings* at Técnico Lisbon.

1.4.2 Evolution of Industry 4.0 research

Chapter 3 addresses RQ 2 and examines how the fast-growing Industry 4.0 literature has evolved with the aim of proposing future research opportunities. The work focuses on the four smart dimensions of Industry 4.0: Smart Manufacturing, Smart Products and Services, Smart Supply Chain, and Smart Working. The machine-learning–based systematic literature review includes 4,973 papers published between 2011 and 2020. I conducted a chronological network analysis considering the growth of these four dimensions and the connections among them. The results show how the four smart dimensions evolved over time and highlight research opportunities for further integrating them. I also analyzed keywords and the main journals publishing work on these four smart dimensions. Scholars can use the findings to understand journals' orientations and identify gaps that can be filled by future research.

This chapter largely builds on an article published in *Technological Forecasting and Social Change* (Meindl et al., 2021a).

1.4.3 Exposure of occupations to technologies of the fourth industrial revolution

Chapter 2 shows that AI moves towards the center of the 4IR landscape and may enable continuing the fast growth of research on human-robot systems. This may have significant impact on the future workforce and in Chapter 3 I suggest, that the are of Smart Working requires future research, e.g., how AI can support workers' decision making. While overall it is not clear how 4IR technologies will impact the future of work, Chapter 4 clarifies which occupations are particularly exposed to technologies of the 4IR. It provides an indicator of technological progress based on patent data, which illuminates the impact of new technologies on jobs and can be used to help companies build up their capabilities accordingly. Using NLP, I compared the content of occupational task descriptions with the text of patent abstracts to calculate similarity scores. These scores identify the patents that best describe the occupational tasks. For each occupation, I calculated patent exposure scores, which represent related technological progress. In order to explore the impact of 4IR, the work differentiates between traditional patent and 4IR patent exposure. This method differs from previous approaches in that it both accounts for the diversity of task-level patent exposures within an occupation and more accurately reflects work activities. The results show which occupations are exposed to 4IR patents and non-4IR patents. These 4IR exposure scores served as input for labor market analysis to achieve insights into the technologies' impacts on job growth. In addition, I published the occupation 4IR patent exposure scores so that other researchers could conduct further, more extensive analyses. The work not only allows the impact of 4IR technologies to be analyzed as a whole but also provides exposure scores for more than 300 technology fields, such as AI and smart office technologies. I also discuss the potential of existing patent-based indicators for future studies (i.e., labor market analyses) and highlight their particular relevance for analyzing current and recent labor market trends. Finally, the work provides a general mapping of patents to tasks and occupations, which will enable future researchers to construct individual exposure measures.

The content of this chapter overlaps with an article uploaded on arXiv (Meindl et al., 2021b). I received valuable feedback when presenting earlier versions of this work at the *IC2S2*, the *Technology, Management, and Policy Consortium (TMP) 2019* conference, and *7th Master Class on EU Cohesion Policy.*

1.4.4 Additional contribution: Binary patent classification methods for few annotated samples

In addition to the three main chapters, I published a conference article on binary classification of text data with small sample sizes. Binary classification algorithms are particularly useful for economic questions, which often use this approach to implement ambiguous and subjective concepts in research areas where human classification is time consuming and sample sizes are accordingly small. This covers examples such as whether workers are susceptible to automation and whether a device is automated. We compared the performance of multinominal naive Bayes (MNB), support vector machines (SVM), random forest, and k-nearest neighbor (k-NN) classifiers with a spaCy convolutional neural network (CNN) model as well as a spaCy CNN model pre-trained on patent data. The results showed that the CNN models had the highest overall accuracy, with significantly improved performance through pre-training. Our analysis suggests that the spaCy pre-trained CNN model provides a highly accurate NLP model that is feasible for implementation without requiring extensive computation capacity. Pre-training was particularly beneficial for small sample sizes. The use of 100 labeled patents already led to an accuracy of 77.2%. The small sample size required may encourage researchers in various fields to use manually labeled patent data to investigate their specific research questions.

This work stems from a project where we intended to classify patents based on whether they are automating tasks or deepening automation (i.e., make machines more efficient). This chapter served to test the classification approach on a generic example: classifying patents into patent classes. The initial application—differentiating between deepening and automation—has not yet been implemented. The annotation process turned out to be highly time consuming, and the annotations are highly skewed toward deepening patents, requiring a larger sample size. However, this method evaluates various text classification approaches and can thus be valuable for future researchers investigating similar questions.

I presented this approach at the *1st Workshop on Patent Text Mining and Semantic Technologies* (PatentSemTech 2019), which was part of the *Semantic 2019* conference in Karlsruhe, Germany. It has been published in the conference proceedings (Meindl et al., 2019).

Chapter 2

Mapping Industry 4.0 technologies[1]

This chapter describes a review of scientific articles in the field of Industry 4.0, using natural language processing (NLP) and network analysis, and shows how Industry 4.0 technologies, clusters, and concepts evolved over time.

2.1 Introduction

Industry 4.0 describes the integration of emerging technologies and digital tools across the manufacturing value chain, building on (cyber-physical systems (CPS)). The term was coined in 2011 by the Industrie 4.0 working group, which produced the report to Secur[e] the future of [the] German manufacturing industry (Kagermann et al., 2013). The concept of Industry 4.0 addresses the significant transformations in various industrial sectors (Roblek et al., 2016) that have created the opportunity to reshape manufacturing (Krzywdzinski et al., 2016) and enable sustainable development (Bai et al., 2020). Aside from the German task force, several other national initiatives have introduced ideas for how to address the technological changes of the fourth industrial revolution (Santos et al., 2017). —for example, "Made in China 2025" (Li, 2018), the New Industrial France program (NIF, 2016), the U.K. Foresight project (Foresight, 2013), and the U.S. Smart Manufacturing Operations Planning and Control Program (NIST, 2014).

Following these initiatives, an increasing amount of research has focused on manufacturing in the context of the fourth industrial revolution. However, a precise characterization of the technological scope of Industry 4.0 is lacking (Zheng et al., 2020). Rüßmann et al. (2015) described, for example, nine high-level technology drivers (core technologies), including autonomous robots and cybersecurity. Chen (2017) presented 10 major technologies, including 3D printing and virtual reality. Culot et al. (2020) used a framework of 13 enabling technologies, while others (Beier et al., 2020) described 15 key technology features. Both Culot et al. (2020) and Beier et al. (2020) reviewed defining elements of previous Industry 4.0 concepts by comparing their frameworks against previous publications. Culot et al. (2020) found, for example, that only one Industry 4.0 article referred to "new materials" as a key enabling technology, whereas the Internet of Things (IoT) was a key technology in many articles.

[1] This chapter largely overlaps with the following article which is currently under review: Meindl, B., Mendonça, J., Mapping Industry 4.0 Technologies: From Cyber-Physical Systems to Artificial Intelligence.

Aside from the term Industry 4.0, some researchers have also used different labels to describe the concepts of emerging manufacturing technologies (Culot et al., 2020). These labels include advanced manufacturing (Cheng et al., 2018; Shah, 1983), smart manufacturing (Kang et al., 2016; Kusiak, 2018; Tao et al., 2018), intelligent manufacturing (Zhong et al., 2017), digital manufacturing (Behrendt et al., 2017; Chryssolouris et al., 2009; Paritala et al., 2017), cloud(-based) manufacturing (Tao et al., 2014; Wu et al., 2013), and Factories of the Future (European Commission, 2013). While all of these concepts describe emerging manufacturing technologies, they are not clearly defined, and there does not seem to be a clear understanding of their differences and boundaries. Thoben et al. (2017) and Mittal et al. (2018), for example, stated that the term "Industry 4.0" is used in Germany, while "smart manufacturing" is used in the United States. Other authors have described smart manufacturing as a sub-concept of Industry 4.0 (Frank et al., 2019a; Lasi et al., 2014; Osterrieder et al., 2020). Research should thus develop a better understanding of Industry 4.0 (Osterrieder et al., 2020), including differences in regional Industry 4.0 profiles (Benitez et al., 2020). This article aims to fill these gaps and provide clarity on the Industry 4.0 technology landscape.

Industry 4.0 is characterized not only by certain technologies, but also by their connections and combinations (Alcácer & Cruz-Machado, 2019; Bai et al., 2020; Culot et al., 2020). Analyzing the relations between technologies complements previous work and provides a better understanding of the evolution of the Industry 4.0 landscape (Meindl et al., 2021a). We build on the approach by Chiarello et al. (2018) to create a network of Industry 4.0 technologies. Building a technology map shows the relevance of technologies, technology clusters, and their interrelations. Whereas Chiarello et al. (2018) based their work on an analysis of Wikipedia articles, we used natural NLP to extract information on technologies from scientific publications. Our work differs from many previous review articles in that it does not rely on a review of conceptual articles but rather builds on the extensive amount of articles about Industry 4.0, including those focusing on specific aspects of the concept. This search yielded more than 14,000 articles published over a period of 10 years. Further, we introduce a timeline in our analysis, which allows us to review the evolution of the concept since 2011. The analysis focuses on articles about Industry 4.0 and related concepts such as smart manufacturing and shows similarities and differences between them. Finally, the obtained technology map provides an extensive dataset of underlying technologies and their relations to each other.

This article makes two main contributions. First, the analysis contributes to a deeper understanding of the various Industry 4.0 and related concepts, shows evolutions and trends, and helps identify further avenues for research to ensure progress in the transformation towards Industry 4.0. As Industry 4.0 is a relatively new, fast-changing concept (Culot et al., 2020; Galati & Bigliardi, 2019), we provide a dynamic view of Industry 4.0 that both provides a snapshot of the technological landscape and shows historical development, enabling it to be updated and to track its progress in the future. Unlike previous articles, we do not rely on predefined Industry 4.0 dimensions. Instead, our approach uses network analysis to create an unbiased bottom-up view of the Industry 4.0 landscape. We identify eight technology clusters with Industry 4.0 at the center. Our trend analysis shows that the importance of artificial intelligence (AI) has seen major growth. We suggest future frameworks to reflect this trend and emphasize the important role of AI in Industry 4.0. Second, this article provides a novel approach for a systematic literature review (SLR). The method extracts relevant information from the Industry 4.0 academic literature using NLP, and we develop a mapping algorithm to create a technology map for visualization and network analysis. The proposed methodology may also be useful in reviewing the evolution of other research fields.

The remainder of the paper is structured as follows. The next section (Section 2.2) presents a review of the literature on Industry 4.0 technologies, after which Section 2.3 presents the approach used. The results (Section 2.4) include an Industry 4.0 technology map; the evolution of technology clusters over time; a review of conceptual terms, such as smart manufacturing; and an analysis of the countries of origin of Industry 4.0 research. A discussion of the results is presented in Section 2.5, and the conclusion follows in Section 2.6.

2.2 The fourth industrial revolution and Indusry 4.0 in the literature

2.2.1 Definitions of Industry 4.0

In their initial article, the Industrie 4.0 working group described Industry 4.0 as a broad industrial paradigm (Kagermann et al., 2013). The smart factory and CPS form the core of this idea, and interact with a smart environment, including smart mobility, smart logistics, smart buildings, smart products, and smart grids. The UK Government supported the Foresight project (Foresight, 2013) has a strong focus on manufacturing and describes a vision for "the future of manufacturing" in terms of four characteristics: They describe technologies as enablers for faster response to customers, more sustainability of new systems, changing skill demands and new business opportunities, e.g., through advanced design opportunities. With a similar focus on manufacturing industries, the U.S. National Institute of Standards and Technology (NIST) supports various manufacturing initiatives to secure the future of manufacturing in the U.S., such as the Smart Manufacturing Operations Planning and Control Program (NIST, 2014). The New Industrial France program (NIF, 2016) aims "to assist French companies move upmarket and position themselves in the markets of the future." It has a broad focus, not only addressing manufacturing, but also other areas, such as eco-mobility and smart cities. Chinas "Made in China 2025" program aims to "enhance industrial capability" and has a particular focus on AI (Li, 2018; Zhou et al., 2018).

Several research articles have focused on the core of Industry 4.0, which the Industrie 4.0 working group (Kagermann et al., 2013) report describes as the "Smart Factory" (Meindl et al., 2021a). A framework developed by Boston Consulting Group (Rüßmann et al., 2015) has frequently served as a reference point for research (Alcácer & Cruz-Machado, 2019; Butt, 2020; Saucedo-Martínez et al., 2018). This concept of Industry 4.0 comprises nine technologies, with a narrow focus on industrial production plants. These technologies include, for example, autonomous robots, augmented reality, and additive manufacturing. Chen (2017) introduced the concept of "integrated and intelligent manufacturing," which incorporates similar technologies as the concept developed by Rüßmann et al. (2015) but focuses on software platforms integrating these technologies.

Lu & Weng (2018) reviewed 30 international case studies to evaluate technology maturity and roadmaps. They also focused on the smart factory aspect of Industry 4.0. The article structured the technologies in four layers to represent a factory infrastructure: An integration layer connects the sensor layer to an intelligent layer that conducts analytics, and the response layer uses information from the intelligent layer to generate actions, such as production and sales prediction. Similarly, other authors have described an Industry 4.0 landscape tightly centered around the smart factory (Brettel et al., 2014; Dalenogare et al., 2018). These articles emphasize the integration of flexible shop floor production and product development processes as a core idea in Industry 4.0. Zhong et al. (2017) described a similar vision, termed "intelligent manufacturing." Their article described intelligent manufacturing as a synonym of smart manufacturing and part of a broader Industry 4.0 paradigm. Bai et al. (2020) also highlighted the role of intelligent systems, describing Industry 4.0 as "a new paradigm of smart and autonomous manufacturing" (p. 2). Unlike most other frameworks, they also considered drones and global positioning systems as key technologies.

Lasi et al. (2014) drew a broader picture of Industry 4.0 technologies. In addition to the factory itself and the development and monitoring processes, their Industry 4.0 vision focuses on logistics, including delivery and suppliers of goods and tools. Their concept thus focuses on a cyber-physical production network, including innovative enterprise resource planning (ERP) systems. Ghobakhloo (2018) presented a framework focusing not only on the smart factory but also on the integration of logistics and customers (e.g., the Internet of People) as key elements of the Industry 4.0 landscape.

Frank et al. (2019a) combined many of these ideas into a framework that describes the Industry 4.0 enterprise as consisting of four "smart" elements. Aside from smart manufacturing (the smart factory), Industry 4.0 comprises a smart supply chain (e.g., platforms with suppliers), smart working (e.g., augmented reality for product development), and smart products (product connectivity). In addition to some underlying base technologies, such as the IoT, they argue that each of the "smarts" is related to different technologies. Nakayama et al. (2020) also highlighted the relevance of Industry 4.0 beyond company boundaries and emphasized the flexibility of decentralized systems as the differentiator from Industry 3.0. Similarly, Alcácer & Cruz-Machado (2019) emphasized the role of Industry 4.0 in decentralization and real-time engagement. Roblek et al. (2016) described Industry 4.0 at a higher level: Aside from the enterprise and supplier level, they also described smart cities and digital sustainability as fundamental components of Industry 4.0.

2.2.2 Reviews of Industry 4.0 concepts in literature

Previous review articles on Industry 4.0 indicate how the concept has been perceived in academia and industry. Table 2.1 summarizes the various literature review approaches to date.

Ghobakhloo (2018) reviewed Industry 4.0 design principles and technology trends in an article that used language processing tools to review 178 articles and book chapters and described Industry 4.0 as a broad concept reaching far beyond the smart factory. Ghobakhloo (2018) identified technology trends—including blockchain, additive manufacturing, and the Internet of People—as well as Industry 4.0 design principles, which include decentralization, integration, individualization, and smart entities. Similarly, in a manual review of 88 articles, Lu (2017) derived a broad concept of Industry 4.0 from the literature, including the smart city, smart grid, and smart home as applications of Industry 4.0. Another manual review was conducted by Saucedo-Martínez et al. (2018), who reviewed 110 articles and provided information on their contributions along with nine key technologies defined by (Rüßmann et al., 2015). They found that horizontal and vertical integration across company boundaries was the pillar of Industry 4.0.

Table 2.1: Overview of Industry 4.0 literature reviews.

Literature review	Articles reviewed	Concepts reviewed	Method
Chiarello et al. (2018)	Wikipedia articles related to Industry 4.0	Industry 4.0	Automated crawling
Culot et al. (2020)	81 academic and 18 non-academic sources	Industry 4.0 and 16 related concepts	Manual
Ghobakhloo (2018)	178 academic and non-academic articles and book chapters	Industry 4.0	IBM language processing and manual review
Hermann et al. (2016)	130 academic and non-academic sources	Industry 4.0	Natural language processing
Kipper et al. (2020)	1,882 academic articles	Industry 4.0, smart manufacturing, fourth industrial revolution (and similar)	Keyword analysis
Liao et al. (2017)	224 academic articles	Industry 4.0	Keyword analysis
Muhuri et al. (2019)	1,619 academic articles (focusing on most-cited articles)	Industry 4.0	Bibliometric review
Saucedo-Martínez et al. (2018)	110 academic articles	Industry 4.0	Manual
Strozzi et al. (2017)	462 academic articles	Smart factory	Bibliometric review
Lu (2017)	88 academic articles	Industry 4.0	Manual
Zheng et al. (2020)	186 academic articles	Industry 4.0	Manual

Unlike the previously described reviews, Liao et al. (2017) undertook a quantitative analysis. They found that Industry 4.0 articles referred most commonly to the manufacturing stage of the product lifecycle. Most common technology keywords referred to information technology (IT), such as modeling, visualization, or big data. Their systematic keyword analysis built on 224 academic articles. Similarly, Hermann et al. (2016) conducted a quantitative text analysis of 130 publications related to Industry 4.0, including scientific articles as well as practical journals and books. They used NLP to create an overview of the most frequent terms in publications and used insights from an expert workshop to identify four Industry 4.0 design principles: technical assistance (e.g., virtual assistance); interconnection (e.g., standards); information transparency (e.g., data analytics); and decentralized decisions. Hermann et al. (2016) provided a vision of Industry 4.0 that centered on the smart factory.

Zheng et al. (2020) conducted a manual review of 186 selected Industry 4.0 articles structured along the business process life cycle. They identified Industry 4.0 applications such as order picking management, collaborative operations with humans, and assembly defect detection. They linked these applications to core technologies, such as IoT, AI , and additive manufacturing. Culot et al. (2020) systematically reviewed 81 academic and 18 non-academic articles on Industry 4.0, addressing the enabling technologies, characteristics, and expected outcomes related to Industry 4.0. Unlike previous reviews, they not only focused on Industry 4.0 but also included different conceptual terms, such as smart manufacturing and cyber manufacturing. They described Industry 4.0 as a broad concept that covers various more specific concepts, such as cloud manufacturing, which focuses on new business models, and social manufacturing, which includes society and consumers as key characteristics. This differentiation among concepts represents a novel contribution to the literature. Like most reviews, Zheng et al. (2020) and Culot et al. (2020) mainly built on articles that aimed to provide a conceptual view of Industry 4.0. Since our work incorporates a broader range of articles on Industry 4.0, we included all articles that refer to Industry 4.0, including case studies and articles about specific technologies (e.g., 5G, 3D printing) or specific contexts (e.g., the supply chain). This allowed us to reflect not only on insights from the Industry 4.0 research community but also from other fields, such as computer science.

Due to the large number of articles referring to Industry 4.0 and related topics, a truly comprehensive manual review is difficult to achieve. Therefore, research may make use of automated methods. For example, Strozzi et al. (2017) and Muhuri et al. (2019) conducted bibliometric reviews of the smart factory and Industry 4.0 literature using structured textual data. Although they mentioned the most frequent keywords and keyword clusters, their main focus was identifying citation networks and the most relevant institutions and authors. Chiarello et al. (2018) reviewed a large number of articles using an automated review to create a network map of Industry 4.0 technologies that provides information on the importance of technology fields, as well as clusters and their interrelations. Their analysis used seed words from scientific articles and industry reports to crawl Wikipedia articles related to Industry 4.0 for further keywords. This approach identified more than 1,200 technologies as an input for their analysis, most of them related to information technologies. The analysis described 11 technology clusters, including big data, programming languages, and computing. Some researchers have identified as a limitation the lack of an agreed-upon Industry 4.0 taxonomy that can be used for a structured review (Zheng et al., 2020). This barrier can be overcome when building a technology network, which relies on articles themselves as a basis for technology clusters or categories, shows the relations between clusters, and is independent of previous frameworks (see Chiarello et al. 2018).

2.3 Methods

We aimed to contribute a better understanding of the Industry 4.0 technology landscape, its evolution, its key technologies, and the relations among those technologies. Therefore, we developed an automated approach to review the literature. First, we retrieved relevant articles from Scopus using its application programming interface (API) and removed duplicate articles and retracted articles. Second, we trained a neural network to identify technology terms in the article abstracts, which we then extracted. Finally, we visualized Industry 4.0 technologies as a technology map and conducted network analysis. The results show connections between technologies, as well as results from network analysis that can better explain the importance of technologies, clusters, trends, and the technological focuses of different conceptual terms (e.g., showing the differing focuses of intelligent and smart manufacturing). Figure 2.1 illustrates the flow of the analysis.

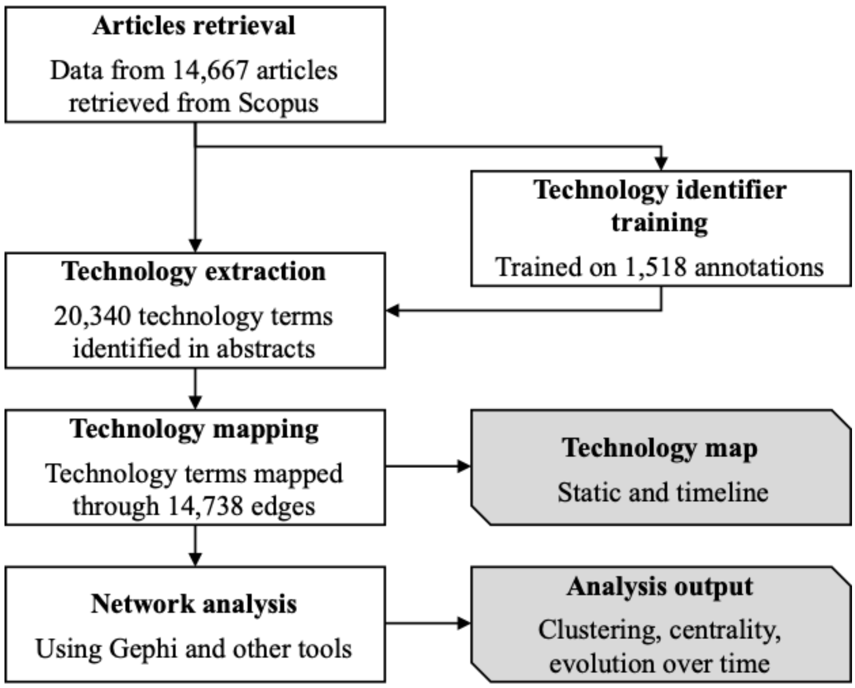


Figure 2.1: Flow diagram representing our analysis to create the Industry 4.0 technology map. First, documents were extracted from Scopus and prepared for analysis. Next, keywords were extracted and mapped, and finally network analysis and visualization were used to show clusters and trends of technologies. White and grey boxes indicate analysis processes and results, respectively.

2.3.1 Data

Our analysis used scientific publication data collected from Scopus, one of the largest repositories of scientific articles (Harzing & Alakangas, 2016; Liao et al., 2017) and a reliable source that has been used in previous reviews (Kipper et al., 2020; Meindl et al., 2021a; Pirola et al., 2020). In addition to article titles, abstracts, and keywords, we retrieved the first author's country to analyze the geographical distribution of research conducted.

Industry 4.0 technologies are described with various conceptual terms; therefore, we created a search query with the most important related terms. Industry 4.0–related terms include: "Industry 4.0"; "Industrie 4.0"; "digital manufacturing"; "smart manufacturing"; "intelligent manufacturing"; "cloud(-based) manufacturing"; "Factory/Factories of the Future"; and "advanced manufacturing." We included articles available in Scopus on March 30 2020, published since 2011 that contained any of these terms in their title, abstract, or keywords. The search yielded 14,667 articles. Figure 2.2 presents the number of papers per year per term and shows that the number of papers related to Industry 4.0 and smart manufacturing increased significantly since 2014. Since 2016, the search term "Industry 4.0" has generated the most results.

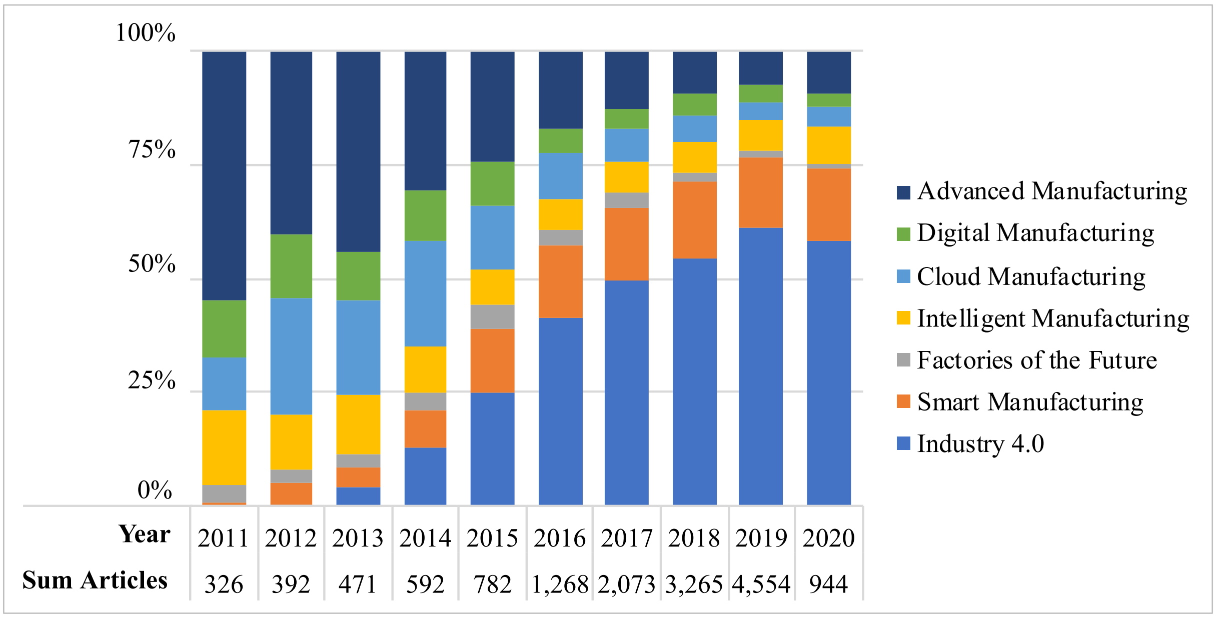


Figure 2.2: Distribution of papers associated with Industry 4.0 conceptual search terms. Column values indicate the share of articles related to a term of all articles published in the same year. Multiple search terms per paper are possible. The sum of articles indicates the total number of articles per year, excluding double counts. Terms are ordered by total growth in number of articles between 2011 and 2019.

We used publications' abstracts for our analysis. Abstracts are particularly useful because they provide the highest ratio of keywords to text (Shah et al., 2003). Similar analyses based on journal abstracts have been conducted to (for example) identify technological fronts (Shibata et al., 2008) or obtain insights into applications of machine learning in smart manufacturing (Sharp et al., 2018).

2.3.2 Definition of technology terms

Previous publications analyzing manufacturing technologies did not provide a definition of "technology" but relied on common understanding (Kang et al., 2016; Lu, 2017; Qin et al., 2016). This paper provides a definition of technology to assure consistency of the work, minimize bias and improve accuracy of the entity recognition algorithm. The definition also offers a basis for further work to build on.

The Oxford Dictionary (Oxford University Press, 2018) describes technology as "The application of scientific knowledge for practical purposes, especially in industry" which can be either machinery and equipment (artifact) or the branch of knowledge (a field). The field, Industry 4.0 related technologies, serves as a filter for our search. Within the results, our definition of a technology refers only to artifact. This is consistent with Aunger (2010), who states that a technology is being made for exogenous and endogenous use, i.e. "to achieve enhanced functionality from behavior without causing the destruction of the artifact." With this definition, a technology may be a robot or a sensor system. Anick et al. (2014) also consider "process/technique" as a technology. They define a process/technique as "the name of a method or process for creating an artifact or doing technical work (e.g., duty cycle control, electron microscopy)." Therefore, also additive manufacturing or cloud computing are considered as a technology term. End products, e.g., a car, is not considered as a technology. Further, Aunger highlights the timeline as a relevant aspect to identify or categorize a technology. For our analysis we only consider post-industrialization technologies, and thus exclude terms such as cylinder, bolt, or hammer.

2.3.3 Technology terms extraction

We used NLP to extract technology terms from documents. Extracting specific entities from a text—such as a location, person, or (in this case) technology term—is called named entity recognition (NER). NER can rely on handcrafted rules for feature engineering (Lample et al., 2016), such as searching for combinations of adjectives and nouns. Analyzing word frequencies (Shibata et al., 2008) has also commonly been used to identify keywords. Zhang et al. (2014), for example, used a word-frequency–based method called term frequency–inverse document frequency (TF-IDF) to identify keywords related to dye-sensitized solar cells, and Prabhakaran et al. (2015) relied on this method to identify paradigm shifts in emerging IT-related fields.

This work takes an approach based on neural networks refined with feature engineering rules. NER methods based on neural networks are superior to other approaches in many cases (Yadav & Bethard, 2019), as they can achieve better accuracy and require less intensive feature engineering. A neural network learns which terms to extract based on manually annotated examples. The algorithm identifies terms not only by remembering words but also using neighboring words. This method enables the algorithm to identify technology terms that have not previously been manually annotated.

This article relies on the spaCy's large neural network model[2] (Honnibal & Montani, 2017), which provides the highest overall accuracy compared to state-of-the-art NLP tools (Jiang et al., 2016; Al Omran & Treude, 2017). We train the model based on 1,518 manual annotations and additional 454 annotations for evaluation of the model. The tool Prodigy[3] was used to annotate technology terms for training the neural network. Prodigy selects phrases for annotation that are expected to have the highest impact on overall accuracy. With this approach we reached an accuracy (precision) rate of 78%. After 968 annotations, accuracy was already at 77%, indicating that additional annotations would not lead to improved accuracy. The accuracy (precision) rate of 78% indicates that most technologies were correctly identified. Evaluation of recall (fraction of technologies identified) indicates similar accuracy (77%). However, NER can generally reach an accuracy of up to 85% (Strubell et al., 2017). Therefore, future researchers could try to increase accuracy through, for example, improved preprocessing or more advanced neural modeling techniques (e.g., training the neural network with the recently introduced method Bidirectional Encoder Representations from Transformers (BERT); Devlin et al. 2018). On the other hand, depending on the complexity of a named entity, the maximum accuracy may be

[2]SpaCy models: en_core_web_lg-2.2.5 retrieved from `https://github.com/explosion/spacy-models/releases/tag/en_core_web_lg-2.2.5`, accessed on 10.01.2020.

[3]`https://prodi.gy/`

lower. The 85% boundary applies for general terms, such as cities, organizations, or persons, whereas technology terms are more complex in structure and difficult to identify. With our 1,518 manual annotations (and also after 968 annotation), we clearly exceeded the precision rate of Anick et al. (2014), who reached a precision rate of 63% for English technology terms based on 3,700 manual annotations. To account for false positives, we included manual cleaning steps (e.g., removing frequent falsely identified non-technology terms). After these steps, a few non-technology terms may still not be filtered out, particularly low-frequency terms. We did not expect a structured bias to result from those words, as they appeared across all types of articles. Therefore, we expected that occasional non-technology terms would not have a significant negative impact on the network but would add some noise.

Since many technology terms consist of multiple words or appear as list, we introduced an approach particularly suitable for identifying those terms. For example, standard NER techniques may identify the expression "flexible and reconfigurable manufacturing systems" as one technology term, but cannot identify both "flexible manufacturing systems" and "reconfigurable manufacturing systems." To overcome this challenge we introduced a two-step algorithm. First, we trained the algorithm to identify only the head words (in this case, "systems") and then using sentence parsing to extract both technology terms. Sentence parsing means, that we used spaCy to identify dependency between words and all words related to the headword via specific dependencies were added to the entity. Relevant dependency types, as described by Marneffe et al. (2014), include adjectival modifiers (amod), compounds, noun phrases as adverbial modifiers (npadvmod), and nominal modifiers (nmod). Figure 2.3 provides an example of dependency parsing.

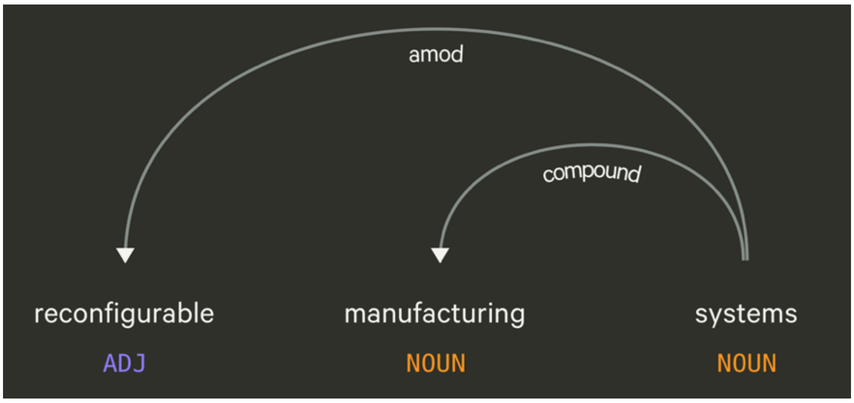


Figure 2.3: Example of word dependencies displayed with the dependency visualizer displaCy from Explosion AI. Image retrieved on Nov. 4, 2018 from https://explosion.ai/demos/displacy.

2.3.4 Mapping technology terms

Our neural network extracted technology terms from each abstract in our analysis. The novel mapping algorithm mapped technologies based on co-occurrence and semantic similarity. Additionally, the document's timestamp was added to each term to allow for an analysis of evolution over time.

Network analysis allows insights from large amounts of data to be connected and has been used to take advantage of the increased availability of data. Kim et al. (2017) used patent network data to research the influence of standards on the convergence of IoT technology. Further, Kwon & Park (2015) proposed a framework that generates social impact scenarios of new technologies based on text mining of online content, while another group of researchers used network analysis to analyze the Dutch innovation system (van Rijnsoever et al., 2015). Network analysis has also been conducted to elucidate Industry 4.0 (e.g., Chiarello et al. 2018; Muhuri et al. 2019; and Meindl et al. 2021a used network analysis to visualize the integration of different aspects of Industry 4.0 over time).

We mapped technology terms based on co-occurrence, which indicates a relatedness between technologies (Chiarello et al., 2018; Shibata et al., 2008; Yoon & Kim, 2012; Zhang et al., 2014). Technologies mentioned in the same article abstract were considered linked. Technologies co-occurring in the same sentence received an additional link, as this indicates close relatedness. The weight of all links within a document adds up to one, ensuring the equal influence of each article on the overall technology map. Additionally, we conducted semantic mapping to ensure that sub-technologies were linked to their base technology. For example, "wireless sensor network" is linked to "wireless sensor." Therefore, we defined the strength of the semantic link between two words such that on average the sum of semantic links per technology term equaled those based on co-occurrence. This ensured that network weights were still mainly driven by technology term frequencies while still reflecting semantic relations.

2.3.5 Network visualization and analysis

The nodes (technology terms) and edges (links between terms) were imported into the tool Gephi (Bastian et al., 2009) for visualization and network analysis. The nodes were arranged for visualization using the ForceAtlas2 algorithm (Jacomy et al., 2014), which allows for high-quality, intuitive mapping. The algorithm arranges the technologies in a force-directed layout where nodes repulse each other while edges (e.g., co-occurrence) act like springs, attracting nodes.

To analyze the importance of a single technology within the technology map, we calculated two measures: the weighted degree and the eigenvector (EV) centrality. The degree describes a node's size. The weighted node degree accounts for the weights of node connections in weighted networks and is defined as the sum of the weights of all connections linked to a node (Opsahl et al., 2010). The weighted degree only accounts for a node's local network (direct connections) and thus does not well represent its importance for the overall network. Therefore, we introduced EV centrality as a second measure. The EV score is higher for nodes connected to other nodes with a high EV (Bonacich, 2007). For the analysis, we used the measures as follows: Weighted node degree served as a filter to select the technologies most frequently mentioned in literature, while EV centrality served as an additional measure for importance—for example, to analyze how the importance of a technology changed over time.

Further, we identified technology clusters. Gephi offers modularity analysis, which divides a network into clusters of closely related nodes (Blondel et al., 2008). We conducted modularity analysis using Gephi and assigned different colors to the technologies in each cluster (Lambiotte et al., 2009). Based on these clusters, we evaluated the relations between clusters. For each technology, we counted the number of connections to technologies in each cluster. Calculating the share of links from a technology to each cluster helps identify cluster-linking technologies: technologies that have strong connections with more than one cluster. In our analysis, we also considered that smaller clusters have, by definition, fewer incoming connections and therefore also normalized the shares of links per cluster by cluster size. Further, we calculated the strengths of connections between clusters by summing the connections of all nodes of a given cluster to the nodes of each remaining cluster. In addition, we created technology profiles for the various concepts. To calculate these profiles, we considered all keywords in the articles related to a given concept and calculated the share of keywords associated with each technology cluster.

2.4 Results

In this section, we present the results of our network analysis. First, we show the overall technology network, clusters, and most important technologies. This analysis includes insights on cluster-bridging technologies and the evolution of the network over time. Second, we analyze the composition of various Industry 4.0 concepts, such as smart manufacturing and digital manufacturing.

2.4.1 Industry 4.0 technology map and trends

Analysis of the scientific corpus resulted in 2,317 technology terms (modeled as nodes in the network). The nodes are connected via 14,560 edges. Cluster analysis identified eight technology clusters of strongly connected technologies, which we describe in detail below. On average, around 80% of a technology's connections occurred within its own cluster. These numbers suggest that clustering provided meaningful results, with clear cluster associations. The clusters represent broad fields of Industry 4.0, and we assigned names to each cluster to best represent its associated technologies. The network comprises a core Industrial Internet of Things (IIoT) cluster and seven outer clusters. IIoT marks the center of the technology map and is the largest cluster. It represents mainly technologies related to the IIoT and communication technologies. In addition, there are seven outer clusters: four related to IT (Algorithms, Cloud Platforms, Management Systems, Sensor Systems), and three related to manufacturing processes (Additive Manufacturing, Computer-aided manufacturing (CAM), Human–Robot Systems). Figure 2.4 visualizes the technology map, with colors indicating clusters. Additionally, the graph data is available for download and to explore online at `https://bmeindl.github.io/technology_network/`.

Table 2.2 provides further information on technologies, such as centrality and degree. It includes the 10 highest-degree technologies of the technology network. We also included the three highest degree technologies per cluster, as well as some of the fastest-growing and fastest-declining technologies per cluster. Finally, the table includes relevant cluster-bridging technologies with strong connections to multiple clusters. The dynamic data allowed us to look into the development of technology clusters over time to identify trending technology fields. We used two measures for timeline analysis. First, Table 2.2 describes the change in importance of single technologies through a centrality measure (EV centrality), which indicates how well connected

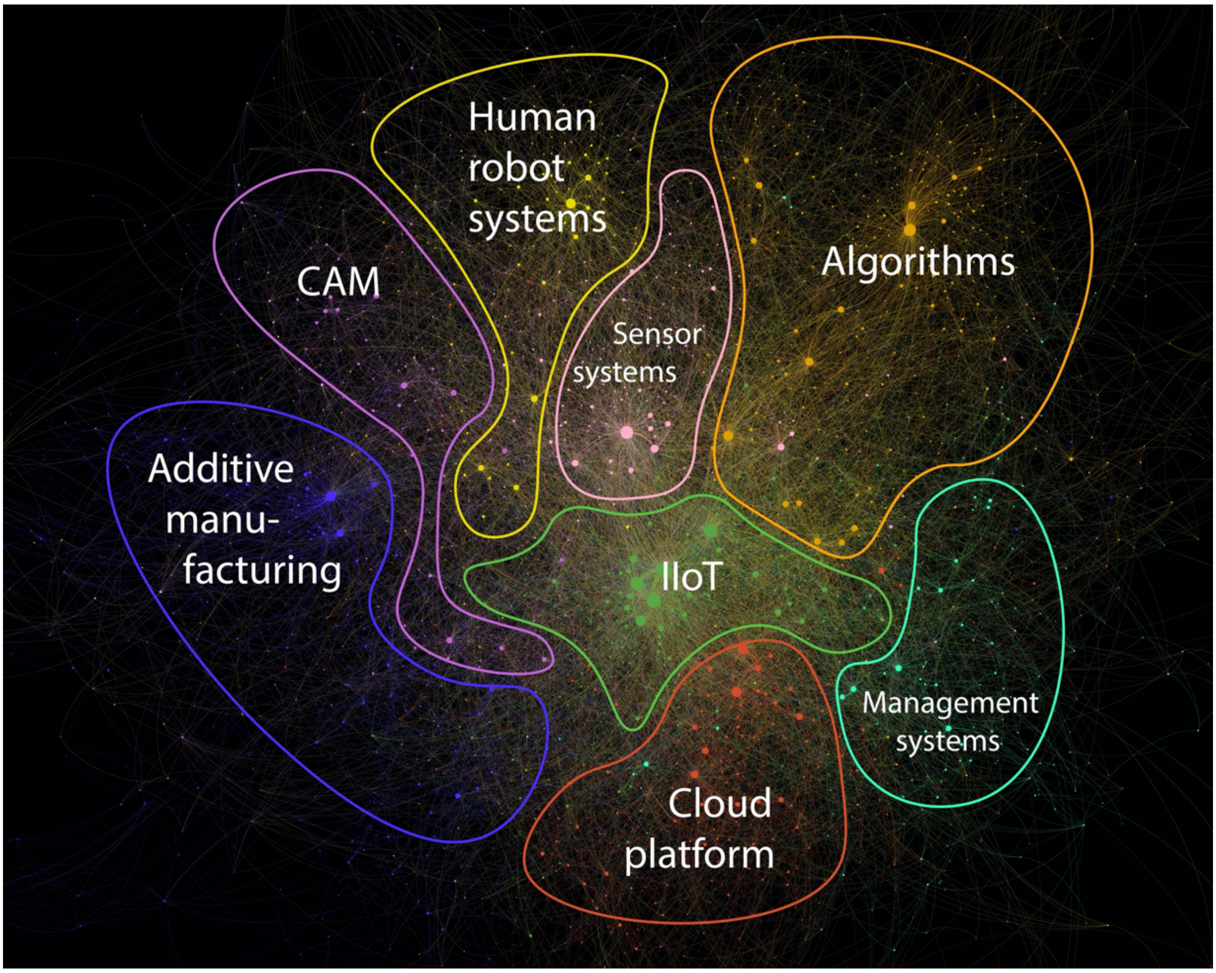


Figure 2.4: Map of Industry 4.0 technologies showing the eight technological clusters. Unassigned areas include technologies from various overlapping clusters. CAM stands for Computer-Aided Manufacturing and IIoT for Industrial Internet of Things.

(and thus relevant) a technology is within the overall network. Second, we described cluster importance using the weighted degree of technologies related to the cluster. Figure 2.5 and Appendix A.1 illustrates the evolution of a cluster's importance in the last decade, measured as the sum of the weighted node degrees within a cluster as a share of the total weighted degree of the network. Figure 2.6 includes additional information on cluster sizes and relations between clusters. The text below includes some information on other technologies not included in Table 2.2.

Table 2.2: This table includes the highest-degree technologies in the technology network and some of the fastest-growing Industry 4.0 technologies.

Technology cluster / Technology	EV centrality (D 2018–2020)	Weighted degree	Related cluster
IIoT: Industrial Internet of Things			
Internet of Things (IoT)	1 (0)	2957	
Cyber-physical system (CPS)	0.84 (-0.03)	2621	
Internet	0.82 (-0.04)	1406	(CP)

Industrial IoT (IIoT)	0.47 (0.10)	1365	
Information and communication technology (ICT)	0.31 (-0.05)	252	(CP)
IoT devices	0.24 (0.05)	266	
Software-defined networking (SDN)	0.16 (0.03)	130	CP
SE: Sensor Systems			
Sensor	0.77 (-0.05)	1558	
RFID	0.35 (-0.05)	344	(IIoT)
Wireless sensor network (WSN)	0.34 (-0.04)	342	(IIoT)
Smart sensor	0.26 (0.03)	162	(IIoT)
Automated guided vehicle (AGV)	0.12 (0.02)	102	(HRS)
Embedded sensor	0.08 (0.06)	36	(IIoT, AM)
AL: Algorithms			
Artificial intelligence	0.52 (0.13)	855	HRS, (IIoT)
Algorithm	0.32 (-0.02)	763	-
Machine learning	0.30 (0.10)	428	-
Analytics	0.19 (-0.06)	115	-
Blockchain	0.18 (0.1)	212	-
Big data analytics	0.17 (0.05)	126	IIoT, (CP)
Genetic algorithm	0.15 (-0.06)	290	-
MS: Management Systems			
Manufacturing execution system (MES)	0.26 (-0.04)	338	-
Enterprise resource planning (ERP)	0.23 (-0.01)	195	-
Product lifecycle management (PLM)	0.14 (0)	174	(CP)
Condition monitoring system (CMS)	0.10 (-0.02)	41	CP
Supply chain management (SCM)	0.09 (0.03)	51	(AL)
CP: Cloud Platform			
Cloud computing	0.66 (-0.07)	678	IIoT, (AL)
Cloud	0.28 (-0.08)	257	
OPC UA	0.27 (-0.01)	362	CAM, (IIoT)
Simulation	0.13 (0.02)	83	AM, HRS, (MS, AL)
Edge computing	0.07 (0.04)	24	AL
Platform	0.07 (0.03)	48	(HRS)
AM: Additive Manufacturing			
Additive manufacturing	0.38 (0.04)	1154	-
3D printing	0.27 (0.08)	393	-
3D printer	0.17 (0.03)	157	CAM
Modeling	0.05 (0.01)	59	CP
CAM: Computer-Aided Manufacturing			
Industrial control system (ICS)	0.16 (0)	100	CP, (IIoT)
Machine tool	0.16 (-0.03)	211	(HRS)
Control system	0.11 (0.06)	107	-
Computerized numerical control (CNC)	0.09 (-0.03)	145	-

Computer-aided design (CAD)	0.05 (0.02)	41	MS, (AM, HRS)
HRS: Human–Robot Systems			
Robotics	0.29 (0.09)	273	(AM)
Augmented reality	0.26 (0.07)	211	AM
Robot	0.21 (0.01)	294	-
Virtual reality	0.19 (0.07)	196	AL
Autonomous robots	0.15 (0.02)	52	AM, (SE, CP)
Collaborative robot/cobot	0.12 (0.03)	93	(AL)

Note: The chapter text refers to additional technology terms. EV means eigenvector, a measure of how central a node is to the overall network. The technologies are ordered by weighted degree, which indicates local importance (within a cluster), whereas EV centrality refers to a central role within the overall network. Information on related clusters includes strong and medium-strong (in parentheses) connections to other clusters.

As illustrated in Figure 2.5, the importance of the IIoT grew significantly in 2015, when it became the most important cluster. The cluster includes the overall most frequent technology terms in the network: IoT and CPS. Within this cluster, many IIoT-related terms increased in importance, including IIoT, (I)IoT device, and (I)IoT system. Some terms related to network infrastructure, such as software-defined networking, smart grid, and 5G, also experienced strong growth. Sensor systems overall grew moderately and have slightly declined since 2015–2016. Some of the fastest-growing technologies include smart or intelligent sensors, embedded sensors, human–machine interfaces, and predictive maintenance. The cluster also includes technologies that heavily rely on sensor networks, such as automated guided vehicle (AGV) and autonomous vehicles, which have been increasing in importance over time.

The second most important cluster is Algorithms. After a decline between 2011–2012 and 2015–2016, this cluster has shown a slight growth in importance. Whereas terms such as genetic algorithm and multi-agent system used to be central within the Algorithms cluster, the recent growth in cluster importance is driven by the fast growth of AI , machine learning, blockchain, and big data analytics. AI and blockchain are the fastest-growing technologies in the network. The importance of the Human–Robot Systems cluster has also increased since 2015. Augmented and virtual reality were some of the key drivers of this growth. Additionally, the importance of other terms related to robotic systems, cobots, and autonomous robots grew in recent years.

Cloud Platforms and CAM are the two clusters whose importance has most declined in the last decade. Within the Cloud Platforms cluster, the Open Platform Communications Unified Architecture (OPC UA) protocol gained the most importance. Among the smaller technologies, edge and fog computing did not play a role until 2018 but since then have shown strong growth in centrality measures. Technologies related to the CAM cluster, such as computerized numerical control (CNC), computer-aided design (CAD), and CAM, have rarely been emphasized in recent publications on Industry 4.0. Similarly, technologies in the Management Systems cluster are infrequently mentioned in recent Industry 4.0. articles. The most important technologies within the Management Systems cluster, such as manufacturing execution system (MES) and condition monitoring systems (CMS), have been declining. Finally, the importance of the Additive Manufacturing cluster also decreased during our analysis period. While key technology

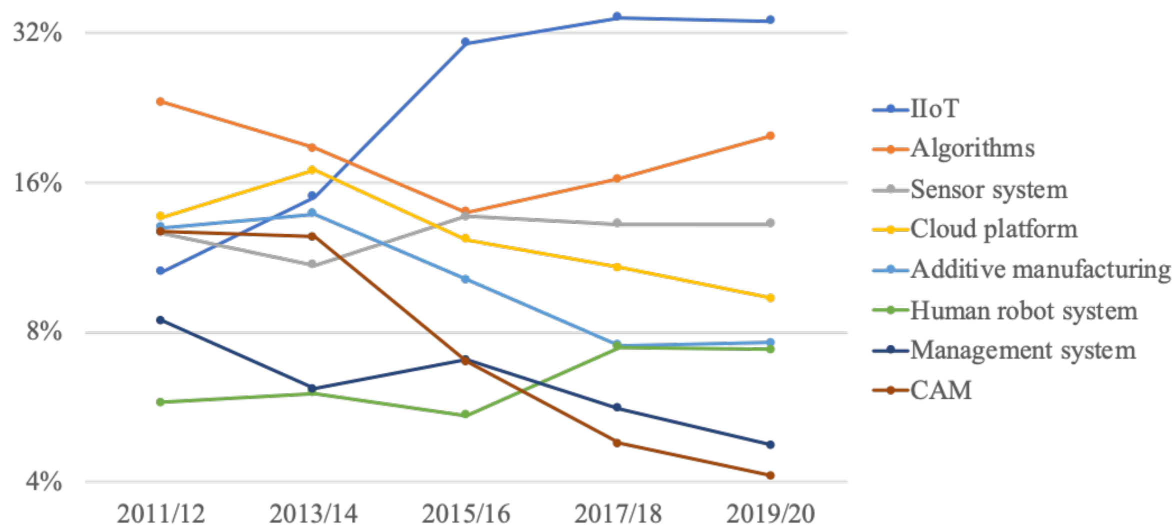


Figure 2.5: Evolution of Industry 4.0 cluster sizes since 2011. We analyzed technology networks based only on the articles within each two-year period to calculate cluster sizes. For each period, we calculated the sum of the weighted degree of all technologies per cluster. The chart indicates the share of cluster degree of the overall size of the network (sum of degrees of all nodes). IIoT refers to the Industrial Interent of Things cluster, and CAM to the computer aided manufacturing cluster.

terms such as 3D printing or additive manufacturing have recently gained importance and are part of many frameworks, the overall cluster (i.e., less relevant keywords) did not grow as much as keywords in other clusters.

The technology map (Figure 2.4) provides some insights into the relations between clusters. Generally, the closer together two clusters are positioned on the map, the stronger the connection between them. However, the map has only two dimensions and thus cannot fully reflect the connections between all clusters. Therefore, we calculated the relative comparative advantage (rca) between all clusters. Rca indicates the number of links between two clusters (i.e., whether a technology in cluster A is connected to a technology in cluster B). The rca value also normalizes for both clusters' total sizes, as there are generally more connections to larger clusters. Figure 2.6 visualizes cluster rca, with darker shades indicating a higher score. For example, even though the IIoT cluster has strong connections to all clusters by total number, its rca is high only for the Sensor Systems and Cloud Platforms clusters. This indicates a particularly strong connection to those clusters relative to IIoT, Sensor Systems, and Cloud Platform's total cluster sizes. There is high rca among Additive Manufacturing, CAM, and Human–Robot Systems, indicating a strong relationship between these clusters. Human–Robot Systems also has high rca for Sensor Systems and Algorithms and low rca for Management Systems. Finally, the Management Systems cluster has high rca for Cloud Platforms and Additive Manufacturing.

Below, we examine cluster-bridging technologies in order to better understand how clusters are connected. These technologies, which have particularly strong connections to technologies outside their own cluster, are indicated in Table 2.2 along with their corresponding clusters. RFID, smart sensors, and wireless sensor networks (all in the Sensor Systems cluster) are tech-

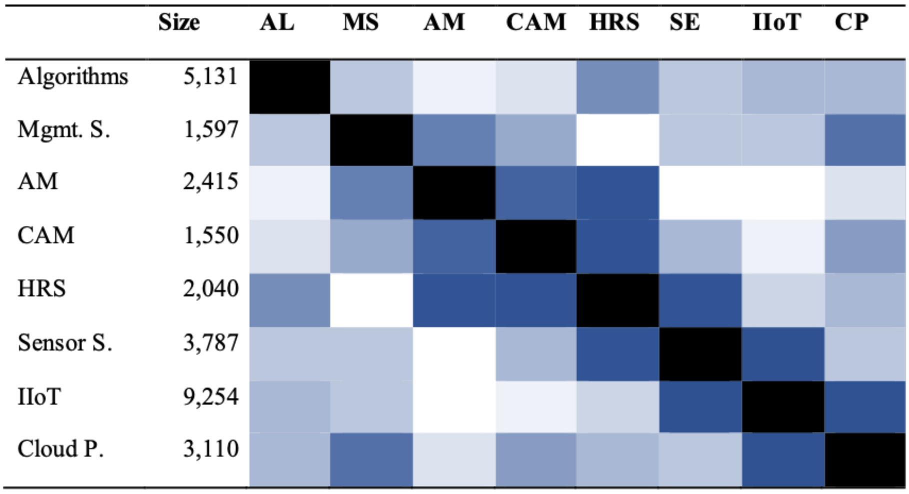


Figure 2.6: Heatmap of relations between Industry 4.0 technology clusters. Numbers indicate the total number of links (measured in weighted degree), and the color code indicates the relative comparative advantage between clusters, with darker colors indicating greater advantage (relative importance).

nologies linking IIoT and Sensor Systems. Terms related to cloud computing and network communications—for example, OPC UA (Cloud Platforms cluster), Internet, and software-defined network (SDN) (IIoT-cluster)—form some of the links between IIoT and Cloud Platforms. The link between IIoT and Algorithms are reflected in the importance of big data analytics and AI (both in the Algorithms cluster) for both clusters. The Human–Robot Systems and Additive Manufacturing clusters have several bridging technologies, such as augmented reality and autonomous robots. Human–Robot Systems also has high relevance to Algorithms; some of the relevant cluster-bridging technologies include virtual reality and cobots (both associated with Human–Robot Systems) and AI (Algorithms). Management Systems is most strongly linked to Cloud Platforms. Product lifecycle management (PLM) and CMS are some of the terms with the highest relevance for both clusters. Cloud Platforms also overlaps with Algorithms; relevant technologies comprise cloud computing, simulation, and big data analytics.

2.4.2 Comparison of Industry 4.0 concepts

To identify differences between the seven Industry 4.0 concepts included in the analysis (advanced, cloud, digital, intelligent, and smart manufacturing; Industry 4.0; Factories of the Future), we reviewed technologies associated with each concept. We evaluated each concept's technology footprint and grouped concepts based on their most important technology cluster (Figure 2.7 presents the technology footprints).

The spider charts (Figure 2.7) show, for each concept, the share of technologies per technology cluster, as described in Section 2.3.5. Industry 4.0, smart manufacturing, and Factories of the Future have a primary focus on IIoT. These concepts also have an overall similar technology footprint, with a secondary focus on the Algorithms, Sensor Systems, and Cloud Platforms clusters. CAM, Additive Manufacturing, and Management Systems play a minor role. Further, digital manufacturing and advanced manufacturing have a strong focus on Additive Manufacturing, but CAM also plays a moderately important role. Sensor Systems and IIoT play only a minor role. The third cluster, with a focus on Algorithms, comprises intelligent manufacturing and cloud manufacturing. This group's main spike is at the Algorithms cluster, but it is not as consistent as the other two groups. While cloud manufacturing has a strong secondary spike at the Cloud Platforms cluster, intelligent manufacturing relates more to CAM, Human–Robot Systems, and Sensor Systems.

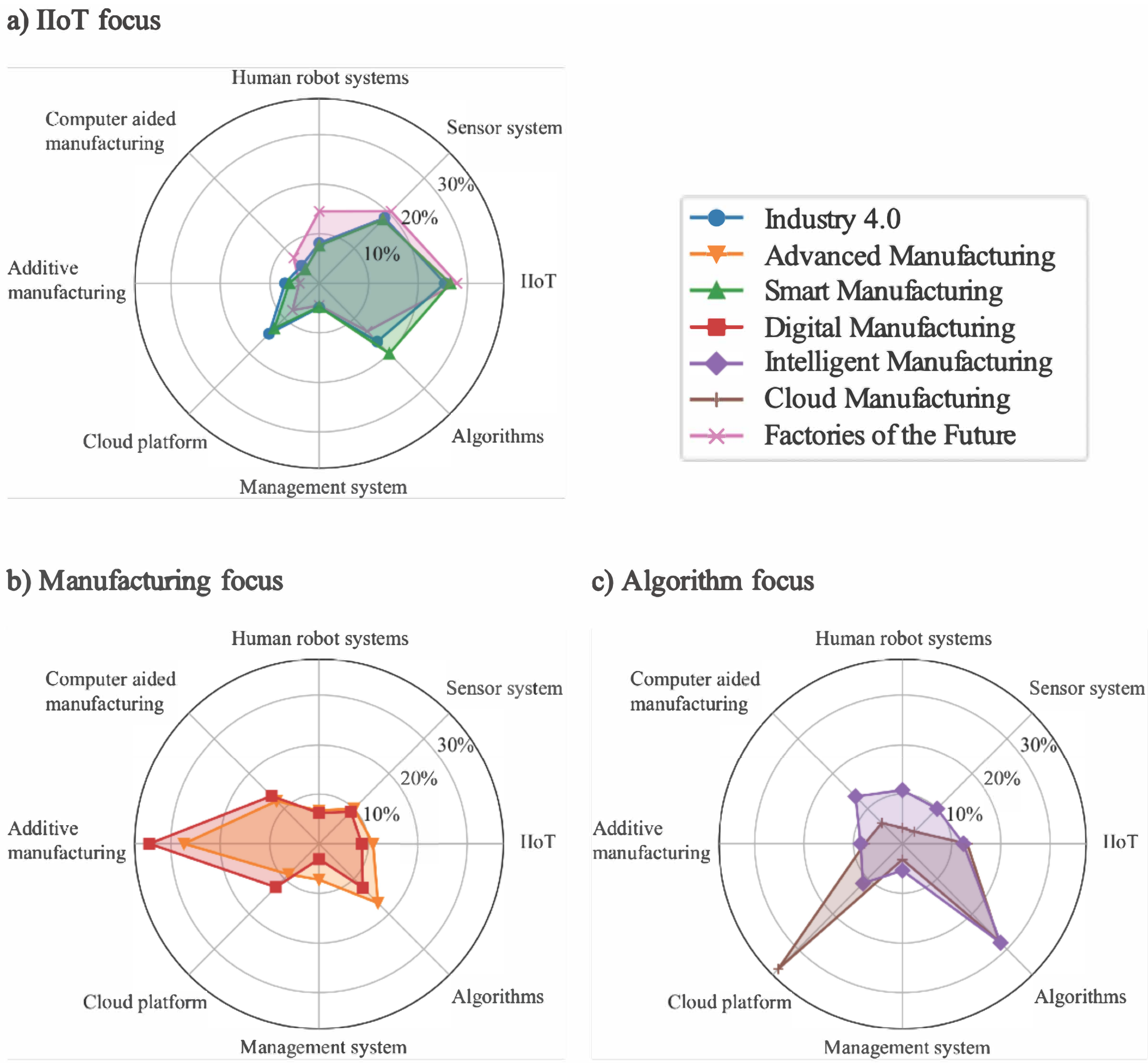


Figure 2.7: Different focus topics per Industry 4.0 concept. Each technology term is associated with a manufacturing concept. For each concept, the radar charts indicate the distribution of associated technologies across clusters. Concepts are grouped into three radar charts based on the similarity of their profiles.

In addition to the different technology focus of each concept, their use also differs by world region. Figure 2.8 shows associations between terms and countries according to the article's country of origin. Most articles originated in China, Germany, the United States, Italy, or the United Kingdom. European researchers mainly use the terms Industry 4.0 and Factories of the Future, whereas U.S. researchers generally refer to digital, smart, and advanced manufacturing. Intelligent manufacturing and cloud manufacturing are almost solely used by Chinese researchers.

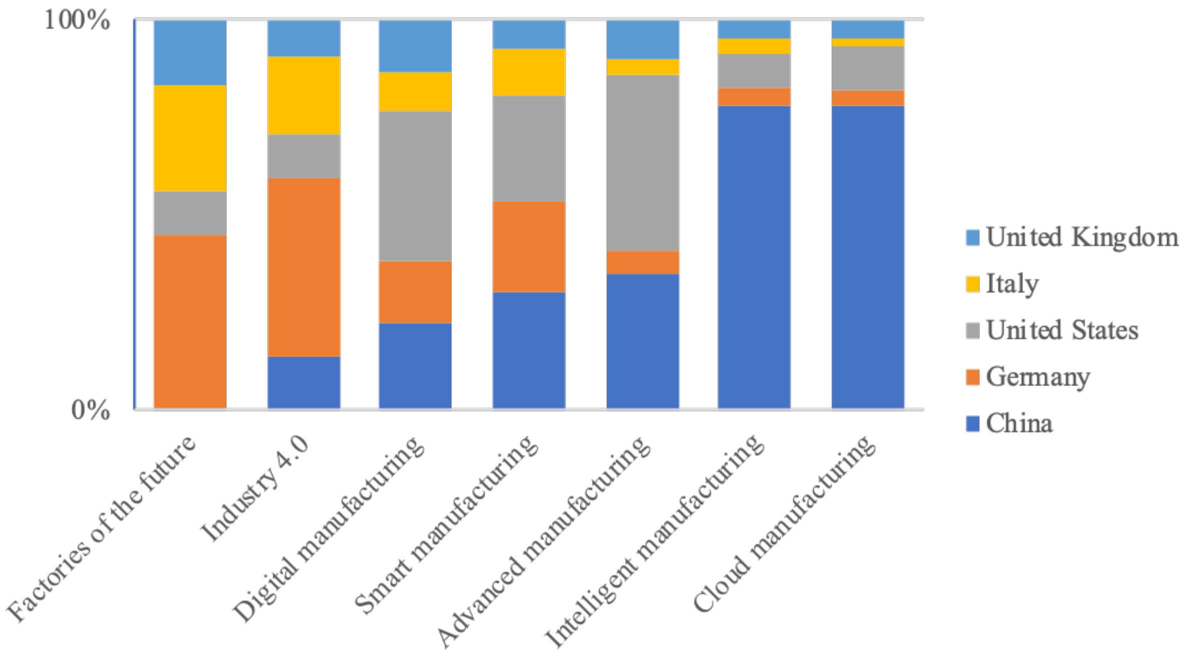


Figure 2.8: Usage of technology concepts in countries with the most research papers in the sample. Values are expressed as the percent of total mentions from selected countries.

Analysis of other countries not shown in Figure 2.8 shows similar patterns. European countries generally focus on IIoT-related topics (e.g., Spain, France, and Portugal have high shares of articles associated with Industry 4.0 and Factories of the Future). Swedish researchers additionally have a strong focus on cloud manufacturing. Other significant contributors to the literature in this analysis generally focus on Industry 4.0, with the exception of South Korea, which mostly addresses smart manufacturing. Research from Taiwan has a relatively high share of articles on smart, intelligent, and cloud manufacturing. Indian researchers have made relatively strong contributions to the advanced manufacturing literature, and Japanese research has a relatively strong focus on smart and digital manufacturing.

2.5 Discussion

We discuss the evolution of the Industry 4.0 landscape and its various concepts, building on technology clusters as defined in Section 2.4.1. These clusters summarize several closely related technologies and help to discuss the technology landscape. Clusters provide suggestions for interpretation, not definite constructs, and different cluster analysis parameters can lead to different numbers of clusters. The present eight clusters offer a level of detail suitable for discussion and lead to homogeneous clusters. Beyond these clusters, our dataset also offers complementary, more detailed insights at the level of individual technologies.

2.5.1 Technology trends

In 2011, Algorithms, Cloud Platforms, and CAM were the most important clusters in the technology landscape. Together with Sensor Systems, Additive Manufacturing, and Management Systems, IIoT was among the less important technologies. Human–Robot Systems did not play a significant role. Our results show that a new paradigm evolved in 2015–2016, when there was a major increase in articles on IIoT-related concepts (Industry 4.0, smart manufacturing, Factories of the Future). IIoT became by far the most important cluster, followed by Algorithms. Management Systems, CAM, and Additive Manufacturing played only minor roles, with each representing less than 6% of technologies within the landscape. With most articles currently related to Industry 4.0 and smart manufacturing, it appears that cluster importance trends have stabilized since 2015–2016.

The IIoT cluster, including CPS, is the largest cluster and forms the core of the Industry 4.0 landscape. The cluster's central position in the technology map indicates its overall relevance (in terms of frequency) and its relatedness to the other technology clusters. IIoT has a particular strong connection to the Algorithms, Cloud Platforms, and Sensor Systems clusters. This is in line with literature that describes the IoT and CPS as the core of Industry 4.0 (Pereira & Romero, 2017). The Industry 4.0 working group (Kagermann et al., 2013) stated that "the fourth industrial revolution [is] based on Cyber-Physical systems" (p. 13), and Roblek et al. (2016) described the IoT as "central to the new industrial revolution" (p. 1). Zheng et al. (2020) found that the IoT was the most important enabling technology of Industry 4.0. Chiarello et al. (2018) and Rüßmann et al. (2015) defined the IIoT as an important element of Industry 4.0, but as one of several pillars rather than the core pillar. Our analysis shows that the term IIoT has recently become more central. Sisinni et al. (2018) described the IIoT as the part of the IoT that focuses on manufacturing, stating that what is "usually addressed as IoT, could be better named as consumer IoT, as opposed to IIoT" (p. 4724), which focuses on manufacturing. The rising importance of IIoT within the technology network may reflect the understanding that, in the manufacturing context, IoT is referred to as IIoT, and future research should increase awareness of this differentiation.

The CAM cluster is the most declining cluster within the technology landscape. While the importance of CAD is growing, the decline of the cluster is driven by lower relevance of terms such as CNC, machine tool, and CAM. Although CNC machines may be a part of the future manufacturing landscape, research no longer explicitly addresses it. Culot et al. (2020) described CAD and CAM as "old" technology and did not include it as a dimension in their review. The CAM cluster also has a high rca to the Management Systems cluster, which likewise contains "old" technologies, such as enterprise information system and ERP (Culot et al., 2020). Finally, this cluster has the strongest links to the Additive Manufacturing cluster, which is also a production-related cluster.

The Management Systems cluster—including MES, PLM, and CMS—has been declining in importance within the Industry 4.0 landscape. Technologies in this cluster are included in some Industry 4.0 frameworks (Dalenogare et al., 2018; Lasi et al., 2014), but many frameworks do not consider them as key technologies (Chiarello et al., 2018; Culot et al., 2020; Ghobakhloo, 2018; Zhong et al., 2017). Supply chain management is a growing technology term in this cluster, and

ERP is only slightly declining. This suggests the growing importance of technologies that go beyond factory boundaries, whereas technologies relating to the factory itself (CMS, MES) are declining in importance. Yu et al. (2015) described integration beyond company boundaries as the particular focus of cloud manufacturing. However, integration also plays an important role in other concepts that describe, for example, horizontal and vertical integration as a key pillar of Industry 4.0 (Frank et al., 2019a; Rüßmann et al., 2015). Saucedo-Martínez et al. (2018) even described it as the most important technology field of Industry 4.0.

The Cloud Platforms cluster has slightly decreased in importance within the manufacturing landscape but remains important. This role is in line with the literature, which frequently describes it as a core Industry 4.0 technology (Frank et al., 2019a; Zheng et al., 2020). At the same time, the slight decline in importance may be due to a decreasing share of articles related to concepts focusing on cloud computing (especially cloud manufacturing). Another reason for this slight decline is that many cloud technologies have become standard and therefore less research is focused on these technologies. For instance, Liao et al. (2017) noted that the OPC UA standard (a technology in the Cloud Platforms cluster) has become standard in machine-to-machine communications. The strong connection of the Cloud Platforms cluster to the IIoT cluster underlines the high interdependency of the two. Cloud Platforms focuses on cloud computing, protocols, and platforms, which are an integral part of the IIoT, as they are related to IoT infrastructure, CPS, and devices. Boyes et al. (2018), for example, described cloud computing as an (optional) part of the IIoT.

The literature increasingly highlights challenges related to cloud computing systems in the context of the IIoT. The challenges of large, highly connected, and centrally controlled networks include high central data accumulation, reliability issues, and high latency (Pan & McElhannon, 2018; Wang et al., 2020). Some of these challenges may be overcome by decentralizing some cloud capabilities into local data centers and intelligent devices (Georgakopoulos et al., 2016). Alcácer & Cruz-Machado (2019) suggested that Industry 4.0 “will be the extinction of the centralized applications used in common manufacturing environments” (p. 915), and Nakayama et al. (2020) described this decentralization as the main difference between Industry 3.0 and 4.0. These decentralized systems are called edge or fog computing—terms that have recently grown significantly in the technology map but are still missing from many frameworks and play only a minor role in the Industry 4.0 literature (Culot et al., 2020; Kipper et al., 2020; Liao et al., 2017; Muhuri et al., 2019; Osterrieder et al., 2020; Rüßmann et al., 2015). However, some terms that also describe distributed intelligence—such as smart sensor, AGV, and autonomous robot—are more present in the literature. Future frameworks could emphasize the distributed nature of networks to facilitate overcoming challenges of network complexity.

Blockchain is among the fastest-growing technologies in our analysis. The technology offers potential benefits related to security, privacy, resilience, and reliability in an increasingly interconnected manufacturing landscape (Lee et al., 2019). Its potential benefits are particularly relevant for decentralized (edge or fog) systems—for example, enabling reliable communication between devices without the need to use a central server for security or reliability. While many Industry 4.0 frameworks (Chen, 2017; Frank et al., 2019a; Kagermann et al., 2013; Rüßmann et al., 2015) did not refer to blockchain technology, future frameworks should follow some recent articles (Bai et al., 2020; Gaiardelli et al., 2021; Zheng et al., 2020) and account for its growing importance within their frameworks.

In general, the Algorithms cluster has been an important element of the manufacturing landscape throughout the timeline of this analysis. Traditional analytics-related terms—such as analytics, genetic algorithm, or multi-agent system—declined in importance in recent years. In contrast, AI , machine learning, and big data analytics are among the fastest-growing terms. While AI already has a high centrality score, its recent strong growth in centrality suggests that it is becoming an increasingly central element of the manufacturing landscape. On the one hand, generally increasing awareness of AI could account for the strong recent AI growth (Frank et al., 2019c). On the other hand, this growth might reflect the increasing number of feasible AI use cases. Chui et al. (2018), for example, identified hundreds of AI use cases already implemented in companies around the world. The share of articles related to algorithm-focused frameworks (intelligent manufacturing, cloud manufacturing) has recently remained constant. This indicates the growing importance of the Algorithms cluster, particularly AI , driven by the growing presence of the topic across all Industry 4.0 concepts. If this trend continues, AI may become the core element of future Industry 4.0 landscapes.

The Human–Robot Systems cluster has been the fastest-growing cluster aside from IIoT. This cluster describes an environment of intelligent robots, which are frequently described in the literature as using sensors to interact with users (Robla-Gomez et al., 2017) and learning from humans through gestures or speech (Du et al., 2018). It is closely linked to Sensor Systems, IIoT, and Algorithms. One of the fastest-growing technologies in this cluster is augmented reality, which can support human workers in future workstations by overlaying machine information with the real-world environment. This can be used to provide assembly instructions for customized products (Mourtzis et al., 2019), train employees (Longo et al., 2017), support teleoperated industrial assembly tasks (Brizzi et al., 2018), or conduct quality control of manufactured parts (Butt, 2020). Through its various capabilities, augmented reality may become a key enabling technology for augmenting workers in a future smart workplace (Frank et al., 2019a; Longo et al., 2017) and experience strong growth in importance in the future (Masood & Egger, 2019).

The importance of the Additive Manufacturing cluster within the Industry 4.0 landscape has been declining since 2011 and currently represents a small share of keywords. This low share reflects the literature, where many articles do not consider additive manufacturing a key Industry 4.0 technology (Hermann et al., 2016; Strozzi et al., 2017). While Liao et al. (2017) and Muhuri et al. (2019) described additive manufacturing as a key Industry 4.0 technology, their reviews did not identify it as one of the most important terms. Similarly, Zheng et al. (2020) defined it as a key technology in their review but found that it played a role in few Industry 4.0 research articles. At the same time, additive manufacturing is included in many other frameworks (Culot et al., 2020; Frank et al., 2019d) and is considered a critical aspect of Industry 4.0 (Kumar, 2018), with an impact across the product lifecycle (Butt, 2020). Two aspects may drive this decline in importance. First, the share of frameworks with a focus on the Additive Manufacturing cluster, such as digital and advanced manufacturing, is declining and thus related terms are declining as well. Second, additive manufacturing was present in the literature before the Industry 4.0 concept was introduced. After the introduction of this concept, IIoT, sensor systems, and human–robot systems became much more relevant in manufacturing research, and consequently the share of articles on additive manufacturing decreased.

Additive Manufacturing has strong connections to the Human–Robot Systems cluster. Some important bridging technologies are robotics and autonomous robots. These links suggest a vision wherein additive manufacturing is fully integrated with the manufacturing environment. This could lower complexity in production and decentralize production (Mehrpouya et al., 2019). However, the application of additive manufacturing in mass production remains limited in the near term (Roca et al., 2017), particularly due to its low throughput (Korner et al., 2020; Mehrpouya et al., 2019). Dalenogare et al. (2018) found that companies currently see the benefits of additive manufacturing in terms of product-related (development, lead time, customization) rather than production-related aspects (costs, productivity, process control). Rapid prototyping has already been established in many companies (Mehrpouya et al., 2019), which is reflected in our analysis, where CAD (in the CAM cluster) has strong links to the Additive Manufacturing cluster. Augmented reality, another strong link between Human–Robot Systems and Additive Manufacturing, can also contribute to the product design process (Butt, 2020). With augmented reality and additive manufacturing, objects can be visualized in a real-life setting and directly manufactured (Mourtzis et al., 2015). The high customization and flexibility of additive manufacturing can be valuable for maintenance activities (Butt, 2020). Ceruti et al. (2019) describe a use case for additive manufacturing wherein it is used to produce spare parts in aerospace, with augmented reality facilitating maintenance work.

2.5.2 Industry 4.0 and other manufacturing concepts

The research focus on IIoT is in line with the quick adoption of the Industry 4.0 concept after its introduction in 2011. Research on Industry 4.0 grew more substantially compared with other concepts (e.g., advanced, digital, and cloud manufacturing), and since 2015, it has been the subject of the most articles within our research space. Until 2019, the share of articles related to Industry 4.0 steadily increased. During this period, research attention shifted towards IoT in manufacturing, not only in IoT-focused frameworks (Industry 4.0, Factory of the Future, and smart manufacturing) but also in other frameworks. Lu & Cecil (2016), for example, described an IoT framework for the advanced manufacturing domain. Industry 4.0 and smart manufacturing have very similar technology footprints, focused on IIoT. This is in line with Frank et al. (2019a), who described smart manufacturing as a sub-concept within Industry 4.0. In some cases, Industry 4.0 and smart manufacturing are used interchangeably, and some authors have reported that Industry 4.0 is called smart manufacturing in the United States (Mittal et al., 2018; Thoben et al., 2017).

Intelligent manufacturing has a strong focus on algorithms. Zhong et al. (2017) describe intelligent manufacturing as an aspect of Industry 4.0 that focuses on intelligent objects, using data analytics for AI -based decision-making. Cloud manufacturing also has a strong focus on algorithms (aside from cloud computing), though the role of algorithms and AI is not as strongly highlighted in cloud manufacturing frameworks compared to intelligent manufacturing (Culot et al., 2020; Xu, 2012; Yu et al., 2015; Zhong et al., 2017). The intelligent manufacturing concept has mainly been used by Chinese authors, which could reflect the country's strong

AI capabilities and the importance of AI in the Chinese Manufacturing 2025 plan (Li, 2018; Zhou et al., 2018). Zhou et al. (2018), for example, suggested that their concept of intelligent manufacturing integrated AI and advanced manufacturing.

Overall, the European-driven concepts of Factories of the Future and Industry 4.0—together with smart manufacturing, which is more dominant in the United States—establish a focus on IIoT in manufacturing research. Recent trends indicate a shift towards AI in the Industry 4.0 landscape. While Chinese researchers have already worked more on AI focused concepts, European and U.S. frameworks may need to further evolve to reflect this trend. Articles introducing Industry 4.0 (Kagermann et al., 2013) and the Factories of the Future roadmap (European Commission, 2013), for example, do not refer directly to AI or machine learning at all. This early focus on IIoT may be one reason AI is still underrepresented within these frameworks.

While most frameworks have a strong focus on IT, the profiles of digital and advanced manufacturing are both related to production technology, focusing on additive manufacturing. The term advanced manufacturing has been used for decades to describe (for example) CAD, CAM, and robotics (Boyer et al., 1997). Rather than a fixed framework, the terms evolved with changing manufacturing technology and were later also used in the context of the IoT (Lu & Cecil, 2016). In addition, digital manufacturing has been framed as a product-oriented framework, focusing on seamless digital processes from design to manufacture (Chryssolouris et al., 2009). Recently, this framework has often been linked to additive manufacturing and referred to as direct digital manufacturing, focusing on a process that directly produces parts based on 3D models without the need for production process planning (Chen et al., 2015; Paritala et al., 2017).

2.6 Conclusions

Over the last 10 years, Industry 4.0 has evolved as a key research topic in management, production, and operations research, but it is still not clearly defined. In this paper, we clarified the concept and boundaries of Industry 4.0, identified the most relevant technology trends, and analyzed its evolution. By creating a technology map, we built on network analysis to isolate technology clusters and identify relations between them. In addition, we compared Industry 4.0 with six related concepts—advanced manufacturing, cloud manufacturing, digital manufacturing, Factories of the Future, intelligent manufacturing, and smart manufacturing—and identified the main differences between them. We use NLP and network analysis to review the technology trends described in more than 14,000 articles referring to Industry 4.0 and related concepts.

With the introduction of Industry 4.0, the IIoT has become a central aspect of the technology map. The IIoT reaches beyond company boundaries, thereby minimizing the role of local production management and control systems. A key aspect of the IIoT is the cloud. Due to the complexity of highly interconnected networks, distributed systems like edge computing and smart sensors and devices will play more important roles in the future. Blockchain technology may be key in enabling the reliability of these systems. These decentralized connected systems enable faster adoption of AI in manufacturing, which has been strongly growing in importance in the overall technology landscape during recent years. These intelligent systems will facilitate human–robot systems, including augmented reality, enabling workers to interact with robots naturally. Due to its high potential to simplify and decentralize production, additive manufacturing may also evolve in relevance from product development, small batch production, and maintenance to mainstream production once productivity further improves.

Industry 4.0, as used in the literature, is frequently still defined around the IIoT and CPS. We suggest accounting for the increasingly central role of AI when defining the fourth industrial revolution. The Industry 4.0 working group (Kagermann et al., 2013) suggested that the fourth industrial revolution is "based on Cyber-Physical Systems" (p. 13). This is a narrow definition compared to how the group describe previous industrial revolutions. For example, they state that the second industrial revolution "follows [the] introduction of electrically-powered mass production based on the division of labor" (p. 13). This description does not limit the revolution to the conveyor belt (a key component of that revolution) but rather its broader—revolutionary—impact: the division of labor. Similarly, the fourth industrial revolution should not be narrowly conceived in terms of IoT-based factories. Instead, a definition could recognize the revolutionary use of AI as a core driver and more natural human–machine interactions as a new way of working across enterprise boundaries and along the product life cycle.

In addition to providing a better understanding of Industry 4.0, this work contributes to the general scientific literature by presenting a new approach for literature reviews. We extracted named entities from article abstracts using NLP, a method that allowed us to evaluate technology terms without relying on article keywords. Future researchers can apply these methods to different fields, and named entity recognition may be trained to recognize not only technologies but also any other concept, such as technical information, chemical formulas, and historical events. We then analyzed the technologies using network analysis. This analysis offers insights into technology clusters, relations, and trends, rather than looking only at word frequencies. While network analysis is commonly used to describe social interactions (social network analysis), we show that this method offers high potential for other fields, such as operations research.

Our work has some limitations and offers potential for future researchers to develop our approach. Our analysis describes the evolution of Industry 4.0 and related concepts. Future work could use additional databases to confirm these trends and gain additional insights. Including Clarivate Web of Science would add other articles to the analysis. Using other types of data, such as Industry 4.0–related patents, could also lead to additional insights and allow for additional analysis of the Industry 4.0 technology landscape (e.g., comparing trends in patents and scientific literature). Our machine learning–based NLP approach offers great flexibility and helped us identify a large range of technologies. With the field of NLP rapidly improving, future researchers may achieve even higher accuracy in identifying technology terms with novel algorithms and future releases of the tools used in our work. Finally, future researchers might build on the method introduced in this article to undertake additional analyses, such as reviewing the business impact of Industry 4.0 or conducting a long-term review of technology concepts and shifts in the landscape. This could further illuminate technological change and create early indicators for future technological shifts.

Chapter 3

Evolution of Industry 4.0 research[1]

This chapter describes a machine learning-supported literature review and shows how the fast growing Industry 4.0 literature has evolved, in order to propose future research opportunities.

3.1 Introduction

Since the term Industry 4.0 was coined in 2011 (Liao et al., 2017), a growing number of studies from different streams of research has been published on this concept (Culot et al., 2020). The literature has devoted special attention to the digital transformation of industries and companies towards an "Industry 4.0" level (Dalenogare et al., 2018). The digital transformation process has been supported by the implementation of four base technologies: the Internet of Things (IoT), cloud computing, big data, and artificial intelligence (AI) (Frank et al., 2019a). These base technologies support the application of several front-end technologies, including product design systems, simulation, augmented and virtual reality, additive manufacturing, and advanced robotics (Dalenogare et al., 2018; Liao et al., 2017). Although Industry 4.0 has started as an industrial policy platform (e.g., Reischauer, 2018; Schwab, 2016), the operations management literature has embraced this topic by considering technologies that can be practically implemented based on maturity models (Mittal et al., 2018) to create different solutions and applications (Benitez et al., 2020; Frank et al., 2019b; Xu et al., 2018). Therefore, Industry 4.0 is considered today a new maturity stage of manufacturing companies in which the combination of advanced technologies – supported by IoT, cloud computing, big data, and AI – allows for the creation of cyber-physical systems (CPS), providing an interconnected level of the company with more deeply integrated processes (Benitez et al., 2020) The use of Industry 4.0-related technologies may have different goals, such as: increasing production efficiency, productivity, and quality; augmenting operational flexibility; integrating the production system with customers and the supply chain; or contributing to workers' safety and operational sustainability (Schuh et al., 2020; Dalenogare et al., 2018). Therefore, Industry 4.0 comprises a set of IoT-driven technologies that may be arranged in different solutions according to the manufacturing goals pursued

[1]This chapter largely overlaps with content from the following article: Meindl, B., Ayala, N. F., Mendonça, J., & Frank, A. G. (2021). The four smarts of Industry 4.0: Evolution of ten years of research and future perspectives. Technological Forecasting and Social Change, 168, 120784

(Benitez et al., 2021).

The literature on Industry 4.0 has mainly focused on the changes in manufacturing systems to create these CPS (Kagermann et al., 2013; Kipper et al., 2020). Such stream has acknowledged Smart Manufacturing as the core of Industry 4.0 (Bueno et al., 2020; Culot et al., 2020; Kipper et al., 2020). However, as the studies on Industry 4.0 evolve, new dimensions related to manufacturing activities have emerged and been integrated, especially considering larger enterprise systems. New disciplines have drawn attention to this issue, including product development (Riel et al., 2017), services (Frank et al., 2019b), ergonomics (Mansfield et al., 2020), and supply chain (Fatorachian & Kazemi, 2020). These different fields of study and application of digital technologies in the Industry 4.0 domain have been summarized by Frank et al. (2019a) in what they called the "Four Smarts model of Industry 4.0," which describes the integration of four dimensions: Smart Manufacturing, Smart Products and Services, Smart Supply Chain, and Smart Working. In this context, Industry 4.0 provides a broad perspective that integrates several domains which converge in the manufacturing system. Smart Manufacturing is still the common root of Industry 4.0. However, new fields of study provide insights contributing to our understanding of the potential of IoT-based solutions for enterprises and entire value chain systems (Bueno et al., 2020).

Although new fields of application have arisen from Industry 4.0 spanning the boundaries of new disciplines, several problems have emerged in the literature, as often is the case with multidisciplinary topics. Firstly, the concept has been studied within the limits of different research fields, which has led to the development of knowledge silos that need to be better integrated (Culot et al., 2020). Operations management, technology management, information systems, innovation management, and industrial policy are some of the fields that have addressed the topic through very different prisms (Liao et al., 2017). This can also lead to unaddressed intersections between topics and, consequently, substantive research gaps. Moreover, this creates a fragmented and sometimes disconnected view of Industry 4.0. For instance: different terms, technologies, and fields of application are proposed for the same concept (Culot et al., 2020). Therefore, as the Industry 4.0 concept completes a decade of existence (2011-2021), it is paramount to seek a clear understanding of its evolution and future avenues in order to guide future research efforts in the different disciplines exploring its potential. Therefore, we use the four smarts perspective, which considers the relationship between Smart Manufacturing, Smart Supply Chain, Smart Products and Services, and Smart Working (Frank et al., 2019a), to analyze the different streams related to the Industry 4.0 concept and provide a picture of research evolution and existing gaps.

Thus, the main objective of this paper is to understand how the Industry 4.0 literature has evolved regarding the four smart dimensions since the term Industry 4.0 was conceived and propose future research opportunities based on an integrative perspective on this field. We aim to analyze the evolution of each of the four smart dimensions, the intersections between them over the last ten years, the key concepts and technologies addressed in each of these dimensions, and how different journals emphasize one or more of these dimensions to the detriment of the others. This will provide an understanding on journal profiles and underexplored fields and topics. We use a Machine Learning-based systematic literature review (SLR) to analyze the chronological evolution during the ten years of research on Industry 4.0 regarding the four smart dimensions, i.e., Smart Manufacturing, Smart Supply Chain, Smart Working and Smart Products and Services, as well as to assess the level of integration between these dimensions. Using a Machine Learning-based SLR, we reviewed 4,973 Industry 4.0-related research papers published between 2011 and 2020 in the leading scientific databases. We conducted a chronological network analysis considering the growth of these four dimensions and of the connections between them, showing how the four smarts of Industry 4.0 have evolved along a decade of research.

Our results show that research started very fragmented, distributed in different disciplines, and highly concentrated around Smart Manufacturing. We show that studies on Industry 4.0 have recently shifted attention towards a mor fe integrative viewpoint, although more research from this perspective is still necessary. We also show that Smart Working is the least explored dimension, with many opportunities for future research. Our findings support the vision of Industry 4.0 as a concept spanning the Smart Manufacturing field, creating opportunities for synergies with other research domains. We also provide a supplementary dataset of the studies analyzed in this paper (4,973 papers), categorized according to the different metrics assessed in this paper, for use in future research.

3.2 The four smart dimensions of Industry 4.0

Since the internet was popularized in the end of the 20th century, the information revolution has impressively boomed and expanded to a new age, the so-called digital age (Brynjolfsson & Mcafee, 2014). Digital transformation has been one of the core issues in industrialized countries (Brynjolfsson & Mcafee, 2014). In this context, the Industry 4.0 concept was coined in 2011 by a German public-private initiative to acknowledge the industrial challenges in this new age and propose a strategic program to develop advanced production systems for German companies (Kagermann et al., 2013). Thenceforth, the concept has spread worldwide, although some countries use different names and lay different emphasis on their industrial policies (Culot et al., 2020).

Several models have been proposed to describe Industry 4.0 and its application. Most of them have a maturity evolution outlook, describing how the implementation of technologies should happen. On the industry side, several models can be found, such as the German Academy of Science and Engineering (ACATECH) Industrie 4.0 Maturity Index created by the German National Academy of Science and Engineering (Schuh et al., 2020), and the Reference Architecture Model Industrie 4.0 – RAMI 4.0, created by the Platform Industrie 4.0 (Hankel & Rexroth, 2015). On the academic side, some models have also been proposed to describe the implementation of this concept, including models for small and medium-sized enterprises (Mittal et al., 2018), models for assessing Industry 4.0 readiness (Schumacher et al., 2016), and models for digital technology roadmap (Sjödin et al., 2018). In the academic literature, the model proposed by Frank et al. (2019a) (Figure 3.1) is one of the most often referenced. This model discriminates between the base and front-end technologies of Industry 4.0, which provides a clearer understanding of the technologies contributing for general purposes and of those oriented to specific activities in the production system. Moreover, this model introduces a broader perspective on Industry 4.0, beyond the manufacturing system. Frank et al. (2019a) proposed different

Industry 4.0 technology application dimensions, all of them related and connected to the manufacturing system as the core of the industrial activity. Thus, we selected this model because the different dimensions it describes can help us explore different streams in the literature. For instance, Frank et al. (2019a) demonstrate that Smart Supply Chains and Smart Products and Services are part of Industry 4.0. This broader view allows us to consider the supply chain and product development literature that has addressed this topic. In this sense, while most other models are concerned about when each level of Industry 4.0 should be implemented, the model by Frank et al. (2019a) emphasizes what should be implemented in terms of technologies and practices in the different dimensions considered, and this is particularly useful to explore the diverse knowledge domains in such a diffuse and rapidly growing literature.

The model proposed by Frank et al. (2019a) is represented in Figure 3.1. The model proposed by these authors has an empirical background. The authors studied different applications of Industry 4.0 reported in the literature and then analyzed and clustered technologies used in companies to implement Industry 4.0 concepts. As the figure shows, Industry 4.0 technologies can be organized in two main levels: the base technology level, and the front-end technology level. Base technologies boost digital transformation in each enterprise dimension and differentiate what Industry 4.0 is regarding previous stages of industrial development. Base technologies comprise the use of IoT, cloud computing, big data, and analytics (including data mining and AI tools). Other works have adopted similar perspectives on the core Industry 4.0 technologies (e.g., Thoben et al., 2017; Zhong et al., 2017. Base technologies support the transformation of a conventional enterprise – where different dimensions are not integrated – into a Smart Enterprise, where the different dimensions are optimally interconnected on an Industry 4.0 level.

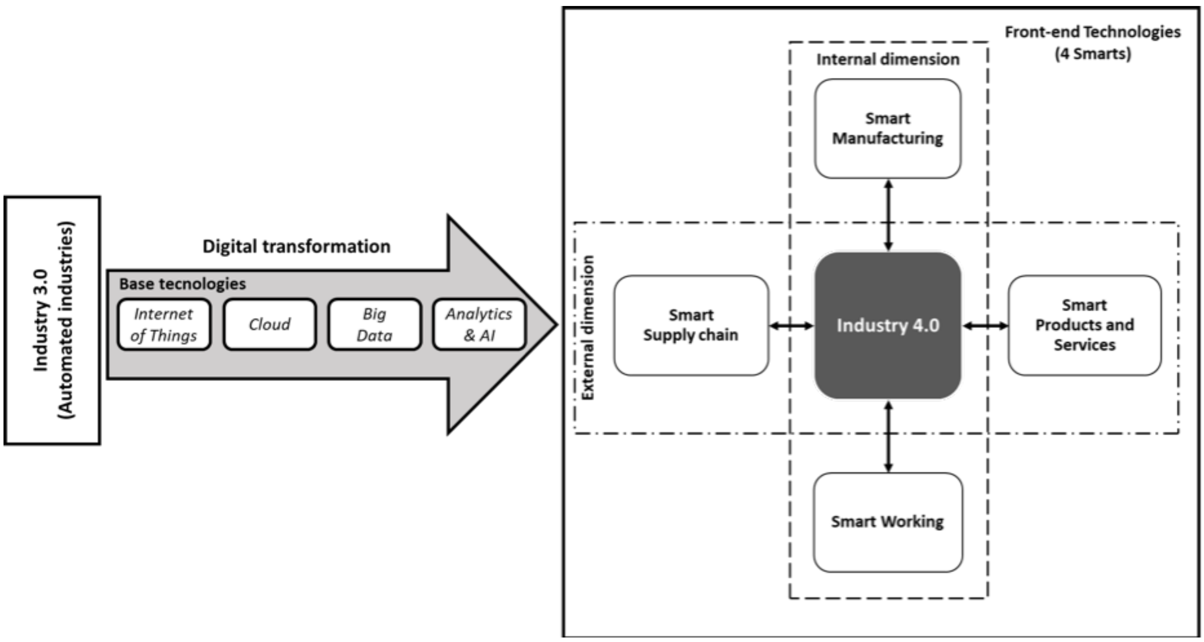


Figure 3.1: Conceptual model of Digital Transformation and the four smarts of Industry 4.0. Adapted from Frank et al. (2019a).

Front-end dimensions (Figure 3.1) comprise a smart enterprise's technologies for specific purposes within and beyond its frontiers. On the one hand, the internal dimensions consider value streams focused on the company's industrial activities: its production processes (Smart Manufacturing) and its workers (Smart Working). The external dimensions, on the other hand, consider value streams that integrate the company's processes with the external environment: its supply chain (Smart Supply Chain) and its customers (Smart Products and Services). These four smart dimensions encompass the full potential of Industry 4.0 technology application, since they cover the main aspects presented by previous frameworks by authors such as Chen (2017), Chiarello et al. (2018) and Roblek et al. (2016). We provide the theoretical background for each of these dimensions and use them as a framework to conduct the SLR. In the following sections, we detail the four smart dimensions.

3.2.1 Smart Manufacturing

The first internal dimension (Figure 3.1), Smart Manufacturing, has been related to the Industry 4.0 concept since the very beginning(Kagermann et al., 2013), with many studies considering them as synonyms. However, with Industry 4.0 describing a broader perspective of the company and the industry, Smart Manufacturing constitutes the core dimension, but not the only one (Schuh et al., 2020). A formal definition is given by (Kusiak, 2018, p.509): "Smart Manufacturing integrates manufacturing assets of today and tomorrow with sensors, computing platforms, communication technology, data-intensive modeling, control, simulation, and predictive engineering. Smart manufacturing utilizes the concepts of the cyber-physical systems, internet of things (and everything), cloud computing, service-oriented computing, AI, and data science." As this definition shows, Smart Manufacturing comprises the use of Industry 4.0 base technologies (IoT, cloud, big data, and AI) on the shop floor to reach cyber-physical manufacturing systems (Tao et al., 2018) and smart production planning and control (Bueno et al., 2020). In this sense, Industry 4.0 technologies such as machine-to-machine communication, vertical integration of information systems, advanced robotics – including collaborative robotics –, are some technologies that can be considered part of the Smart Manufacturing dimension (Dalenogare et al., 2018; Wang et al., 2016). Such technologies also support the production planning and control process using AI and real-time data to better organize the manufacturing activity (Bueno et al., 2020). Smart Manufacturing also considers smart maintenance based on AI to predict potential failures and anticipate equipment shutdowns (Bokrantz et al., 2020).

Besides considering the technologies used to manufacture products, Smart Manufacturing also includes technologies for other activities in the manufacturing process. Advanced technologies to better manage energy consumption are also an essential aspect of Smart Manufacturing (Kusiak, 2018). It also considers product design technologies (sometimes called "smart design") used to meet customer requirements and increase manufacturing effectiveness. Such technologies include virtual and augmented reality for product design and manufacturing assembly, advanced CAD/CAE tools such as generative design, 3D prototyping, and product lifecycle management systems, among others (Dalenogare et al., 2018). Thus, Smart Manufacturing considers the end-to-end engineering principle of Industry 4.0, where engineering design is integrated with the manufacturing system to work as a single mechanism in the production system (Dalenogare et al., 2018).

3.2.2 Smart Working

The second internal dimension, Smart Working – sometimes also called Smart Work – considers the way technologies are used to support workers in a company's activities (Figure 3.1). It acknowledges that workers play a critical strategic role in manufacturing activities and that they should be enhanced rather than replaced (Kaasinen et al., 2020). Recently there has been much debate about the human role in the Industry 4.0 context, and some studies have proposed a new worker profile called the "Operator 4.0" or "Smart Operator" (Romero et al., 2016, 2020; Cimini et al., 2020). While several studies point out that autonomous machines may replace operational and low value-added activities, the most significant potential of Industry 4.0 is to provide support for workers (operators as well as other hierarchical levels) to perform their work smarter. Such work is based on the human cognitive capacity to add value to the production system (Cohen et al., 2019; Fantini et al., 2018). Therefore, Smart Working considers how to take the best from workers' potential by using advanced technologies to support decision-making processes (Segura et al., 2020; Zolotová et al., 2020), manage knowledge (Kaasinen et al., 2020; Pinzone et al., 2020), foster creativity and design (Fantini et al., 2018), and increase workers' safety and satisfaction (Fletcher et al., 2020). In this sense, following Frank et al. (2019a), we adopt the "Smart Working" terminology to consider both the operational activities performed by smart operators and the flexible and remote activities involving a broader scope of workers including managers, engineers and supervisors, who perform the cognitive activities of the manufacturing processes.

Different technologies have been described in the literature to enhance and empower workers Frank et al. (2019a). Virtual reality enables the safe use of hazardous equipment and enhanced learning of procedures, while augmented reality augments the workplace with relevant information useful for the execution of tasks (Segura et al., 2020). AI allows managers to quickly and efficiently analyze datasets to support real-time decision making applied to predictive maintenance and production planning (Cohen et al., 2019). Smart glasses can help workers make rapid decisions on maintenance and quality control (Dalenogare et al., 2019). Other wearables such as eye trackers and biosensors can allow the integration of human-related data to better understand how people are effectively working, how they move, and how they use tools and resources (Peruzzini et al., 2020).

Regarding ergonomics and physical effort, smart exoskeletons use algorithms that automatically adjust these devices to human body motion, enabling workers to handle heavy loads (Huysamen et al., 2018). Collaborative robots (Cobots) are powerful devices that can actively cooperate with operators during specific tasks (Cohen et al., 2019). All these are examples of technologies used in the Industry 4.0 context that have a significant impact on the way work is performed by people and on required capabilities (Szalavetz, 2019).

3.2.3 Smart Supply Chain

The first external dimension of Industry 4.0 is the Smart Supply Chain (Figure 3.1). This concept consolidates previous definitions such as Supply Chain 4.0 (Frederico et al., 2020), Digital Supply Chain (Büyüközkan & Göçer, 2018) and Logistics 4.0 (Strandhagen et al., 2017). Smart Supply Chain considers the support of Industry 4.0 base technologies to improve supply chain information flows Frank et al. (2019a). New opportunities emerge due to connectivity and mass storage of data shared in real time between different stakeholders in the supply chain (Frederico et al., 2020). Industry 4.0 introduces technological changes that help to improve supply chain visibility, allowing comprehensive disruption risk management by mapping the supply chain from end to end (Ivanov et al., 2016). Technologies applied to integrity control (e.g., sensors, big data analytics, decentralized agent-driven control) can ensure the right products, at the right time, place, quantity and condition, and at the right price, along the supply chain (Barreto et al., 2017; Ivanov et al., 2016). At the physical logistics level, the Smart Supply Chain dimension also comprises warehouse handling by autonomous robots and vehicles and tracking and decision-making systems for inventory control (Strandhagen et al., 2017). This also involves the "smart" handling of raw materials (input of the production line) and manufactured outputs on the shop floor. Such handling can be supported by the use of robotic sensing technologies, including automated guided vehicles (AGVs) and autonomous mobile robots Frank et al. (2019a).

On the downstream side, Smart Supply Chain considers the digitization of supply chain operational processes, mainly through two different approaches: platform-based crowdsourcing of standard processes, and on-demand provision of customized services (Hahn, 2020). On the one hand, platform-based crowdsourcing of standard processes includes activities such as the monetization of warehouse excess capacity (Hahn, 2020) or of transport logistics, the "uberisation" of the freight transport offer in order to connect idle capacity with demand (Monios & Bergqvist, 2019). It also includes the use of AI and machine learning solutions to manage and integrate the supply chain with the demand (Agrawal et al., 2018). On the other hand, on-demand provision of customized services is deeply connected to the Smart Products and Services dimension. Meeting customers' demand in real time and in a customized manner is possible due to smart devices (i.e., IoT-based products) and smart services through apps, web platforms, or IoT solutions embedded in the smart devices (Frank et al., 2019b). Additionally, real-time big data analytics of vehicles, products, and facilities locations allows manufacturers and distributors to find optimal routing for material and product transportation (Strandhagen et al., 2017). Finally, the current democratization of additive manufacturing is allowing for on-site, on-demand, rapid manufacturing that reduces the need for storing products (Ivanov et al., 2016; Strandhagen et al., 2017).

3.2.4 Smart Products and Services

Smart Products and Services, the second external dimension of the Industry 4.0 framework (Figure 3.1), comprises two kinds of provisions that can be separated but are usually integrated into a bundled solution. Smart products are artifacts that, besides their physical components, are supported by Industry 4.0 base technologies (IoT, cloud, big data analytics, and AI) to collect, monitor, control and optimize user data (Hofmeister Kahle et al., 2020; Porter & Heppelmann, 2014). Smart services, in turn, consider firms employing digital technologies in order to offer services to their users, such as cloud services, remote assistance and monitoring, and AI-based attendance (Ardolino et al., 2018; Cenamor et al., 2017). These services can be offered as independent services to support customers in their use of products, or the product itself can be offered as a service in a pay-per-use system (Ayala et al., 2017).

Manufacturing companies are witnessing a fast-growing servitization process, which means including service provision as part of the manufacturing business model (Ayala et al., 2019). In a recent study, Frank et al. (2019b) explained the link between servitization and Industry 4.0 as two different industry streams that can converge and create synergy. According to them, servitization is connected with Industry 4.0 when the manufacturer provides digital services that create value for customers and, simultaneously, provide feedback to the manufacturing and engineering system. Smart solutions can evolve into integrated smart product-service systems when products and services are designed to work jointly, leveraging new IoT-enabled business models (Paschou et al., 2020; Lu et al., 2019). Such new business models can be supported by advanced data analytics, such as predictive analytics, to reduce the risk and cost of assuming operations' performance (Grubic & Jennions, 2018).

3.3 Research method

We aim to analyze the evolution of the four smart dimensions of Industry 4.0 over the almost ten years of existence of the Industry 4.0 concept. We seek to understand how the intersections between these concepts are evolving and identify research topics in the early stages and the profile of different journals exploring this topic, especially those that are taking a holistic perspective on Industry 4.0 as a manifestation of the integration of the four smarts. Therefore, we performed a SLR followed by a network analysis of the papers. We used a machine learning-based approach to remove irrelevant articles in search queries. SLR is a useful research approach when the aim is to understand the evolution and trends of a specific research field and identify patterns in the research topics addressed in different research fields.

3.3.1 Data collection procedures

We created our search queries to cover a broad range of relevant Industry 4.0 papers but excluding irrelevant articles that might otherwise distort our findings. Therefore, we first downloaded potentially relevant articles and filtered only non-retracted journal articles. We then performed a machine learning-based review to remove irrelevant articles, and finally, we identified which of the four smarts were related to each article. Figure 3.2 provides an overview of the approach, which we describe in detail in the following section.

Search and download of relevant articles

Our research aims to analyze articles from a broad range of sources to ensure a comprehensive description of the Industry 4.0 research landscape. In order to be effective, such an analysis required a structured and curated, high-quality bibliometric data source. Therefore, we used the Scopus (Elsevier) database, one of the largest databases of research articles (Harzing & Alakangas, 2016; Liao et al., 2017) and considered a reliable source for previous bibliometric research projects (Kipper et al., 2020). Scopus offers an application programming interface (API) for researchers, enabling efficient automated search and retrieval of articles.

To identify relevant articles, we defined search terms indicating whether an article content was related either to Industry 4.0 in general or to any of the four smart dimensions specifically,

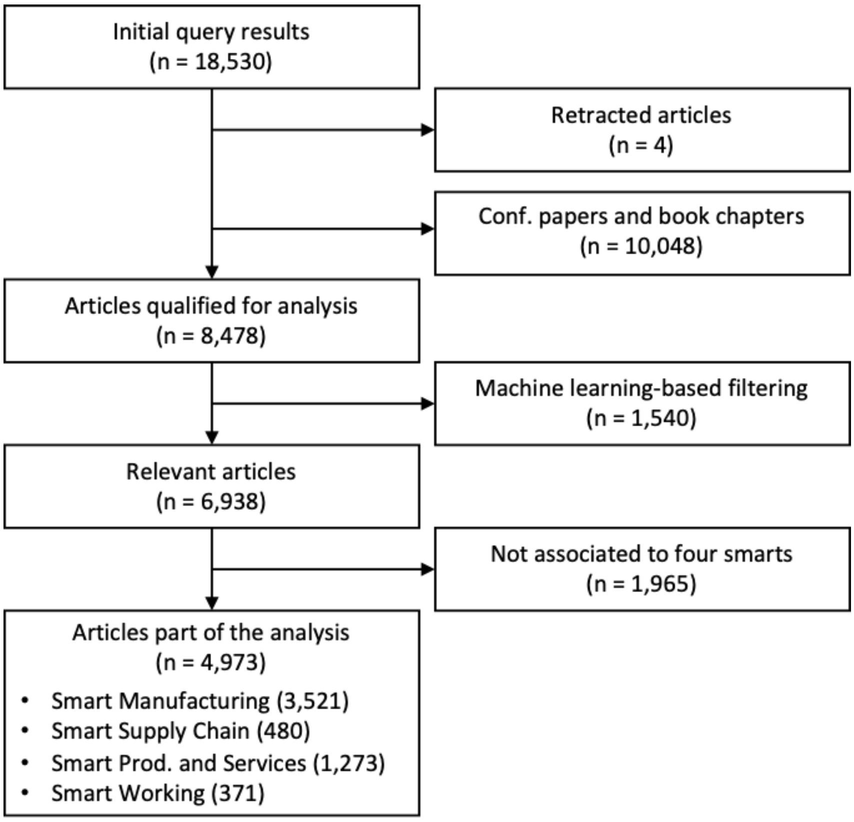


Figure 3.2: The flow diagram shows the steps in our literature search along the four smarts of Industry 4.0. The numbers of articles related to each smart add up to more than 4,973 because some articles are related to multiple smarts.

as described in the literature section of this article. Therefore, we extracted potential search terms from articles included in the literature review. We used the Scopus web interface to test the relevance of those search terms, i.e., we tested whether the search terms led to relevant articles without a large number of unrelated articles. From those results, we extracted further potential keywords and reviewed their relevance. We ran an initial search query in April 2019 to assess the reliability of our results and adjust the keywords. Final adjustments were made in March 2020. Table 3.1 (Part A) summarizes the resulting keywords associated with each of the search topics.

Table 3.1: Search queries terms.

	Terms - Part A	Terms - Part B	Sources
Industry 4.0	“industry 4.0” OR “industrie 4.0” OR “industrial internet” OR “industrial internet of things” OR “IIoT” OR “industrial IoT” OR “fourth industrial revolution” OR “4th industrial revolution”		Benitez et al. (2021, 2020); Chiarello et al. (2018); Culot et al. (2020); Kagermann et al. (2013); Kipper et al. (2020); Thoben et al. (2017)
Smart Manufacturing	“smart factory” OR “digital factory” OR “intelligent factory” OR “smart manufacturing” OR “digital manufacturing” OR “digitalized manufacturing” OR “intelligent manufacturing” OR “advanced manufacturing” OR “cyber manufacturing” OR “factory 4.0” OR “cyber physical production system” OR “cyber physical manufacturing” OR “factory of the future” OR “factories of the future” OR “cloud based manufacturing” OR “cloud manufacturing”	“Production” OR (“manufacturing” AND NOT “manufacturing firm” OR manufacturing industry* OR manufacturing enterprise) OR (“manufacture” AND NOT “manufacturer” OR “factory”	Benitez et al. (2020); Chen et al. (2017); Chen (2017); Dalenogare et al. (2018); Fantini et al. (2018); Frank et al. (2019a); Ghobakhloo (2018); Ivanov et al. (2016); Jeschke et al. (2017); Kang et al. (2016); Kipper et al. (2020); Kuo & Wang (2012); Pinzone et al. (2020); Riel et al. (2017); Romero et al. (2020); Tao et al. (2018); Zhong et al. (2017)
Smart Products and Services	“smart product” OR “smart connected products” OR “smart PSS” OR “smart product service” OR “servitization” OR “product service system”	“products” OR (“services” AND NOT “manufacturing service” OR “business service” OR “knowledge service” OR “micro service” OR “microservice) OR “pss”	Ayala et al. (2019); Dalenogare et al. (2019); Frank et al. (2019a); Gao et al. (2011); Geum & Park (2011); Hofmeister Kahle et al. (2020); Paschou et al. (2020); Riel et al. (2017)

Smart Supply Chain	“smart logistics” OR “smart supply chain” OR “intelligent logistics” OR “intelligent supply chain” OR “digital supply chain” OR “digital value chain” OR “logistics 4.0” OR “supply chain 4.0”	“Logistics” OR “supply chain” OR “value chain”	Barreto et al. (2017); Bowles & Lu (2014); Büyüközkan & Göçer (2018); Fareri et al. (2020); Frank et al. (2019a); Galati & Bigliardi (2019); Lee et al. (2018); Strandhagen et al. (2017)
Smart Working	“work 4.0” OR “operator 4.0” OR “cyber physical human system” OR “human cyber physical system” OR “human centric manufacturing” OR “human machine collaboration”	*In title, abstract or keyword:* “human machine interaction” OR “human machine interface” OR “human computer interaction” OR “human centric” OR “smart work*” *Only in keywords:* “workplace” OR “work place” OR “work area” OR (“work design” AND NOT “network design”) OR “worker” OR (“human” AND NOT “humanitarian”) OR “employee” OR” assisted work” OR “work environment”	Kaasinen et al. (2020); Frank et al. (2019a); Klumpp et al. (2019); Krugh & Mears (2018); Peruzzini et al. (2020); Pinzone et al. (2020); Romero et al. (2020)

We ran the final queries on March 30, 2020, using the Scopus API. The results are limited to documents of the type "article," published since 2011, when the term Industry 4.0 was introduced (Liao et al., 2017). The search, excluding duplicates, resulted in a total of 18,530 articles. Four articles were removed because they were retracted. We also eliminated 10,048 conference papers and book chapters remaining in the general search, leaving only journal articles in our dataset. A total of 8,478 articles resulted from this stage for further processing.

Machine learning-based article filtering

The search terms above were defined so that they would cover as many relevant articles as possible. However, some of those search terms also occur in articles that are not directly relevant for our research on the Industry 4.0 field. Several articles focus on a particular technical aspect. However, they do not relate their work to a broader level of any of the four smart dimensions of Industry 4.0. For example, some articles referring to "Advanced Manufacturing" describe the properties of metal alloys or very specific parameters of manufacturing technologies, such as selective laser melting. Besides, some articles refer to one of the keywords but are actually focused on a different topic. We proceeded to exclude those articles, as they do not provide insights into the evolution of the concept of Industry 4.0 in academic literature.

With more than 18,000 search results and 8,478 potentially relevant articles after removing non-journal articles and retracted articles, a manual review of each article's relevance would be highly resource-intensive, and reproducing and updating the dataset would hardly be feasible. Therefore, we used machine learning for the classification of articles and exclusion of irrelevant articles.

For the classification task, we transformed (unstructured) text into structured data. Recent approaches in natural language processing (NLP) use text embeddings, where text is represented as multidimensional vectors, representing its meaning (Li et al., 2018; Mikolov et al., 2018). Those embeddings are created through neural networks, which learn the meaning of words by processing large corpora of text. Mikolov et al. (2018), for example, trained their widely used model on more than 630 million words of text. The resulting vector representations enable calculations with word meanings; for instance, by subtracting the vector of "man" from the vector of "king," and then adding the vector of "woman," the vector of "queen" should result. Li et al. (2018) used word embeddings to train a neural network for patent classification. The use of neural networks, in combination with word embeddings, has shown to outperform alternative approaches for text classification (Zaghloul et al., 2009). Our approach builds on the SpaCy library (Honnibal & Montani, 2017) for word embedding and text classification, which provides a high-performance algorithm combined with high processing speeds. The model represents the meaning of words as 300-dimensional word vectors. The SpaCy text categorizer uses the word vectors as an input to train its convolutional neural network (CNN) for text categorization. Based on manually annotated data, the neural network learns to assign appropriate labels to texts, in this case, relevant or non-relevant article. Meindl et al. (2019) showed that this approach could achieve valuable results with low numbers of annotated samples.

We manually annotated the relevance of 495 articles for our analysis, based on a manual review of article titles and abstracts. Training and annotation were executed using the Prodigy[2] interface, which enabled a fast and straightforward workflow. Prodigy provides an active learning feature, meaning it uses previous annotations to select the most relevant samples for annotation, the ones it expects will contribute most to high accuracy in text categorization. Of all annotations, 396 served as input for training the neural network, and 99 were used for evaluation. SpaCy's neural network classification algorithm led to an accuracy of 96%. The trained classification algorithm assigns a score to each article, indicating its probability of being relevant. We considered 6,938 articles – all with a relevance score equal to or above 0.5 – as being relevant for our work and excluded 1,540 articles with a lower score. The vast majority of articles had scores either close to one or close to zero (see Appendix B.1 for a visualization of relevance scores), which corroborates the high level of reliability of the algorithm.

Association of the articles to the four smarts

To evaluate the evolution of research related to the four smarts, we tagged the associated smarts for each article. First, we associated articles to smarts based on the search query results described in Table 3.1 (Part A). For example, articles resulting from the smart manufacturing query, with keywords such as "digital factory," were tagged as "Smart Manufacturing." Second, we created additional associations based on each article's description and metadata. Some search terms strongly indicated an association with Smart Working but did not qualify as a search term for the initial query. As they are also commonly used in other areas, this would lead to a high number of irrelevant search results. We searched for those terms in title, abstract, and keywords (see the complete list of articles in the chapter 3 appendix file, or the article supplementary materials[3]). For example, "human-machine interface" is a search term relevant for Smart Working, but it is also used in several other contexts, like computer gaming or neuroscience. Table 3.1 (Part B) summarizes those terms. Additionally, we identified search terms that can indicate an association with one of the smarts but may also appear in abstracts that do not refer to the smart, such as the term "workplace." We only associate those articles to one of the four smarts if the search terms are listed as keywords. Searching only within keywords ensures that the article truly focuses on the topic, for instance, the workplace. See Table 3.2 for a complete list of those terms.

3.3.2 Data analysis

We conducted three types of analysis to explore the research field of Industry 4.0. First, we reviewed the articles to identify research overlap in the four smarts at a general level. Based on article publication dates, we showed a chronological evolution of the Industry 4.0 landscape. Second, we considered the keyword level to explore Industry 4.0 research topics related to each of the four smart concepts. Finally, we considered the journal level to identify the most relevant journals for Industry 4.0 and each of the smart concepts. The analysis comprises methods of network analysis and visualization and calculation of importance scores for keywords and articles. Network analysis is a useful approach to identify interrelations between different units

[2] `https://prodi.gy/`
[3] `doi:10.1016/j.techfore.2021.120784`

of analysis and allows to identify proximities between these units and ways they interconnect. Bibliometric studies have frequently used this technique showing its robustness for literature analysis, as considered in this paper (Fahimnia et al., 2015; Machado et al., 2020).

Visualization of research related to Smart concepts

We visualized the Industry 4.0 landscape development through a network map with the four smarts as central hubs. Each hub is linked to several nodes, each presenting one research article and lines representing the links. Edges are created based on the association of an article to each of the smarts. For instance, if an article relates to Smart Manufacturing and Smart Working, it is linked to both hubs accordingly (via two edges). We implemented the visualization using Gephi (Bastian et al., 2009) and build on the ForceAtlas2 algorithm for the layout (Jacomy et al., 2014). The algorithm enables an intuitive visualization of the network by relying on physical principles. Nodes repel each other, like equal poles placed together, whereas edges attract poles, acting similarly to a spring. Finally, we introduced a timeline component. Article publication dates served as a timestamp, allowing for analysis and the creation of snapshots for specific dates and timeframes, which was useful to understand the evolution of the concepts.

Keyword importance per smart

We wanted to identify the typicality of each keyword per smart. Therefore, we calculated a relevance score, indicating this typicality per keyword per smart dimension. We called this score relative comparative advantage (rca). The rca is calculated in two steps. First, we calculate keyword importance scores for each of the smarts, importance(kw) in Equation (3.1). To do so, we look at articles related to each smart separately. For each keyword, we divide the count of articles with the keyword by the total number of articles.

$$importance(kw) = \frac{articles(s, kw)}{\sum_{kw' \epsilon KW} s, kw'} \tag{3.1}$$

Second, Equation (3.2) calculates the rca(s,kw) by normalizing the importance(kw) across all smarts. Therefore, we divide it by a score indicating the overall importance of a keyword within the whole research landscape. For each keyword, we calculate the score as the number of all articles containing the keyword, divided by the total number of all articles.

$$rca(s, kw) = \frac{importance(kw)}{\frac{\sum_{s' \epsilon S} articles(s', kw)}{\sum_{s' \epsilon S, kw \epsilon KW} articles(s', kw')}} \tag{3.2}$$

The calculation provides rca scores that indicate the relevance of a keyword per each of the smarts (and related to unassociated Industry 4.0 articles). These scores provide the foundation to associate clusters to search queries. Therefore, we summed the importance scores for each keyword within a cluster, weighted by keyword count. This led to an overall importance score per cluster and smart dimension.

Visualization of Industry 4.0 research landscape

We summarized our overall insights as an Industry 4.0 research landscape. The landscape should include most important keywords of all articles and reflect their importance for each of the smart dimensions. Therefore, we define an eleven-dimensional framework to which we assign the keywords. Four dimensions represent the four smarts, three dimensions the overlap of smart manufacturing with each of the three remaining smarts and, three dimensions describing the overlap between the three remaining smarts including smart manufacturing. We choose this layout, which describe the research space of the four smarts around smart manufacturing, as most keywords have a moderate to high association to smart manufacturing and smart manufacturing is by far the largest dimension of the framework. The eleventh dimension describes keywords which are relevant for all smarts. Those are defined by having a variance of rca score across all smarts of below 0.15. This value indicates a very equally distributed importance score which is not highly typical for one of the smarts. Additionally, we consider keywords with an rca with a variance score of below 0.5 rca of 0.5 per smart as central technologies. The rca score above 0.5 indicates a moderate importance of a keyword for a smart and the variance is still moderately low. We only consider the most frequent keywords and group together very similar terms, such as “big data” and “big data analytics,” or “cloud computing” and “cloud.” Terms which are not meaningful for the four smart framework, such as “evaluation,” or “technology” are also excluded.

Evaluation of Journals in the field of Industry 4.0

We also evaluated the contribution of journals to the research topic in order to provide an overview of the scope of journals in the Industry 4.0 domain, considering how much they emphasize each of the smarts and the systemic integration of such dimensions. To this aim, we selected journals listed in the Academic Journal Guide ranking ranked with 2 to 4 stars (ranking is from 1 to 4 stars, and we excluded the less qualified journals with 1 star) and contributing with at least ten articles to the analysis. We display the percentage of articles within a Journal that are related to each smart. Further, we calculated the importance of the journals for each smart dimension, similarly to the rca score described in this section. Instead of the keyword variable, we used the journal as an input parameter.

3.4 Results

We analyzed the evolution of articles referring to the four smart dimensions, the keywords per smart dimension, and the relevance of different journals. Figure 3.3 shows an overview of the articles per smart per year. Herein we provide a supplementary data file (see Chapter 3 appendix file) with the full list of articles used in our analysis, including their associations to the smart dimensions and the relation of each paper to the main keywords.

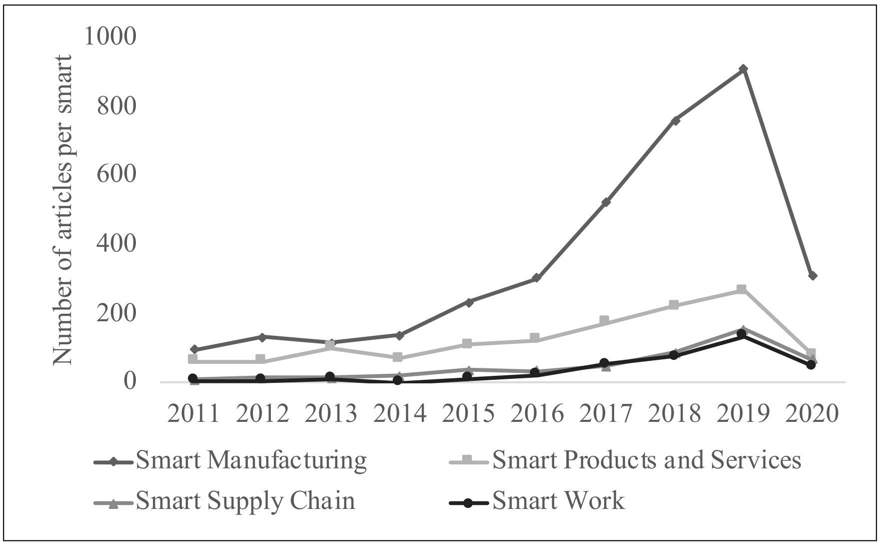


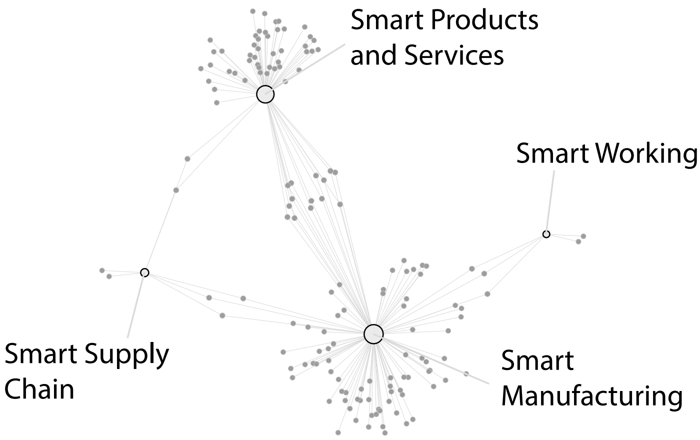
Figure 3.3: Timeline of articles per smart and year. The graph shows the total number of articles related to each smart, where one article can relate to multiple smarts.

3.4.1 Evolution of the four smart dimensions of Industry 4.0

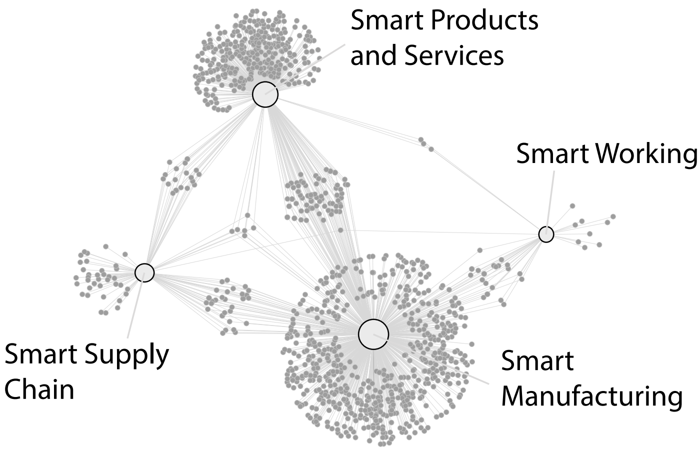
Figure 3.4 summarizes the evolution of research in each of the Industry 4.0 dimensions. It shows that the general research activity in the field of Industry 4.0 has continuously grown since 2011. Most articles are related to Smart Manufacturing (3,521), which represents 71% of the sample collected. Smart Manufacturing, Smart Supply Chain, and Smart Working have exponentially grown since 2015, although the former has slightly slowed down since 2018, suggesting that this sub-field is achieving stability and consolidation. On the other hand, Smart Products and Services are showing a linear and stable growth, with its overall share of articles declining since 2013. In 2012, there was a substantial increase in articles on Smart Manufacturing topics, which led to a decrease in the share of articles on Smart Products and Services, although the output related to the topic remained stable.

Figure 3.4 also provides insights about the links between the four smarts through a network visualization. Figure 3.4 shows the articles per year, for each smart. The hubs (larger white dots) represent the four smarts, while each of the small dark grey dots represents an article. The lines (edges) indicate the relatedness of an article to the smart. The evolutionary graph shows that in an initial phase of Industry 4.0, there were some connections of Smart Products and Services with Smart Manufacturing. At that time, there was almost no research on Smart Working and Smart Supply Chain. In 2015, the Smart Supply Chain field grew with some connections to Smart Products and Services and Smart Manufacturing. The field of Smart Working started to grow in 2015, although it remains mainly related to Smart Manufacturing. By 2020, Figure 3.4 shows that intersections between Smart Products and services and Smart Working, Smart Supply Chain and Smart Products and Services, and the holistic integration of three or four of these smarts are fields still largely unexplored. Moreover, this figure shows that there is generally a reasonable connection between the four smarts in the Industry 4.0 context, which reinforces the view of Industry 4.0 as a larger field that comprehends these four dimensions.

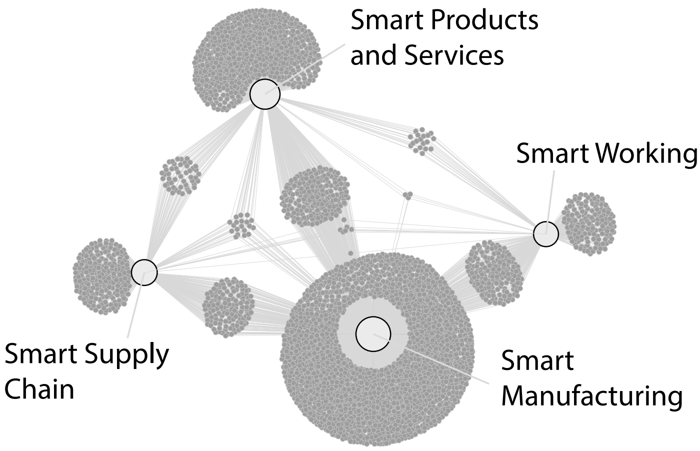
The network visualization (Figure 3.4) evidences that most Industry 4.0-related research



(a) 2011



(b) 2011 – 2015



(c) 2011 – 2020

Figure 3.4: Evolution of the network of articles from 2011 to 2020 (March). The white dots represent the smart dimensions. The size of those circles represents the logarithm of the number of edges (relevant articles until the year). Each of the grey dots represents an article. The lines (edges) indicate a relation between a smart and an article.

has been focusing on Smart Manufacturing. Since 2011, the share of articles related to Smart Manufacturing has remained around 70%. Around 16% of the articles on Smart Manufacturing are also related to other smarts. We chose a conservative approach to identifying associations with any of the smarts. Therefore, we do not consider an article related to, e.g., Smart Working, if the article focuses on manufacturing and the impact on workers is briefly mentioned. Only if the article considers Smart Working as a key topic, and therefore describes it in the abstract or includes relevant keywords, do we count it as associated to Smart Working. The results indicate that Smart Manufacturing is the heart of Industry 4.0, which has been built around this concept from its beginning. Therefore, Smart Manufacturing is also the smart dimension most often connected with other smarts, meaning that Industry 4.0 is reckoned predominantly as a matter of manufacturing activities connected to other dimensions.

In 2011, the literature on Industry 4.0 was mainly related to Smart Manufacturing and Smart Products and Services. Around 10% of the articles relate to both smarts. In the following years, this share remained mostly constant, and articles related to each of the smarts and overlapping articles had a similar growth rate. Articles related only to Smart Products and services cover, for example, servitization (Rabetino et al., 2018), usage-focused business models, and circular economy (Bressanelli et al., 2018), or service composition in Industry 4.0 (Viriyasitavat et al., 2020). A few recent articles also cover other smarts.

Moreover, around 10% of the articles in our review are related to Smart Supply Chain. This share has been increasing in recent years. As seen in Figure 3.4(c), the area is strongly integrated with other smarts, with almost half the articles relating to multiple smarts. There has been a strong overlap with Smart Manufacturing, suggesting that the literature in operations management has a broader view of Industry 4.0 and considers manufacturing and supply chain activities in the Industry 4.0 framework. Regarding Smart Working, Figure 3.4 shows that over the decade of research on Industry 4.0, this topic was mostly unrelated to any other smarts, with the exception of Smart Manufacturing. Only recently has the number of articles relating Smart Working to Smart Products and Services increased. However, the relationship of Smart Working with Smart Supply Chain is still neglectable in the literature. Only six papers refer to Smart Supply Chain and Smart Working, including Hahn (2020), who reviews Industry 4.0 in the Smart Supply Chain context and identifies a human-centric approach as a key finding. Another article related to both Smart Supply Chain and Smart Working is Liboni et al. (2019), who review the impact of Industry 4.0 on human resource management in supply chain management. Only the article by Klumpp et al. (2019) refers to Smart Supply Chain and Smart Working along with Smart Manufacturing. They review human-computer interaction in production logistics.

3.4.2 Keyword analysis for the smart dimensions of Industry 4.0

Table 3.2 provides an overview of the most common keywords in the analysis. A comprehensive list of keywords, counts, and importance scores is available in the supplementary data file to this chapter (appendix file for Chapter 3). Our keyword analysis provides insights into the research topics considered in each of the smart dimensions of Industry 4.0. The most common keywords present in the four dimensions relate to connected and intelligent information systems, particularly the IoT, CPS, and big data. Additionally, some general keywords are observed which describe the environment in which Industry 4.0 is discussed. Those keywords include sustainability, innovation, and digitalization.

Table 3.2: List of the most frequent keywords by Industry 4.0 smart dimension.

		Share of keywords per smart (rel. importance)			
Keyword	Count	SM	SSC	SPS	SW
Internet of things (IoT)	475	62% (1.0)	15% (1.8)	18% (0.8)	6% (0.8)
Cyber physical systems	374	77% (1.2)	6% (0.8)	10% (0.4)	8% (1.0)
Big data	152	73% (1.2)	15% (1.9)	8% (0.4)	4% (0.5)
Cloud computing	147	79% (1.3)	7% (0.9)	10% (0.5)	3% (0.5)
Sustainability	125	34% (0.6)	11% (1.4)	53% (2.4)	2% (0.2)
Additive manufacturing	115	88% (1.4)	9% (1.1)	3% (0.1)	1% (0.1)
Digital twin	110	85% (1.4)	1% (0.1)	10% (0.5)	4% (0.5)
Simulation	102	74% (1.2)	12% (1.5)	10% (0.4)	5% (0.7)
Artificial intelligence (AI)	86	66% (1.1)	9% (1.2)	6% (0.3)	19% (2.5)
Circular economy	85	19% (0.3)	8% (1.0)	71% (3.2)	1% (0.2)
Machine learning	77	75% (1.2)	4% (0.5)	10% (0.5)	10% (1.4)
Ontology	72	75% (1.2)	6% (0.7)	17% (0.7)	3% (0.4)
Business model	70	21% (0.3)	7% (0.9)	69% (3.1)	3% (0.4)
Radio frequency ident. (RFID)	69	49% (0.8)	33% (4.1)	10% (0.5)	7% (1.0)
Innovation	67	46% (0.7)	15% (1.9)	31% (1.4)	7% (1.0)

Note: Percentage values indicate the share of each keyword related to each smart (SM – Smart Manufacturing, SSC – Smart Supply Chain, SPS – Smart Products and Services, SW – Smart Work) the value in brackets and coloring indicates the relative importance per keyword per smart (see section 3.3.2). White color indicates a relative importance score below 1, scores between 1 and 1.5 being light grey, and scores above 1.5 being dark grey.

Most articles in our analysis are linked to Smart Manufacturing. Therefore, most keywords appear in articles exploring this concept. The relative importance score accounts for the frequency of concepts. For example, AI is most often present in Smart Manufacturing articles; however, relative to the overall number of articles, it is highly present in Smart Working articles. Therefore, its relative importance is higher for Smart Working. The Smart Products and Services dimension has low overall importance scores, except for a high importance score for sustainability and circular economy. This suggests that many topics common to the other Industry 4.0 fields could be further explored in this field. For instance, it is surprising that the Smart Products and Services dimension has few keywords ($\leq$ 10%) on big data and cloud computing, which have been proposed as enabling technologies for the digitization of products and services (Hofmeister Kahle et al., 2020). Regarding Smart Supply Chain, this concept has high overall importance scores for the most relevant keywords. However, it is worth noting that few studies explore new business models enabled by a Smart Supply Chain.

To describe the overall insights from the keyword analysis, we created a simplified Industry 4.0 research landscape (see Figure 3.5). Each keyword is placed on the landscape based on the relative importance score, e.g., additive manufacturing is mainly relevant for smart manufacturing and smart supply chain and is therefore placed in between both concepts (see Section 3.3.2 for a more detailed description). This indicates, that the topic is researched in articles related to both topics. In turn, it does not mean that many articles related to additive manufacturing

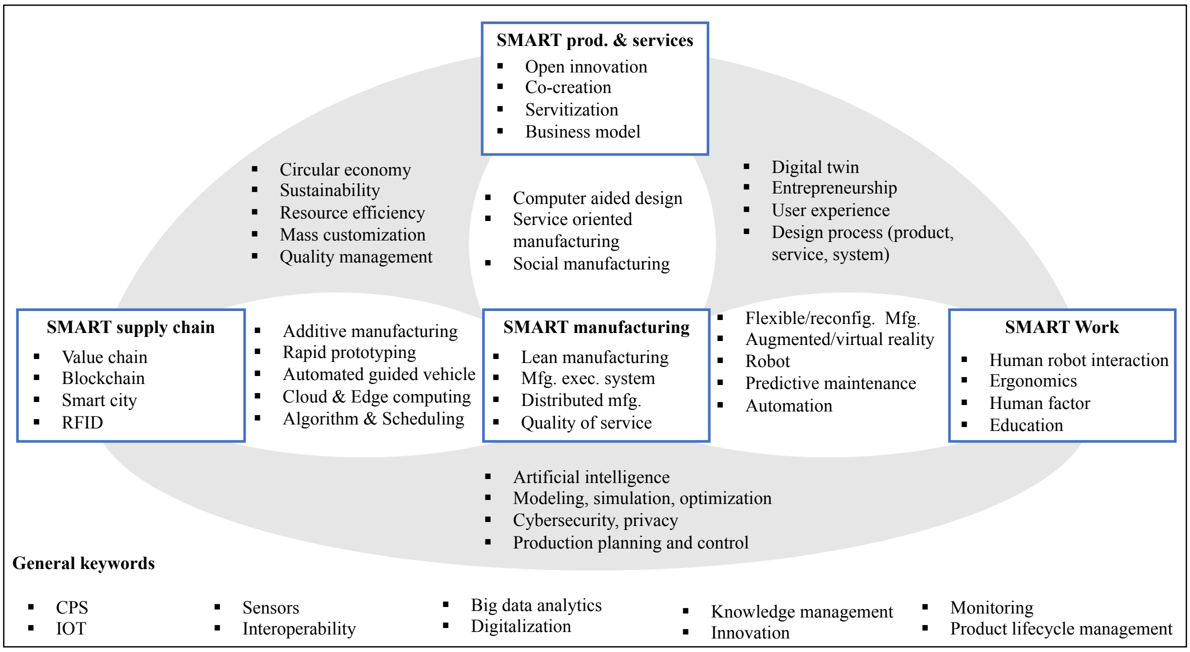


Figure 3.5: The Industry 4.0 research landscape shows keywords associated to the four smarts. Associations are identified through rca values. Terms with a high relevance for multiple smarts are placed in between those.

are both, addressing smart manufacturing as well as smart supply chain. Topics, such as CPS, Sensors, and ditigalization, which are relevant for all smarts are identified as general keywords.

3.4.3 Journal analysis for the smart dimensions of Industry 4.0

We also evaluated which journals are relevant for the different aspects of Industry 4.0. Therefore, we evaluated which Journals were most relevant and calculated a journal importance score for each smart dimension (see procedures in Section 3.3.2). Table 3.3 provides an overview of the analysis for ABS ranked journals[4] with a rating of two stars or higher[5]. The analysis shows differences between journals. In the Operations Management category, the International Journal of Production Research is the one that provides most articles to our analysis among these journals (157 articles), with greater concentration on Smart Manufacturing. The International Journal of Production Economics, Computers in Industry, Production Planning and Control, Expert Systems with Applications, and the International Journal of Operations and Production Management show a balance among three smarts (Smart Manufacturing, Smart Supply Chain, and Smart Products and Services). The latter journal has fewer articles on Smart Manufacturing and Smart Supply Chain but is the leading one on Smart Products and services with a very high share (85%). Only Computers and Industrial Engineering contains a relevant number of articles on Smart Working, and several of them are also related to Smart Manufacturing. Table 3.3 shows that journals from the business category (BUSS) are mainly concentrated around one of the smart dimensions. In this category, Technological Forecasting and Social Change is the most

[4] `https://charteredabs.org/academic-journal-guide-2018/`

[5] A more comprehensive list of journals, including additional metrics, is available in the supplementary file to this chapter, see Chapter 3 appendix file.

influential journal for Industry 4.0, since it covers the four smart dimensions. This journal is also one of the most integrative of the four smarts, since it has a significant share of articles related to all smart dimensions, while it also covers the growing Smart Working dimension, which is sometimes ignored in other Industry 4.0-related journals. Other journals in the business category showing high relative importance of Smart Supply Chain management also have high relevance scores for Smart Products and services, i.e., both are highly correlated.

Table 3.3: List of the most frequent journals by Industry 4.0 smart dimension.

		Share of articles related to topic per smart			
Category: Journal	Total	SM	SSC	SPS	SW
OM: Int. J. of Production Research	157	68%	17%	29%	5%
OM: Int. J. of Computer Integrated Manufacturing	97	93%	4%	13%	5%
OM: Computers in Industry	77	68%	12%	25%	6%
OM: Computers and Industrial Engineering	66	71%	8%	18%	20%
OM: Int. J. of Production Economics	60	48%	17%	55%	2%
OM: Production Planning and Control	36	50%	28%	58%	0%
OM: Int. J. of Operations and Prod. Management	33	24%	12%	85%	0%
OM: Industrial Management and Data Systems	14	79%	7%	29%	0%
OM: Expert Systems with Applications	10	30%	30%	60%	0%
BUSS: Journal of Cleaner Production	122	25%	5%	80%	0%
BUSS: Industrial Marketing Management	31	3%	0%	100%	0%
BUSS: Enterprise Information Systems	27	100%	0%	0%	4%
BUSS: Journal of Business Research	20	5%	0%	100%	0%
BUSS: Technological Forecasting and Social Change	19	68%	21%	21%	5%
BUSS: Journal of Business and Industrial Marketing	18	11%	11%	94%	6%
BUSS: Research Technology Management	16	13%	13%	88%	0%
BUSS: Business Process Management Journal	12	17%	42%	42%	0%

Note: Journals listed in the Academic Journal Guide 2018 (only those ranked as 2 to 4 stars with 10 or more relevant articles). Journal categories: OM – Operations Management; BUSS – Business and Management. Shadings indicate the relative importance of a journal per smart, as described in Section 3.3.2. Percentage values can sum up to more than 100%, as each article can be associated with more than one smart (SM – Smart Manufacturing, SSC – Smart Supply Chain, SPS – Smart Products and Services, SW – Smart Work).

3.5 Discussion

The initial definition of Industry 4.0, introduced by the German consortium in 2011 (Kagermann et al., 2013) and the following model of Industry 4.0 implementation proposed by the ACATECH (Schuh et al., 2020) describe a comprehensive landscape of the future production systems with smart factories, integrated supply chains, connected products, and enhanced workplaces. However, as shown in our results, the academic literature focused mainly on Smart Manufacturing, while other dimensions of Industry 4.0 were largely overlooked. The high share of papers on Smart Manufacturing is not surprising because manufacturing is indeed at the core of Industry 4.0 (Dalenogare et al., 2018). Nevertheless, our results show an opportunity for future research considering a more balanced approach with equal focus on the multiple smart dimensions of

Industry 4.0, which can also benefit other stakeholders in the field of Industry 4.0, including the academic journals interested in this topic and practitioners that may want to apply or promote Industry 4.0. We discuss these opportunities in the following subsections.

3.5.1 Opportunities for scholars to conduct future research on Industry 4.0 in the next decade

Based on our findings, one of the key priorities for future research should be integrating Smart Supply Chain and Smart Working, since we identified only six articles covering this intersection of fields. In addition, human capital is a critical element for technological change and can constitute a competitive base for companies and regions (Teece, 1998). Hahn (2020) reviewed Industry 4.0 studies in the Smart Supply Chain context and identified the human-centric approach as one of the key contributions for the digital supply chain needing more attention in future research. In this sense, as manufacturing activities become more autonomous in the Industry 4.0 context, more opportunities may be created in complementary business functions such as external logistics and distribution. Considering the less explored keywords in these fields (Table 3.2 and Figure 3.5), we suggest some potential future research topics, including (i) further exploring how AI can support workers' decision-making in supply chain decisions; (ii) studying how simulation tools and digital twins can enable workers to better understand the effects of changes in the supply chain on production activities; (iii) defining how workers should be trained in additive manufacturing to enhance the provision of products and components in distributed logistics; and (iv) defining how new workers' skills could help to create new business models in the supply chain, based on digital solutions. These are but a few examples of the opportunities that emerge from crossing the less explored keywords in this intersection between Smart Working and Smart Supply Chain, but other topics may emerge from our analysis.

Our findings also identified a strong research field in Smart Products and Services, but this field has been largely independent from the other smart dimensions. Recent literature has stressed the connections of product-service systems with the smart factory and different examples of how this can happen in practice (Frank et al., 2019b). Most studies today consider only how Smart Manufacturing enhances Smart Products and Services by using technologies and concepts such as computer-aided design, service-oriented manufacturing, or social manufacturing as a service (Figure 3.4(c)). However, as Frank et al. (2019b) described, future research may explore this intersection by also explaining how Smart Manufacturing can benefit from real-time data from product usage. Knowing exactly how products are being used in the market can enhance the production planning process and product improvement activities. Table 3.3 also shows the current lack of deep connections between Smart Products and Services and Smart Manufacturing. Topics with high relevance for one smart are of little relevance for the other. Future research may create new intersections. One of these topics is research on digital technology-driven change of business model in manufacturing. The topic is motivated by research on Smart Products and Services but has low relevance in the field of Smart Manufacturing. Furthermore, the integration of Smart Products and services with Smart Supply Chains is also a growing field that deserves more attention. When looking at the share of keywords (Table 3.2), it is possible to see concepts such as big data, cloud computing, and AI that were little investigated in these fields as compared to the others. Thus, the use of connected products and services to increase supply chain efficiency is a field with much potential to grow. Finally, as with the other smart dimensions, the intersection of Smart Products and Services with Smart Working is a small field that has the potential to grow. Capabilities of the Operator 4.0, such as enhanced strength through exoskeletons, enhanced view through augmented reality, or enhanced decisions based on AI tools (Romero et al., 2020) can be useful for a better offer of services or to increase productivity in service provision. Also, augmented reality is being introduced to manufacturing industries (Abraham & Annunziata, 2017; Gay-Bellile et al., 2015). The use of AR has a place not only in processes integrated in the value chain, such as design, sales support or production, in order to make the businesses more efficient and cost-effective, but can also be integrated in the product itself, e.g., as driving assistance, in order to enhance the perception of the driving environment (Gay-Bellile et al., 2015). Therefore, this research stream may be highly relevant.

3.5.2 Opportunities for journals to explore the new frontiers on Industry 4.0 in the next decade

Our results can also help scholars to understand the profile of the leading scientific journals regarding Industry 4.0 related dimensions (Table 3.3). Similarly, it can be helpful for journals interested in Industry 4.0 to focus on new aspects, as a way to expand the frontiers on this topic. For instance, our findings show that the Smart Working dimension has been little addressed in the operations management literature. This could be due to the German origin of the concept, which has a highly autonomous production system as the aim of Industry 4.0 (Schuh et al., 2020). In recent years, such an automation-centered view has been questioned, leading to the proposition of an Industry 5.0 concept, placing the worker at the center of digital transformation (Kumar et al., 2020). However, Industry 4.0 technologies can enhance workers' capabilities instead of just replacing them (Romero et al., 2020), and there are many opportunities for scholars and journals to explore this angle. In this sense, leading countries in the digital transformation have started to acknowledge that emerging technologies can significantly impact jobs and that the new context provided by Industry 4.0 may also demand new skills and knowledge that deserves further study. Therefore, an anthropocentric view of technology has been embraced by initiatives such as the MIT Work of the Future of the Massachusetts Institute of Technology in the United States, and the Future of Work campaign of the Federal Ministry of Education and Research in Germany. Following this trend, journals interested in Industry 4.0 should consider how the changes in work due to the implementation of "smart" concepts can contribute to or change other aspects of Industry 4.0.

3.5.3 Opportunities for practitioners to adopt and promote Industry 4.0 in the next decade

For companies, our analysis brings valuable information on the status of research in critical industry 4.0 topics, which may lead them to focus on emerging streams of research to develop innovative solutions. In addition, start-ups may use our work to identify untapped opportunities in the gaps identified herein, such as the use of connected products and services to increase supply chain efficiency, the intersection of Smart Products and Services with Smart Working, and the capabilities of the Operator 4.0. For policy makers, this analysis provides a valuable tool anchoring the design of public policies for research and innovation in academic research. Besides, our results may serve as a measure of the efficiency and outputs of past and current policy and incentives.

Regarding the development of Industry 4.0 in different countries, the literature has identified different maturity levels and technological needs in developed and emerging countries (Dalenogare et al., 2018). Since our study was focused on providing perspectives to the international community, we did not investigate such differences of application – our study only considered what has been studied but not where this has been applied. We believe that the concepts presented in our results are applicable in both developed and emerging countries. Furthermore, it is worth noting that the model by Frank et al. (2019a), which we used as a guideline for our literature review, has been developed based on the Brazilian manufacturing industry and has been accepted by the international community of developed countries, mostly in Europe, as a reference on the field. Although the smart dimensions would not change, policy directions might change, especially regarding the role of workers and the cost of technologies (Autor et al., 2020). Consequently, policymakers should consider which of the gaps highlighted in our study are more relevant to their specific context. For countries in the early stage of adoption, our work may provide useful content to decide where to catch up. Further analysis could provide more detailed insights into the knowledge being produced and absorbed in different world regions.

3.6 Conclusion

Our results show that the four-smarts framework provides a comprehensive description of the Industry 4.0 concept and helps to understand different research fields and their intersections. Using this framework, we could review the development of the field along its ten years of existence. We showed that Industry 4.0 is mostly centered around Smart Manufacturing. More research on the use of advanced technologies in work, including how work is enhanced and transformed, is still needed. We showed that Smart Working is a dimension that deserves further integration with the other smarts (Smart Manufacturing, Smart Products and Services, and Smart Supply Chain). More integrative research is needed since few studies have adopted a holistic perspective, i.e., integrating all the four dimensions of Industry 4.0. The intersections between these different Industry 4.0 dimensions have multiplied, but there is still a lot of potential to consider several less explored issues and keywords. Thus, this work provides valuable insights into future research avenues, helping researchers frame their studies holistically and emphasizing the relevance of a broader Industry 4.0 landscape, rather than only focusing on manufacturing concerns.

Although this study mainly focuses on providing insights for future research in the field, our findings also provide new perspectives for the application of Industry 4.0 and the development of best practices. In this sense, practitioners can learn to envision Industry 4.0 from an integrative point of view and develop practices and a technology strategy. Our results reinforce the view of Industry 4.0 as a connection of at least four "smart dimensions" as previously proposed by Frank et al. (2019a). The Industry 4.0 journey should be planned, focusing on integrating these dimensions in order to obtain more benefits from this concept. Our study also highlights some gaps in research that practitioners influenced by academia may underemphasize. Therefore, we call attention to these aspects, which need to be considered in practice. For instance, we showed that Smart Working is still in the early stages of investigation in the context of other dimensions such as Smart Supply Chains. Thus, practitioners should consider whether this should also be further developed in their companies.

This study has some limitations that can create opportunities to improve future methodological procedures. One of the limitations is that our machine learning-based article filtering approach was used only on the metadata of articles in the Scopus database, such as titles, abstracts and keywords. Therefore, we have not considered a deeper level of content analysis in the body of the manuscripts. Future research could use our dataset (see supplementary Chapter 3 appendix file) to perform a content analysis of these articles to obtain a deeper understanding of the literature. Expanding our search to other scientific datasets besides Scopus, which may include other publishers and books, can also help to increase the number of studies taken into consideration and, thus reach a broader perspective. Moreover, the dataset that we explored provides opportunities for many other detailed analyses like the ones discussed in this paper. Future research may investigate, for example, how the relations between single technologies (represented by keywords) evolve. Future research may also develop methods to increase accuracy in the association of articles with the smarts, especially by refining the NLP tool to analyze textual context and create more refined filters. The study of how specific Industry 4.0 technologies evolve in each smart dimension can also provide important insights into the trends in this topic.

Another limitation is related to the theoretical framework adopted. As we used the four smart dimensions framework for Industry 4.0 proposed by Frank et al. (2019a), our investigation was limited to the domains of such dimensions. The recent literature on Industry 4.0 has emphasized the sustainable aspect of operations (de Sousa Jabbour et al., 2018), which is not explicitly included in that framework. Frank et al. (2019a) considered sustainability as a requirement included in each of these dimensions but including this as an additional topic would help enlighten future research paths.

Chapter 4

Exposure of occupations to technologies of the fourth industrial revolution[1]

This chapter describes the creation of a patent-based score that shows which occupations are particularly exposed to technologies of the Fourth Industrial Revolution (4IR).

4.1 Introduction

Technological progress continuously impacts the economic environment. The current wave of technological progress is driven by digitalization and the adoption of artificial intelligence (AI), and is often referred to as the 4IR. AI might enable machines to become increasingly able to perform tasks that previously only humans could perform. Whereas previously machines were mainly able to perform clearly defined, repetitive, routine tasks (Acemoglu & Autor, 2011), future automation might cover much more diverse tasks, for example, some requiring emotional intelligence (Brynjolfsson & Mitchell, 2017). These new automation patterns create fears of machines making workers obsolete and creating unemployment. However, technological change has various effects on the labor market, and previous waves of automation did not result in long-lasting technology-induced rising unemployment (Mokyr et al., 2015; Autor, 2015; Bessen, 2019).

Acemoglu & Restrepo (2019a) observe that automation not only decreases labor demand but also has a productivity effect, which increases labor demand. Further, they describe the reinstatement effect, where automation leads to newly-created tasks carried out by humans. The relative size of these effects and their interaction determine the overall effect of automation on the labor market. Therefore, automation causes changes in the task content of occupations due to lower demand for some tasks, higher demand for remaining tasks, and the creation of new tasks.

To evaluate these effects and prepare for future shifts in the labor market, researchers, e.g., Frey & Osborne (2017) and Brynjolfsson & Mitchell (2017), construct measures of automation

[1]This chapter largely relies on work from an article I am currently preparing for submission: Meindl, B., Morgan, R.F., Mendonça, J., Exposure of occupations to technologies of the fourth industrial revolution.

potential of occupations. These measures can provide valuable insights for future research in terms of overall automation potential. Our approach does not aspire to predict the share of automated jobs, but aims to reflect actual technology maturity (diffusion), which is not covered by the aforementioned indicators (Arntz et al., 2020). For example, scores by Brynjolfsson & Mitchell (2017) are based on expert assessments of "what can machine learning do?" We use patent data as an indicator for technological progress; patents actually document what existing technology can currently do. The McKinsey Global Institute (MGI) follows a similar objective and focuses on what actual automation might look like until 2030, acknowledging that there is a much higher automation potential in the long term (Manyika et al., 2017). They provide estimates of automation potentials per occupation, which they expect to be implemented until 2030.

Linking patent data to occupation activities offers a direct indicator of the exposure of occupations to technology. There exist patent occupation mappings at an industry (Silverman, 2002) and occupation level (Kogan et al., 2020; Webb, 2019) which have been used for economic analyses (Mann & Püttmann, 2017; Acemoglu et al., 2020). Webb (2019) found, for example, that exposure to previous automation technologies had a negative impact on employment at an occupation level, and Mann & Püttmann (2017) identified an overall positive impact of automation patents on employment.

We build on the approach of Kogan et al. (2020) and refine for improved accuracy and to account for task-level differences. Each occupation relates to several tasks, and technology exposure may vary among different tasks within an occupation (Brynjolfsson & Mitchell, 2017). The task level, as the "unit of work that produces output," is a highly insightful level of detail for evaluating the impact of technologies on jobs (Acemoglu & Autor, 2011). Our approach has two main benefits. On the one hand, it allows accounting for a specific technology exposure for each task, which is ignored when looking at occupations as a whole. On the other hand, our task-level approach increases the accuracy of the mapping, as it identifies patents specific to each activity, rather than patents which have many words in common with the overall occupation description. For example, our approach might avoid associating a robot engineer mainly with robot patents in general (e.g., improved efficiency of assembly robots), but rather with patents which describe innovations that help to "plan robot path", "debug robot programs", and "maintain robots." Further, we introduce a measure of technology exposure; we therefore differentiate between technologies of the 4IR (4IR patents) and other patents (non-4IR patents) for creating technology exposure scores. These scores indicate patent exposure at the task and occupation level. Our analysis includes patent data since 1970 and thus allows us to review developments over time, e.g., when 4IR technologies have been developed and how long it takes them to impact the labor market.

Various researchers identify the lack of high-quality data on technological progress of key 4IR technologies as a key barrier to better understanding the impact of those technologies on the workforce (Frank et al., 2019c; Mitchell & Brynjolfsson, 2017). With this chapter, we address this issue by providing a mapping of patents to occupational tasks and introducing a 4IR technology exposure score per occupation.

4.2 Patents as an indicator for technological change

Several studies build on patent data as an indicator of technological progress or innovative activity. Silverman (2002), for example, creates a concordance table of technologies used and produced per industry to analyze the impact of technological resources on corporate diversification. The work builds on manually-annotated patent data from the Canadian patent office. Mann & Püttmann (2017), for example, build on this dataset, to evaluate the effects of automation technologies per sector of use. The dataset is generally interesting but builds on data from 1990–1993 and only provides a patent category-to-industry mapping. Therefore, it is not suitable for a fine-grained mapping of patents to occupations, particularly for newer technologies.

Additionally, Dechezleprêtre et al. (2020) use patent data to evaluate the relation of wages and automation innovations. They focus on advanced manufacturing patents, as defined by Aschhoff et al. (2010) combined with their own patent search terms. They link those patents to industry sectors according to a concordance table provided by Lybbert & Zolas (2014). This patent-industry mapping is based on industry-specific terms. The mapping does not account for activities conducted by workers within industries; for example, industry descriptions contain terms related to the output, e.g., "manufacture cars," and not the task, such as "cut sheet metal."

Webb (2019) proposed an approach to link patent data more directly to occupations. He extracts verb-noun pairs from patent titles and task descriptions, and uses those as a basis for mapping. For example, if a doctor's task is "diagnose a patient's condition," the verb-noun pair is "diagnose condition." The analysis thus aims to identify patents describing similar actions as task descriptions. Similar tasks can be described with different words, e.g., "diagnose condition" may be very similar to "diagnose disease." Webb partly overcomes this challenge by using word hierarchies to identify more general words for the matching (higher-level terms per word). The previous examples would both become "diagnose state."

Kogan et al. (2020) also uses natural language processing for calculating similarity scores of patents and occupation descriptions. Instead of relying on word hierarchy information, they use text embeddings. They represent words through vectors, which are trained on large amounts of text data. Words that are more similar are represented as vectors that are more similar. For example, "king" minus "man" plus "woman" leads to a vector similar to "queen." Word vectors allow the calculation of similarity scores of words, and a more fine-grained differentiation is possible. Further, the approach does not rely only on verb-noun pairs (such as Webb 2019), but on the full-text data. This enables the approach to account better for context descriptions. For example, it differentiates between "diagnose *patients* condition" and "diagnose *machine* condition."

We aim to build on this general idea of mapping patent data to occupations through text embedding, while implementing the mapping at a task level. This has two key advantages. First, Kogan et al. (2020) use task descriptions for an occupation and combine them into one text. They compare this block of text with patent texts to identify the most relevant patents. However, technology exposure may vary for different tasks within an occupation (Brynjolfsson & Mitchell, 2017). Therefore, the task level, as the "unit of work that produces output," is a highly insightful level of detail for evaluating the impact of technologies on jobs (Acemoglu & Autor, 2011). For example, a robot technician does "attach wires between controllers," and "develop robotic path motions." The tasks are very different and probably have exposure to different technologies. Second, the occupation-level approach has a chance of identifying general context patents, ignoring the broad range of patents related to the different activities of an occupation. For example, a barber performs tasks such as "clean and sterilize scissors," "recommend and sell lotions," and "shampoo hair." Stringing together all task descriptions of a barber will lead to a text with many words such as shampoo, hair, and lotion. Therefore, it is likely that there is a high similarity score to patents describing hair care products. Looking at task statements independently accounts better for words describing the actual activity and identifies patents related to the activity itself, e.g., *sell* lotions. Further, if there are many tasks related to general activities, such as working with hair, and few tasks related to secondary activities, such as bookkeeping, the occupation level search is likely to ignore the secondary activities.

4.3 Mapping patents to occupations

To map occupations to patents, we compare task descriptions with patent abstracts. Therefore, we use natural language processing (NLP) methods. Figure 4.1 provides an overview of the approach.

4.3.1 Data

We compare patent texts with descriptions of activities conducted in an occupation. We build on occupation descriptions from O*NET, which describes each occupation as 20 to 40 tasks (more than 20k tasks). This approach helps to identify the most relevant patents per occupation and occupational task. All task descriptions contain information about the activity conducted. Also, some task descriptions contain information about the tools and technologies used. We focus on activities only, and thus remove references to tools used. For example, for the task "Monitor geothermal operations, using programmable logic controllers," we would remove "using programmable logic controllers." This ensures that the search includes all patents related to monitoring geothermal operations, independently of the technology they use.

Our analysis builds on the PATSTAT patent dataset[2], which is commonly used in research (Kang & Tarasconi, 2016). We include all patents since 1970 with an English abstract in our analysis. While a long-term analysis was was not the objective of this work, the 1970 cutoff should enable us to compare patterns before and druing the 4IR. Identical patents registered in multiple patent offices share a common patent family ID. We only select one patent per family to avoid double counting. For each patent family, we choose patents in the following order: preferably we use United States Patent and Trademark Office (USPTO) patents and those of other English-speaking countries' patent offices, followed by World Intellectual Property Organization (WIPO) and European Patent Office (EPO) patents. Appendix C.1 provides an overview of the number of patents per authority and year used for our analysis. For the task similarity scores we rely on patent abstracts. Abstracts are highly suitable for this task and comprise the densest information on inventions (Benzineb & Guyot, 2011). Our dataset includes

[2] `https://www.epo.org/searching-for-patents/business/patstat.html`

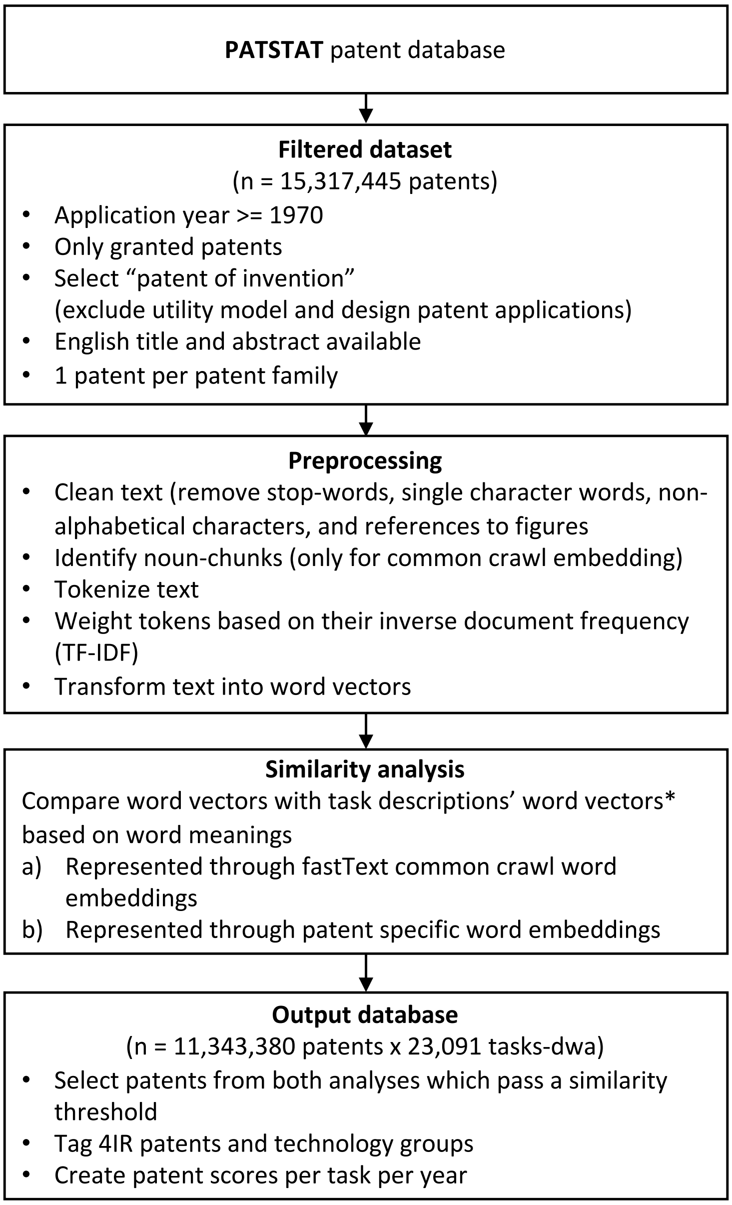


Figure 4.1: Description of our analysis to identify relevant patents per task. This chapter provides a more detailed description of the processing steps.
* We conducted a similar preprocessing for task descriptions.

15,317,445 patents filed between 1970 and 2020. Figure 4.2 shows that the number of patents increased until 2013, then sharply decreased, particularly after 2017. The recent decline is due to the time lag between the filing date and publication in the PATSTAT database. Further, the total number of patents decreased in 1976, 1991-1992, and in 2004. Due to the substantial decline in patents after 2015, our analysis offers its most valuable insights from 1970 to 2015.

4.3.2 Patent task similarity scores

We identify the most relevant patents per task by comparing the texts of task descriptions and patent abstracts. Our algorithm builds on word embeddings, where multidimensional word vectors represent words. Vector representation identifies text similarity, even if there are no

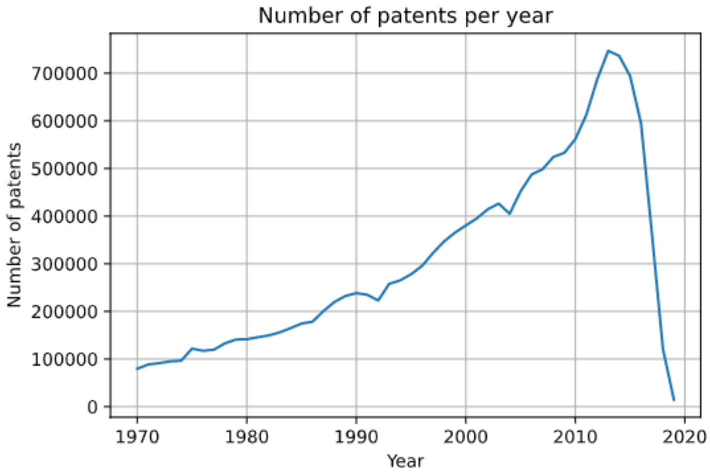


Figure 4.2: Number of patents in our dataset per year. Our dataset includes patents with application dates between 1970 and 2019. Overall, the number of patents increased until 2013. The decline afterwards is possibly due to unpublished patents and patents not yet included in the PATSTAT dataset. Dips in patenting activity are likely due to economic decline, e.g., after the Dotcom crises in 2000.

words in common. A similar approach has been used by Kogan et al. (2020), who map patent texts to occupation descriptions. The approach yields the best results when the queries comprise at least a few sentences. For very short queries (such as task statements), there is a risk that single words bias the results. For example, the task "Develop engineering specifications or cost estimates for automotive design concepts" describes the activity (cost estimates) and the context (automotive design). There is a risk that results may include patents only related to the context, if there is a high word overlap, e.g., a patent describing "cost-efficient automotive design." We try to avoid a bias towards the context, and combine task statements that describe similar activities for our queries. Therefore, we build on a high level task category by O*NET, the detailed work activity (DWA). O*NET clusters task statements into 2,067 DWAs, such as "estimate operational costs." The DWA "estimate operational costs" includes, for example, the tasks mentioned above as well as the task "analyze, estimate, or report production costs." On the one hand, combining both statements increases the weight of the activity (cost estimation), as this is common in all task statements. On the other hand, combining these statements reduces the weight of the context (automotive design), as many tasks describe a slightly different context. The different context words are valuable for narrowing the broader scope of the activity. In this case, the context relates to (cost estimation in) a production-related setting and, e.g., is not related to software development or services. For our queries, we combine all tasks related to one DWA. We create a unique query for each task by over-sampling its task statement and thus giving higher weights to the actual task description.

In order to work with the text data, we conduct some preprocessing steps. We remove expressions in brackets from patent abstracts, such as numeric references or abbreviations, which do not add relevant information for comparison with task descriptions. Further, we remove non-alphanumeric characters and single-character expressions. Also, we remove stop words, which are frequent words, such as "is" or "and", and which do not add much information to the text.

Using only the most relevant words per text is commonly performed to increase accuracy when working with text. Finally, we transform all characters to lower case.

We transform text data for analysis in three steps. First, we count word frequencies per document and weigh the frequency with the word frequencies across all texts. This method is called term frequency–inverse document frequency (TF-IDF). TF-IDF assigns higher weights to infrequent words, emphasizing those words that are most specific for a document. Second, we create a word similarity matrix using word embeddings. Word embeddings are created based on large corpora, where algorithms assign vectors based on word relations to neighboring words. Those word vectors represent the meaning of words and allow for comparisons and calculation of text, without relying on the presence of similar words. In our example, word vectors enable the identification of a similarity, e.g., for "steer a car" and "drive the vehicle," even though there are no common words. Third, we calculate similarity scores for each patent task combination using TF-IDF scores and the word similarity matrix. This leads to a task-patent similarity matrix of 23,091 task-DWA combinations and 15,317,445 patents, with similarity scores ranging from 0 to 1, with 1 representing similar documents. We reduce the matrix size for faster processing by setting irrelevant scores below a certain threshold (see below) to zero.

We conduct this approach with two different word embedding algorithms to account for patent-specific terms and general language, and identify a high number of relevant documents. On the one hand, we use fastText word embedding, which has been trained on 600 billion tokens (words) from the common crawl corpus[3] (Mikolov et al., 2018). The embedding provides 300-dimensional vectors for around 2 million words, n-grams (combinations of words), and sub-words (parts of words, which can be used to construct vectors for unknown words). The common crawl corpus comprises texts from various sources available on the web, and therefore the embeddings represent various aspects of language. However, the structure and words used in patent data are more technical than standard language. This leads to noise in our results when we compare task descriptions and patents. We only consider patents related to a task if the similarity score is above a certain threshold. The mean similarity score is 0.020, and many highly relevant patents are in the long tail of higher relevance scores. Manual review shows that above a threshold of 0.197, or 9 standard deviations above the mean, there is a high density of relevant patents. With this approach, we identify hundreds of relevant patents per task.

We repeat the approach with another word embedding, which has been trained on the text of five million patents (over 38 billion tokens) (Risch & Krestel, 2019). This embedding accounts for patent-specific language. It therefore helps to identify patents which use unusual words and would not have been identified though the common language approach. In turn, it has a higher noise caused by the language used in task descriptions, which is not necessarily similar to patent language. Similarly to the other embedding, the threshold value of nine standard deviations above the mean (0.170) proves to be a practical cutoff value to exclude most of the irrelevant patents.

Using this dual approach increases the number of patents identified and helps to validate the robustness of the approach. Combining both approaches allows us to sustain a high cutoff value and thus reduce noise through irrelevant patents. At the same time, both searches com-

[3] `https://commoncrawl.org/`

plement each other and increase the overall number of patents mapped to tasks. We find that both approaches show identical overall results patterns. For example, both embeddings lead to a similar distribution of patents per task category (see Appendix C.2 for detailed evaluation results). These identical patterns indicate that the approach is robust enough to ensure that the choice of word embedding did not systematically bias the results. Simultaneously, relying on both word embeddings significantly contributed to covering a broad range of patents and enabled the mapping of 75% of patents in our sample to at least one task.

We also conducted our analysis with cutoff values of ten and twelve standard deviations. Those additional analyses showed a similar evolution of number of patents per year and similar distribution patterns of total patent exposure across occupations. This suggests that the nine standard deviation cutoff value is sufficient, and higher cutoff values do not change the mapping significantly. Each standard deviation of higher cutoff value reduced the total coverage of patents associated with tasks by around 10 percent; we therefore choose to conduct our analysis with the nine standard deviation cutoff value.

4.3.3 Patents of the fourth industrial revolution

The current wave of digitization is often described as the 4IR. Kagermann et al. (2013) describe the 4IR as merging physical and virtual environments into cyber-physical systems. Since then, a large amount of research has evolved in this area. Several definitions exist and the literature frequently describes the Internet of Things (IoT) as the core of the 4IR, as it enables smart technologies to transform traditional enterprises (Meindl et al., 2021a). Similarly, Ménière et al. (2020) describe the 4IR as the "full integration of information and communication technologies (ICT) in the context of manufacturing and application areas such as personal, home, vehicle, enterprise and infrastructure [...] towards a fully data-driven economy." Ménière et al. build on this definition and leverage the expertise from the EPO to identify patents related to the 4IR across 350 4IR technology fields.

Ménière et al. describe technology fields along two dimensions. First, they describe application domains, such as "healthcare" and "industrial." Their second dimension describes technologies, such as "software", "connectivity", and "core AI"[4]. For each of the fields, they identify multiple ranges of patent classification classes, describing these technologies. Each patent is assigned to at least one classification. We consider a patent to be a 4IR patent if any of its classifications falls within one of the 4IR classifications defined by Ménière et al. (2020).

4.3.4 4IR exposure score calculation

We calculate the 4IR exposure scores per task using all the associated patents, with a similarity score above the nine standard deviation threshold. We weight those patents using the similarity scores for aggregating, i.e., a patent with a similarity score of 0.7 has double count than one with a score of 0.35. With this approach we increase the weight of patents where we are most confident it represents the activity. The task level exposure score is calculated as logarithm of

[4] They identify three core technology fields, IT hardware, software, and connectivity, as well as eight enabling technology fields, including data management, user interfaces, core AI, geo-positioning, power supply, data security, safety, three-dimensional support systems. Application domains include consumer goods, home, vehicles, services, industrial, infrastructure, healthcare, and agriculture

the sum of all related patent similarity scores, in order to account for the log-normal distribution of patents per task. These task level 4IR exposure scores serve as a basis to calculate occupation or activity level exposure scores.

In the following sections, various analyses refer to the aggregated number of patents, e.g., average patents per task related to an occupation. Aggregated exposure scores, such as DWA or work activity, are calculated as the mean value of associated task exposure scores (i.e., the log values). Relying on log values allows us to minimize the impact of outliers, i.e., tasks with a very high number of associated patents, and reflects whether there are generally many tasks with high or low patent counts. When aggregating task counts to occupation level, we weight the task scores with the task importance score for an occupation, as provided by O*NET, ranging from above 0 to 1. This leads to less important tasks having less impact on an occupation's exposure score. A similar approach has been followed by Brynjolfsson & Mitchell (2017) when calculating suitability for machine learning (SML) scores for occupation. For aggregation by occupation cluster, we first aggregate at an occupation level and next on a cluster level to avoid bias through the number of tasks per occupation.

Acemoglu & Autor (2011) provide a classification of six task types (e.g., routine manual, cognitive analytical), based on work activity and work context variables, as described by O*NET. We apply their categorization in order to assign task types to each tasks. They define *work activity descriptions* and *work context variables* for categorization of occupations. The work activity descriptions, e.g., "4.A.2.a.4 Analyzing data/information" map directly to DWAs and thus to task descriptions. A direct mapping is possible. There is no direct mapping of tasks to wor context variables. Therefore, we link those variables to task descriptions as follows. O*Net provides a mapping of work context variables, e.g., "4.C.3.b.7 Importance of repeating the same tasks" map to both *skills* and *abilities* – for both of which a mapping to work activities is provided. Therefore, we map work context variables via the intermediate variables skills and abilities to the task descriptions. Through this mapping, a task may be related to multiple categories. We calculate z-scores per measure (e.g., many task might have 1 connection to the above mentioned work context variable, whereas task A might have 5 connections; this may lead to task A having a positive z-score and other tasks a negative z-score.) and aggregate those z-scores across task categories. For the analysis, we associate all tasks where the sum of z-scores is above 0 to a category, which may lead to some tasks being associated with more than one category.

4.4 Patent task mapping results

The previous section identifies a mapping that shows the most relevant patents for each occupational task. Overall, our analysis associates 11,343,380 patents to tasks, thus 74% of all patents in our dataset are associated with tasks. Not all technology clusters are equally present in our results. On the one hand, chemistry or biotechnology patents only map to few occupations, which is not surprising, as those inventions mostly do not describe activities, but rather formulas or materials. On the other hand, computer technology and IT methods for management patents are highly relevant for many tasks.

To evaluate the quality of the patent-task mapping, we manually investigate the relevance of patents with a high similarity score. Table 4.1 shows some example tasks and patents.

Table 4.1: Most similar patents per tasks.

Similarity	Patent title
Operate diagnostic or therapeutic medical instruments or equipment: Assemble and use equipment, such as catheters, tracheotomy tubes, or oxygen suppliers.	
0.42	Fast trachea incising device
0.40	Device and method for electrocardiography on freely moving animals
0.38	Method and apparatus for weaning ventilator-dependent patients
0.38	Method and apparatus for weaning ventilator-dependent patients
0.37	Devices, systems and methods for using and monitoring tubes in body passageways
Prepare scientific or technical reports or presentations: Write up or orally communicate research findings to the scientific community, producers, and the public.	
0.52	Research product automatic register service method and system
0.48	System and method for medical image interpretation
0.47	Research collection and retention system
0.46	Method and apparatus for the design and analysis of market research studies
0.45	A System for Joint Research on the Internet
Evaluate quality of materials or products: Perform visual inspections of finished products.	
0.40	Mobile unit for express-control of oil products characteristics
0.40	Automatic monitoring method for coated products and fabrication process
0.38	Method and apparatus for inspecting manufactured products for defects in response to in-situ monitoring
0.37	Automatic quality inspection method, device and system
0.37	Optical fiber sensing system consistency test method

Note: Multicolumn lines include detailed work activities (bold) and task descriptions. Our approach maps patents not only based on the actual task description, but includes information from detailed work activities. Two different word embedding methods have been used; the table includes the average similarity score of both embeddings.

The example results indicate a good match of task descriptions and patent results. Our analysis is conducted at a task level and also contains information on other tasks within the same DWA. Therefore, the results may include patents not directly related to the task but capturing the activity described by the DWA.

Appendix C.3 provides additional information on the presence of patents per technology field and task type. In the following section, we review patent task mapping quality, task exposure scores, and evolution over time.

4.4.1 Patents over time

Figure 4.3 shows the mean number of patents associated per task, per year. The mean number of patents per task increased over time, mostly in line with the overall number of patents in our sample. An exception is the period between 2000 and 2004, where the number of patents per task remained nearly constant, whereas the total number of patents in our raw data decreased only in 2004 and otherwise increased. Another plateau is in the years before 2008, with a substantial increase after that, whereas the curve of the total number of patents in our raw data is more smooth. Finally, the overall number of patents declined in recent years. This is because not all of these patent applications have yet been published in the PATSTAT dataset (see Figure 4.2 for total patents in our dataset).

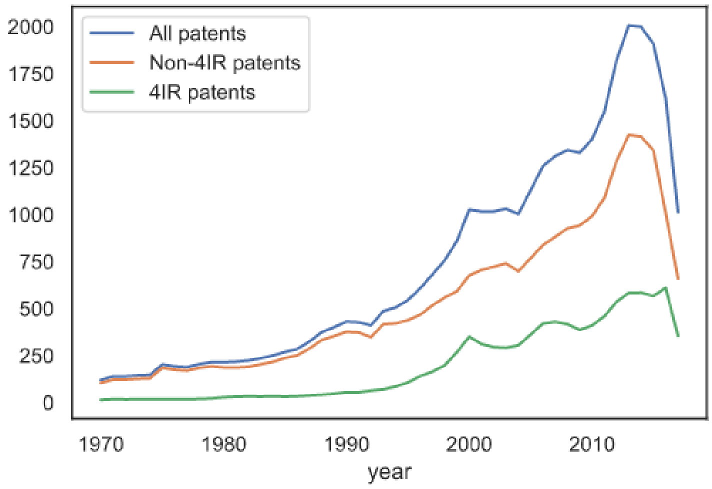


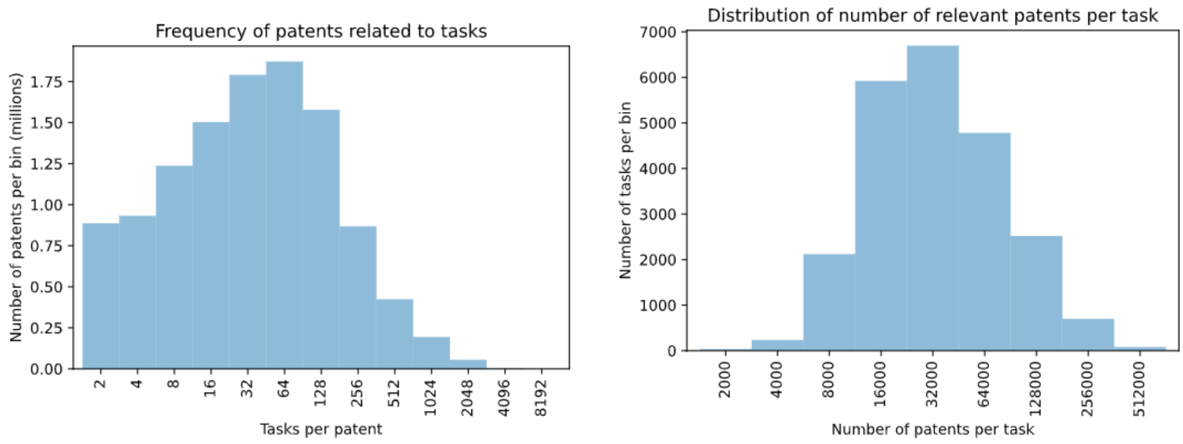
Figure 4.3: Average number of patents per task and year. The figure shows a plateau between 2000 and 2004 after strong growth in the previous years. The plateau in overall patent exposure between 2000 and 2004 is driven by a decline in 4IR patents during this period.

Ménière et al. (2020) identify technologies of the 4IR, which include, e.g., virtual reality, machine learning algorithms, and 3D modeling software. Some of those patents already existed decades ago, e.g., machine learning is a technology introduced decades ago, but which reached market maturity only recently. We compare exposure to 4IR patents with traditional (non-4IR) exposure. Our findings show that exposure to Industry 4.0 patents grew more than overall exposure to patents. Our findings indicate that 4IR patents drove the strong growth in patent exposure between 1990 and 2000. The share of Industry 4.0 patents of all patents in the results is 29%, ranging from 12% in 1976 to 38% in 2016. A drop in Industry 4.0 patents after the year 2000 also explains the plateau of overall patents per task, whereas the number of non-Industry 4.0 patents continued its constant growth. Interestingly, other research articles also observe this change in patenting behavior. Kelly et al. (2021), for example, identify "breakthrough patents", which relate to novel technologies and differ in text content from previous patents. They observe a strong rise of breakthrough patents between 1980 and 2000, with a steep drop afterwards.

4.4.2 Patents per task

Most of the 11 million patents in our results map to multiple tasks. On the one hand, task descriptions share similar content, and DWA-level information forms part of the search queries, which leads to a patent being likely to be mapped to multiple patents within a DWA. On the other hand, a patent may describe an invention which is relevant for different tasks. Figure 4.4(a) shows the frequency of patents in our results. The x-axis indicates how many tasks a patent is associated with (out of more than 23,000 tasks). Most patents relate to around 30 to

60 unique task/DWA combinations.



(a) Distribution of number of tasks per patent.

(b) Distribution of number of patents per task.

Figure 4.4: Patent frequencies and number of patents per task.

Next, we look into the differences of relevant patents per task. Figure 4.4(b) shows that the number of patents per tasks ranges from around 2,000 to more than 500,000 patents. The distribution of the number of patents per tasks is log-normal, with most tasks associated to 32,000 patents. To account for the logarithmic distribution of patents per task, we use the logarithm of the number of patents for further analysis, where we aggregate numbers of patent per task, to avoid bias through tasks with high numbers of patents. Tasks with the lowest number of relevant patents are related to the activities “prepare operational budgets for green energy or other green operations,” “collaborate with others to determine technical details of productions,” and “manage budgets for personal services operations.” Tasks with the highest number of relevant patents are related to the activities “thread wire or cable through ducts or conduits,” “record research or operational data,” and “prepare data for analysis.”

4.4.3 Evaluation per task type

To better understand the patent exposure of tasks over time, we categorize tasks into six different task types (e.g., routine, physical, cognitive), based on the classification provided by Acemoglu & Autor (2011). Our analysis provides a patent count per task and year. For each task, we evaluate the number of patents over time. We aggregate the log number of patents per task type to see the overall evolution of patents per task type. Figure 4.5 shows the evolution of patents over time.

Section 4.4.1 showed that there is a strong growth in the overall number of patents (for all task types); this section shows each task type’s share of total patent exposure per year. Non-routine manual tasks have the highest number of related patents, but with the rise of 4IR technologies their share of total patents decreased. They have the lowest exposure to 4IR patents. In contrast, analytical tasks experienced a strong growth in patent exposure, driven by a high share of 4IR patents. This is in line with the exposure to breakthrough technologies, as defined by Kogan et al. (2020). In addition, exposure scores to breakthrough patents increase

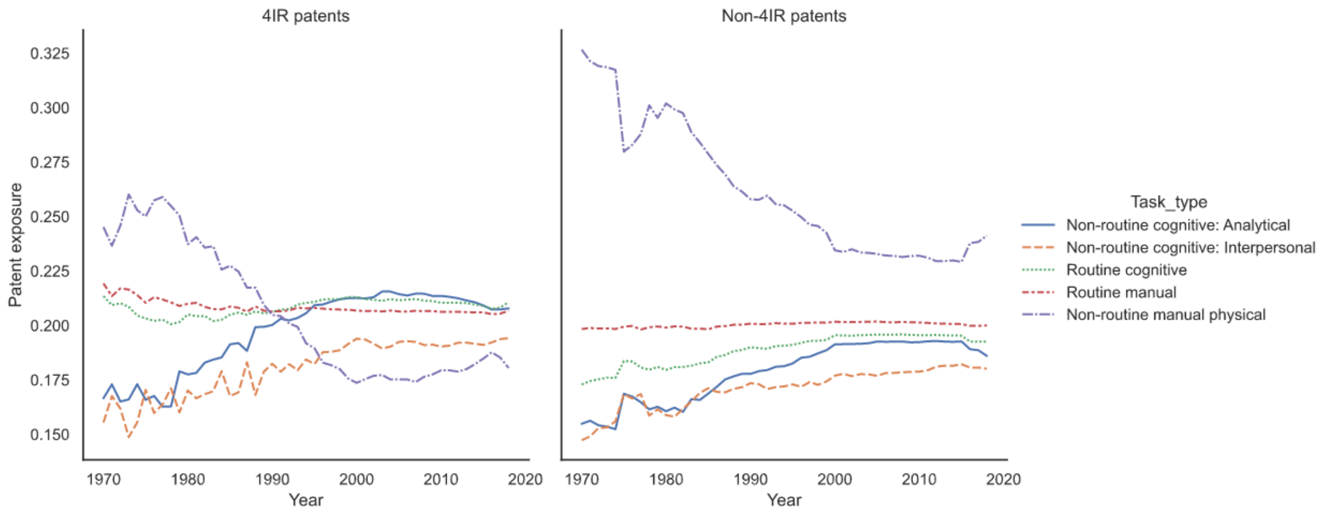


Figure 4.5: Number of patents per task type over time. The graphs show total exposure per task and exposure to 4IR patents and non-4IR patents. The exposure values indicate the share of patents per year and task type.

for routine cognitive tasks and decrease for routine manual tasks, which does not occur for 4IR patents. Aside from the general difference in methodology, Kogan et al. use an additional dimension, "non-routine manual interpersonal tasks," which is not considered in our analysis and thus might impact exposure to other task types. Kogan et al. conducted an analysis of patent exposure for subsets of patents, based on technology type, such as electronics patents. Future researchers can build on our patent task mapping and create similar exposure scores for a more detailed comparison of exposure scores. The results also show changes around the year 2000, when both 4IR and non-4IR patent shares stabilized compared with previous years. Similarly, Kogan et al. (2020) observed strong changes in exposure shares before 2000 and a stabilization in the years after.

4.5 Patent exposure score

The 4IR represents technologies which have been adopted in the current wave of technological change. These patents include, e.g., AI, machine vision, and autonomous robots. These technologies may change the way we work, even in areas that have previously been less impacted by automation technologies (Brynjolfsson & Mitchell, 2017).

Previous research introduced indicators on the automation potential for different occupations (Frey & Osborne, 2017; Brynjolfsson & Mitchell, 2017). These indicators help to better understand which occupations and industries might be particularly impacted by the 4IR. A number of inventions play together in this progress, and diffusion might differ across occupations. However, current measures do not address the gap of technical feasibility and diffusion (Arntz et al., 2020). We address this gap by introducing a patent-based measure of exposure to 4IR technologies. Two criteria for the granting of patents includes novelty and usefulness; therefore, building on patent data should reflect technological progress Strumsky & Lobo (2015).

In the following section we describe the patent exposure scores and provide some examples.

Section 4.6 compares patent exposure scores with other automation and AI indicators, and Section 4.7 shows their potential use for labor market analyses.

4.5.1 Description of exposure scores

We construct the 4IR exposure scores by calculating the mean of log 4IR patents per task. The 4IR exposure is higher if a task is mapped to many 4IR patents. Occupation scores are derived from the scores associated with its tasks, weighted by task importance. Therefore, occupations with many tasks, and with a high number of related 4IR patents (see Section 4.3.3), have a high 4IR exposure score. Activities such as interacting with computers and recording information have a high 4IR exposure score. Staffing organizational units and negotiating with others have the lowest 4IR exposure score. Appendix C.12 provides more details on task exposure scores. Occupation-level 4IR exposure scores range from a value of below three to above six. The lowest 4IR exposure scores are for occupations such as Meat, Poultry & Fish cutters/trimmers and Floor Sanders & Finishers. High 4IR exposure scores are for occupations such as Credit Authorizers, Data Entry Keyers, Computer Network Support Specialists, and Statistical Assistants. Table 4.2 provides an overview of 4IR exposure scores per SOC Career Clusters[5] from Career Technical Education (CTE) and the supplementary file for Chapter 4 contains exposure scores per occupation and task.

Table 4.2: 4IR exposure score per SOC career cluster.

SOC Career Clusters	Mean 4IR exposure
Information Technology	5.42
Finance	5.32
Marketing	5.17
Business Management & Administration	4.87
Government & Public Adminstration	4.68
Science, Technology, Engineering & Mathematics	4.63
Transportation, Distribution & Logistics	4.58
Health Science	4.53
Arts, Audio/Video Technology & Communications	4.51
Human Services	4.50
Education & Training	4.49
Hospitality & Tourism	4.40
Law, Public Safety, Corrections & Security	4.28
Agriculture, Food & Natural Resources	4.17
Manufacturing	4.15
Architecture & Construction	3.82

Note: A full table of exposure scores per occupation is provided in the supplementary data file for Chapter 4.

On average, low-skilled occupations have a lower 4IR exposure score than medium- to high-skilled occupations, whereas exposure to overall patents is higher for lower-skilled occupations (Appendix C.6 and Appendix C.7 provide more information on skill levels and exposure scores).

[5] `https://careertech.org/sites/default/files/Perkins_IV_Crosswalk_Table_5_SOC-ONET-Nontrad-Cluster-Pathway.xls`

The 4IR exposure scores grew fastest during the 1990s, and growth was particularly strong for medium- and high-skilled occupations. In recent years (between 2007 and 2013), 4IR exposure grew faster for occupations with currently lower scores (thus, some occupations caught up on 4IR exposure). This trend is true, particularly for low-education occupations (less than 20% of workers with college degrees). Nearly all occupations within the 50% of occupations with the highest exposure to non-4IR patents are low-education occupations. Within 4IR exposure scores, there is less differentiation between low and high-skilled occupations, even though most of the lower quintile occupations are low-educated. This supports the findings of Brynjolfsson & Mitchell (2017), that the impact of machine learning differs from previous waves of automation, impacting many tasks previously considered not automatable.

4.5.2 Patent exposure sub-scores and non-4IR exposure

Based on our patent occupation mapping it is possible to calculate patent exposure scores for various technology groups, as long as there is a definition available for which scores are related to which technology. Those exposure scores could include, for example, AI, or robot patents, as used by Webb (2019). For this chapter, we calculated sub-scores of the 4IR exposure scores. Our definition of 4IR technologies describes 367 distinct technologies (see Section 4.3.3 for an overview and Ménière et al., 2020 for detailed information on those technologies), and we provide technology exposure scores for each of the technologies. These technologies include categories such as healthcare, software, and user interfaces. Each of these categories comprises one or more technologies. A full dataset of 4IR exposure scores and sub-scores is available and Appendix C.4 provides some examples of occupation exposure to 4IR exposure sub-scores.

In addition we calculate exposure to (traditional) non-4IR patents. The analysis showed that 4IR exposure and non-4IR exposure are inversely related, and thus, that future technology impact likely differs from the past. Construction, manufacturing, transportation, agriculture, and hospitality occupations are mainly exposed to traditional technologies. Occupations in marketing, finance, administration, education, and law, for example, have a relatively strong exposure to 4IR technologies. Appendix C.5 provides an overview of 4IR and non-4IR exposure per SOC Career Cluster.

4.6 Direct comparison of 4IR exposure and other technology and automation indicators

The 4IR exposure score provides a direct indicator of an occupation's exposure to technologies of the 4IR. Whereas our patent-based 4IR exposure scores and the AI exposure by Webb (2019) account for existing technological capabilities, Felten et al. (2021) provides a forward-looking indicator of occupation exposure to AI technologies based on scientific progress. In both cases, high exposure scores might indicate which occupations are particularly impacted by new technologies. On the one hand, this impact may include automation, which is a replacement of tasks conducted by humans, with machines. On the other hand, there is a reinstatement effect, leading to the creation of new tasks. Additionally, tasks may be augmented through new technologies. Existing exposure scores do not differentiate between these types of change. There exist automation exposure scores which indicate how likely technologies are to perform tasks currently performed by humans. Such automation exposure scores include SML by Brynjolfsson & Mitchell (2017), computerization probability (CP) by Frey & Osborne (2017), and automation potential by Manyika et al. (2017). This section explores the relationship of the 4IR exposure score with the aforementioned scores in order to provide additional context to the 4IR exposure. Table 4.3 provides an overview of these indicators.

Table 4.3: Overview of automation, AI, and 4IR indicators.

	4IR exposure (this article)	AI exposure (Felten et al., 2021)	SML (Brynjolfsson & Mitchell, 2017)	CP (Frey & Osborne, 2017)	Automation potential (Manyika et al., 2017)	AI exposure (Webb, 2019)
Description	Exposure to technologies of the 4IR (most of which enable the introduction of AI)	Exposure to artificial intelligence technology	Suitability for machine learning technologies	Risk of being automated through AI and robotics	Share of activities potentially being automated	Exposure to AI patents
Level of analysis	Tasks conducted by occupations (O*Net)	Abilities required by occupations (O*Net)	Tasks conducted by occupations (O*Net)	Occupation level, scaled through abilities (O*Net)	18 performance capabilities (based on O*Net abilities)	Tasks conducted by occupations (O*Net)
Index basis	Relevant patents per task	Scientific AI advances, as described by the EFF, per ability	Online survey of SML characteristics per task	Expert evaluation of selected occupations automation potential	Expert evaluation of technology performance per capability	Relevant patents per task
Measure interpretation	Existing technology 4IR capabilities based on patents (current diffusion)	Existing theoretical AI capabilities based on scientific articles (possibly near future diffusion)	Potential future capability of ML (possible future diffusion)	Potential future capabilities of technology (possible future diffusion)	Potential future capabilities of technology (possible future diffusion)	Existing technology AI capabilities based on patents (current diffusion)

Most of the scores in our comparison refer to AI or machine learning, whereas the 4IR exposure score relates to technologies of the 4IR. 4IR technologies, as defined in our work, are not limited to machine learning patents. Instead, they include a range of technologies, including "core technology fields" IT hardware, software, and connectivity technologies, as well as "applications" technologies, such as remote health monitoring, predictive maintenance, and smart ATMs. However, even though not all those technologies directly relate to machine learning, they allow data generation, transmission, and analysis, thus contributing to an environment that enables the application of "smart" AI technologies. Chapter 2 explains that, with the increasing maturity of the 4IR, the concept increasingly centers around AI. Additionally, Frank et al. (2019a) explain in their widely used framework of Industry 4.0, that several base technologies (e.g., cloud computing) support the implementation of "smart" (or intelligent) front end technologies. Therefore, depending on the definition of AI, we consider that 4IR exposure scores may be comparable to the AI exposure indices. We decide not to rely on the "core AI" 4IR sub-exposure score (see Section 4.5.2 for more background on 4IR sub scores) for this comparison, which mainly relates to AI algorithms itself, but may not reflect the full potential impact of AI, e.g., through smart office solutions or autonomous vehicles.

4.6.1 Comparison of 4IR exposure and AIOI index by Felten et al.

Felten et al. (2021) create an occupation level AI occupational exposure (AIOE) score. They identify the most relevant AI capabilities based on the Electronic Frontier Foundation (EFF) AI Progress Measurement project and link those capabilities to occupational abilities, as described by O*NET. They calculate the AIOE as a relative measure, accounting for all abilities related to an occupation.

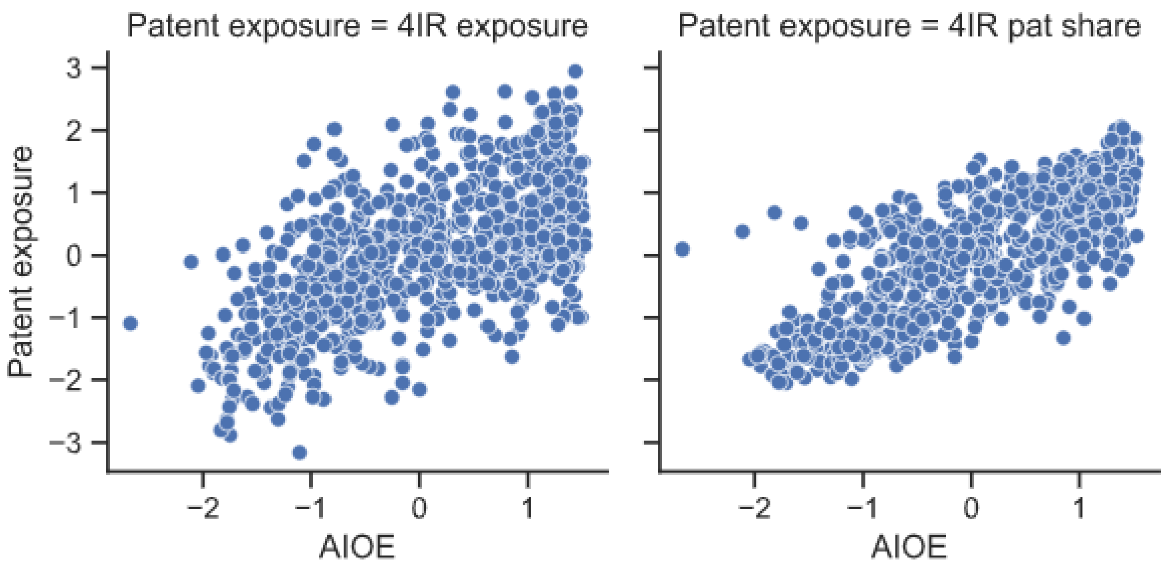


Figure 4.6: The graph maps the AI occupation exposure scores of Felten et al. (2021) to our patent exposure scores. Aside from the 4IR exposure (left), we include the 4IR patent share (right), which is the share of 4IR patents to all patents. All patent exposure scores are z-scores. Both graphs show some correlation with the following correlation values: Pearson r (4IR exposure) = 0.58 and Pearson r (4IR pat share) = 0.74.

Figure 4.6 shows a correlation between AIOE and both 4IR exposure and the share of 4IR patents within all patents related to an occupation. The correlation is stronger for AIOE and 4IR patent share. The strong AIOE and 4IR patent share correlation is not surprising, as both measures describe occupations which mainly relate to abilities and tasks which can only be conducted using AI or 4IR technologies. A high 4IR exposure score indicates that there is a high number of total patents related to the occupation. This could indicate that the 4IR technologies are already in a more advanced state. A high 4IR patent share does not necessarily indicate a high number of total patents, but only relates to the share of 4IR patents of all related patents. Some occupations, such as lawyers and compliance officers, have many tasks for which overall few relevant patents exist, and within those, many patents are about information and communication technologies related to the 4IR. Similarly, high AIOE occupations require many abilities which can be addressed by AI (but not by traditional technologies), independently if these technologies are already diffused. In addition, even though AI technologies have certain capabilities (such as image recognition), this does not necessarily mean that solutions have already been invented which can help workers conduct their tasks. Accordingly, Felten et al. (2021) describe their exposure score as being forward-looking.

Following this argumentation, the 4IR patent share describes the degree to which occupational tasks can be conducted or supported by 4IR technologies, whereas 4IR patent exposure takes into account the actual capabilities of current 4IR technologies.

4.6.2 Combining 4IR exposure score and suitability for machine learning

Brynjolfsson & Mitchell review occupational task descriptions with regard to whether they are potentially suitable for machine learning. A high SML indicates a high risk of automation (due to the various effects described; this does not necessarily mean decreasing labor demand of related tasks). A low SML may comprise two types of tasks: first, "old" tasks, which are not automatable; second, "new" tasks, which themselves result from the introduction of machine-learning technologies such as "Develop or apply data mining and machine learning algorithms."

We consider the 4IR exposure to describe the diffusion of technologies enabling the (front end) application of machine-learning technologies. The SML score presents the general automation potential once technologies are fully implemented. Therefore, SML and 4IR exposure scores can complement each other. Figure 4.7 maps occupations based on their 4IR exposure score and their SML score. Overall, there is a correlation, which indicates that high automation suitability is also related to a high number of 4IR patents. However, various occupations are outside this pattern; therefore we look at the graph along its four quadrants for interpretation of potential changes that occupations might undergo. We define quadrants along the mean values of the SML and the 4IR scores.

Quadrant I comprises occupations highly suitable for machine learning, and exposed to many automation patents. Through technology availability and automation suitability, these occupations might undergo significant changes in the future. They include consumer services, banking services, professional sales, web/digital communications, and travel/tourism occupations. Work activities which relate to this quadrant include "interacting with computers," "perform administrative activities," "selling and influencing others," and "analyzing data or information."

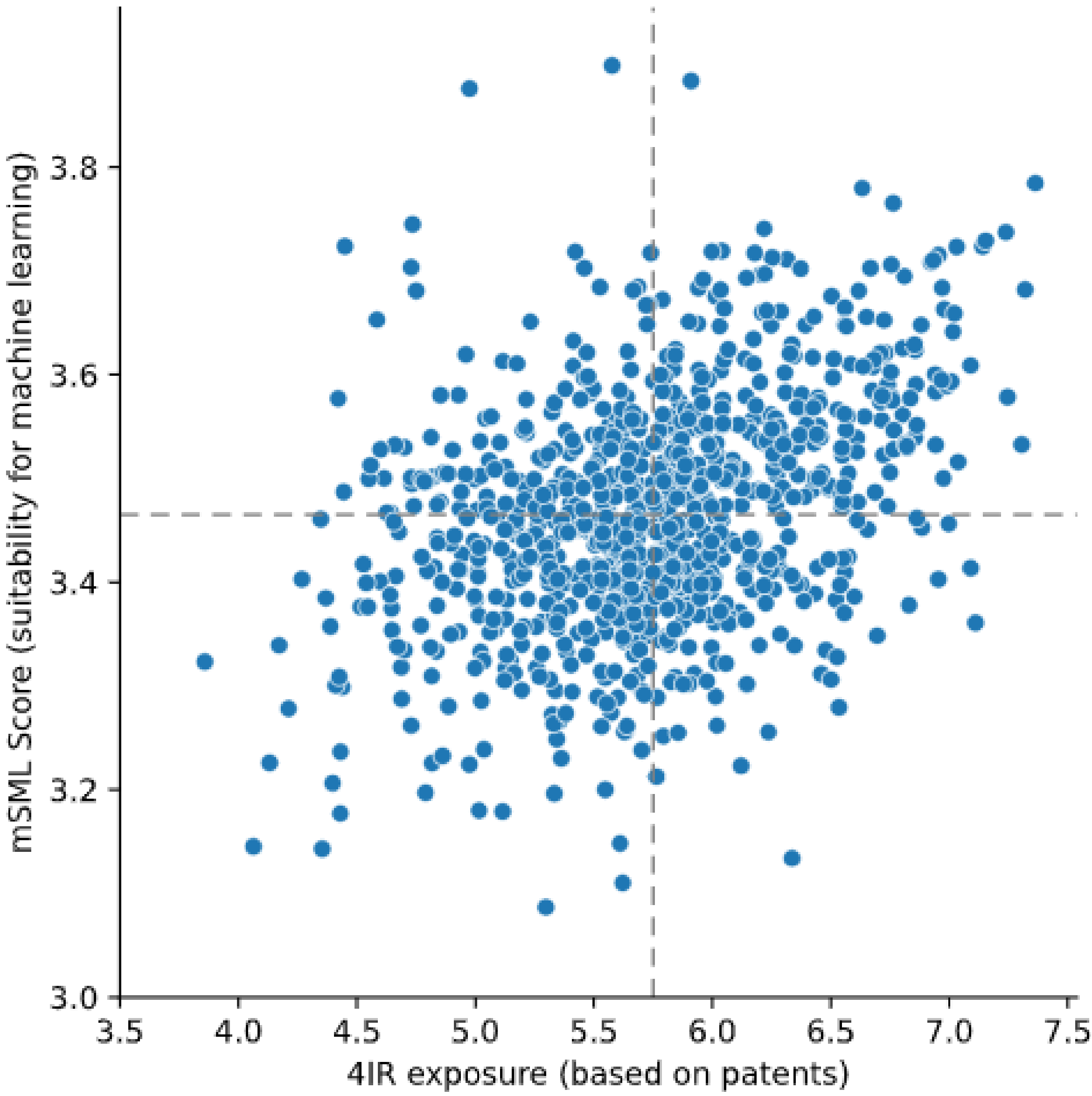


Figure 4.7: The graph maps 4IR exposure scores and suitability for machine learning (SML). The graph shows a slight correlation with Pearson's r = 0.36.

Quadrant II describes occupations highly suitable for automation but with a low number of patents. These occupations might include many tasks which require more innovation before automation can happen. This group includes early childhood development and services, power, structural & technical systems, lodging, agribusiness systems, and human resource management occupations. Work activities in this quadrant include "judging the quality of things, services, or people"; "establishing or maintaining interpersonal relationships"; and "guiding, directing, and motivating subordinates."

Quadrant III comprises occupations which might not undergo large changes through the 4IR. They only have a few tasks suitable for automation and only a few automation patents associated with those tasks. This quadrant includes construction, maintenance operations, and performing arts occupations. Work activities in this quadrant include "updating and using relevant knowledge"; "coaching and developing others"; and "resolving conflicts and negotiating with others."

Quadrant IV describes occupations exposed to a high number of 4IR patents but with many tasks with low SML. These occupations might either already have undergone changes due to the 4IR or mainly benefit from 4IR technologies through newly-created tasks or labor augmentation effects. This quadrant includes programming and software development, transportation systems or infrastructure planning, marketing communications, and professional support services occupations. Work activities in this quadrant include: "monitoring and controlling resources"; and "performing for or working directly with the public."

The four quadrants could also help to describe the pressure for change to the occupations. In Quadrant I, where there is a high 4IR exposure and a high SML, there might be a high pressure on the occupations to adapt to new technologies, change tasks, reskill, or upskill. In Quadrant IV, there is also a high 4IR exposure, but due to the low SML score, there might be a lower pressure to adapt to the new technology as activities might not become obsolete as quickly, but rather benefit through augmentation. Even though there is a high SML potential for Quadrant II occupations, the pressure on those occupations is lower due to the slow diffusion of 4IR technologies in those areas. Finally, Quadrant III is impacted least by current 4IR technologies. Further analysis revealed that since 2001 the 4IR exposure grew more for occupations in the Quadrants II and III than in the highly exposed Quadrants I and IV. This delayed 4IR exposure growth suggests that technology diffusion is slower for those occupations.

Overall there is a correlation between 4IR exposure and SML scores. This correlation is strongest within the least-educated occupations (less than 10% of workers with a college degree). The Pearson's r for low education is 0.40 vs. 0.29 for others). Low education occupations are primarily in Quadrant III and there are hardly any in Quadrant IV. High-education occupations (more than 80% college degree) mostly have average 4IR exposure scores and are distributed across all quadrants. Medium education occupations are slightly more present in Quadrants I and IV. Therefore, there is only a small share of low-education occupations in Quadrant IV, which possibly describes occupations benefiting from 4IR technologies through labor augmentation.

Overall, the four quadrants could be described as follows in the context of the 4IR. Quadrant IV includes 4IR augmented occupations, which are likely benefiting most from 4IR occupations and have a high chance of being suitable to use labor-augmenting 4IR technologies. A relatively high share of these are highly educated. Quadrant I is currently undergoing big changes, and Quadrant II will change when more technology is developed. Quadrant III occupations have many tasks which are little impacted by the 4IR.

4.6.3 Comparison of exposure scores to computerization probabilities by Frey and Osborn

A highly-cited risk of automation index has been provided by Frey & Osborne (2017). They asked a group of experts whether selected occupations may be automatable and used those ratings as a basis to calculate CP scores for all occupations. They consider 47% of US automation to have a high computerization probability. Frey & Osborne label high CP scores if their calculation leads to a risk of automation above 80%. The results were bimodally distributed, with most occupations on the high and low extremes. Arntz et al. (2017) show that accounting for job heterogeneity within occupations leads to a more normal distribution with more medium CP jobs, and thus much fewer "high-risk" occupations.

We compare the CP score with SML (Brynjolfsson & Mitchell, 2017) and find there is no correlation (see Appendix C.8 for the analysis). This is surprising, as both indices aim to estimate future automation potential for occupations. Whereas the SML score is based on a structured analysis of the suitability for machine learning at a task level, the CP has been calculated based on expert evaluations of overall occupations. Unlike SML scores, it appears that CP scores are closely related to education level. Arntz et al. (2020) describe that this represents the idea of skill-biased technological change, which was observed before the 1980s and has been replaced by routine-replacing technological change. One possible explanation of this bias towards low-skilled automation by Frey & Osborne is the method they use for calculating the CP scores. They identify automation bottlenecks comprising nine skills and capabilities (features) to calculate CP scores. However, theory suggests that task content has a higher relevance for predicting automation potential than the skills a machine can perform, as skills do not directly translate into automatable tasks (Acemoglu & Autor, 2011). The SML score is calculated based on task-level indicators, and Brynjolfsson & Mitchell (2017) find that AI-driven automation might impact a much broader range of tasks and occupations than described by previous models. Brynjolfsson & Mitchell found that, for example, even tasks from occupations that require social interaction, such as sales and customer interaction or family and community services, are suitable for machine learning. Those occupations have low CP scores, even when they are low-educated. In turn, there are high CP scores for manual tasks such as construction, production, and maintenance, which have low or medium SML scores. Finally, CP scores are very low for most high-skilled education, whereas Brynjolfsson & Mitchell (2017) find that those occupations also include a number of tasks that are suitable for machine learning. Information technology occupations, for example, have the lowest CP scores and medium SML scores.

Next, we explore the relationship of 4IR exposure and CP scores. Figure 4.8 shows that (similarly to the comparison of CP and SML scores) there is no overall correlation of 4IR exposure and CP. We evaluate the graph along three groups (highlighted as boxes A, B, and C) which comprise most occupations. In group A, we find that low CP occupations, which are also high-education occupations[6], generally having low exposure to non-4IR patents. These occupations include the fields of information technology; science, technology, engineering, mathematics; education and training; human services; health sciences; and arts, video technology & communications. Next, we describe occupations in groups B and C. Those have a high CP, are mostly low-skilled, and the 4IR exposure is inversely related to the exposure to non-4IR patents. Group B comprises those occupations with high 4IR exposure and medium non-4IR exposure, including mainly low-skilled administrative and clerical occupations in the fields of finance, marketing, and business management & administration. Group C comprises low 4IR exposure and high non-4IR exposure occupations, which are mainly manual occupations, including the fields of transportation and logistics, hospitality & tourism; agriculture, food & natural resources; and

[6]We consider an education level high if more than 80% of workers in an occupation hold a college degree. Results are similar if skill level is measured according to O*NET Job Zones.

manufacturing, and construction.

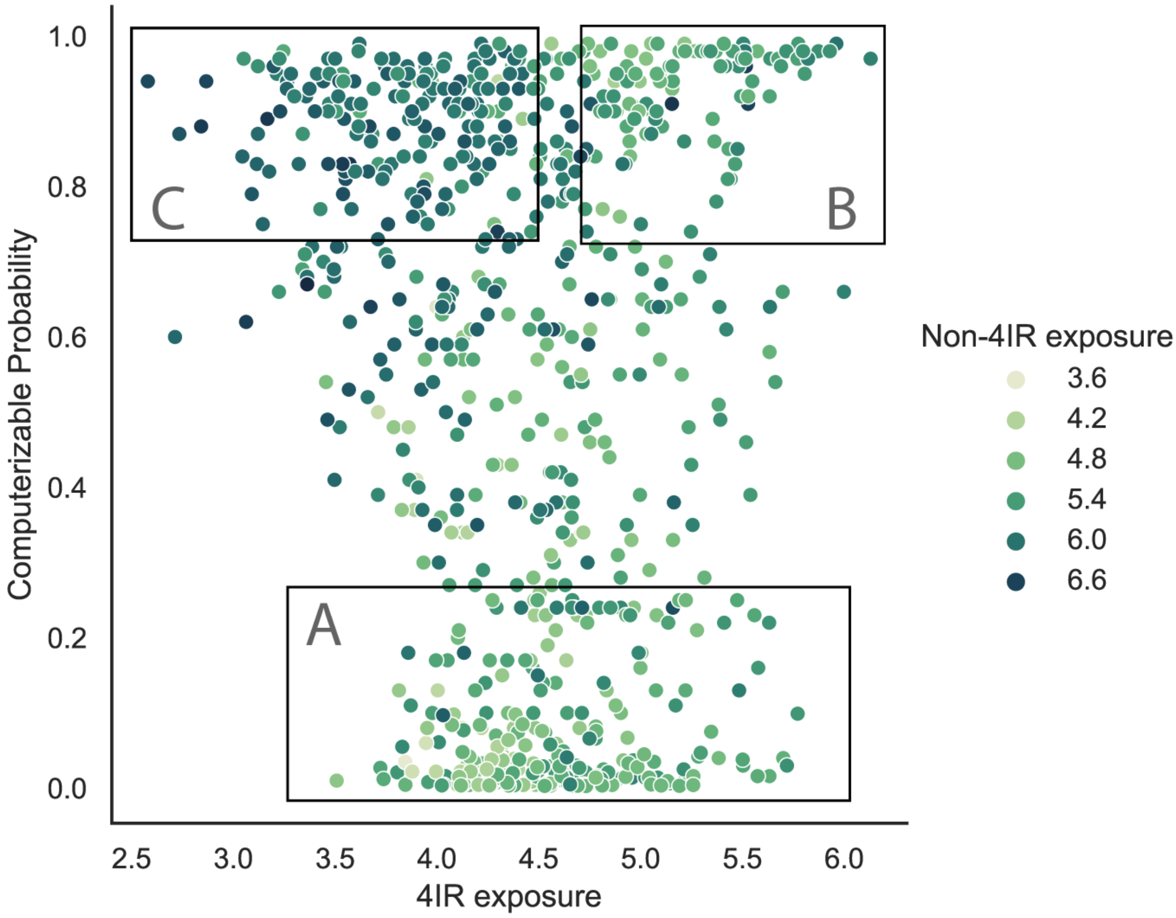


Figure 4.8: 4IR exposure scores and Frey Osborn automation probabilities. Section A comprises mainly high-skilled occupations. Section B comprises mainly low-skilled occupations. Section C comprises exclusively low-skilled occupations. Overall CP scores have a weak correlation with 4IR exposure (Pearson's r = -0.13) but stronger correlation to Non-4IR exposure (Pearson's r = 0.50).

Overall, the two main predictors for high CP scores are low education and a medium or high exposure to non-4IR patents. While an interpretation of this observation is difficult, it raises the question of whether the expert evaluation of automation probability is biased towards skill-biased technological change and ignores the disruptive impact of machine learning, which might significantly differ from previous waves of automation.

4.6.4 Comparison of exposure scores to MGI automation scores.

The MGI Manyika et al. (2017) conducted a detailed evaluation of the automation potential of occupations, in which they identified which sets of capabilities are required for each of over 2000 work activities (O*NET DWAs). For each of these capabilities they evaluated the technological readiness to automate these capabilities until 2030, which served as a basis for the evaluation of which share of jobs can be automated per occupation.[7]

We compare patent exposure scores with MGI automation potential estimates in Figure 4.9. The correlation shows that exposure to non-4IR patents relates to higher automation potential, whereas automation potential is lowest for medium to high 4IR exposure occupations. Whereas SML and CP scores describe general automation potential, the MGI describes which automation might actually be implemented by 2030. The results suggest that most automation potential relies on non-4IR technologies, whereas occupations with higher exposure to 4IR patents (except for very high exposure) seem more prone to short-term automation. They include administrative support, sales and service, quality assurance, and health informatics occupations, and they also have high SML scores (see Appendix C.9 for more details). These high 4IR and occupations are possibly the first to feel the automation impact of 4IR technologies. If we assume that high 4IR exposure is due to an advanced development of associated technologies, this is in line with Kogan et al. (2020), who highlight that some time is required before new technologies have an impact on jobs.

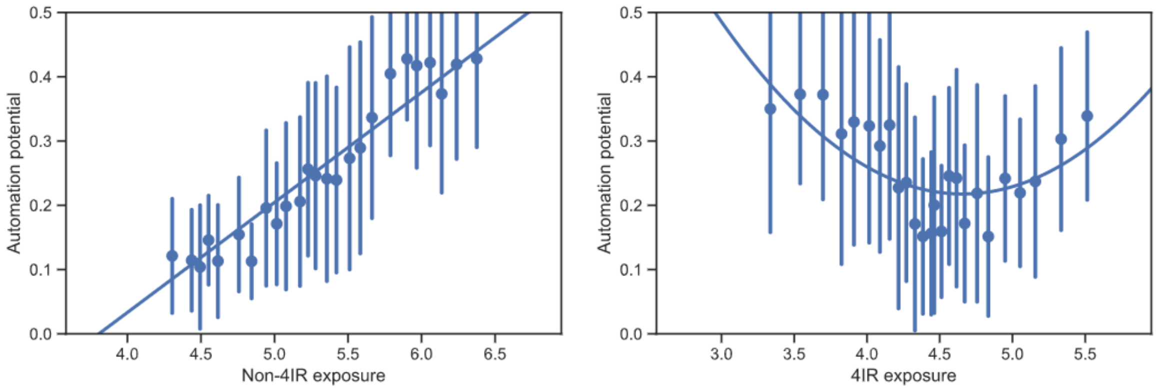


Figure 4.9: Patent exposure score compared with MGI automation potentials. The right graph shows that exposure to non-4IR patents is related to higher MGI automation potential (Pearson's r = 0.64). The left side shows that medium 4IR exposure relates to lowest automation potential. The graphs show data for 545 occupation in 25 bins. The vertical lines indicate the standard deviation of occupations per bin.

4.6.5 Comparison of 4IR exposure and Webb's exposure scores

Webb (2019) created a mapping of occupations to AI, software, and robot patents. Webb uses a different NLP approach for mapping patents to jobs than the one described in this chapter (for a detailed comparison see section 4.2). Additionally, the method for construction of the exposure

[7]The dataset can be explored through a Tableau public dataset at https://public.tableau.com/profile/mckinsey.analytics#!/ vizhome/AutomationandUSjobs/USAutomationlandscape, accessed on April 10, 2021.

score differs. Whereas Webb (2019) creates separate exposure scores for 3 technologies, this chapter provides one main exposure score to patents of the 4IR and additionally more than 300 exposure sub-scores for a broad range of 4IR technologies, such as computer-aided design (CAD), NLP, and smart office solutions.

Figure 4.10 compares Webb's exposure scores with our 4IR and Non-4IR patent exposure. The results show that there is no correlation for AI patents and only a slight correlation between software exposure and exposure to non-4IR patents. Also, considering 4IR sub-scores related to AI and software (see Appendix C.4.2), we cannot find a strong correlation of Webb's scores and our 4IR exposure scores. There are various reasons for the non-correlation. First, the fundamentally different algorithm of mapping patents and tasks can be a reason why the exposure scores per occupation differ. Second, the scores created (4IR, vs. AI, software, robot) are different. Whereas our 4IR definition is very broad, Webb's definition relates to specific technologies. Further, Webb selects technology patents via text search, while our approach relies on CPC classifications of patents. Third, the aggregation method for task scores to occupation scores differs between the two approaches.

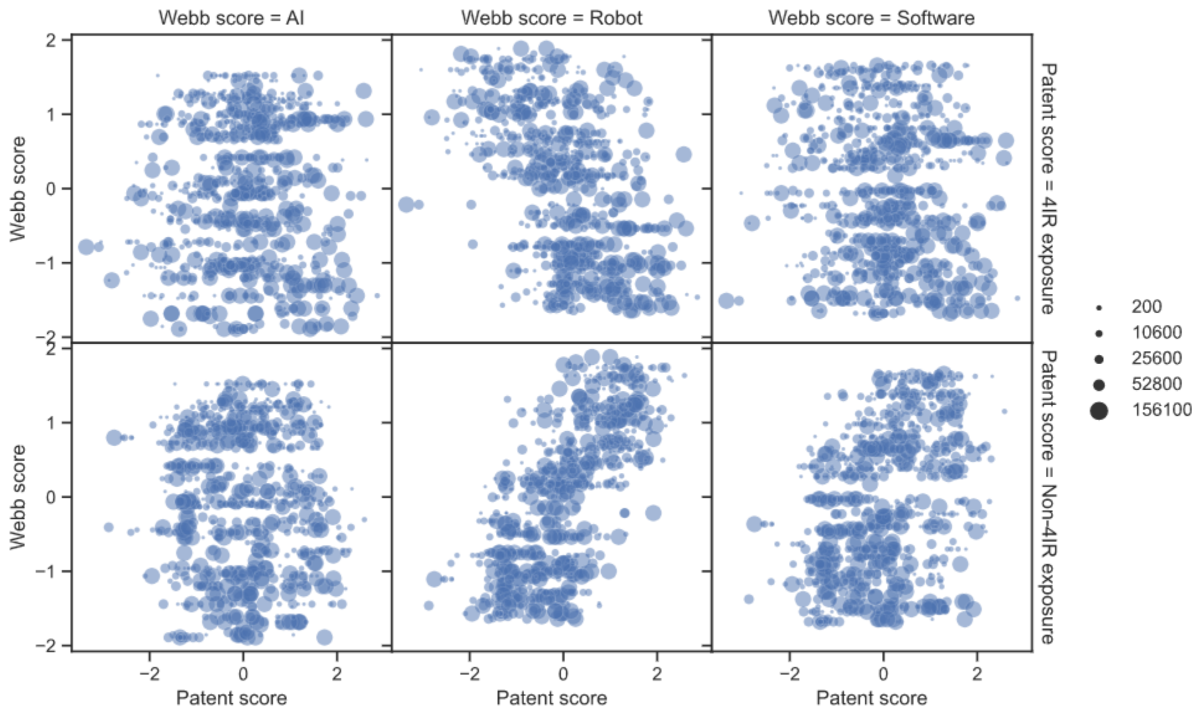


Figure 4.10: Patent exposure scores compared with Webb (2019) exposure scores. For comparison we use z-scores of patent exposure scores and Webb's percentile scores per technology. Bubble sizes indicate numbers of jobs. Pearson's r for the first row are 0.05, 0.44, -0.84, for the second row 0.05, 0.71, 0.42.

4.7 Impact of technologies on the labor market and patent exposure

The previous section described different scores indicating exposure to AI or 4IR technologies, some of which have been useful for labor market analyses. This section describes how the 4IR exposure scores can be used for analyses in order to contribute to the debate on the impact of 4IR technologies on the labor market.

In order to explain the impact of technologies on jobs, a leading hypothesis has been that new technologies are biased in favor of skilled workers, which is called skill-biased technological change (SBTC). When the SBTC hypothesis became unable to explain all major changes in the labor market, Acemoglu & Autor (2011) and Autor & Dorn (2013) developed a more refined explanation, the routine replacing technological change (RRTC) hypothesis. RRTC builds on the idea that computers are particularly efficient at performing clearly defined "routine tasks", and this leads to a decline in demand for human labor to perform these tasks. As the share of routine tasks is highest among middle-paid occupations, computerization was accompanied by a hollowing-out of the wage structure, with declining shares of middle-paid jobs, known as employment polarization (Goos et al., 2014; Autor et al., 2006; Oesch & Menés, 2011).

However, the RRTC theory does not explain fully the impact of new technologies on jobs, as automation potentials do not necessarily translate into employment losses, due to various macroeconomic adjustment processes. Using regional-level information for the US, Autor & Dorn (2013), for example, find no net negative employment effects of computerization. Also, Gregory et al. (2016) find that computerization in Europe did not reduce employment but increased it. They show that significant replacement effects exist, which are overcompensated by productivity effects. So, the net employment effect is positive, despite large capital-labor substitution. Acemoglu & Restrepo (2022) show in their seminal theory that the effects of new technologies crucially depend on the type of technological progress. They differentiate between different types of technological progress: (1) automation, i.e., machines learn to perform tasks which previously only humans could do, (2) deepening, i.e., machines become better at tasks already automated, and (3) reinstatement, i.e., humans take over new tasks. In addition, the diffusion of new technologies creates demand for workers who produce or maintain those technologies (capital accumulation effect). The relative size of these effects and their interaction determine the overall effect of automation on the labor market.

Patent-based exposure scores can provide a valuable indicator for technological change, and have been used in recent labor market analyses. Mann & Püttmann (2017) use patent exposure per industry and identify a positive overall impact of automation patents on employment. Webb (2019) found a negative impact at an occupation level. Kogan et al. (2020) showed that recent breakthrough technologies are more related to cognitive tasks than previous waves of technological change. Acemoglu et al. (2020) find that establishments with occupations exposed to AI patents posted more AI vacancies but fewer non-AI vacancies.

In the following section we show basic correlations and compare our patent exposure scores with other patent-based indicators to help researchers better understand the potential use of the 4IR exposure score.

4.7.1 Patent exposure per wage percentile and education

We analyze patent exposure per education level and per wage percentile to compare patent indicators with theory on SBTC and RRTC. Our analysis shows that exposure to (traditional) non-4IR patents follows expected patterns, whereas 4IR patents show different patterns.

For the analysis of the exposure per education level we rely on two indicators. First, we show the exposure per education level of workers in occupations. Therefore we rely on O*NET data on the education level for workers per occupation. Second, we use O*NET information on Job Zones, an indicator comprising the education and experience required to work in an occupation. Our analysis shows that the exposure to non-4IR patents is highest for low-education skilled occupations (see Appendix C.6) and is in line with the theory of SBTC. In contrast, 4IR patent exposure is highest for medium-to-high-skilled occupations.

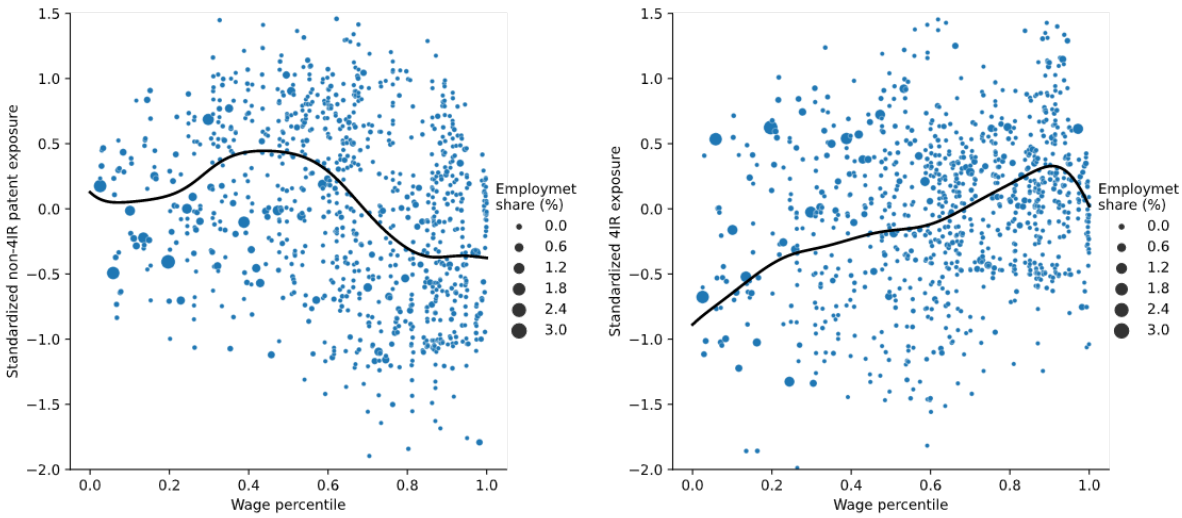
Next, we analyze patent exposure score per wage percentile. Therefore, we build on U.S. Bureau of Labor Statistics (BLS) occupation and employment statistics data[8] to extract occupation shares of workforce and wage data. Figure 4.11 shows exposure to non-4IR patents and to 4IR patents per wage percentile. Non-4IR exposure is highest for middle-wage occupations. This supports the ideas of RRTC, where (traditional, non-4IR) technologies mainly address routine-heavy medium-wage occupations. Interestingly, this observation is not confirmed at a task level. Section 4.4.3 shows, that routine cognitive tasks have higher non-4IR exposure than non-routine cognitive tasks, but non-routine manual physical tasks have higher exposure scores than routine manual tasks. On the one hand, this may indicate that RRTC theory is particularly appropriate for cognitive tasks, where computers have mainly been programmed to conduct routine tasks. On the other hand, this could indicate that routine intensive occupations are more likely being automated by technologies, whereas non-routine occupations are more likely to benefit from technologies, e.g., through augmentation. The 4IR patent exposure follows a different pattern and correlates with mean income.

Overall this analysis shows different patterns of technology exposure for 4IR patents than for non-4IR patents. This suggests that the impact of 4IR technologies may differ from previous industrial revolutions, and thus confirms the findings of Brynjolfsson & Mitchell (2017), who found that future waves of automation might affect different occupations which have been considered non-automatable in the past. The exposure to wage varied for different 4IR exposure sub-scores. 4IR-manufacturing patents showed a curve similar to non-4IR patents, with highest exposure for medium-wage occupations and lower exposure for low- and high-wage occupations. 4IR-software patents showed a curve similar to the 4IR exposure curve, and 4IR-AI patents had an even stronger skew towards high-wage occupations (see Appendix C.10 for an overview of 4IR exposure sub-scores exposure per wage percentile). This shows that 4IR sub-scores can be valuable for detailed assessments of the impact of technologies on occupations.

4.7.2 Patent exposure and job growth

In this section we calculate patent exposure scores based on 2012 task descriptions and 2012 patent data. We evaluate the relation of these exposure scores to job growth between 2012 and

[8] https://www.bls.gov/oes/



(a) Exposure to Non-4IR patents

(b) Exposure to 4IR patents

Figure 4.11: Exposure to patents per wage percentile. We apply a locally smoothed regression following Acemoglu & Autor (2011), using a bandwidth of 0.08 with 100 observations.

2018. The analysis described in this section should provide an initial idea of potential impact patterns of 4IR patents on jobs, in order to compare 4IR exposure scores with other patent exposure scores. For more reliable insights, an analysis with micro data needs to be conducted.

The analysis builds on BLS occupation and employment statistics data from 2012 and 2018 at an occupation level. We controlled for occupation industry shares in 2012, average education level per occupation, and mean annual wage. Figure 4.9 showed a convex relationship of 4IR exposure and MGI estimated automation potential. Therefore, we included the quadratic value of 4IR exposure scores as an independent variable.

Time is required for an invention to have an impact on the labor market. Webb (2019) argue that it can even take decades for new technologies to impact the labor market, and Kogan et al. (2020) find the largest impacts of patents on the labor market five to 20 years after patent filing. Our analysis supports these findings. We conduct an analysis with exposure scores from different years and find that coefficients were lower for more recent patent data (see Appendix C.11 for regression results for different exposure scores). Patent exposure from 1992, which is 20 years before the start date of the labor market data we use, showed the highest coefficients for job growth in 2012–2018. This accounts for the time required for technologies to mature.

Regression analysis (see Table 4.4) suggests an overall negative and concave relationship of patent exposure and job growth. The overall negative relation is in line with the literature, where Kogan et al. (2020) identified a negative impact of breakthrough patents on job growth and Webb (2019) identified a negative correlation of exposure to robot and software patents and job growth. Our analysis builds on occupation-level labor market data, and the results have to be confirmed with micro data analysis. In addition, the regression does not account for various effects of automation on the labor market. Therefore, even though our analysis shows a negative relation of 4IR exposure on job growth, there may be a different overall impact, considering, for example, the impact of deepening automation and capital accumulation.

Table 4.4: Exposure to 4IR patents squared and change in employment 2012-2018.

	(1)	(2)	(3)	(4)	(5)	(6)
4IR $exposure_{92}$		-0.17***	-0.11***	-0.13***	-0.05	-0.05
		(-4.61)	(-2.85)	(-3.00)	(-1.18)	(-1.21)
4IR $exposure_{92}{}^2$			-0.09***		-0.11***	-0.11***
			(-3.54)		(-4.23)	(-4.29)
Job Zone	0.24***			0.17***	0.16***	0.21***
	(3.00)			(2.90)	(2.81)	(2.67)
LOG Wage	-0.02					-0.06
	(-0.36)					(-0.93)
Industry share	Yes	No	No	Yes	Yes	Yes
Adjusted R^2	0.083	0.028	0.044	0.095	0.117	0.117

Note: Analysis of 704 observations based on BLS labor market data from 2012 and 2018. Industry share relates to the industry share of the occupation in 2012. 4IR exposure scores are based on 1992 patent exposure per occupation. * p<0.10, ** p<0.05, *** p<0.01.

4.7.3 Patent exposure scores as indicators for labor market analysis

This section discusses the potential use of our 4IR exposure scores and patent exposure scores by Webb (2019) and Kogan et al. (2020), for labor market analysis. Webb (2019) divided their analysis into three parts. They calculated exposure scores for robot, software, and AI patents. Kogan et al. (2020) created an occupation-level exposure score for breakthrough patents, which differ in text content from previous patents. These categories are different from our 4IR patent exposure scores, making a direct comparison impossible.

Figure 4.10 indicates a slight correlation of robot and software exposure by Webb (2019) to non-4IR patent exposure. Comparison with labor market indicators provides a similar overall picture. Exposure to software, robots, and 4IR scores is inversely related to education level (see Appendix C.6) and shows a concave relationship to exposure per wage percentile, with highest scores for medium-wage occupations. The AI exposure score by Webb has, like the 4IR exposure score, highest exposure for high wage occupations. Exposure per education differs slightly: whereas AI exposure correlates with education level, 4IR exposure is concave, with highest values for medium-high education levels. Even though some of the results overlap and suggest that the overall mapping approaches follow similar patterns, the 4IR exposure has a different scope than the patent exposure score provided by Webb (see section 4.6.5).

Similar to our approach, Kogan et al. (2020) uses text embeddings for mapping patents to occupations. Breakthrough patents have a similar exposure per wage percentile than non-4IR patents, with highest exposure for medium-wage occupations, and task-level exposure scores show similar patterns to 4IR patent exposure scores for some task types (see Section 4.4.3 for analysis of exposure per task type). The vast majority of patents in the 1980 to 2002 period are non-4IR patents (see Figure 4.3), so it is not surprising that the exposure per wage percentile is similar to the exposure to non-4IR patents. However, a comparison of both scores needs to be reviewed with caution, as the approaches vary in a number of aspects. On the one hand, Kogan et al. present a different exposure score. They define breakthrough patents, which differ in text content from previous patents, and thus the nature of breakthrough patents changes over time. This changing scope may be appropriate for the long time frame of their analysis. Our score relates specifically to a set of selected 4IR patents. Some 4IR technologies were introduced decades ago and may no longer fall under the definition of breakthrough patents. On the other hand, Kogan et al. (2020) calculate the exposure score at an occupation level, whereas our analysis is conducted at a task level. As described in Section 4.3.2, this leads to a number of differences, such as that the task-level mapping accounts better for the high variety of tasks conducted within an occupation.

In general, it is difficult to compare the exposure scores of the different studies, as they have a different focus. However, our analysis provides a patent occupation mapping which allows us to calculate not only 4IR exposure scores but also exposure to any other technology. Aside from the 4IR exposure and 4IR exposure sub-scores, researchers can, for example, calculate exposure to AI, software, robot, or breakthrough patents, as long as patent technology concordances are available.

Both, Webb (2019) and Kogan et al. (2020) use their exposure scores to analyze the impact of technologies on the labor market. They analyze time frames from 1980 to 2010 and 1850 to 2010 respectively. They find an overall negative impact of technologies on the number of jobs. Their analyses are based on occupation descriptions from a given time, e.g., Webb (2019) used O*NET task descriptions from 2017. While those descriptions offer accurate insights into the tasks conducted by an occupation in 2017, and thus allow us to identify the relevant patents for those tasks, they might not appropriately reflect the task descriptions of these occupations some decades ago. Therefore, the further the analysis goes back in time, the more caution is needed when reading the results. Similarly, our analysis is based on current task descriptions by O*NET (in addition we calculate exposure scores based on 2012 task descriptions for the regression analysis). When using historic patent exposure scores, e.g., from 1992, we consider these scores valuable for explaining when technologies that shape current jobs were invented (see also Section 4.7.2). As the patent occupation mappings are based on current task descriptions, it is possible that exposure scores do not accurately reflect patent exposure of that occupation in 1992. Therefore, we believe these exposure scores, which are based on current task descriptions, are most useful for analyzing the recent changes in the labor market. For the regression analyses of job growth 2012 to 2018 we calculated patent exposure scores based on 2012 task descriptions (our general exposure scores are based on 2020 task descriptions provided by O*NET).

4.8 Conclusions and future work

The aim of this chapter is to better understand the exposure of occupations to technologies of the 4IR. Several existing indicators describe the theoretical automation potential or future exposure potential of occupations. We introduce an indicator reflecting actual technology diffusion, based on patent data. This paper presents a method for mapping patents to tasks and introduces an occupation and task-level indicator of exposure to patents of the 4IR (4IR exposure score). We refine existing approaches to better account for task-level differences in patent exposure and the context in which an activity is conducted (e.g., diagnose *machine* condition vs. diagnose *patient* condition). We therefore consider that this approach offers a highly valuable contribution towards mapping patents to tasks and occupations.

Occupations with higher exposure scores may, for example, be more impacted by 4IR technologies. The analysis shows that ratio of exposure to 4IR and non-4IR patents differs per occupation. Occupations with many manual tasks, such as manufacturing and construction, have high non-4IR exposure and low 4IR exposure, whereas many non-manual occupations, such as finance and marketing occupations, have a higher ratio of 4IR exposure.

The 4IR exposure score is also valuable as a complementary score to other technology or automation scores. For example, comparing theoretical and actual technology exposure can provide insights into which occupations might undergo changes through current technologies versus future diffusion.

This direct measure of technological progress can provide highly valuable data for further exploration of the impact of technological change on employment (Mitchell & Brynjolfsson, 2017) and may serve as a source for labor market analysis to explore impact patterns of technologies on jobs.

We compared our 4IR exposure scores with labor market indicators and found that exposure to non-4IR patents is highest for medium-wage occupations, and that 4IR exposure is highest for high-wage occupations. Further, regression analysis showed a negative (concave) relation of 4IR exposure to job growth. Patent exposure 10 and 20 years ago showed higher coefficients on the impact on job growth than more recent patent exposure. The gap may reflect the time between invention and technology diffusion and is in line with findings of Kogan et al. (2020). Further analysis with micro data is required to confirm these findings. To estimate the overall impact on the labor market, more complex modeling is required, e.g., considering the effect of deepening of automation or capital accumulation.

Acemoglu & Restrepo (2019a) observed that different technologies may have different impact patterns on the labor markets. Therefore, differentiating between 4IR technologies may offer additional value for labor market analyses. Researchers can build on our mapping for technology-level analysis. On the one hand, we provide technology-specific exposure scores (e.g., CAD, augmented reality for surgery, and smart office technologies). On the other hand, our mapping of patents to tasks allows researchers to build any other exposure scores, such as robots, or breakthrough patents, as long as a patent technology mapping is available. Also, patent data is available at firm level and allows for time-varying measures.

Our work provides an occupation (and task)-level indicator of 4IR patent exposure. Patents describe inventions, and not all inventions have an equal impact. Future work could thus further improve the indicator by accounting for a patent's impact. The count of patent citations is frequently discussed as potential measure for novelty and social usefulness, but its validity is ambiguous (Strumsky & Lobo, 2015). Another approach is described by Kelly et al. (2021), who describe "breakthrough patents" which significantly differ in text content from previous patents and thus might have particularly high impact.

Our approach builds on occupation and task description data provided by O*NET. We take advantage of its extensive and hierarchical descriptions of occupational activities and tasks. Future work could rely on additional information provided by O*NET. For example, at a task and occupation level, the dataset indicates which technologies and tools are used, such as word processing software or programmable logic controllers. Building on this information may provide information on inventions related specifically to labor augmentation. The O*NET database describes occupations in the context of the US labor market. There exist concordance tables, which can help to use the patent occupation mapping in other contexts. These might provide additional accuracy to directly map patents to those regional occupation descriptions, if regional databases with similar hierarchical structures exist.

Chapter 5

Binary patent classification methods for few annotated samples [1]

5.1 Introduction

New technologies play a key role for economic development and wealth Acemoglu (2009). This covers a large and currently very active debate on the effects of automation technologies on the labor market (Mokyr et al., 2015; Autor, 2015). The economic debate often relies on binary classifications to analyze the effects of new technologies on the economy. For example, economists study how automation potentials of new technologies affect workers (Mann & Püttmann, 2017), whether workers are susceptible or non-susceptible to automation (Frey & Osborne, 2017), how innovation vs. imitation affects the economy (Segerstrom, 1991), or the role of process vs. product innovations for firms (Bartel et al., 2007). Patent texts are well recognized indicators to describe the technological state of the art. As such, patents contain relevant information to measure the mentioned concepts, e.g., by classifying patents that refer to automats vs. non-automats (Mann & Püttmann, 2017). This is often complex due to the ambiguity of the concepts and the similarity of patents that refer to distinct categories. Being able to assign patents to unique categories allows linking them to other economic data. Until now there only exist few and very broad concordances that allow assigning patents either to technologies (Schmoch, 2008) or to industries (Van Looy et al., 2015). But these classifications are rather broad.

In this chapter, we compare binary patent classifiers, which may be used for analyzing technological change. The main challenge not only lies in the complexity and ambiguity of the concepts but also in the sample size, because human coders often require significant time for classifying such cases. These algorithms may be applied to other cases with complex and ambiguous binary classes and few training data.

The rest of this chapter is organized as follows: Section 5.2 provides a description of the underlying patent data and Section 5.3 our machine learning algorithms. We present and discuss our insights in Sections 5.4 and 5.5. Section 5.6 concludes.

[1] This chapter largely represents the following article: Meindl, B., Ott, I., Zierahn, U.; Binary Patent Classification Methods for Few Annotated Samples. 1st Workshop on Patent Text Mining and Semantic Technologies. 2019

5.2 Patent data

We aim at developing a classifier which is able to handle cases with high ambiguity / large overlap. Additionally, it should provide sufficient precision even with low numbers of examples, as hand-classification is costly when human coders have to read large parts of a patent to classify it. In order to develop algorithms which are suited for such cases, we focus on data which contains a binary outcome variable with ambiguous classes. In particular, we rely on patent data, which is particularly suited to study technological change. Moreover, we focus on two selected Cooperative Patent Classification (CPC) classes as our outcome variable to analyze a binary outcome. We focus on two CPC classes which are potentially hard to differentiate for an algorithm in order to train algorithms which are suited for ambiguous cases.

We motivate the choice of our patent sample by the recent interest in robot technologies and the widespread interest this technology field receives in current public and economic debate (e.g., Acemoglu & Restrepo 2019b; Dauth et al. 2017; Graetz & Michaels 2018). The United States Patent Classification (USPC) class 901 - robot - has been mapped to the CPC with the most recent update being from 2012[2]. Most statistically relevant CPC classes related to the USPC class 901 are G 05D, A 61B, G 05B, B 25J, B 23K, B 06B, and G 01N. Most similar from a technological perspective are thus CPC classes G 05B and G 05D.[3]

We thus restrict our sample to the two sub-classes G 05D and G 05B and use these two classes as a natural delineation to train binary classifiers. G 05D refers to systems for controlling or regulating non-electric variables, e.g., for welding, pressure control, and so on. G 05B relates to control and regulating systems which are "clearly more generally applicable." The fact that G 05B refers to systems which are more generally applicable, whereas G 05D refers to those that control or regulate only non-electric variables, creates a certain ambiguity. Such an abiguity is often present in the economic examples noted above: Without a sufficient training it is often hard to assess for a human, whether a patent is sufficiently generally applicable to be classified as G 05B instead of G 05D. this challenge is similar to the economic samples described in the introduction, such as Mann & Püttmann (2017) who define an automat as a device that carries out a process *independently*. Their classification task (i.e., automats vs. non-automats) involves ambiguity, as devices typically require at least some kind of human involvement, so that the interpretation of *independence* remains a subjective assessment of the human coders.

Another objective of the algorithm is to achieve high accuracy with low sample data, as hand-classification is costly when human coders have to read large parts of a patent to classify a patent. Mann & Püttmann (2017), for example, build their analysis of patents describing "automats" on 560 hand classified patents. We will compare our algorithms for different sample sizes, to evaluate requirements on sample sizes for potential annotation tasks. We start with the smallest sample size of 100 patents only, which may be mainly relevant for early validation of the feasibility of an idea, and as an input for active learning, which is an early training of the model to select further patents for more efficient classification. Next, we include datasets with 250 and 500 patents. We expect 500 patents to be a potential minimum sample size for analysis, e.g., similar to Mann & Püttmann (2017). Finally we build larger datasets of 1,500 and 5,000

[2]USPC has been deprecated in favor of CPC.
[3]compare https://www.uspto.gov/web/patents/classification/ cpc/pdf/us901tocpc.pdf.

patents, to evaluate the benefit of higher investment of resources for annotation.

We draw our sample data from the USPTO-2m patent abstract dataset (Li et al., 2018), which is commonly used for patent classification benchmarking. For each dataset, we draw 50% each G 05D and G 05B examples, whereas patents with both labels are considered as G 05D. For evaluation, we use 250 randomly drawn patents of each category.

5.3 Patent classification algorithms

In our analysis, we compare different approaches for patent classification. Mann & Püttmann (2017) use a multinominal naive Bayes (MNB) algorithm to identify patents describing an "automat." Based on 560 manual annotations, they achieve a correct prediction of 80% of patents. One valuable feature of MNB is the ability to interpret results. Mann & Püttmann (2017), for example, extract tokens typical for "automats." Support vector machiness (SVMs) may outperform MNB (Joachims, 1998; Fall et al., 2003) or other approaches such as k-nearest neighbor (Krier & Zacc, 2002) for text classification, and also allow for feature extraction. Benites et al. (2018) performed best at the ALTA 2018 patent classification task, using a method based on SVM.

Further approaches for patent classification are based on neural network models (Abbas et al., 2014). Grawe et al. (2017) and Li et al. (2018) describe the potentially high precision of neural networks for patent classification and Zaghloul et al. (2009) find that they may outperform SVMs, particularly for shorter texts. Some recent advances in the field of natural language processing rely on pre-training and fine-tuning neural network models, e.g., Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) and Universal Language Model FIne-Tuning (ULMFiT) (Howard & Ruder, 2018). Lee & Hsiang (2020) outperformed previous approaches of patent classification using patent data to pre-train a BERT convolutional neural network (CNN) model.

Pre-training models such as BERT require extensive computational resources. Therefore, Li et al. (2019); Kumar & Tsvetkov (2019) describe alternative models, achieving a significant reduction in computational resource requirements with nearly similar performance. A similar model, called Language Modelling with Approximate Outputs (LMAO) is implemented in the spaCy library[4].

For our analysis, we want to a compare binary classification performance of a pre-trained CNN with alternative approaches. Naive Bayes has been used as a baseline for similar efforts (Mollá & Seneviratne, 2018). We use a Bernoulli naive Bayes (BernoulliNB) classifier as a baseline for our work, which accounts particularly for the binary decision. Further, we evaluate an SVM based model, which has been successfully used for various patent classification tasks. Also, we implement a random forest classifier (RandomForest) and a k-nearest neighbor (k-NN) for comparison.

BernoulliNB, SVM, RandomForest, and k-NN classifiers are implemented using Scikit-learn. Therefore, we lemmatize words (using NLTK[5]), remove stopwords, and extract the most relevant words per document through term frequency–inverse document frequency (TF-IDF), using

[4]https://spacy.io/

[5]https://www.nltk.org

unigrams as well as bigrams. (D'hondt et al., 2013) finds that TF-IDF analysis using bigrams (instead of unigrams only) may lead to higher accuracy, as it accounts for complex multi-word expressions. We use the Scikit-learn model selection, GridSearchCV, for optimization of model parameters.

We implement a CNN based classifier using spaCy, which allows LMAO pre-training. SpaCy's combination of high accuracy and speed is especially relevant for patent classification, as it enables research on large patent data sets with reasonable resources. Our analysis includes two spaCy based approaches. First, we use the default large English language model. Second, we use the same model pre-trained with patent data ($\text{spaCy}_{\text{pre}}$). To assure high contextual relevance of pre-training, we use the 25,212 patents in the class G 05 from the USPTO-2m dataset. The algorithm ran 200 passes over the dataset until the loss function did not further decrease. In addition, we run the same models with the software prodigy[6]. Prodigy builds on spaCy and allows for straightforward implementation of natural language processing analysis and annotation. It provides a simple application programming interface (API) requiring only basic knowledge in programming. We want to evaluate whether using the tools compromises performance compared to a manual implementation of spaCy.

5.4 Results

A comparison of the different algorithms shows that the CNN model outperforms remaining models (see table 5.1) for each sample size. The results show the advantage through pre-training decreases with sample size and almost disappeared for the large dataset.

The regular spaCy model performs second best for all sample sizes. From the remaining models, the BernoulliNB classifier performed best for all sample sizes but the largest one. The performance of the SVM model fluctuated strongly for different sample sizes, and did even decrease, e.g., comparing the 1,500 dataset with the 250 dataset. RandomForest and k-NN were within lowest performing classifiers for all sample sizes, however, they reach a reasonable accuracy for the largest dataset.

In addition to the results shown in the table, we ran the spaCy models through the Prodigy software. The results were similar to both spaCy models and are thus not listed in Table 5.1.

5.5 Discussion

The results show a significant increase in performance through pre-training with patent data. The benefits are strongest for small sample sizes, where 100 annotations led to accuracy scores of 77.2%. This score suggests, that pre-trained neural network may be well suitable for active learning, which aims at increasing the efficiency of annotations through active learning (Tong & Koller, 2001).

With sample sizes of 500 and 1500 patents an accuracy of 0.832 and 0.866 has been achieved. This accuracy scores may be appropriate for a number of further analyses and may encourage future researchers to use labeled patent data for their analyses.

[6]https://prodi.gy

Table 5.1: Comparison of patent classification performance.

Model	**Sample size**				
	100	250	500	1,500	5,000
BernoulliNB	0.706	0.776	0.798	0.808	0.842
SVM	0.612	0.536	0.794	0.774	0.858
RandomForest	0.590	0.668	0.752	0.770	0.836
k-NN	0.598	0.704	0.716	0.772	0.838
spaCy	0.726	0.786	0.806	0.858	0.872
spaCy$_{\mathbf{pre}}$	**0.772**	**0.800**	**0.832**	**0.866**	**0.874**

The models implemented are Bernoulli naive Bayes (BernoulliNB), support vector machine (SVM), random forest, k-nearest neighbour, spaCy large English model, and a spaCy model pre-trained with patent data. The models have been tested with different sample sizes, of 100, 250, 500, 1,500, and 5,000 patents in categories G 05D, and G 05B. Scores relate to recognition of G 05D.

The spaCy LMAO pre-training does not require extensive computation capacity. Therefore, the described methods may be suitable for a broad range of researchers, providing high accuracy and enabling efficient implementation. However, future research may evaluate, whether more expensive pre-training methods provide even stronger models.

5.6 Conclusions

Patent classification, in general, is a highly relevant research field. Besides pre-classification of patent applications, which is highly relevant for patent offices Krier & Zacc (2002), also other fields may benefit from advances in this area. Particularly economists may benefit from improved methods of patent analyses. Frank et al. (2019c), for example, describe the lack of high-quality data and empirically informed models as a key challenge for a better understanding of automation technologies. Patent data may be a rich source of data to address this challenge.

Our work contributes to patent as well as natural language processing (NLP) research by evaluating a powerful pre-trained CNN based approach for binary patent classification. The proposed method offers a fast, high accuracy tool enabling a broad range of researchers conducting patent classification or other text classification tasks.

Chapter 6

Conclusion

6.1 Overall summary

My dissertation contributes to navigating the Fourth Industrial Revolution (4IR). The work was motivated by two challenges facing researchers, policymakers, practitioners, and workers. First, researchers and practitioners need to understand 4IR trends and research opportunities in order to direct their research and implementation efforts. Second, with more technologies being implemented, it is becoming increasingly important to understand the impact of technological change on the workforce in order to prepare for a successful and sustainable transformation. This dissertation contributes to this field through the three research questions (RQs). RQ 1 aims to contribute to the field by exploring the evolution of technological trends and concepts, RQ 2 aims to identify research directions to assure an integrated holistic evolution of Industry 4.0[1], and RQ 3 focuses on analyzing the exposure of occupations to 4IR technologies. Chapters 2 through 5 each discusses its findings and highlights its specific contributions. This chapter summarizes the key insights from the previous chapters.

Chapter 2 addresses RQ 1. The chapter shows that Industrial Internet of Things (IIoT) technologies have become the center of Industry 4.0 technology map. This finding is in line with the initial definitions of Industry 4.0, which centered on the IIoT. Given the recent pronounced growth in the importance of artificial intelligence (AI) technologies, the work suggests accounting for AI's fundamental role in Industry 4.0 and understanding the 4IR as an AI-powered natural collaboration between humans and machines. For companies navigating the 4IR space these insights might mean accordingly that they might initially focus on implementing the IIoT in their manufacturing landscape as a basis for implementing AI solutions in a next step. Industry 4.0 technology trends also point toward a more decentralized infrastructure with several smart components. The results enable researchers, industry, and policymakers to better navigate the large corpus of work; reveal the differences between concepts such as advanced and intelligent manufacturing; and highlight trends and research gaps with the intent of helping these actors reap the benefits of digital transformations.

Chapter 3 addresses RQ 2. The chapter reviews the Industry 4.0 literature based on the

[1] This dissertation uses the term *4IR* as a general framing of the research field. I also treat *Industry 4.0* as a general conceptual term, albeit more narrowly focused on the field of operations and manufacturing research. For Chapters 2 and 3 (and other associated paragraphs, such as those in the conclusion), I mostly use the term "Industry 4.0," as this phrase is more widely used among the intended audiences of these articles.

four smarts framework developed by Frank et al. (2019a). The findings show that the existing literature has mainly been devoted to the study of Smart Manufacturing, although attention to the other smart dimensions has grown in recent years. Smart Working is the least explored dimension. The findings support the vision of Industry 4.0 as a concept transcending the Smart Manufacturing field, thus creating opportunities for synergies with other related fields. The intersection of Smart Manufacturing and Smart Supply Chain has been the most heavily explored, and more research is required on the remaining links. In particular, the intersection of Smart Working and Smart Supply Chain requires more attention. Further, there is a large body of research on Smart Products and Services but relatively little integration with other smart dimensions.

Chapter 4 addresses RQ 3. The chapter introduces a patent-based measure of technology exposure by occupation and shows that exposure to 4IR technologies differs from traditional technology exposure. Manual tasks—and, accordingly, occupations such as construction and production—have been exposed mainly to traditional (i.e., non-4IR) patents and have low exposure to 4IR patents. The analysis suggests that 4IR technologies may have a negative impact on job growth that appears 10 to 20 years after patent filing. Researchers could validate the findings through further analyses using micro-level data, and my dataset can serve as a source for more complex labor market analyses. I also compared the 4IR exposure score to other automation and AI exposure scores. While many measures refer to theoretical automation potential, my patent-based indicator reflects actual technology diffusion. The chapter shows that a combination of 4IR exposure and other automation measures may provide additional insights. For example, near-term automation might be driven by non-4IR technologies. My work not only allows for analyses of the impact of 4IR technologies as a whole but also provides exposure scores for more than 300 technology fields, such as AI and smart office technologies. Finally, the work provides a general mapping of patents to tasks and occupations, which enables future researchers to construct individual exposure measures.

The dissertation also contributes to the research methodology literature. Chapter 2 introduces a novel method for reviewing literature wherein natural language processing (NLP) is used to extract technology terms from scientific articles. These terms serve as a basis for network analysis, which is used to show technology clusters, relations, and trends. Using the tool Gephi, I exported the network graph and made it available online[2]. Chapter 3 uses NLP to filter relevant articles for a literature review, replacing the step of manual scanning of thousands of articles. Chapter 4 refines existing NLP approaches to calculate similarity scores for patent and task descriptions, which serve as a basis for calculating the exposure of occupations to 4IR technologies. The data are available for download. A Tableau Public file[3] enables an interactive exploration of the dataset.

Overall, the dissertation provides dynamic views of the evolution of 4IR technologies, research trends, and the exposure of various occupations to 4IR technologies. These contributions can help researchers, practitioners, workers, and policymakers navigate the 4IR. The key findings answer the three RQs as follows:

[2] `https://bmeindl.github.io/technology_network/`

[3] Available online at Tableau Public at `https://public.tableau.com/app/profile/benjamin.meindl/viz/4IR_tech/Landing`, as well as for download to explore offline.

RQ 1	Based on the initial definitions, the IIoT is currently the core of the 4IR technology landscape, but strong growth in the field of AI suggests that AI may become the new core technology.
RQ 2	In the four smarts framework, Smart Working—the least explored dimension to date—offers many opportunities for future research. Further opportunities lie in investigating the intersections of the various smart dimensions.
RQ 3	Exposure to 4IR technologies differs from traditional technology exposure. Occupations with many manual tasks, such as construction and production, are exposed mainly to traditional (non-4IR) technologies but have low exposure to 4IR technologies.

6.2 Directions for future research

This dissertation provides directions for future research in terms of methodological contributions, the insights derived from the analyses, and the established datasets.

In terms of research methodology, Chapter 2 presents a method of reviewing textual data for certain entities (in this case, 4IR technologies). Researchers can build on the proposed method in two ways. First, they can extend the analysis to additional datasets, such as patent data or industry reports, to create an even broader review of technology trends. Second, the method can be applied in different fields (e.g., reviewing historical events or researching trends in other fields). Similarly, Chapter 3 presents a method for reviewing the 4IR literature. The NLP-based filtering allows for faster screening of articles, and researchers can build on this framework to conduct systematic literature reviews (SLRs) in different research areas.

The insights into 4IR trends and research opportunities offer guidance for future research. I show that, in the past decade, Industry 4.0 research has focused on establishing infrastructures for smart factories (i.e., the IIoT). This work suggests that, going forward, AI may be seen as the core of Industry 4.0 and that future frameworks should account for this. Further, with the establishment of the IIoT and the further diffusion of AI technologies, Industry 4.0 systems will become more decentralized (e.g., through edge computing, smart sensors, and smart devices). Researchers and practitioners should consider this trend, which is currently not well reflected in many frameworks. The evaluation of Industry 4.0 research along the four smarts perspective (Smart Manufacturing, Smart Supply Chain, Smart Products and Services, Smart Working) additionally showed that more interdisciplinary research is required at the intersections of the smart dimensions. In particular, more research is required at the intersection of Smart Working and Smart Supply Chain as well as that of Smart Products and Services and any other smart dimension. This could include, for example, how data from smart products can help improve the manufacturing process. Finally, the field of Smart Working is currently the least researched of the four smarts. More research could focus how smart workplaces (e.g., worker augmentation) may appear in production and along the supply chain. In addition, future researchers could combine the approaches described in Chapters 2 and 3 to describe a more in depth evolution of technologies in the context of the four smart dimensions. This can be implemented, e.g., considering for each appearance of a technology term, which of the four smarts the containing article is related to.

While Smart Working is a relevant field within the Industry 4.0 manufacturing environment, 4IR technologies will also have a more general labor market impact. This dissertation provides insights and datasets that can serve as a basis for future work in this area. The work provides an index of 4IR technology exposure by occupation. A high exposure score does not necessarily mean that jobs are becoming automated but rather that work activities might change significantly and reskilling may be required. Future researchers could evaluate how highly exposed occupations might change and how workers can adapt to these changes. More generally, researchers can use the 4IR exposure scores (as well as the patent–occupation mapping itself) to evaluate the general impact of 4IR technologies (or any other group of technologies) on the labor market. As this dissertation provides only simple regression models to evaluate the impact of technologies on the labor market, future researchers may account for more complex labor market adjustment mechanisms and rely on micro-level data for their analyses.

Bibliography

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, *37*, 3–13. doi:10.1016/j.wpi.2013.12.006.

Abraham, M., & Annunziata, M. (2017). Augmented reality is already improving worker performance. *Harvard Business Review*, *13*, 1–5.

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Princeton University Press.

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, *4*, 1043–1171. doi:10.1016/S0169-7218(11)02410-5.

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2020). AI and Jobs: Evidence from Online Vacancies. *NBER Working Paper*, *No. w28257*. doi:10.3386/w28257.

Acemoglu, D., & Restrepo, P. (2019a). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, *33*, 3–30. doi:10.1257/jep.33.2.3.

Acemoglu, D., & Restrepo, P. (2019b). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, *128*, 2188–2244. doi:10.1086/705716.

Acemoglu, D., & Restrepo, P. (2022). Demographics and automation. doi:10.1093/restud/rdab031.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press.

Al Omran, F. N. A., & Treude, C. (2017). Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments. *IEEE International Working Conference on Mining Software Repositories*, (pp. 187–197). doi:10.1109/MSR.2017.42.

Alcácer, V., & Cruz-Machado, V. (2019). Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Engineering Science and Technology, an International Journal*, *22*, 899–919. doi:10.1016/j.jestch.2019.01.006.

Anick, P., Verhagen, M., & Pustejovsky, J. (2014). Identification of Technology Terms in Patents. In *LREC* (pp. 2008–2014).

Ardolino, M., Rapaccini, M., Saccani, N., Gaiardelli, P., Crespi, G., & Ruggeri, C. (2018). The role of digital technologies for the service transformation of industrial companies. *International Journal of Production Research*, *56*, 2116–2132. doi:10.1080/00207543.2017.1324224.

Arntz, M., Gregory, T., & Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. *OECD Social, Employment and Migration Working Papers*, *2*, 47–54. doi:10.1787/5jlz9h56dvq7-en.

Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, *159*, 157–160. doi:10.1016/j.econlet.2017.07.001.

Arntz, M., Gregory, T., & Zierahn, U. (2020). Digitalization and the future of work: Macroeconomic consequences. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics* ZEW Discussion Papers. Cham: Springer. doi:10.2139/ssrn.3413653.

Aschhoff, B., Crass, D., Cremers, K., Grimpe, C., Rammer, C., Brandes, F., Diaz-Lopez, F., Woolthuis, R. K., Mayer, M., & Montalvo, C. (2010). European Competitiveness in Key Enabling Technologies. *European Competitiveness in KETs*, .

Aunger, R. (2010). Types of technology. *Technological Forecasting and Social Change*, *77*, 762–782. doi:10.1016/j.techfore.2010.01.008.

Autor, D., & Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, *103*, 1553–1597. doi:10.1257/aer.103.5.1553.

Autor, D., Katz, L., & Kearney, M. (2006). The Polarization of the U.S. Labor Market. *The American economic review*, *96*, 189–194. doi:10.1257/000282806777212620.

Autor, D., Mindell, D., & Reynolds, E. (2020). *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines*. Technical Report. URL: https://workofthefuture.mit.edu/research-post/the-work-of-the-future-building-better-jobs-in-an-age-of-intelligent-machines/.

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

Ayala, N. F., Gerstlberger, W., & Frank, A. G. (2019). Managing servitization in product companies: the moderating role of service suppliers. *International Journal of Operations & Production Management*, *39*, 43–74. doi:10.1108/IJOPM-08-2017-0484.

Ayala, N. F., Ghezzi, A., & Frank, A. G. (2017). Knowledge sharing dynamics in service suppliers' involvement for servitization of manufacturing companies. *International Journal of Production Economics*, *193*, 538–553. doi:10.1016/j.ijpe.2017.08.019.

Bai, C., Dallasega, P., Orzes, G., & Sarkis, J. (2020). Industry 4.0 technologies assessment: A sustainability perspective. *International Journal of Production Economics*, *229*, 107776. doi:10.1016/j.ijpe.2020.107776.

Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., Artemova, E., Tutubalina, E., & Chowell, G. (2021). A large-scale covid-19 twitter chatter dataset for open scientific research—an international collaboration. *Epidemiologia*, *2*, 315–324. doi:10.3390/epidemiologia2030024.

Barreto, L., Amaral, A., & Pereira, T. (2017). Industry 4.0 implications in logistics: an overview. *Procedia Manufacturing*, *13*, 1245–1252. doi:10.1016/j.promfg.2017.09.045.

Bartel, A., Ichniowski, C., & Shaw, K. (2007). How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills. *Quarterly Journal of Economics*, *122*, 1721–1758. doi:10.1162/qjec.2007.122.4.1721.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third International AAAI Conference on Weblogs and Social Media*, *3*, 361–362. doi:10.1136/qshc.2004.010033.

Behrendt, A., Kadocsa, A., Kelly, R., & Schirmers, L. (2017). Digital Manufacturing Capturing sustainable impact at scale. *McKinsey & Company*, (pp. 1–24).

Behrendt, A., Kelly, R., Rettig, R., & Stoffregen, S. (2018). Digital Manufacturing – escaping pilot purgatory. *McKinsey & Company*, .

Beier, G., Ullrich, A., Niehoff, S., Reißig, M., & Habich, M. (2020). Industry 4.0: How it is defined from a sociotechnical perspective and how much sustainability it includes – A literature review. *Journal of Cleaner Production*, (p. 120856). doi:10.1016/j.jclepro.2020.120856.

Benites, F., Malmasi, S., & Zampieri, M. (2018). Classifying Patent Applications with Ensemble Methods. *Proceedings ofAustralasian Language Technology Association Workshop*, (pp. 89–92).

Benitez, G. B., Ayala, N. F., & Frank, A. G. (2020). Industry 4.0 innovation ecosystems: An evolutionary perspective on value cocreation. *International Journal of Production Economics*, *228*, 107735. doi:10.1016/j.ijpe.2020.107735.

Benitez, G. B., Ferreira-Lima, M., Ayala, N. F., & Frank, A. G. (2021). Industry 4.0 technology provision: the moderating role of supply chain partners to support technology providers. *Supply Chain Management*, . doi:10.1108/SCM-07-2020-0304.

Benzineb, K., & Guyot, J. (2011). Automated Patent Classification. In L. M., M. K., T. J., & T. A. (Eds.), *Current Challenges in Patent Information Retrieval. The Information Retrieval Series, vol 29.* (pp. 239–261). Springer Berlin Heidelberg. doi:10.1007/978-3-642-19231-9_12.

Bessen, J. E. (2019). Automation and Jobs: When Technology Boosts Employment. *Economic Policy*, *34*, 589—-626. doi:10.1093/epolic/eiaa001.

Blondel, V. D., Guillaume, J.-l., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*, P10008. doi:10.1088/1742-5468/2008/10/P10008.

Bokrantz, J., Skoogh, A., Berlin, C., Wuest, T., & Stahre, J. (2020). Smart Maintenance: a research agenda for industrial maintenance management. *International Journal of Production Economics*, *224*, 107547. doi:10.1016/j.ijpe.2019.107547.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social networks*, *29*, 555–564. doi:10.1016/j.socnet.2007.04.002.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*, 892–895. doi:10.1126/science.1165821.

Bowles, M., & Lu, J. (2014). Removing the blinders: A literature review on the potential of nanoscale technologies for the management of supply chains. *Technological Forecasting and Social Change*, *82*, 190–198. doi:10.1016/j.techfore.2013.10.017.

Boyer, K. K., Leong, G. K., Ward, P. T., & Krajewski, L. J. (1997). Unlocking the potential of advanced manufacturing technologies. *Journal of Operations Management*, *15*, 331–347. doi:10.1016/S0272-6963(97)00009-0.

Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in Industry*, *101*, 1–12. doi:10.1016/j.compind.2018.04.015.

Bressanelli, G., Adrodegari, F., Perona, M., & Saccani, N. (2018). Exploring how usage-focused business models enable circular economy through digital technologies. *Sustainability (Switzerland)*, *10*. doi:10.3390/su10030639.

Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2014). How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective. *FormaMente*, *12*, 47–62. doi:10.1016/j.procir.2015.02.213.

Brizzi, F., Peppoloni, L., Graziano, A., Di Stefano, E., Avizzano, C. A., & Ruffaldi, E. (2018). Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment. *IEEE Transactions on Human-Machine Systems*, *48*, 197–206. doi:10.1109/THMS.2017.2782490.

Broniatowski, D. A., Paul, M. J., & Dredze, M. (2014). Twitter: Big data opportunities. *Science*, *6193*, 148. doi:10.1126/science.345.6193.148-a.

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

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

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 23–57). University of Chicago Press. URL: http://www.nber.org/papers/w24001.pdf.

Bueno, A., Godinho Filho, M., & Frank, A. G. (2020). Smart production planning and control in the Industry 4.0 context: A systematic literature review. *Computers and Industrial Engineering*, *149*, 106774. doi:10.1016/j.cie.2020.106774.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., & Trench, M. (2017). Artificial Intelligence the Next Digital Frontier? *McKinsey Global Institute*, . URL: https://www.mckinsey.com/~/media/McKinsey/Industries/AdvancedElectronics/OurInsights/Howartificialintelligencecandeliverrealvaluetocompanies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

Bughin, J., Seong, J., Manyika, J., Hämäläinen, L., Windhagen, E., & Hazan, E. (2019). *Notes from the AI frontier, Tackling Europe's Gap in Digital and AI*. Technical Report February McKinsey Global Institute. URL: https://www.mckinsey.com/~/media/mckinsey/featuredinsights/artificialintelligence/tacklingeuropesgapindigitalandai/mgi-tackling-europes-gap-in-digital-and-ai-feb-2019-vf.ashx.

Butt, J. (2020). Exploring the Interrelationship between Additive Manufacturing and Industry 4.0. *Designs*, *4*, 13. doi:10.3390/designs4020013.

Büyüközkan, G., & Göçer, F. (2018). Digital Supply Chain: Literature review and a proposed framework for future research. *Computers in Industry*, *97*, 157–177. doi:10.1016/j.compind.2018.02.010.

Cenamor, J., Rönnberg Sjödin, D., & Parida, V. (2017). Adopting a platform approach in servitization: Leveraging the value of digitalization. *International Journal of Production Economics*, *192*, 0–1. doi:10.1016/j.ijpe.2016.12.033.

Ceruti, A., Marzocca, P., Liverani, A., & Bil, C. (2019). Maintenance in aeronautics in an Industry 4.0 context: The role of Augmented Reality and Additive Manufacturing. *Journal of Computational Design and Engineering*, *6*, 516–526. doi:10.1016/j.jcde.2019.02.001.

Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., & Yin, B. (2017). Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges. *IEEE Access*, *6*, 6505–6519. doi:10.1109/ACCESS.2017.2783682.

Chen, D., Heyer, S., Ibbotson, S., Salonitis, K., Steingrímsson, J. G., & Thiede, S. (2015). Direct digital manufacturing: Definition, evolution, and sustainability implications. *Journal of Cleaner Production*, *107*, 615–625. doi:10.1016/j.jclepro.2015.05.009.

Chen, Y. (2017). Integrated and Intelligent Manufacturing: Perspectives and Enablers. *Engineering*, *3*, 588–595. doi:10.1016/J.ENG.2017.04.009.

Cheng, Y., Matthiesen, R., Farooq, S., Johansen, J., Hu, H., & Ma, L. (2018). The evolution of investment patterns on advanced manufacturing technology (AMT) in manufacturing operations: A longitudinal analysis. *International Journal of Production Economics*, *203*, 239–253. doi:10.1016/j.ijpe.2018.06.019.

Chiarello, F., Trivelli, L., Bonaccorsi, A., & Fantoni, G. (2018). Extracting and mapping industry 4.0 technologies using wikipedia. *Computers in Industry*, *100*, 244–257. doi:10.1016/j.compind.2018.04.006.

Chryssolouris, G., Mavrikios, D., Papakostas, N., Mourtzis, D., Michalos, G., & Georgoulias, K. (2009). Digital manufacturing: History, perspectives, and outlook. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *223*, 451–462. doi:10.1243/09544054JEM1241.

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute*, . URL: https://www.mckinsey.com/~/media/mckinsey/featuredinsights/artificialintelligence/notesfromtheaifrontierapplicationsandvalueofdeeplearning/notes-from-the-ai-frontier-insights-from-hundreds-of-use-cases-discussion-paper.ashx.

Cimini, C., Lagorio, A., Romero, D., Cavalieri, S., & Stahre, J. (2020). Smart Logistics and The Logistics Operator 4.0. In *21st IFAC World Congress* April. Berlin, Germany.

Cohen, Y., Naseraldin, H., Chaudhuri, A., & Pilati, F. (2019). Assembly systems in Industry 4.0 era: a road map to understand Assembly 4.0. *The International Journal of Advanced Manufacturing Technology*, *105*, 4037–4054. doi:10.1007/s00170-019-04203-1.

Culot, G., Nassimbeni, G., Orzes, G., & Sartor, M. (2020). Behind the definition of Industry 4.0: Analysis and open questions. *International Journal of Production Economics*, *226*, 107617. doi:10.1016/j.ijpe.2020.107617.

Dalenogare, L. S., Baseggio, M. M., Ayala, N. F., Dain, M.-A. L., & Frank, A. G. (2019). The contribution of Smart Glasses for PSS. *Procedia CIRP*, *83*, 318–323. doi:10.1016/j.procir.2019.03.307.

Dalenogare, L. S., Benitez, G. B., Ayala, N. F., & Frank, A. G. (2018). The expected contribution of Industry 4.0 technologies for industrial performance. *International Journal of Production Economics*, *204*, 383–394. doi:10.1016/j.ijpe.2018.08.019.

Dauth, W., Findeisen, S., Südekum, J., & Wößner, N. (2017). German Robots - The Impact of Industrial Robots on Workers. *CEPR Discussion Papers*, (p. 57). URL: http://doku.iab.de/discussionpapers/2017/dp3017.pdf.

Dauth, W., Findeisen, S., Suedekum, J., & Woessner, N. (2018). Adjusting to Robots: Worker-Level Evidence. *Opportunity and Inclusive Growth Institute, Federal Reserve Bank of Minneapolis*, . doi:10.21034/iwp.13.

de Sousa Jabbour, A. B. L., Jabbour, C. J. C., Foropon, C., & Filho, M. G. (2018). When titans meet – Can industry 4.0 revolutionise the environmentally-sustainable manufacturing wave? The role of critical success factors. *Technological Forecasting and Social Change*, *132*, 18–25. doi:10.1016/j.techfore.2018.01.017.

Dechezleprêtre, A., Hémous, D., Olsen, M., & Zanella, C. (2020). Automating Labor: Evidence from Firm-level Patent Data. doi:10.2139/ssrn.3508783.

Devlin, J., Chang, M.-w., Kenton, L., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, .

D'hondt, E., Verberne, S., Koster, C., & Boves, L. (2013). Text Representations for Patent Classification. *Computational Linguistics*, *39*, 755–775. doi:10.1162/COLI.

Du, G., Chen, M., Liu, C., Zhang, B., & Zhang, P. (2018). Online Robot Teaching With Natural Human–Robot Interaction. *IEEE Transactions on Industrial Electronics*, *65*, 9571–9581. doi:10.1109/TIE.2018.2823667.

European Commission (2013). *Factories of the Future: Multi-Annual Roadmap for the Contractual PPP under Horizon 2020*. doi:10.2777/29815.

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, *162*, 101–114. doi:10.1016/j.ijpe.2015.01.003.

Fall, C. J., Törcsvári, A., Benzineb, K., & Karetka, G. (2003). Automated Categorization in the International Patent Classification. *Acm Sigir Forum*, *37*, 10–25.

Fantini, P., Pinzone, M., & Taisch, M. (2018). Placing the operator at the centre of Industry 4.0 design: Modelling and assessing human activities within cyber-physical systems. *Computers & Industrial Engineering*, *139*, 105058. doi:10.1016/j.cie.2018.01.025.

Fareri, S., Fantoni, G., Chiarello, F., Coli, E., & Binda, A. (2020). Estimating Industry 4.0 impact on job profiles and skills using text mining. *Computers in Industry*, *118*, 103222. doi:10.1016/j.compind.2020.103222.

Fatorachian, H., & Kazemi, H. (2020). Impact of Industry 4.0 on supply chain performance. *Production Planning & Control*, (pp. 1–19). doi:10.1080/09537287.2020.1712487.

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses. *Strategic Management Journal*, (pp. 1–23). doi:10.1002/smj.3286.

Fletcher, S. R., Johnson, T., Adlon, T., Larreina, J., Casla, P., Parigot, L., Alfaro, P. J., & Otero, M. d. M. (2020). Adaptive automation assembly: Identifying system requirements for technical efficiency and worker satisfaction. *Computers & Industrial Engineering*, *139*, 105772. doi:10.1016/j.cie.2019.03.036.

Foresight (2013). The Future of Manufacturing: A new era of opportunity and challenge for the UK. *The Government Office for Science, London*, (pp. 1–20). doi:10.1049/tpe.1971.0034.

Frank, A. G., Dalenogare, L. S., & Néstor, F. A. (2019a). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, *210*, 15–26. doi:10.1016/j.ijpe.2019.01.004.

Frank, A. G., Mendes, G. H., Ayala, N. F., & Ghezzi, A. (2019b). Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. *Technological Forecasting & Social Change*, *141*, 341–351. doi:10.1016/j.techfore.2019.01.014.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moroa, E., Wang, D., Youn, H., & Rahwan, I. (2019c). Toward understanding the impact of artificial intelligence on labor. *PNAS*, *116*, 6531–6539. doi:10.1073/pnas.1900949116.

Frank, M. R., Wang, D., Cebrian, M., & Rahwan, I. (2019d). The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence*, *1*, 79–85. doi:10.1038/s42256-019-0024-5.

Frederico, G. F., Garza-Reyes, J. A., Anosike, A., & Kumar, V. (2020). Supply Chain 4.0: concepts, maturity and research agenda. *Supply Chain Management*, *25*, 262–282. doi:10.1108/SCM-09-2018-0339.

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

Gaiardelli, P., Pezzotta, G., Rondini, A., Romero, D., Jarrahi, F., Bertoni, M., Wiesner, S., Wuest, T., Larsson, T., Zaki, M., Jussen, P., Boucher, X., Bigdeli, A. Z., & Cavalieri, S. (2021). Product-service systems evolution in the era of Industry 4.0. *Service Business*, *15*, 177–207. doi:10.1007/s11628-021-00438-9.

Galati, F., & Bigliardi, B. (2019). Industry 4.0: Emerging themes and future research avenues using a text mining approach. *Computers in Industry*, *109*, 100–113. doi:10.1016/j.compind.2019.04.018.

Gao, J., Yao, Y., Zhu, V. C., Sun, L., & Lin, L. (2011). Service-oriented manufacturing: A new product pattern and manufacturing paradigm. *Journal of Intelligent Manufacturing*, *22*, 435–446. doi:10.1007/s10845-009-0301-y.

Gay-Bellile, V., Bourgeois, S., Larnaout, D., & Tamaazousti, M. (2015). Applications of Augmented Reality in the Automotive Industry. In *Fundamentals of Wearable Computers and Augmented Reality* (pp. 433–456). CRC Press.

Georgakopoulos, D., Jayaraman, P. P., Fazia, M., Villari, M., & Ranjan, R. (2016). Internet of Things and Edge Cloud Computing Roadmap for Manufacturing. *IEEE Cloud Computing*, *3*, 66–73. doi:10.1109/MCC.2016.91.

Geum, Y., & Park, Y. (2011). Designing the sustainable product-service integration: A product-service blueprint approach. *Journal of Cleaner Production*, *19*, 1601–1614. doi:10.1016/j.jclepro.2011.05.017.

Ghobakhloo, M. (2018). The future of manufacturing industry: a strategic roadmap toward Industry 4.0. *Journal of Manufacturing Technology Management*, *59*, 602–614. doi:10.1108/EL-01-2017-0019.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining Job Polarization: Routinization and Offshoring. *American Economic Review*, *104*, 1–20. doi:10.1257/aer.104.8.2509.

Graetz, G., & Michaels, G. (2018). Robots at work. *The Review of Economics and Statistics*, *C*, 753–768. doi:10.1162/rest.

Grawe, M. F., Martins, C. A., & Bonfante, A. G. (2017). Automated Patent Classification Using Word Embedding. *16th IEEE International Conference on Machine Learning and Applications*, . doi:10.1109/ICMLA.2017.0-127.

Gregory, T., Salomons, A., & Zierahn, U. (2016). Racing With or Against the Machine? Evidence from Europe. doi:10.2139/ssrn.2815469.

Grubic, T., & Jennions, I. (2018). Remote monitoring technology and servitised strategies–factors characterising the organisational application. *International Journal of Production Research*, *56*, 2133–2149. doi:10.1080/00207543.2017.1332791.

Hahn, G. J. (2020). Industry 4.0: a supply chain innovation perspective. *International Journal of Production Research*, *58*, 1425–1441. doi:10.1080/00207543.2019.1641642.

Hankel, M., & Rexroth, B. (2015). The Reference Architectural Model Industrie 4.0 (RAMI 4.0). *ZWEI: Die Elektroindustrie*, .

Harzing, A.-w., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*, 787–804. doi:10.1007/s11192-015-1798-9.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust-The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, *105*, 105–120. doi:10.1016/j.techfore.2015.12.014.

Hermann, M., Pentek, T., & Otto, B. (2016). Design principles for industrie 4.0 scenarios. In *2016 49th Hawaii international conference on system sciences (HICSS)* (pp. 3928–3937). IEEE. doi:10.1109/HICSS.2016.488.

Hofmeister Kahle, J., Marcon, É., Ghezzi, A., & Frank, A. G. (2020). Smart Products value creation in SMEs innovation ecosystems. *Technological Forecasting and Social Change*, *156*, 120024. doi:10.1016/j.techfore.2020.120024.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, . URL: spacy.io.

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. *Proceedings ofthe 56th Annual Meeting ofthe Association for Computational Linguistics (Volume 1: Long Papers)*, (pp. 328–339).

Huysamen, K., Bosch, T., de Looze, M., Stadler, K. S., Graf, E., & O'Sullivan, L. W. (2018). Evaluation of a passive exoskeleton for static upper limb activities. *Applied Ergonomics*, *70*, 148–155. doi:10.1016/j.apergo.2018.02.009.

Ivanov, D., Dolgui, A., Sokolov, B., Werner, F., & Ivanova, M. (2016). A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0. *International Journal of Production Research*, *54*, 386–402. doi:10.1080/00207543.2014.999958.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, *9*, 1–12. doi:10.1371/journal.pone.0098679.

Jeschke, S., Brecher, C., Meisen, T., Özdemir, D., & Eschert, T. (2017). Cyber-Physical Systems Engineering for Manufacturing In Industrial Internet of Things. *Industrial Internet of Things. Springer Series in Wireless Technology. Springer, Cham*, (pp. 3–19). doi:10.1007/978-3-319-42559-7.

Jiang, R., Banchs, R. E., & Li, H. (2016). Evaluating and Combining Named Entity Recognition Systems. In *Proceedings of the Sixth Named Entity Workshop* (pp. 21–27). doi:10.3115/1572392.1572430.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *European conference on machine learning*, (pp. 137–142).

Kaasinen, E., Schmalfuß, F., Özturk, C., Aromaa, S., Boubekeur, M., Heilala, J., Heikkilä, P., Kuula, T., Liinasuo, M., Mach, S., Mehta, R., Petäjä, E., & Walter, T. (2020). Empowering and engaging industrial workers with Operator 4.0 solutions. *Computers & Industrial Engineering*, *139*, 105678. doi:10.1016/j.cie.2019.01.052.

Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative INDUSTRIE 4.0*. Technical Report April Forschungsunion. doi:10.13140/RG.2.1.1205.8966.

Kang, B., & Tarasconi, G. (2016). Patstat revisited: Suggestions for better usage. *World patent information*, *46*, 56–63.

Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., Kim, B. H., & Noh, S. D. (2016). Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering and Manufacturing - Green Technology*, *3*, 111–128. doi:10.1007/s40684-016-0015-5.

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, *3*, 303–20. doi:10.1257/aeri.20190499.

Keynes, J. M. (1930). Economic Possibilities for our Grandchildren. In *Essays in persuasion* (pp. 358–373). New York: W. W. Norton & Co, 1963. URL: https://www.marxists.org/reference/subject/economics/keynes/1930/our-grandchildren.htm.

Kim, D.-h., Lee, H., & Kwak, J. (2017). Standards as a driving force that influences emerging technological trajectories in the converging world of the Internet and things: An investigation of the M2M/IoT patent network. *Research Policy*, *46*, 1234–1254. doi:10.1016/j.respol.2017.05.008.

Kipper, L. M., Furstenau, L. B., Hoppe, D., Frozza, R., & Iepsen, S. (2020). Scopus scientific mapping production in industry 4.0 (2011–2018): a bibliometric analysis. *International Journal of Production Research*, *58*, 1605–1627. doi:10.1080/00207543.2019.1671625.

Klumpp, M., Hesenius, M., Meyer, O., Ruiner, C., & Gruhn, V. (2019). Production logistics and human-computer interaction—state-of-the-art, challenges and requirements for the future. *International Journal of Advanced Manufacturing Technology*, *105*, 3691–3709. doi:10.1007/s00170-019-03785-0.

Kogan, L., Papanikolaou, D., Schmidt, L., & Seegmiller, B. (2020). Technological Change and Occupations over the Long Run. *SSRN Electronic Journal*, . doi:10.2139/ssrn.3585676.

Korner, M. E. H., Lambán, M. P., Albajez, J. A., Santolaria, J., Corrales, L. D. C. N., & Royo, J. (2020). Systematic literature review: Integration of additive manufacturing and industry 4.0. *Metals*, *10*, 1–24. doi:10.3390/met10081061.

Krier, M., & Zacc, F. (2002). Automatic categorisation applications at the European patent office. *World Patent Information*, *24*, 187–196.

Krugh, M., & Mears, L. (2018). A complementary Cyber-Human Systems framework for Industry 4.0 Cyber-Physical Systems. *Manufacturing Letters*, *15*, 89–92. doi:10.1016/j.mfglet.2018.01.003.

Krzywdzinski, M., Jürgens, U., & Pfeiffer, S. (2016). The Fourth Revolution: The Transformation of Manufacturing Work in the Age of Digitalization. *WZB Report*, (pp. 22–25).

Kumar, A. (2018). Methods and Materials for Smart Manufacturing: Additive Manufacturing, Internet of Things, Flexible Sensors and Soft Robotics. *Manufacturing Letters*, *15*, 122–125. doi:10.1016/j.mfglet.2017.12.014.

Kumar, S., Mookerjee, V. S., Choi, T.-M., & Yue, X. (2020). Call for Papers: Special Issue of Production and Operations Management Journal on "Disruptive Technologies and Operations Management". *Production and Operations Management*, . doi:10.1111/poms.12398.

Kumar, S., & Tsvetkov, Y. (2019). Von Mises-fisher loss for training sequence to sequence models with continuous outputs. *7th International Conference on Learning Representations, ICLR 2019*, (pp. 1–15).

Kuo, C. F., & Wang, M. J. J. (2012). Motion generation and virtual simulation in a digital environment. *International Journal of Production Research*, *50*, 6519–6529. doi:10.1080/00207543.2011.653698.

Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, *56*, 508–517. doi:10.1080/00207543.2017.1351644.

Kwon, H., & Park, Y. (2015). Exploring Social Impact of Emerging Technologies from Futuristic Data. *International Journal of Industrial and Manufacturing Engineering*, *9*, 754–757. doi:10.5281/zenodo.1100154.

Lambiotte, R., Delvenne, J., & Barahona, M. (2009). Laplacian Dynamics and Multiscale Modular Structure in Networks, . (pp. 1–29). doi:10.1109/TNSE.2015.2391998.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural Architectures for Named Entity Recognition. doi:10.18653/v1/N16-1030.

Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business and Information Systems Engineering*, *6*, 239–242. doi:10.1007/s12599-014-0334-4.

Lee, C. K., Lv, Y., Ng, K. K., Ho, W., & Choy, K. L. (2018). Design and application of internet of things-based warehouse management system for smart logistics. *International Journal of Production Research*, *56*, 2753–2768. doi:10.1080/00207543.2017.1394592.

Lee, J., Azamfar, M., & Singh, J. (2019). A blockchain enabled Cyber-Physical System architecture for Industry 4.0 manufacturing systems. *Manufacturing Letters*, *20*, 34–39. doi:10.1016/j.mfglet.2019.05.003.

Lee, J.-S., & Hsiang, J. (2020). Patent classification by fine-tuning bert language model. *World Patent Information*, *61*, 101965.

Leyer, M., & Schneider, S. (2021). Decision augmentation and automation with artificial intelligence: Threat or opportunity for managers? *Business Horizons*, *64*, 711–724. doi:10.1016/j.bushor.2021.02.026.

Li, L. (2018). China's manufacturing locus in 2025: With a comparison of "Made-in-China 2025" and "Industry 4.0". *Technological Forecasting & Social Change*, *135*, 66–74. doi:10.1016/j.techfore.2017.05.028.

Li, L. H., Chen, P. H., Hsieh, C.-J., & Chang, K.-W. (2019). Efficient Contextual Representation Learning Without Softmax Layer, .

Li, S., Hu, J., Cui, Y., & Hu, J. (2018). DeepPatent: patent classification with convolutional neural networks and word embedding. *Scientometrics*, *117*, 721–744. doi:10.1007/s11192-018-2905-5.

Liao, Y., Deschamps, F., Loures, E. d. F. R., & Ramos, L. F. P. (2017). Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. *International Journal of Production Research*, *55*, 3609–3629. doi:10.1080/00207543.2017.1308576.

Liboni, L. B., Cezarino, L. O., Jabbour, C. J. C., Oliveira, B. G., & Stefanelli, N. O. (2019). Smart industry and the pathways to HRM 4.0: implications for SCM. *Supply Chain Management*, *24*, 124–146. doi:10.1108/SCM-03-2018-0150.

Llopis-Albert, C., Rubio, F., & Valero, F. (2021). Impact of digital transformation on the automotive industry. *Technological Forecasting and Social Change*, *162*, 120343. doi:10.1016/j.techfore.2020.120343.

Longo, F., Nicoletti, E., & Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & Industrial Engineering*, *113*, 144–159. doi:10.1016/j.cie.2017.09.016.

Lu, D., Lai, I., & Liu, Y. (2019). The Consumer Acceptance of Smart Product-Service Systems in Sharing Economy: The Effects of Perceived Interactivity and Particularity. *Sustainability*, *11*, 928. doi:10.3390/su11030928.

Lu, H.-p., & Weng, C.-i. (2018). Smart manufacturing technology, market maturity analysis and technology roadmap in the computer and electronic product manufacturing industry. *Technological Forecasting and Social Change*, *133*, 85–94. doi:10.1016/j.techfore.2018.03.005.

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, *6*, 1–10. doi:10.1016/j.jii.2017.04.005.

Lu, Y., & Cecil, J. (2016). An Internet of Things (IoT)-based collaborative framework for advanced manufacturing. *International Journal of Advanced Manufacturing Technology*, *84*, 1141–1152. doi:10.1007/s00170-015-7772-0.

Lybbert, T. J., & Zolas, N. J. (2014). Getting patents and economic data to speak to each other: An 'Algorithmic Links with Probabilities' approach for joint analyses of patenting and economic activity. *Research Policy*, *43*, 530–542. doi:10.1016/j.respol.2013.09.001.

Machado, C. G., Winroth, M. P., & Ribeiro da Silva, E. H. D. (2020). Sustainable manufacturing in Industry 4.0: an emerging research agenda. *International Journal of Production Research*, *58*, 1462–1484. doi:10.1080/00207543.2019.1652777.

Mann, K., & Püttmann, L. (2017). Benign Effects of Automation: New Evidence From Patent Texts. *Available at SSRN 2959584*, .

Mansfield, N., Naddeo, A., Frohriep, S., & Vink, P. (2020). Integrating and applying models of comfort. *Applied Ergonomics*, *82*, 102917. doi:10.1016/j.apergo.2019.102917.

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). *A future that works: AI, automation, employment, and productivity*. Technical Report McKinsey Global Institute. URL: https://www.mckinsey.com/~/media/McKinsey/FeaturedInsights/DigitalDisruption/Harnessingautomationforafuturethatworks/MGI-A-future-that-works_Executive-summary.pdf.

Marneffe, M.-c. D., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., & Manning, C. D. (2014). Universal Stanford Dependencies : A cross-linguistic typology. *LREC*, *14*, 4585–4592. URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1062_Paper.pdf.

Masood, T., & Egger, J. (2019). Augmented reality in support of Industry 4.0—Implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing*, *58*, 181–195. doi:10.1016/j.rcim.2019.02.003.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*, 61–67. doi:00475394.

Mehrpouya, M., Dehghanghadikolaei, A., Fotovvati, B., Vosooghnia, A., Emamian, S. S., & Gisario, A. (2019). The potential of additive manufacturing in the smart factory industrial 4.0: A review. *Applied Sciences (Switzerland)*, *9*. doi:10.3390/app9183865.

Meindl, B., Ayala, N. F., Mendonça, J., & Frank, A. G. (2021a). The four smarts of Industry 4. 0: evolution of ten years of research and future. *Technological Forecasting and Social Change*, *168*, 120784. doi:10.1016/j.techfore.2021.120784.

Meindl, B., Frank, M. R., & Mendonça, J. (2021b). Exposure of occupations to technologies of the fourth industrial revolution. arXiv:2110.13317.

Meindl, B., & Mendonça, J. (2021). Mapping industry 4.0 technologies: From cyber-physical systems to artificial intelligence. arXiv:2111.14168.

Meindl, B., Ott, I., & Zierahn, U. (2019). Binary Patent Classification Methods for Few Annotated Samples. In L. Andersson, H. Aras, F. Piroi, & A. Hanbury (Eds.), *Proceedings of The 1st Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech 2019)* (pp. 13–17). Karlsruhe. doi:10.34726/pst2019.4.

Mendonça, J., & Heitor, M. (2016). The changing patterns of industrial production: How does it play for the Iberian Peninsula? *Technological Forecasting and Social Change*, *113*, 293–307. doi:10.1016/j.techfore.2016.07.042.

Ménière, Y., Philpott, J., Rodríguez, J. P., Rudyk, I., Wewege, S., & Wienold, N. (2020). *Patents and the Fourth Industrial Revolution (EPO)*. December. Munich: European Patent Office.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2018). Advances in Pre-Training Distributed Word Representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Mitchell, T., & Brynjolfsson, E. (2017). Track how technology is transforming work. *Nature*, *544*, 290–292. doi:10.1038/544290a.

Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2018). A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). *Journal of Manufacturing Systems*, *49*, 194–214. doi:10.1016/J.JMSY.2018.10.005.

Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different? *Journal of Economic Perspectives*, *29*, 31–50. doi:10.1257/jep.29.3.31.

Mollá, D., & Seneviratne, D. (2018). Overview of the 2018 ALTA Shared Task: Classifying Patent Applications. *Proceedings of the Australasian Language Technology Association Workshop 2018*, .

Monios, J., & Bergqvist, R. (2019). The transport geography of electric and autonomous vehicles in road freight networks. *Journal of Transport Geography*, *80*, 102500. doi:10.1016/j.jtrangeo.2019.102500.

Mourtzis, D., Papakostas, N., Mavrikios, D., Makris, S., & Alexopoulos, K. (2015). The role of simulation in digital manufacturing: Applications and outlook. *International Journal of Computer Integrated Manufacturing*, *28*, 3–24. doi:10.1080/0951192X.2013.800234.

Mourtzis, D., Zogopoulos, V., & Xanthi, F. (2019). Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling. *International Journal of Advanced Manufacturing Technology*, *105*, 3899–3910. doi:10.1007/s00170-019-03941-6.

Muhuri, P. K., Shukla, A. K., & Abraham, A. (2019). Industry 4.0: A bibliometric analysis and detailed overview. *Engineering Applications of Artificial Intelligence*, *78*, 218–235. doi:10.1016/j.engappai.2018.11.007.

Nakayama, R. S., de Mesquita Spínola, M., & Silva, J. R. (2020). Towards I4.0: A comprehensive analysis of evolution from I3.0. *Computers and Industrial Engineering*, *144*, 106453. doi:10.1016/j.cie.2020.106453.

Narayanamurthy, G., & Tortorella, G. (2021). Impact of COVID-19 outbreak on employee performance – Moderating role of industry 4.0 base technologies. *International Journal of Production Economics*, *234*, 108075. doi:10.1016/j.ijpe.2021.108075.

NIF (2016). New Industrial France: Building France's industrial future. URL: https://www.economie.gouv.fr/files/files/PDF/web-dp-indus-ang.pdf.

NIST (2014). Smart Manufacturing Operations Planning and Control. URL: https://www.nist.gov/system/files/documents/2017/05/09/FY2014_SMOPAC_ProgramPlan.pdf.

Oesch, D., & Menés, J. R. (2011). Upgrading or polarization? Occupational change in Britain, Germany, Spain and Switzerland, 1990-2008. *Socio-Economic Review*, *9*, 503–531. doi:10.1093/ser/mwq029.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks : Generalizing degree and shortest paths. *Social Networks*, *32*, 245–251. doi:10.1016/j.socnet.2010.03.006.

Osterrieder, P., Budde, L., & Friedli, T. (2020). The smart factory as a key construct of industry 4.0: A systematic literature review. *International Journal of Production Economics*, *221*, 107476. doi:10.1016/j.ijpe.2019.08.011.

Oxford University Press (2018). Oxford Dictionary. URL: http://en.oxforddictionaries.com/definition/technology.

Pan, J., & McElhannon, J. (2018). Future Edge Cloud and Edge Computing for Internet of Things Applications. *IEEE Internet of Things Journal*, *5*, 439–449. doi:10.1109/JIOT.2017.2767608.

Paritala, P. K., Manchikatla, S., & Yarlagadda, P. K. (2017). Digital Manufacturing- Applications Past, Current, and Future Trends. *Procedia Engineering*, *174*, 982–991. doi:10.1016/j.proeng.2017.01.250.

Paschou, T., Rapaccini, M., Adrodegari, F., & Saccani, N. (2020). Digital servitization in manufacturing: A systematic literature review and research agenda. *Industrial Marketing Management*, *89*, 278–292. doi:10.1016/j.indmarman.2020.02.012.

Pereira, A. C., & Romero, F. (2017). A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manufacturing*, *13*, 1206–1214. doi:10.1016/j.promfg.2017.09.032.

Peruzzini, M., Grandi, F., & Pellicciari, M. (2020). Exploring the potential of Operator 4.0 interface and monitoring. *Computers & Industrial Engineering*, *139*, 105600. doi:10.1016/j.cie.2018.12.047.

Pinzone, M., Albè, F., Orlandelli, D., Barletta, I., Berlin, C., Johansson, B., & Taisch, M. (2020). A framework for operative and social sustainability functionalities in Human-Centric Cyber-Physical Production Systems. *Computers & Industrial Engineering*, *139*, 105132. doi:10.1016/j.cie.2018.03.028.

Pirola, F., Boucher, X., Wiesner, S., & Pezzotta, G. (2020). Digital technologies in product-service systems: a literature review and a research agenda. *Computers in Industry*, *123*, 103301. doi:10.1016/j.compind.2020.103301.

Plitsos, S., Repoussis, P. P., Mourtos, I., & Tarantilis, C. D. (2017). Energy-aware decision support for production scheduling. *Decision Support Systems*, *93*, 88–97. doi:10.1016/j.dss.2016.09.017.

Porter, M. E., & Heppelmann, J. E. (2014). How Smart, Connected Product Are Transforming Competition. *Harvard Business Review*, (pp. 64 – 89). doi:10.1017/CBO9781107415324.004.

Prabhakaran, T., Lathabai, H. H., & Changat, M. (2015). Detection of paradigm shifts and emerging fields using scientific network: A case study of Information Technology for Engineering. *Technological Forecasting & Social Change*, *91*, 124–145. doi:10.1016/j.techfore.2014.02.003.

Qin, J., Liu, Y., & Grosvenor, R. (2016). A Categorical Framework of Manufacturing for Industry 4.0 and beyond. *Procedia CIRP*, *52*, 173–178. doi:10.1016/j.procir.2016.08.005.

Rabetino, R., Harmsen, W., Kohtamäki, M., & Sihvonen, J. (2018). Structuring servitization-related research. *International Journal of Operations & Production Management*, *38*, 350–371. doi:10.1108/IJOPM-03-2017-0175.

Reischauer, G. (2018). Industry 4.0 as policy-driven discourse to institutionalize innovation systems in manufacturing. *Technological Forecasting and Social Change*, *132*, 26–33. doi:10.1016/j.techfore.2018.02.012.

Riel, A., Kreiner, C., Macher, G., & Messnarz, R. (2017). Integrated design for tackling safety and security challenges of smart products and digital manufacturing. *CIRP Annals - Manufacturing Technology*, *66*, 177–180. doi:10.1016/j.cirp.2017.04.037.

van Rijnsoever, F. J., van Den Berg, J., Koch, J., & Hekkert, M. P. (2015). Smart innovation policy: How network position and project composition affect the diversity of an emerging technology. *Research Policy*, *44*, 1094–1107. doi:10.1016/j.respol.2014.12.004.

Risch, J., & Krestel, R. (2019). Domain-specific word embeddings for patent classification. *Data Technologies and Applications*, *53*, 108–122. doi:10.1108/DTA-01-2019-0002.

Robla-Gomez, S., Becerra, V. M., Llata, J. R., Gonzalez-Sarabia, E., Torre-Ferrero, C., & Perez-Oria, J. (2017). Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments. *IEEE Access*, *5*, 26754–26773. doi:10.1109/ACCESS.2017.2773127.

Roblek, V., Meško, M., & Krapež, A. (2016). A Complex View of Industry 4.0. *SAGE Open*, *6*. doi:10.1177/2158244016653987.

Roca, J. B., Vaishnav, P., Mendonça, J., & Morgan, M. G. (2017). Getting past the hype about 3-D printing. *MIT Sloan Management Review*, *58*, 57–62. URL: http://mitsmr.com/2l0tcXn.

Romero, D., Stahre, J., & Taisch, M. (2020). The Operator 4.0: Towards socially sustainable factories of the future. *Computers & Industrial Engineering*, *139*, 106128. doi:10.1016/j.cie.2019.106128.

Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-berglund, Å., & Gorecky, D. (2016). Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies. In *Proceedings of the international conference on computers and industrial engineering (CIE46), Tianjin, China* (pp. 29–31).

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*. Technical Report The Boston Consulting Group.

Santos, C., Mehrsai, A., Barros, A. C., Araújo, M., & Ares, E. (2017). Towards Industry 4.0: an overview of European strategic roadmaps. *Procedia Manufacturing*, *13*, 972–979. doi:10.1016/j.promfg.2017.09.093.

Saucedo-Martínez, J. A., Pérez-Lara, M., Marmolejo-Saucedo, J. A., Salais-Fierro, T. E., & Vasant, P. (2018). Industry 4.0 framework for management and operations: a review. *Journal of Ambient Intelligence and Humanized Computing*, *9*, 789–801. doi:10.1007/s12652-017-0533-1.

Schmoch, U. (2008). *Concept of a Technology Classification for Country Comparisons*. Technical Report June World Intellectual Property Organisation (WIPO).

Schuh, G. G., Anderi, R., Gausemeier, J. J., Anderl, R., Gausemeier, J. J., ten Hompel, M. M., Wahlster, W. E., Ander, L., Gausemeier, J. J., ten Hompel, M. M., & Wahlster, W. E. (2020).

Industrie 4.0 Maturity Index. Managing the Digital Transformation of Companies. *Acatech Study*, (p. 64). URL: www.acatech.de/publikationen.www.acatech.de/publikationen.

Schumacher, A., Erol, S., & Sihn, W. (2016). A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises. In *Procedia CIRP* (pp. 161–166). volume 52. doi:10.1016/j.procir.2016.07.040.

Schwab, K. (2016). *The Fourth Industrial Revolution*. (1st ed.). World Economic Forum.

Seager, C. (2016). Will jobs exist in 2050? URL: https://www.theguardian.com/careers/2016/oct/13/will-jobs-exist-in-2050.

Segerstrom, P. S. (1991). Innovation, Imitation, and Economic Growth. *Journal of Political Economy*, *99*, 807–827. doi:10.1086/261779.

Segura, Á., Diez, H. V., Barandiaran, I., Arbelaiz, A., Álvarez, H., Simões, B., Posada, J., García-Alonso, A., & Ugarte, R. (2020). Visual computing technologies to support the Operator 4.0. *Computers & Industrial Engineering*, *139*, 105550. doi:10.1016/j.cie.2018.11.060.

Shah, P. K., Perez-Iratxeta, C., Bork, P., & Andrade, M. A. (2003). Information extraction from full text scientific articles: Where are the keywords? *BMC Bioinformatics*, *4*, 1–9. doi:10.1186/1471-2105-4-20.

Shah, R. (1983). Advanced manufacturing systems in Modern Society. *Computers in Industry*, *4*, 31–46. doi:10.1016/0166-3615(83)90051-9.

Sharp, M., Ak, R., & Hedberg, T. (2018). A survey of the advancing use and development of machine learning in smart manufacturing. *Journal of Manufacturing Systems*, *48*, 170–179. doi:10.1016/j.jmsy.2018.02.004.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, *28*, 758–775. doi:10.1016/j.technovation.2008.03.009.

Silverman, B. S. (2002). *Technological resources and the logic of corporate diversification*. London: Routledge.

Sisinni, E., Saifullah, A., Han, S., Jennehag, U., & Gidlund, M. (2018). Industrial internet of things: Challenges, opportunities, and directions. *IEEE Transactions on Industrial Informatics*, *14*, 4724–4734. doi:10.1109/TII.2018.2852491.

Sjödin, D. R., Parida, V., Leksell, M., & Petrovic, A. (2018). Smart Factory Implementation and Process Innovation. *Research Technology Management*, *61*, 22–31. doi:10.1080/08956308.2018.1471277.

Strandhagen, J. O., Vallandingham, L. R., Fragapane, G., Strandhagen, J. W., Stangeland, A. B. H., & Sharma, N. (2017). Logistics 4.0 and emerging sustainable business models. *Advances in Manufacturing*, *5*, 359–369. doi:10.1007/s40436-017-0198-1.

Strozzi, F., Colicchia, C., Creazza, A., & Noè, C. (2017). Literature review on the 'smart factory' concept using bibliometric tools. *International Journal of Production Research*, *55*, 1–20. doi:10.1080/00207543.2017.1326643.

Strubell, E., Verga, P., Belanger, D., & McCallum, A. (2017). Fast and Accurate Entity Recognition with Iterated Dilated Convolutions. doi:10.18653/v1/D17-1283.

Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, *44*, 1445–1461. doi:10.1016/j.respol.2015.05.008.

Szalavetz, A. (2019). Industry 4.0 and capability development in manufacturing subsidiaries. *Technological Forecasting and Social Change*, *145*, 384–395. doi:10.1016/j.techfore.2018.06.027.

Tao, F., Cheng, Y., Xu, L. D., Zhang, L., & Li, B. H. (2014). CCIoT-CMfg : Cloud Computing and Internet of Things-Based Cloud Manufacturing Service System. *IEEE Transactions on Industrial Informatics*, *10*, 1435–1442. doi:10.1109/TII.2014.2306383.

Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, *48*, 157–169. doi:10.1016/j.jmsy.2018.01.006.

Teece, D. (1998). Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. *California Management Review*, *40*, 55–79. doi:oi.org/10.2307/41165943.

Thoben, K.-D., Wiesner, S., & Wuest, T. (2017). "Industrie 4.0" and Smart Manufacturing – A Review of Research Issues and Application Examples. *International Journal of Automation Technology*, *11*, 4–16. doi:10.20965/ijat.2017.p0004.

Tong, S., & Koller, D. (2001). Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, *2*, 45–66.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, *4*, 178–185. doi:10.1074/jbc.M501708200.

Van Looy, B., Vereyen, C., & Schmoch, U. (2015). *Patent Statistics: Concordance IPC V8 – NACE REV.2 (version 2.0)*. Technical Report Eurostat.

Viriyasitavat, W., Da Xu, L., Bi, Z., & Sapsomboon, A. (2020). Blockchain-based business process management (BPM) framework for service composition in industry 4.0. *Journal of Intelligent Manufacturing*, *31*, 1737–1748. doi:10.1007/s10845-018-1422-y.

Wang, S., Wan, J., Zhang, D., Li, D., & Zhang, C. (2016). Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. *Computer Networks*, *101*, 158–168.

Wang, T., Luo, H., Jia, W., Liu, A., & Xie, M. (2020). MTES: An Intelligent Trust Evaluation Scheme in Sensor-Cloud-Enabled Industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, *16*, 2054–2062. doi:10.1109/TII.2019.2930286.

Wang, X. V., & Wang, L. (2021). A literature survey of the robotic technologies during the COVID-19 pandemic. *Journal of Manufacturing Systems*, *60*, 823–836. doi:10.1016/j.jmsy.2021.02.005.

Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. *Available at SSRN 3482150*, .

Wu, D., Greer, M. J., Rosen, D. W., & Schaefer, D. (2013). Cloud manufacturing: Strategic vision and state-of-the-art. *Journal of Manufacturing Systems*, *32*, 564–579. doi:10.1016/j.jmsy.2013.04.008.

Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: state of the art and future trends. *International Journal of Production Research*, *56*, 2941–2962. doi:10.1080/00207543.2018.1444806.

Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing*, *28*, 75–86. doi:10.1016/j.rcim.2011.07.002.

Yadav, V., & Bethard, S. (2019). A Survey on Recent Advances in Named Entity Recognition from Deep Learning models, .

Yoon, J., & Kim, K. (2012). TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents. *Expert Systems with Applications*, *39*, 2927–2938. doi:10.1016/j.eswa.2011.08.154.

Yu, C., Xu, X., & Lu, Y. (2015). Computer-Integrated Manufacturing, Cyber-Physical Systems and Cloud Manufacturing - Concepts and relationships. *Manufacturing Letters*, *6*, 5–9. doi:10.1016/j.mfglet.2015.11.005.

Yun, J. H. J., Jeong, E. S., & Park, J. S. (2016). Network analysis of open innovation. *Sustainability (Switzerland)*, *8*, 1–21. doi:10.3390/su8080729.

Zaghloul, W., Lee, S. M., & Trimi, S. (2009). Text classification: neural networks vs support vector machines. *Industrial Management & Data Systems*, *109*, 708–717. doi:10.1108/02635570910957669.

Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C. (2014). "Term clumping" for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change*, *85*, 26–39. doi:10.1016/j.techfore.2013.12.019.

Zheng, T., Ardolino, M., Bacchetti, A., & Perona, M. (2020). The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review. *International Journal of Production Research*, *0*, 1–33. doi:10.1080/00207543.2020.1824085.

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering*, *3*, 616–630. doi:10.1016/J.ENG.2017.05.015.

Zhou, J., Li, P., Zhou, Y., Wang, B., Zang, J., & Meng, L. (2018). Toward New-Generation Intelligent Manufacturing. *Engineering*, *4*, 11–20. doi:10.1016/j.eng.2018.01.002.

Zolotová, I., Papcun, P., Kajáti, E., Miškuf, M., & Mocnej, J. (2020). Smart and cognitive solutions for Operator 4.0: Laboratory H-CPPS case studies. *Computers & Industrial Engineering*, *139*, 105471. doi:10.1016/j.cie.2018.10.032.

Appendix A

Appendix for Chapter 2

A.1 Graphical view of the evolution of the Industry 4.0 technology network

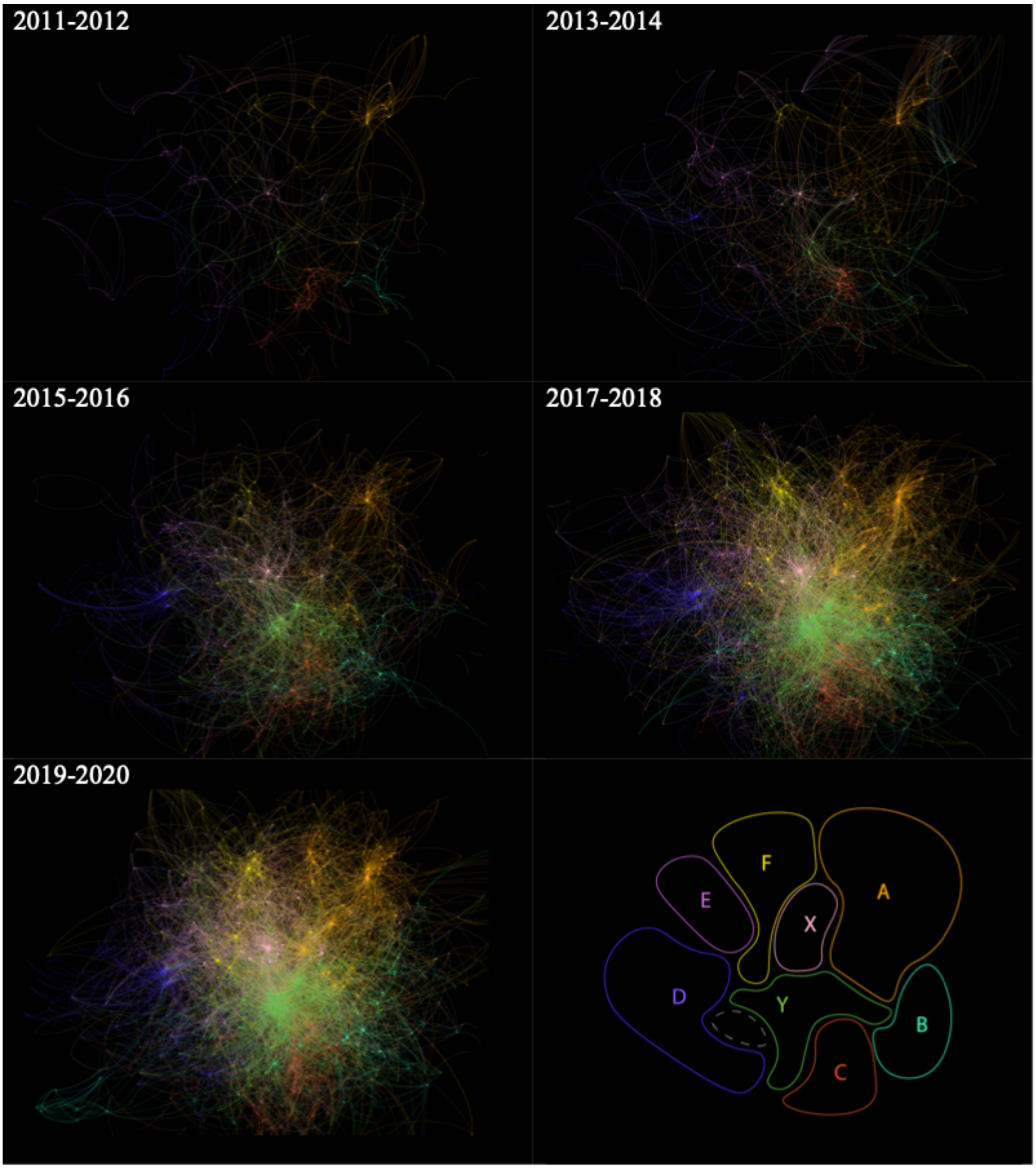


Figure A.1: Development of the Industry 4.0 technology map between 2011 and 2020. Each picture shows the technology map, based on the research articles published in the according years.

Appendix B

Appendix for Chapter 3

B.1 Relevance scores of Industry 4.0 articles

For selection of relevant articles for Chapter 3, I use a machine learning-based filtering method. Figure B.1 shows the relevance scores associated to each article, based on the method described in Section 3.3.1.

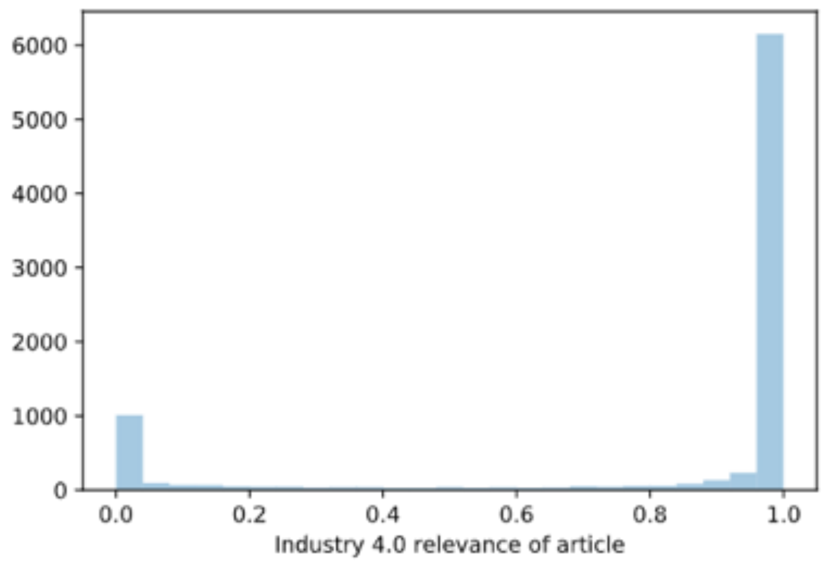


Figure B.1: Frequency of articles' Industry 4.0 relevance scores. Each bar represents a bin of 0,04 which means, e.g., that more than 6000 articles had a relevance score of 0.96 or higher. Overall 6,938 articles have a relevance score of equal or above 0.5. 1,540 articles with a score below 0.5 are excluded in the further analysis.

Appendix C

Appendix for Chapter 4

C.1 Deep-dive on patent structure in the sample

Our dataset builds on the PATSTAT database. There, we only select granted patents, where English abstracts and titles are available. Similar patents filed in different offices are grouped into patent families. We only include one patent per family in our sample to avoid double-counting of inventions. Our analysis builds on English-language text; we therefore prioritize patents from countries with English as a first language, followed by European Patent Office (EPO) and World Intellectual Property Organization (WIPO) patents, and then remaining countries. This leads to all relevant United States Patent and Trademark Office (USPTO) patents forming part of our dataset. In addition, many patents from the Japanese, Chinese and Korean offices are included, followed by other patent offices (see Figure C.1).

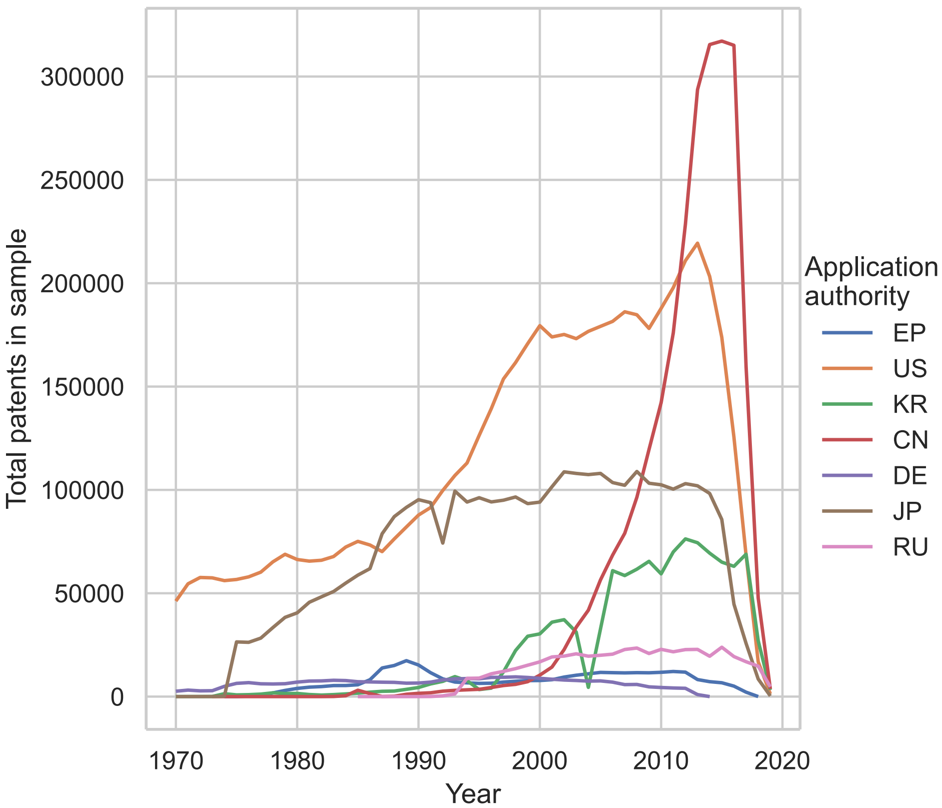


Figure C.1: Number of total patents in our sample by patent authority. Only includes authorities contributing more than 300,000 patents, those are authorities from Europe (EP), the US, Korea (CR), China (CN), Germany (DE), Japan (JP) and Russia (RU). Patents filed in multiple offices have been prioritized as described in section 4.3.1

C.2 Validity check through comparing both word embeddings.

Our approach compares texts of patent abstracts and task descriptions. Calculation of text similarity relies on word embedding (see Section 4.3.2). Our results comprise relevant patents identified through two different embeddings. One embedding represents patent-specific language; the other embedding builds on a more general language (common crawl embeddings). Our approach identifies almost twice as many patents through the patent-specific embeddings. Only 35% of the patent-task relations identified through the general language embeddings overlap with the patent-specific embeddings.

We compare the results of both embeddings, to validate our overall results. First, we evaluate, if an embedding biases results toward certain technologies. Table C.1 shows that the results per technology sector only differ slightly between both approaches. Therefore, we do not expect a bias towards certain technologies. Second, we examine whether there is a bias towards certain task types, e.g., if more patents are mapped to certain tasks. Table C.2 shows no indication of a bias of embeddings towards a task type. Those identical patterns indicate that the approach is robust enough so that the choice of word embedding did not systematically bias the results.

Table C.1: The share of patents per technology type is highly similar for both word embeddings.

	Share of patents per technology sector	
	General corpus	Patent corpus
Chemistry	8%	10%
Electrical engineering	50%	46%
Instruments	18%	19%
Mechanical engineering	16%	17%
Other fields	8%	9%

Note: Percentages do not sum up due to rounding

Table C.2: The share of patents per task type, is highly similar for both word embeddings.

	Share of patents per task type	
	General corpus	Patent corpus
Information input	15%	14%
Interacting with Others	14%	16%
Mental processes	14%	12%
Work output	58%	58%

Note: Percentages do not sum up due to rounding

C.3 Patent per technology fields and task type

PATSTAT provides a classification of patents into technology clusters, based on Schmoch (2008). These PATSTAT technology groups are different technology groups than the 4IR technology groups that are described, e.g., in Section 4.5.2. We use the PATSTAT technology classification to evaluate our results and show to which technology clusters our results refer. These results are particularly relevant for validating our mapping. The results show that patents in the field of information technology (IT) methods for management are, on average, relevant for most tasks. Also control, computer technology, and digital communication technologies are frequently linked to occupations. Other fields, such as chemistry or nanotechnology do not provide many direct links to tasks. Further details on total numbers of patents are provided in C.3.

Our mapping provides links of patents to occupations at a task level. In Table C.3 we explore which tasks are exposed to which technology fields. Therefore, we group tasks into four high-level activity categories and evaluate the share of patents per activity category associated with each of the technology fields.

Table C.3: Technology clusters per activity group.

Technology field	Information input	Interacting with Others	Mental processes	Work output
Electrical machinery, apparatus, energy	3%	2%	2%	4%
Audio-visual technology	4%	3%	4%	5%
Telecommunications	4%	5%	4%	3%
Digital communication	6%	9%	8%	5%
Computer technology	16%	16%	26%	12%
IT methods for management	5%	17%	12%	4%
Semiconductors	2%	1%	1%	2%
Optics	2%	1%	1%	2%
Measurement	24%	3%	8%	4%
Control	5%	10%	7%	3%
Medical technology	4%	5%	3%	2%
Handling	2%	3%	2%	5%
Machine tools	2%	1%	2%	8%
Engines, pumps, turbines	2%	1%	1%	2%
Other special machines	2%	2%	2%	5%
Transport	3%	4%	2%	4%
Furniture, games	2%	5%	2%	2%
Other consumer goods	1%	2%	1%	2%
Civil engineering	2%	2%	2%	7%

Note: O*Net clusters tasks into four broad activity groups. The values indicate the shares of patents of technology clusters comprising broad activity categories. The figure includes only technology fields representing more than 1 % of patents; thus, not all columns sum up to 100%.

Information input tasks have a particularly high share of measurement patents; mental process tasks and interaction with other tasks have a particularly high share of computer technology, IT, and communication and control patents. Finally, work-output-related tasks have a high share of machine tool and civil engineering patents. Those findings confirm expectations on task-technology links and suggest the validity of our mapping.

Table C.4 shows the total size of each technology field (the number of patents per cluster in our mapping) and indicates for each cluster the mean frequency (number) of tasks to which a patent is related. The results show that patents in the field of IT methods for management are, on average, relevant for most tasks. Also control, computer technology, and digital communication technologies are frequently linked to occupations. Other fields, such as chemistry or nanotechnology, do not provide many direct links to tasks.

Table C.4: Patents per technology field.

Technology field	Patent count [10^6]	Frequency per patent
Electrical machinery, apparatus, energy	34	32
Audio-visual technology	48	73
Telecommunications	38	86
Digital communication	63	128
Computer technology	161	156
IT methods for management	70	612
Semiconductors	18	31
Optics	17	28
Measurement	80	98
Control	50	180
Medical technology	32	60
Materials, metallurgy	17	39
Handling	36	74
Machine tools	52	99
Engines, pumps, turbines	20	46
Textile and paper machines	17	49
Other special machines	36	63
Thermal processes and apparatus	17	59
Mechanical elements	22	45
Transport	39	56
Furniture, games	28	64
Civil engineering	53	77

Note: Bars indicate total number of patents per field. Coloring indicates the number of tasks an average patent in a given field is associated with. Only the most important technology fields, covering 90% of patents, are included. Excluded fields are mainly related to nanotechnology, biotechnology, and chemistry.

Next, we evaluate results at an occupation level. We therefore group occupations based on national career clusters[1] from Career Technical Education (CTE), comprising occupations with overlapping activities. For each career cluster, Table C.5 shows the composition of patents, as shares of patents associated with one of five high-level technology fields. (High-level technology fields comprise technology fields as, for example, used in Figure C.4).

[1] `https://careertech.org/sites/default/files/Perkins_IV_Crosswalk_Table_5_SOC-ONET-Nontrad-Cluster-Pathway.xls`

Table C.5: Technology clusters per career path.

	Technology fields				
Career clusters	Chemistry	Electrical engineering	Instruments	Mechanical engineering	Other fields
Agriculture, Food & Natural Resources	17%	36%	15%	25%	7%
Architecture & Construction	12%	26%	13%	28%	21%
Arts, Audio/Video Technology & Communicatio	4%	61%	13%	11%	11%
Business Management & Administration	4%	68%	15%	9%	5%
Education & Training	3%	59%	22%	7%	8%
Finance	1%	77%	15%	3%	4%
Finance; Finance	2%	77%	13%	4%	4%
Finance; Human Services	1%	80%	14%	3%	3%
Government & Public Administration	4%	62%	19%	11%	5%
Government & Public Administration; Science,	5%	67%	15%	7%	6%
Health Science	15%	38%	34%	7%	5%
Health Science; Science, Technology, Engineeri	6%	57%	24%	10%	3%
Hospitality & Tourism	15%	41%	14%	14%	18%
Human Services	10%	50%	17%	11%	12%
Law, Public Safety, Corrections & Security	4%	58%	19%	11%	8%
Manufacturing	12%	31%	17%	31%	9%
Marketing	2%	71%	15%	6%	6%
Science, Technology, Engineering & Mathemati	9%	56%	21%	10%	4%
Transportation, Distribution & Logistics	7%	33%	18%	34%	8%

Note: The values indicate the shares of patents of technology clusters comprising work activity categories. A mapping of occupations to technology groups is also available.

C.4 4IR exposure sub scores

C.4.1 Example exposure of occupation to 4IR sub-scores

Table C.6: Example exposure of occupations to 4IR exposure sub scores.

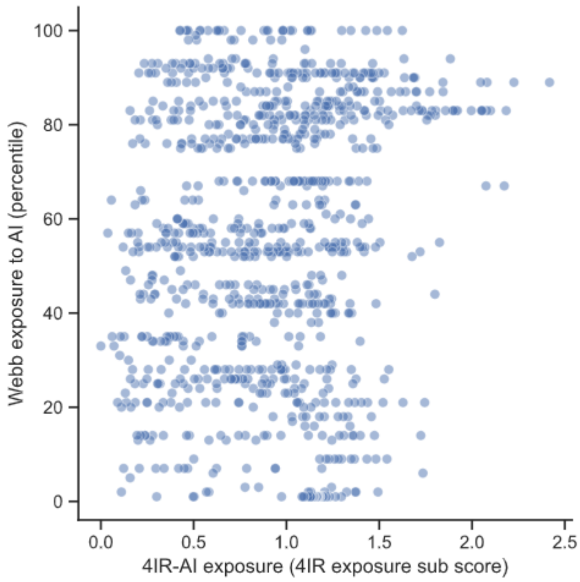
% of total exposure per technology

0.0000% 0.4000%

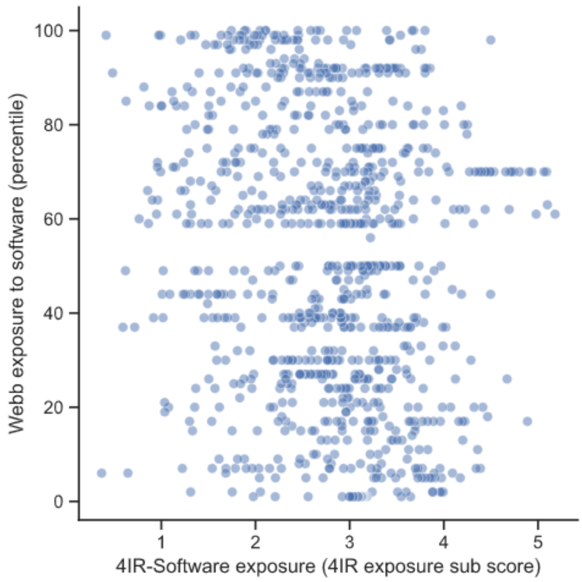
SOC Occupation Title	Agriculture	Artificial Intelligence	Augmented reality for surgery and diagnosis	CAD	Load transporting vehicles	Marketing	Natural language processing (NLP)	Smart Office
Farmworkers and Laborers, Crop	0.17	0.02	0.00	0.05	0.11	0.01	0.01	0.05
Graphic Designers	0.00	0.03	0.00	0.76		0.04	0.03	0.05
Human Resources Managers	0.01	0.05	0.00	0.01	0.00	0.06	0.02	0.22
Laborers and Freight, Stock, and Material Movers, Hand	0.01	0.01	0.00	0.04	0.58	0.01	0.00	0.03
Lawyers	0.00	0.07		0.02		0.04	0.05	0.12
Management Analysts	0.00	0.12	0.00	0.02		0.09	0.03	0.12
Nursing Assistants	0.01	0.03	0.03	0.02	0.02	0.03	0.01	0.07
Retail Salespersons	0.00	0.03	0.00	0.02	0.01	0.17	0.02	0.05

Note: Table scores indicate the exposure of the occupation to the relevant technology. The shading refers to the share of the occupation-technology exposure of the overall exposure of the relevant technology (e.g., what share of CAD patents is related to designers).

C.4.2 4IR exposure sub scores and Webb patent exposure



(a) Exposure to 4IR-AI patents



(b) Exposure to 4IR-Software patents

Figure C.2: Comparison of 4IR exposure sub-scores and Webb scores. The figures show 4IR-exposure sub-scores for AI patents and software patents to compare to Webb's exposure to AI and software patents.

C.5 Exposure to non-4IR patents

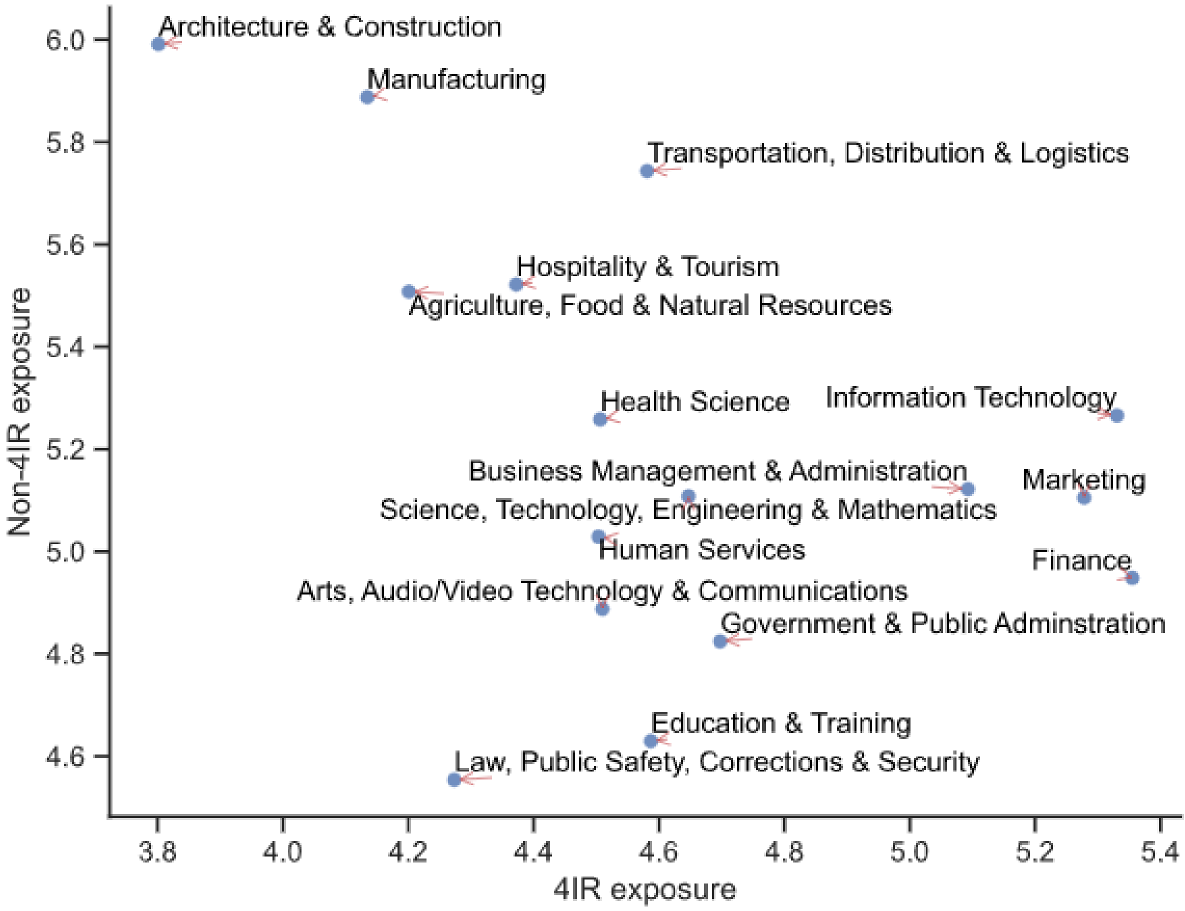
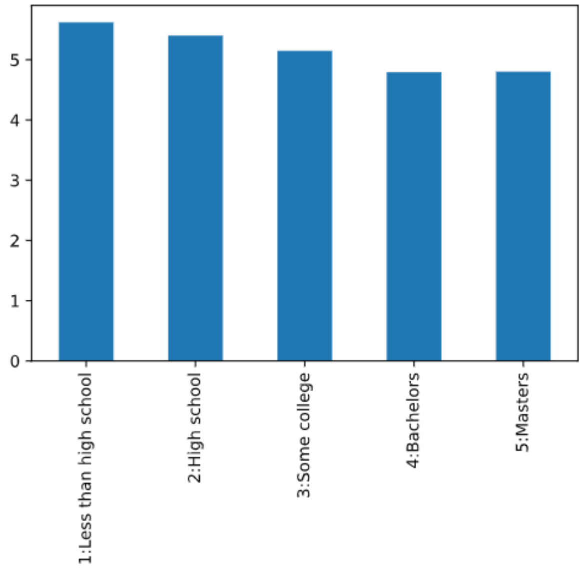


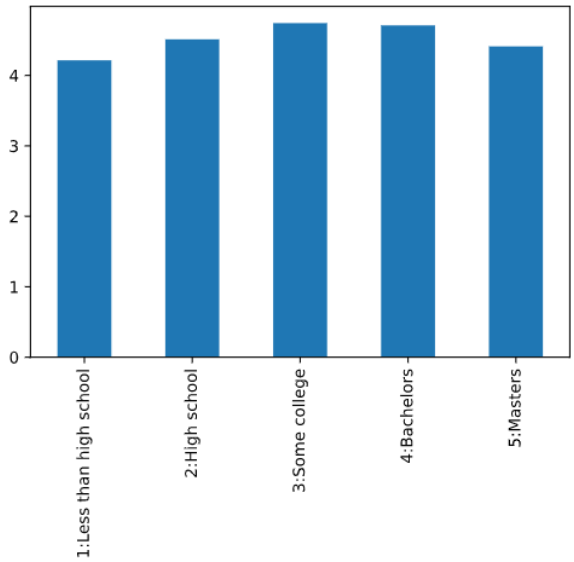
Figure C.3: Exposure to 4IR vs. non-4IR patents.

C.6 Education and patent exposure

Figure C.4 describes the exposure to patents per education level for 4IR patents and non-4IR patents. The analysis shows that non-4IR patents show an inverse correlation to education level, whereas 4IR exposure is highest for medium-to-high education levels. C.7 provides similar insights based on an analysis of patent exposure per job zone.



(a) Non-4IR patents.



(b) 4IR patents

Figure C.4: Exposure to patents per education level. Education level is derived from O*Net occupation data and occupations are aggregated weighted by share of the total jobs, as provided by BLS.

Webb (2019) found that exposure to robot and software patents is inversely correlated to education level and is lowest for low-education occupations. This pattern is similar to our exposure scores to non-4IR patents. Webb's AI patent exposure is higher for higher-educated occupations, whereas 4IR exposure scores are highest for medium-to-high occupations.

C.7 Total number of patents clustered along job zones.

We group occupations based on the skill level required. Therefore, we use the O*Net job zones, which reflect the experience, education, and skills required to conduct a job. Job zone 1 describes low-skilled occupations and job zone 5 the highest skill level. The analysis shows that the exposure of high-skilled jobs to patents increased more strongly than for lower-skilled jobs. Particularly in the 1990s, when 4IR patents increased, the exposure of high-skilled occupations increased more strongly than low-skilled exposure. Whereas low-skilled occupations have a higher overall patent exposure, high-skilled occupations are more exposed to 4IR patents.

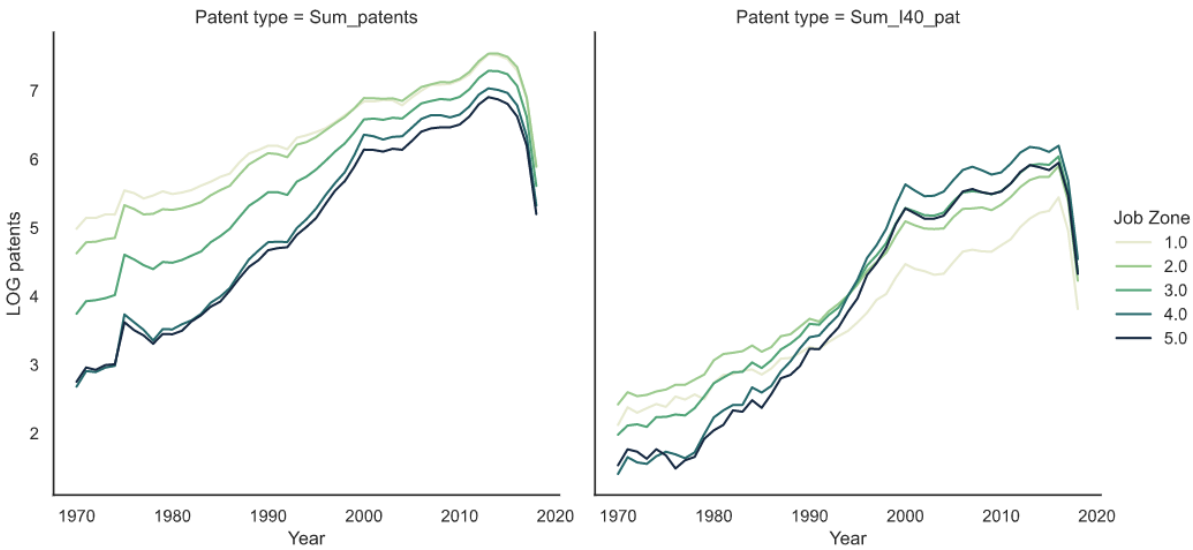


Figure C.5: Total number of patents associated with occupations, clustered along job zones. The left graph shows total number of patents; the right graph only includes 4IR patents (4IR patent exposure).

C.8 Comparison of SML and CP scores

Figure C.6 compares computerization probability (CP) and SML scores. We group occupation scores by SOC career clusters for better readability.

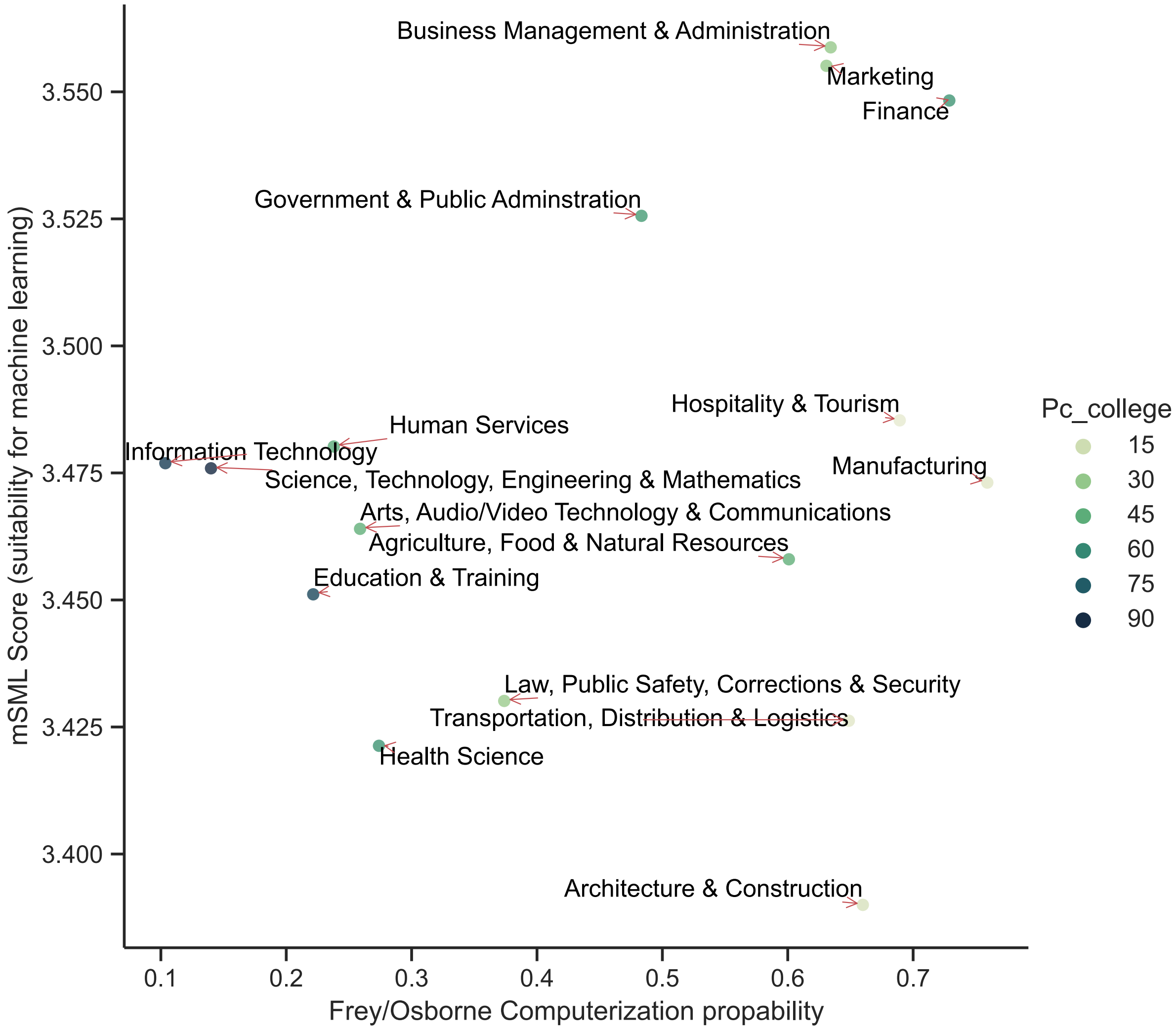


Figure C.6: Comparison of SML (Brynjolfsson & Mitchell, 2017) and CP scores (Frey & Osborne, 2017). Color coding indicates the mean share of workers with college degree per SOC Career Cluster.

C.9 MGI score, SML score and 4IR exposure

Figure C.7 shows that McKinsey Global Institute (MGI) automation estimates are highest for low 4IR exposure occupations and for those occupations with the highest 4IR exposure scores and high suitability for machine learning (SML) scores. The MGI score describes automation potential until 2030. Therefore, the graph indicates that MGI automation is mainly driven by non-4IR technologies, with the exception of those occupations with highest 4IR exposure scores and SML scores. The highest 4IR/SML occupations are the first ones likely to feel the impact of 4IR automation. They include administrative support, sales and service, quality assurance, and health informatics occupations. Additionally, transportation systems and infrastructure planning occupations have high MGI scores and 4IR exposure. However, their SML scores are lower than for the other examples; thus, these occupations may be among the first to see benefits through labor augmentation.

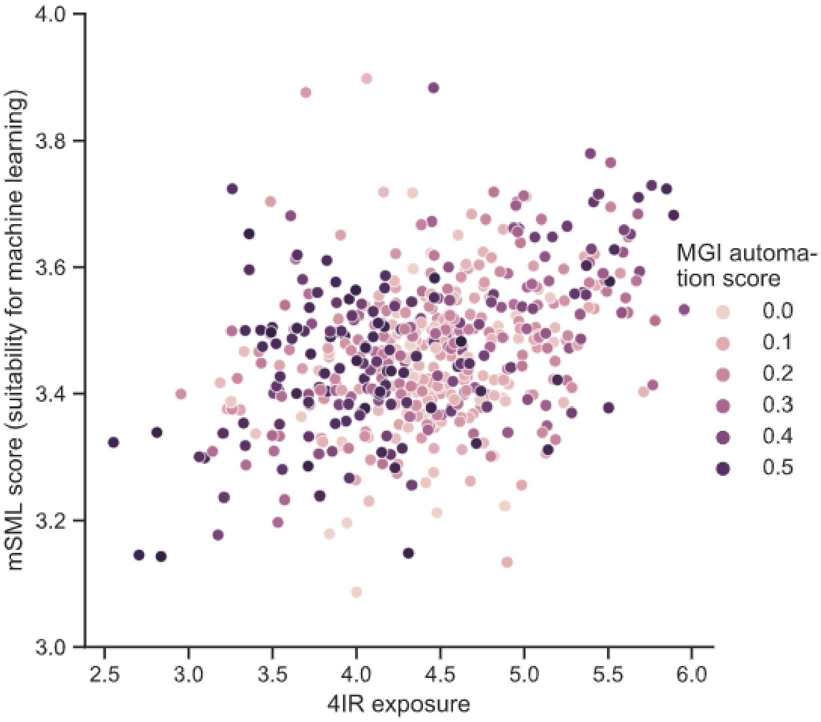
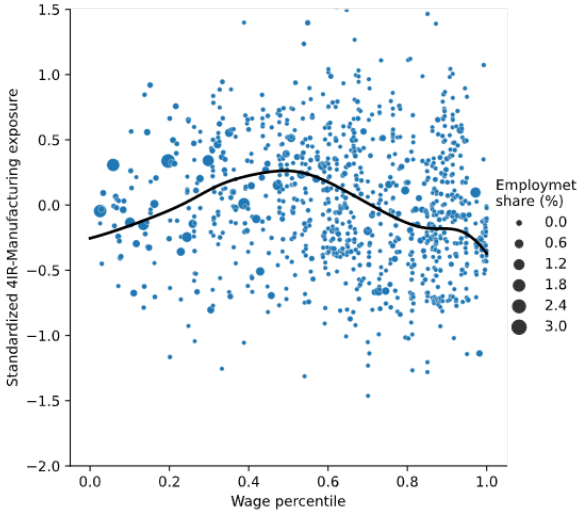
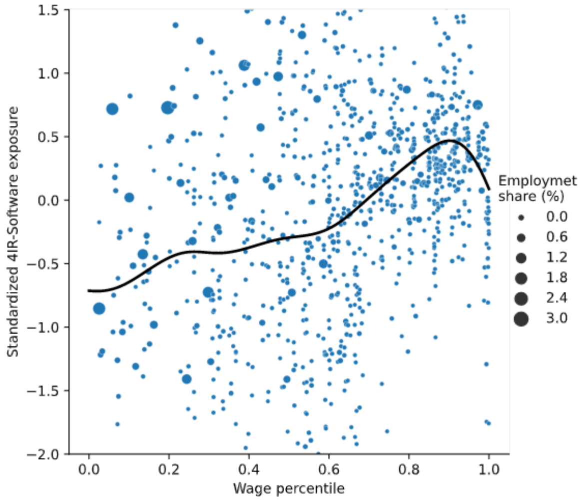


Figure C.7: The graph shows the matrix of suitability of machine learning and 4IR patent exposure scores; coloring indicates MGI automation scores.

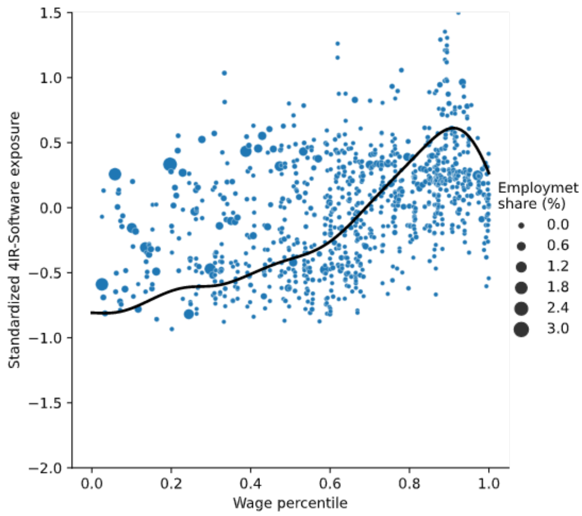
C.10 4IR exposure sub-scores per wage percentile



(a) Exposure to 4IR-Manufacturing patents



(b) Exposure to 4IR-Software patents



(c) Exposure to 4IR-AI patents

Figure C.8: 4IR sub-score exposure per wage percentile.

C.11 Impact of historic patent exposure on jobs

Table C.7: Exposure to historic 4IR patents and change in employment 2012-2018.

	(1)	(2)	(3)	(4)	(5)
4IR exposure 1982	-0.10**				
	(-2.00)				
4IR exposure 1982^2	-0.07**				
	(-2.31)				
4IR exposure 1992		-0.05			
		(-1.18)			
4IR exposure1992^2		-0.11***			
		(-4.23)			
4IR exposure 2002			-0.07		
			(-1.42)		
4IR exposure 2002^2			-0.12***		
			(-3.76)		
4IR exposure 2007				-0.07	
				(-1.38)	
4IR exposure 2007^2				-0.12***	
				(-3.70)	
4IR exposure 2012					-0.07
					(-1.44)
4IR exposure 2012^2					-0.11***
					(-3.60)
Industry Education	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.102	0.117	0.107	0.105	0.103

Note: Analysis of 720 observations based on BLS labor market data from 2011 and 2018. Controls IE indicates that industry and education (Job Zone) controls were included. Industry share relates to the industry share of the occupation in 2011. * p<0.10, ** p<0.05, *** p<0.01.

Table C.8: Exposure to historic 4IR patents and change in employment 2012-2018.

	(6)	(7)	(8)	(9)	(10)
4IR exposure 1982	-0.15***				
	(-3.18)				
4IR exposure 1992		-0.13***			
		(-3.00)			
4IR exposure 2002			-0.10**		
			(-2.20)		
4IR exposure 2007				-0.10**	
				(-2.03)	
4IR exposure 2012					-0.09*
					(-1.83)
Industry Education	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.096	0.095	0.090	0.089	0.088

Note: Analysis of 704 observations based on BLS labor market data from 2011 and 2018. Controls IE indicates that industry and education (Job Zone) controls were included. Industry share relates to the industry share of the occupation in 2011. * p<0.10, ** p<0.05, *** p<0.01.

Table C.9: Exposure to 4IR patents and change in employment 2012-2018, no control variables.

	(11)	(12)	(13)	(14)	(15)
4IR exposure 1982	-0.23***				
	(-6.33)				
4IR exposure 1992		-0.17***			
		(-4.61)			
4IR exposure 2002			-0.07*		
			(-1.79)		
4IR exposure 2007				-0.05	
				(-1.35)	
4IR exposure 2012					-0.05
					(-1.44)
Industry Education	No	No	No	No	No
Adj. R^2	0.053	0.028	0.003	0.001	0.002

Note: Analysis of 720 observations based on BLS labor market data from 2011 and 2018. Controls IE indicates that industry and education (Job Zone) controls were included. Industry share relates to the industry share of the occupation in 2011. * p<0.10, ** p<0.05, *** p<0.01.

C.12 Task- and activity-level 4IR exposure scores

The analysis provides exposure scores for more than 20k tasks. This section provides some example scores at a task level as well as aggregated scores at a work activity level. Table C.10 shows the highest and lowest exposure scores at a work activity level. C.11 shows exposure scores for randomly-selected detailed work activities.

Table C.10: Highest and lowest 4IR exposure scores per work activity, based on an aggregation of task-level exposure scores.

Work Activities Element Name	Mean 4IR exposure
Interacting With Computers	5.76
Documenting/Recording Information	5.30
Selling or Influencing Others	4.98
Processing Information	4.97
...	...
Controlling Machines and Processes	4.10
Repairing and Maintaining Mechanical Equipment	4.05
Staffing Organizational Units	3.92
Resolving Conflicts and Negotiating with Others	3.82

Table C.11: Exposure scores of randomly-selected detailed work activities (DWA).

DWA Title	4IR exposure
Develop data analysis or data management procedures.	6.03
Measure equipment outputs.	5.70
Correspond with customers to answer questions or resolve complaints.	5.54
Modify software programs to improve performance.	5.42
Manage financial activities of the organization.	5.40
Process library materials.	4.99
Develop plans for programs or services.	4.77
Monitor construction operations.	3.98
Inspect operations of green energy facilities.	3.96
Develop organizational goals or objectives.	3.66