GROWINPRO

Growth Welfare Innovation Productivity

Working Paper

Generation, Diffusion and Productivity Effects of Industry 4.0 Technologies

Vanessa Behrens
ZEW-Leibniz Centre for European Economic Research Mannheim

Florence Blandinieres
ZEW-Leibniz Centre for European Economic Research Mannheim

Bettina Peters
ZEW–Leibniz Centre for European Economic Research Mannheim,
MaCCI, University of Luxembourg

16/2021 May

This project has received funding from the European Union Horizon 2020 Research and Innovation action under grant agreement No 822781

Generation, Diffusion and Productivity Effects of Industry 4.0 Technologies*

Vanessa Behrens[◊] , Florence Blandinieres[△] and Bettina Peters[§]

April 2021

Abstract

This paper explores the role of 4.0 technologies in three dimensions. First, we examine the patterns and trends in the emergence of 4.0 technologies in Europe, differentiating between technology groups (core, enabling and twin technologies). A novel feature is that we specifically focus on the digital intensity of 4.0 patents, as they show strong heterogeneity in terms of how large the share of 4.0-related technical features of the protected technology is. Second, we study the diffusion of 4.0 technologies by means of forward citations. Finally, we examine the impact of the adoption of AI technologies on firm performance. The novelty of our study is to examine the complementarity between the adoption of AI and investments in complementary intangible assets, more specifically data infrastructure. In summary, our results show that the adoption of AI does not automatically lead to productivity gains and that only firms that invest in internal complementary intangible assets do so, regardless of whether AI technologies are developed internally or external AI solutions are used. However, productivity gains are almost twice as high when internal data infrastructure investments are combined with internal AI development compared to external AI development Firms that develop AI themselves increase their productivity when they simultaneously invest in an internal data infrastructure, but not when they combine their in-house development strategy with external data infrastructure.

Keywords: Patent, Industry 4.0, Productivity, AI, Data Infrastructure

JEL Classification: O31, O33,O34

[*]**Acknowledgments**: This paper is an outcome of the GrowInPro project which was funded by the European Commission, Research Directorate General as part of the European Union Horizon 2020 Research and Innovation Action under grant agreement No 822781. Further details can be found at: www.growinpro.eu

[◊]ZEW-Leibniz Centre for European Economic Research Mannheim. vanessa.behrens@zew.de

[△]ZEW-Leibniz Centre for European Economic Research Mannheim. florence.blandinieres@zew.de

[§]ZEW-Leibniz Centre for European Economic Research Mannheim, MaCCI, University of Luxembourg. bettina.peters@zew.de

1 Introduction

The Fourth Industrial Revolution (4IR), also known as Industry 4.0, is inspiring numerous discussions in industry and politics about its various economic and organizational consequences. Behind the buzzword Industry 4.0 lies the ongoing process of automation and digitisation in manufacturing through modern smart technologies. 4IR represents the fourth major wave of technological and structural change in the production of goods and services. The first wave was characterised by mechanisation in connection with the spread of steam and water power, the second by mass production related to electricity, and finally the third wave was by automation in the course of the spread of information and communication technologies (ICTs).

Techno-optimists like Brynjolfsson and McAfee see the digital revolution and its applications in production still in its infant installation phase and expect major future productivity improvements to come (see Brynjolfsson and McAfee 2011, 2014; Brynjolfsson et al. 2017). However, unlike previous waves, the 4.0 revolution has also received much skepticism regarding its impact on productivity. Some authors argue that digitisation is responsible for the current productivity slowdown coming from innovations with diminishing returns (Gordon 2012). Other studies suggest that the overall decline in productivity growth is the result of two trends in the adoption of digital technologies: on the one hand, frontier firms are better equipped to invest in digital technologies than laggards and achieve positive productivity gains, especially in the ICT sector (see Andrews et al. 2016), and on the other hand, laggards face decreasing productivity gains because the diffusion of these new digital technologies has slowed down. Gal et al. (2019) argue that digitisation is therefore accelerating the divergence between frontier and laggard firms. The diffusion of digital technologies does not only involve investing in them, but also adapting the organization of production systems and business models (Haskel and Westlake 2018), stemming from numerous trials and errors in experimenting with the use of digital technologies (Brynjolfsson et al. 2017). The adoption of digital technologies might therefore be at a too early stage to have a positive impact on productivity.

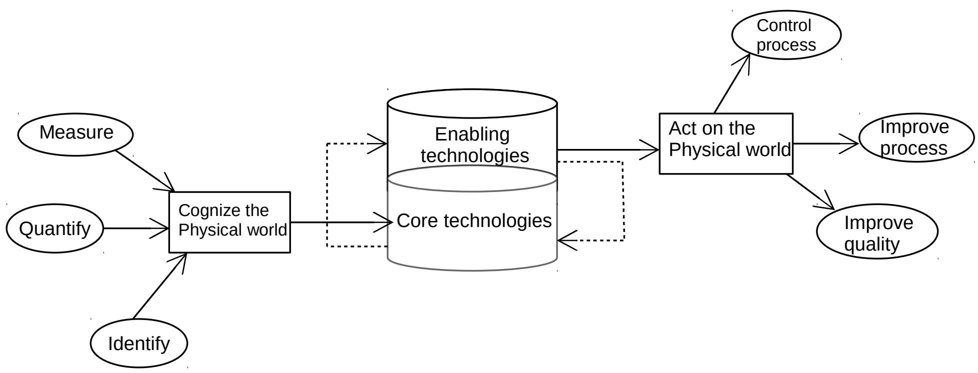
However, little is known about this diffusion process, even if empirical evidence shows a very uneven pattern across sectors (Calvino et al. 2018). This paper aims at filling this gap by assessing with different data sources (i.e., patent and survey data) and level of analysis (i.e. sectorial and firm level) the pattern of generation and diffusion of different types of 4.0 technologies as well as its impact on productivity. Section 2 starts with a categorization of 4.0 technologies, followed by a description of our data and how we measure 4.0 technologies using the recently developed classification of industry 4.0 patents in Europe. Section 3 examines patterns and trends in the generation of 4.0 technologies in their entirety and differentiated by technology groups. In section 4, we study the diffusion of 4.0 technologies. We measure diffusion by looking at European 4.0 patents and their respective pattern of forward citations across sectors. Section 5 finally focus on the productivity impact of 4.0 technologies. More specifically, we focus on artificial intelligence (AI) due to data constraints. But AI is considered to be the most important technology within 4.0 technologies. In addition to patent data, we use survey data from the German CIS 2018 for this analysis. This study is novel as it not only investigate the role of AI adoption but also investigates the complementarity between AI adoption and data infrastructure investments. Section 6 concludes.

2 Industry 4.0 Technologies

2.1 Categorization of 4.0 Technologies

The main feature of the Fourth Industrial Revolution lies in its capacity to automatize decision-making by reducing human involvement. This includes both human-to-human and human-to-machine interaction. The communication between physical and virtual technologies via the transmission of large amount of data aims at automatizing the decision making regarding the production process steps. In order to do so, one can distinguish two main types of technologies, those operating in the physical world to gather and transmit data (*core technologies*) and those linked to the exploitation of data in a broad sense (*enabling technologies*). Both combined create a system to digitalize the physical world of production into virtual characteristics (see Figure 1). "Cognizing" the features of the physical world lies at the core of digitization (Qi et al. 2019).

Figure 1: Digital Manufacturing System: Functions and Relationships Between Core and Enabling Technologies



Source: Own representation.

In this sense, virtual machines are considered "smart" by continuously learning about a given production process, leading to the term "smart manufacturing". These smart virtual machines refer to a wide range of technologies (i.e. machine learning, artificial intelligence) that aim at modelling a given phenomenon to monitor, simulate, predict, and validate the different steps of the production process.

This smart production process relies on complementary technologies (e.g. physical and virtual technologies) to achieve the data collection and treatment to be used as inputs by smart machines. These complementary technologies can be distinguished based on their respective purposes (data collection and management vs transmission) and physical features.

Core technologies. They represent established ICT fields inherited from the previous industrial revolution. Under the umbrella of core technologies, we can distinguish three categories: *hardware*, *software* and *connectivity* (see Ménière et al. 2017). *IT Hardware* technologies comprise physical components of computers or devices that aim at collecting data. Examples of hardware technologies include sensors, advanced memories, processors, adaptive displays, cameras, QR codes, and so on. The second category consists of *software* infrastructure which aims at collecting data as well. Software technologies include for example intelligent cloud storage and computing structures, adaptive databases, mobile operating systems, virtualisation, Application Programming Interfaces (APIs), web crawlers, search engines or blockchain technologies. Finally, the last category, known as *connectivity*, aims at connecting the different physical components to the enabling technologies. Connectivity technologies includes network protocols for massively connected devices, wireless technologies (WiFi, Bluetooth) or other cable transmissions (see Qi et al. 2019, for more details).

Enabling technologies. The fourth industrial revolution is additionally characterized by enabling technologies that build upon and complement these core technologies (Martinelli et al. 2020). Enabling technologies bear a high transformative potential for the productive system in which they are used for a variety of applications (Teece 2018), and therefore play an important role in providing inputs for smart manufacturing systems. The group of enabling technologies consist of a wide range of technologies that aim at improving the performance of physical entities in the physical world (Qi et al. 2019). Enabling technologies can act in *carrying a specific action* (e.g. providing energy to a machine in an intelligent way, known as power technologies), or in *improving the communication* between the physical components and the virtual machines. They intervene in enhancing the connection between physical components and data generation (GPS, visual recognition to identify pattern for example), data transmission (interface, security), or even exploitation of data (data analytics for big data, machine learning). The specialized literature in digital technologies stresses the interdependence and complementarity between the different components (e.g. physical and virtual representing the core and enabling technologies) (Qi et al. 2019; Fuller et al. 2020). Besides the technological infrastructure between core and enabling technologies, the authors also highlight the importance of having access to a large number of standardized data (Qi et al. 2019; Fuller et al. 2020). Data storage (in cloud for example) and data security represent also two key dimensions to foster the diffusion and use of digital technologies in the economy (Qi et al. 2019; Fuller et al. 2020).

2.2 Data and Measurement of 4.0 Technologies

We measure the generation of industry 4.0 technologies using European patent data. Patents have been criticized as measure for innovation as not all inventions can be patented and not all inventions are patented even if they could since firms can use other measures to protect their inventions. Furthermore not all inventions lead to marketable innovations. For our research question, we have additionally take into account that the use of patents might differ across 4.0 technologies and might lead to an underestimation of the corresponding development. In this regard, we expect that some enabling technologies might be more protected with copyrights due to the nature of the underpinning technologies (e.g. algorithms). Nevertheless, the availability of internationally comparable data over a long period of time and its richness in terms of technological information currently make patent data the best proxy available to study the development of 4.0 technologies as stressed by Martinelli et al. (2020).

In the following, we mainly focus on the development and diffusion of three main technological areas: 4.0 patents protect technologies related to the fourth industrial revolution. Among them, we further differentiate between enabling technology patents and core technology patents. The delineation is based on the CPC (Cooperative Patent Classification) classes of a given patent following the categorization of the European Patent Office (EPO) (Ménière et al. 2017). A patent is considered as a **4.0 patent** (also called 4IR patent) if at least one of its CPC classes belongs to one of the *320 4IR-CPC technology field ranges* classified by EPO patent examiners as being relevant for industry 4.0 inventions. Relevant in this context means in which CPC classes they would assign 4IR inventions (Ménière et al. 2017). These 4IR-CPC classes encompass the core technologies and enabling technologies explained in section 2.1, but also 4.0 technologies developed for final applications in various domains of the economy, including smart manufacturing, vehicles, infrastructure, home and personal devices for individuals. Noteworthy, core technologies, enabling technologies and smart application domains are not mutually exclusive and some CPC classes may belong to more than one of the three technology domains. Furthermore, patents may have more than one CPC class that belong to different technology domains, for instance to both enabling technologies and application domains, and they may also have been assigned to non-4IR-CPC classes. Thus, we also assess for each 4.0 patent its digital intensity. The digital intensity is measured by its **digital score** which is calculated as the share of 4IR-CPC classes in the total number of CPC classes. A higher value of the digital score reflects a stronger digital orientation of the patent.

Similarly, a patent is classified as an **enabling technology patent** if at least one of its CPC classes falls into the list of CPC classes identified by EPO as relevant for enabling technologies. Following the idea of the digital score, we also compute the **enabling score** which measures the relative share of CPC classes related to enabling technologies. Within the group of enabling technology patents, we follow the EPO classification and further distinguish between the following seven enabling technology fields:

- *data management* which covers technologies to create value from data like diagnostic and analytical systems for massive data, monitoring functions, planning and control systems;
- *user interfaces* which enable the display and input of information like virtual reality, augmented reality or speech recognition;
- *core AI* which enables machine understanding and covers for example technologies for machine learning, neural networks, statistical and rule-based systems and AI platforms;
- *Geo-positioning* which improves the determination of the position of objects; (v) *power supply* which enables intelligent power handling;
- *data security* which improves the security of data;
- *safety* which provides technological solutions for the safety of physical objects like intelligent safety systems for theft and failure prevention and
- *3D printing* which enable the realisation of physical or simulated 3D systems like 3D printers and scanners, automated 3D design or 3D user interfaces.

Conversely, we call a patent a *non-enabling technology patent* if it is a 4.0 patent but none of its CPC classes belong to CPC classes relevant for enabling technologies.

We apply the corresponding methodology to identify patents belonging to **core technologies** and their **core score**. Within the group of core technology patents, we further distinguish between (i) *hardware*, (ii) *software*, and (iii) *connectivity.* Doing so, we also study the set of patents that simultaneously belong to core and enabling technologies which are also known as **twin technologies** (Fuller et al. 2020; Qi et al. 2019)). The latter are characterized by a positive share of CPC classes in both enabling and core technology categories.

Our analysis relies on patent application data extracted from Patstat 2020 Spring Edition. The initial sample covers all patent applications filed at the EPO with earliest filing date between 1980 and 2017. As the last year of data is still not complete because of the publication lag in patent data, we only consider patent applications until 2016 and leave out the final year 2017. Overall, this makes up a sample of 3,459,374 unique patent applications at the EPO during this time period, representing 3,228,343 unique patent families. Our analyses are based on the patent family level. For studying the evolution of 4.0 patenting over time, we only select those patents for our analysis in section 3 that have a positive share of CPC classes in 4.0 technologies. We end up with a final sample that is composed of 513,880 distinct patent families (560,677 unique patent applications) related to 4.0 technologies. This represents 15.9%. In the following we will refer to them as 4.0 patents.

Table 1 summarizes the amount of patents in each category. 64.5% of the patents are related to core technologies and 43.7% of the patents belong to enabling technologies.

These numbers show that the technological domains are not mutually exclusive but can overlap. At 37.6%, the lion's share of our sample consists of patents with an orientation only towards core technologies, followed by twin technology patents having components in core and enabling technologies at 27.0%. 16.8% of the patents are pure enabling technologies, while 18.5 % of the 4.0. patents have no orientation in core and enabling technologies and can be categorized as smart application patents only.

Table 1: Composition of the Sample

Type of 4.0 patent	Number	%	Definition
Digital patents	513,880	100.0	Share of digital CPC classes positive
Core technologies	331,691	64.5	Share of core CPC classes positive
Enabling technologies	224,644	43.7	Share of enabling CPC classes positive
Core technology only	193,163	37.6	Share of core CPC classes positive & share of enabling CPC classses zero
Enabling technology only	86,116	16.8	Share of enabling CPC classes positive & share of core CPC classses zero
Twin technologies	138,528	27.0	Share of core and enabling CPC classes positive
Smart applications only	96,073	18.7	Share of digital CPC classes positive but share of core and enabling CPC classes zero

Notes: Notes: All patents applied for at the European Patent Office during 1980 and 2016 with at least one CPC class falling into the range of CPC classes relevant for 4.0 technologies. A patent is defined as patent family. Source: PatStat Spring 2020 edition. Own calculation.

In the next sections, we compare the respective trends regarding the generation and diffusion of different 4.0 technologies. Doing so, we give a specific emphasis to the distinct trends occurring among enabling and core technological fields which represent the key technologies within the 4IR revolution.

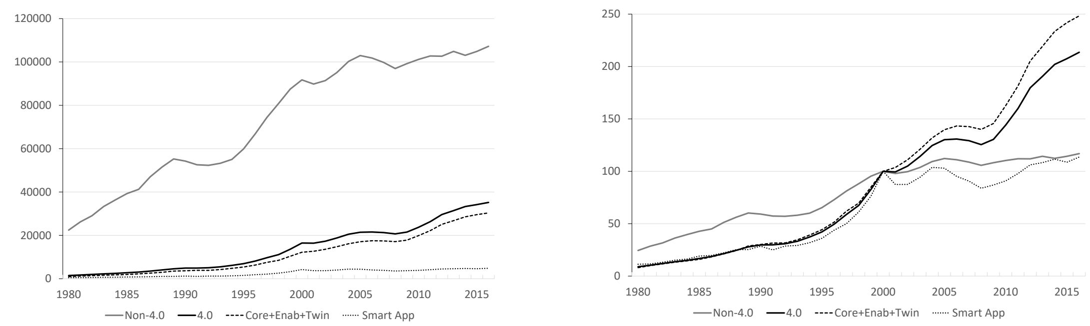
3 Development of 4.0 Patenting over Time

3.1 Development of the Number of 4.0 Patents over Time

Figure 2 shows the trend in 4.0 and non-4.0 patenting since 1980. The left graph depicts the absolute number of both types of patents, while the right graph provides the index time series of the number of patents with the baseline year 2000=100. The latter allows us to infer the growth in patenting over time. The left graph in Figure 2 shows an increasing trend in both 4.0 and non-4.0 patenting over the last three decades, despite a small slow down during the Great Recession in 2008. The right graph illustrates that the growth in patenting 4.0 technologies has been much larger over this period than for other technologies. Compared to the baseline year 2000, the annual number of 4.0 patents in 2016 has more than doubled from 16,481 to 35,199, which corresponds to a growth rate of 114%. Non-4.0 patents have only grown by roughly 17% in the same period. The dynamics in the patenting of 4.0 technologies is mainly due to core, enabling and twin technologies. Compared to the baseline year 2000, the annual number of patent applications for these technologies has grown by 148% in 2016. Much of this growth has taken place in the current decade. From 2010 to 2016, core, enabling and twin technologies have grown by 53%. In contrast, the number of patents related only to smart applications has remained comparatively small, with a rather modest increase of 14% since 2000.

Figure 2: Number and Growth of 4.0 and Non-4.0 Patents over Time

(a) Absolute number (b) Growth (baseline year 2000)



Notes: Patents are measured based on patent applications. The term patent denotes a patent family.
Source: PATSTAT 2020 - spring edition, own calculation.

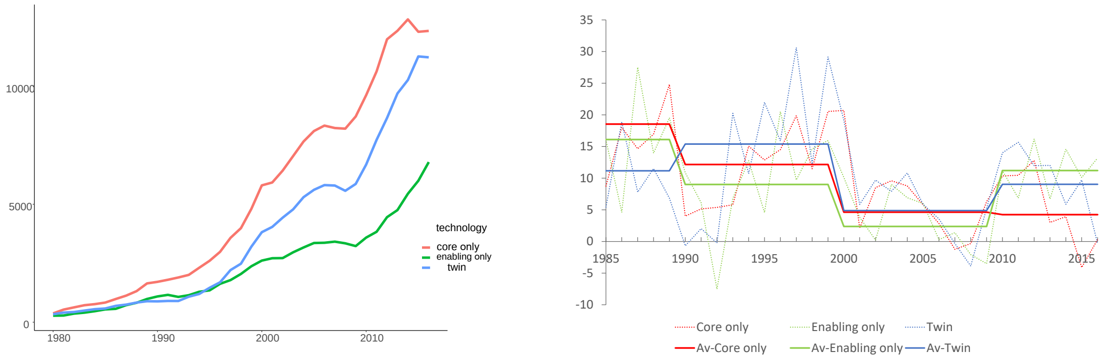
Figure 3 depicts the development of patenting activity over time by 4.0 technological areas. From here onwards, we focus on 4.0 patents in core and enabling technology areas and remove 4.0 patents purely related to smart applications as they cover different types of technologies. To avoid double counting, we split the patents into three mutually exclusive groups: core technology patents only, enabling technology patents only and twin technology patents. Figure 3 (a) shows the absolute number of patents by 4.0 technology area over time, while Figure 3 (b) depicts the corresponding annual growth rates (dotted lines) and average annual growth rates (solid lines) for the four periods 1985-1989, 1990-1999, 2000-2009 and 2010-2016. For all three technological areas we see a rather steady increase in the annual number of patents applied for, except for the beginning of the nineties and the period around the financial crisis 2008. The annual growth rate for each 4.0 technological area is rather volatile, but it follows a similar time pattern for all three groups. While patenting of twin technologies had the highest average annual growth rate in the decades 1990-1999 and 2000-2009, and patenting of enabling technologies had the lowest growth rate, this pattern has changed for the most recent period 2010-2016, where the growth of enabling technologies has overtaken that of the twin technologies. Especially since 2012, growth in enabling technologies seem to have decoupled from growth in core technology only and twin technologies, which have since shown a declining trend in annual growth.

3.2 Development of Digital Intensity in 4.0 Patenting over Time

Technical features of an invention are captured by the CPC classes on the patent document. An invention might consist of both, features related to 4.0 technologies and to non-4.0 technologies. In the previous subsection, we have equally weighted each patent,

Figure 3: Number and Growth of 4.0 Patents over Time, by Technological Area

(a) Absolute number (b) Annual growth rate



Notes: In Figure (b) dotted lines depict the annual growth rates while solid lines depict the average annual growth rate for the period 1985-1989, 1990-1999, 2000-2009, and 2010-2016, respectively. The average annual growth rate has been calculated as geometric growth.
Source: PATSTAT 2020 - spring edition, own calculation.

independently of its 4.0 digital intensity. Each patent that has at least one CPC class belonging to a CPC class relevant for industry 4.0, has been counted as a 4.0 patent. However, the digital intensity varies substantially across 4.0 patents as Table 2 reveals. On average, a little more than half of the CPC classes on a 4.0 patent document belong to 4.0 relevant CPC classes (53.3%), while 46.7% of the CPC classes describe non-4.0 related technical features.[1] The percentiles in Table 2 furthermore show that the digital intensity is rather equally distributed across all 4.0 patents. On average, 39.4% of the CPC classes on a 4.0 patent describe core technology features, while 1 out of 4 CPC classes (24.6%) describe a feature associated with enabling technologies. Thus, 4.0 patents show a stronger core technology content then enabling technology content. Most frequent are CPC classes related to hardware technology (22.4%), followed by technical feature related to connectivity (12.8%) and analytics (12.2%).

Table 3 provides the digital intensity separately for patents classified as core technology only, enabling technology only and twin technology. For all three groups of patents the average digital intensity is similar, varying between 52% for core only and 56% for enabling technology only. However, the Table also reveals substantial heterogeneity by technology score. On the one hand, comparing twin technology patents with core technology patents only, we observe a much stronger orientation of twin technology patents towards connectivity and to a lesser extent also to hardware, while core technology only patents score higher on software. On the other hand, comparing twin technology patents with enabling technology patents only, twin technology patents mainly score higher on technical features related to security, while patents counted as enabling technology more frequently protect technical feature in the area of analytics, GPS and AI. Overall, the AI score is still rather small. On average, 3% of the CPC classes of enabling technology only patents are relevant for AI.

[1]Due to the overlap between CPC classes in enabling and core categories, we standardize the respective share in both scores in order to be able to interpret the results in percentages.

Table 2: Digital Intensity of 4.0 Patents

Share of ...	Mean	St. Dev.	Min	p25	p75	Max
Non-digital CPC	0.467	0.310	0.000	0.200	0.750	0.994
Digital CPC	0.533	0.310	0.006	0.250	0.800	1.000
Core CPC	0.394	0.344	0	0.1	0.7	1
Hardware CPC	0.224	0.310	0	0	0.4	1
Software CPC	0.071	0.208	0	0	0	1
Connectivity CPC	0.128	0.238	0	0	0.2	1
Enabling CPC	0.246	0.320	0.000	0.000	0.417	1.000
Analytics CPC	0.122	0.250	0.000	0.000	0.125	1.000
Security CPC	0.060	0.185	0.000	0.000	0.000	1.000
AI CPC	0.008	0.068	0.000	0.000	0.000	1.000
GPS CPC	0.036	0.151	0.000	0.000	0.000	1.000
Power CPC	0.007	0.049	0.000	0.000	0.000	1.000
3D CPC	0.001	0.013	0.000	0.000	0.000	1.000
Interface CPC	0.010	0.066	0.000	0.000	0.000	1.000

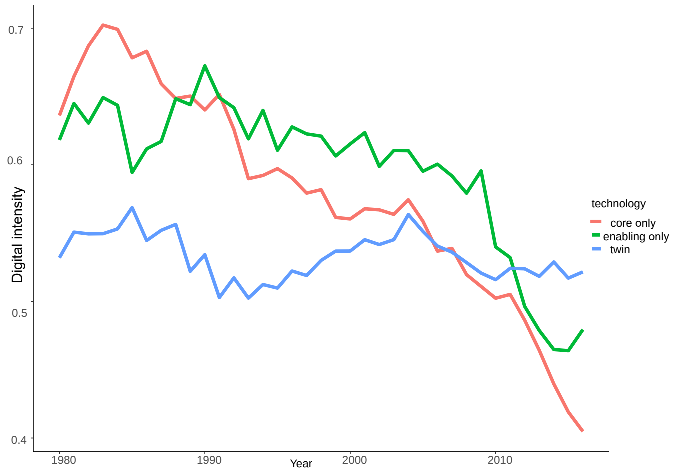
Notes: Population: 471,807 4.0 patents with a positive score in core or enabling technologies. Patents related to smart applications only are excluded.
Source: PatStat Spring 2020 edition. Own calculation.

Table 3: Digital Intensity of 4.0 Patents, by Technological Area

	Core Tech only	Enabling Tech only	Twin Tech
Share of ...	Mean	Mean	Mean
Non-digital CPC	0.48	0.44	0.47
Digital CPC	0.52	0.56	0.53
Core CPC	0.52	–	0.46
Hardware CPC	0.31	–	0.54
Software CPC	0.12	–	0.05
Connectivity CPC	0.11	–	0.42
Enabling CPC	–	0.52	0.42
Analytics CPC	–	0.28	0.20
Security CPC	–	0.03	0.16
AI CPC	–	0.03	0.01
GPS CPC	–	0.15	0.02
Power CPC	–	0.001	0.01
3D CPC	–	0.003	0.001
Interface CPC	–	0.02	0.02

Notes: Population: 193,163 core technology patents only, 86,116 enabling technology patents only and 138,528 twin technology patents.
Source: PatStat Spring 2020 edition. Own calculation.

Figure 4: Digital Intensity of 4.0 Patents over Time, by Technological Area



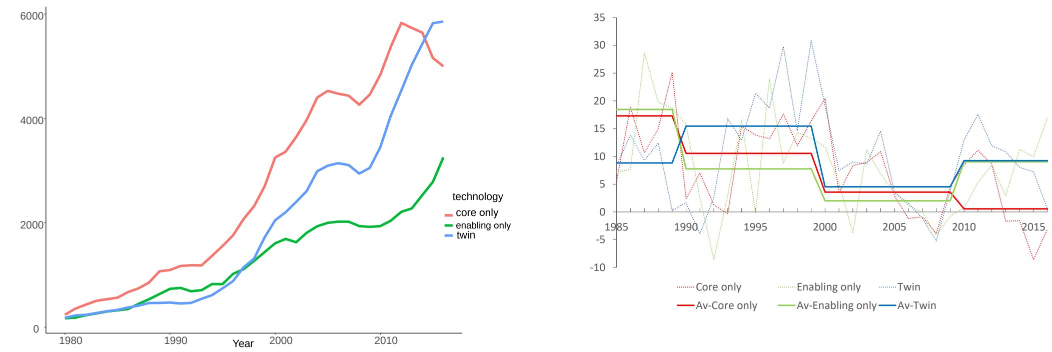
Notes: Depicted is the mean digital intensity by year and technological area.
Source: PATSTAT 2020 - spring edition, own calculation.

Finally, Figure 4 shows how the digital intensity of 4.0 patents has evolved over time. Two developments stand out: First, the digital intensity of twin technology patents has been fairly stable over the entire period since 1980, centered at around 53% over time. In contrast, the digital intensity of patents relating only to either core or enabling technologies shows steadily decreasing trend over time. This decline has become especially strong after 2005 for core technologies and after 2008 for enabling technologies. Our findings imply that while the absolute number of core and enabling technology patents has sharply increased over time, the average digital content of these patents has become smaller over time. Most likely this reflects the diffusion of 4.0 technologies in the economy. In recent years, these technologies have been more widely used for inventions which combine 4.0 and non-4.0 technological features. We will dig deeper into the diffusion of 4.0 technologies using citation analysis in section 4.

Given the strong variation in the digital content of 4.0 patents and its varying evolution over time by technological areas, we propose to use fractional counts in addition to the absolute number of patents. Fractional counts imply that we weight each 4.0 patent by its digital intensity. Similar to Figure 3, Figure 5 (a) shows the number of 4.0 patents over time using fractional counts and (b) the corresponding growth rates by technological area. Given that the digital intensity of twin technologies has been fairly stable over time, we see a similar pattern for twin technologies as in Figure 3, albeit at a lower level in terms of the number of patents. In contrast, differences emerge for core and enabling technologies. The sharp decline in digital intensity for core technologies has more than offset the increase in the absolute number of patents, so that we see a decline in core technologies after 2012. For the entire period 2010-2016 the average growth rate is close to zero. For enabling technologies, the increase in the number of patents was stronger than the decline in digital intensity, so we still see an increasing trend for enabling technologies in the period 2010-2016.

Figure 5: Number and Growth (Fractional Counts) of 4.0 Patents over Time, by Technological Area

(a) Fractional count (b) Annual growth rate

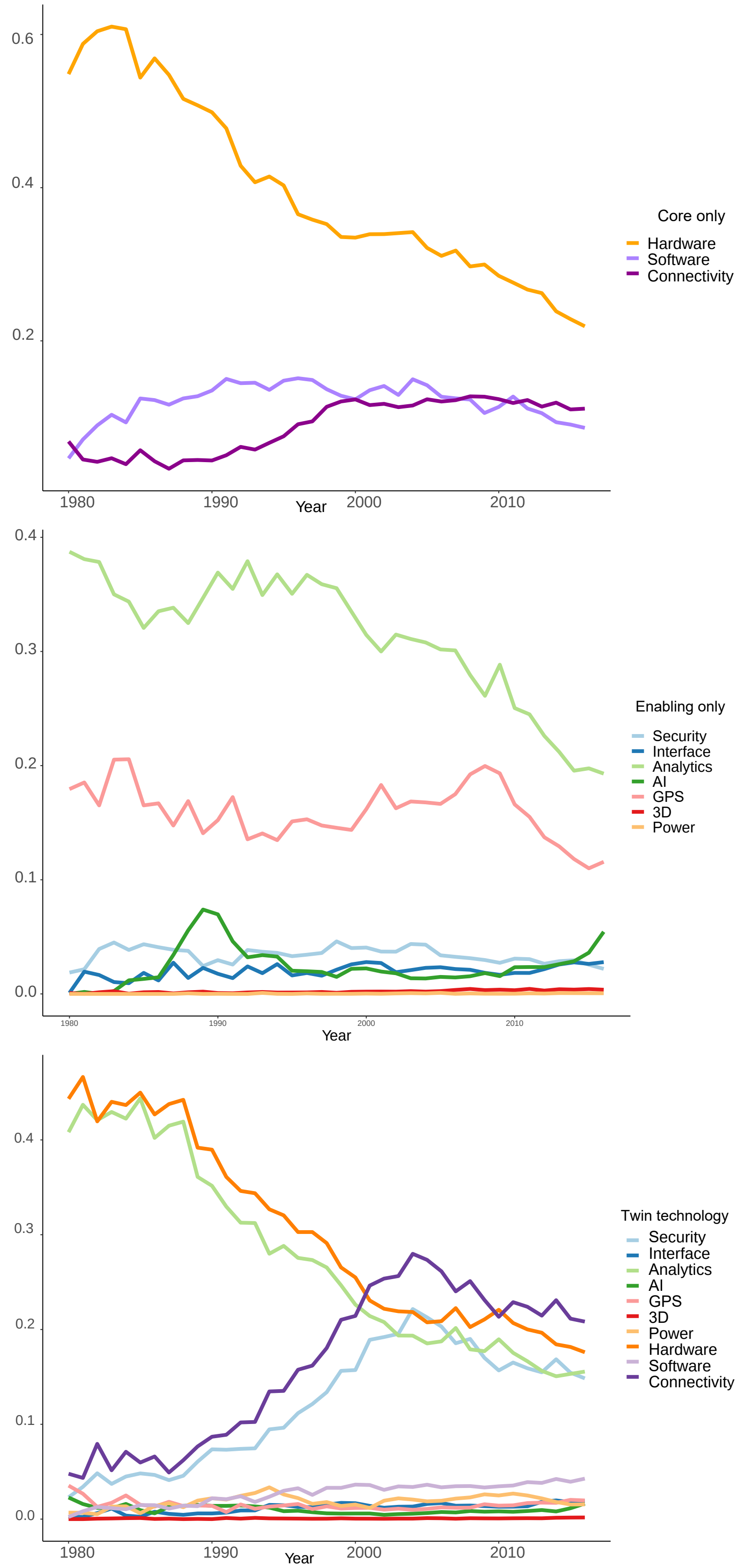


Notes: Dotted lines depict the annual growth rates while solid lines depict the average annual growth rate for the period 1985-1989, 1990-1999, 2000-2009, and 2010-2016.
Source: PATSTAT 2020 - spring edition, own calculation.

We further break down the analysis by looking at the evolution of different technologies within technological areas in order to better understand whether specific technologies are the main drivers of the development we see for core, enabling and twin technology patents. Figure 6 shows for every technology its average intensity. The hardware intensity is calculated as the average share of CPC classes related to hardware technology over all CPC classes in a given year. The declining trend for core technologies in digital intensity shown in Figure 6 is to a very large extent driven by the declining importance of CPC classes related to hardware features. This share has fallen from about 60% in mid of the eighties to roughly 20% in 2016. In contrast, technical features related to connectivity increased strongly until 2000 and have remained stable at around 12% thereafter. The software intensity has been fairly stable over the entire period at around 12%. The falling trend in digital intensity of purely enabling technology patents is mainly driven by a decline in data analytics and since 2008 also in GPS technology which represent the lion share of enabling technological developments. In contrast, technologies related to security, interface, 3D and power haven been fairly stable over time. Outstanding is the development of AI within purely enabling technology patents. Its AI intensity has more than tripled since 2004. In 2016, on average 5.4% of the CPC classes on enabling technology patents describe AI technologies compared to 1,35% in 2004. Finally, looking at patents for twin technology, whose overall digital intensity has been fairly stable over time, Figure 6 shows that the mix of technologies has substantially changed over time. While the intensity of hardware and data analytics has steadily fallen, feature related to connectivity and security have significantly increased and in case of connectivity even surpassed hardware and analytics since the beginning of the 2000. Since the mid of 2000, however, we see a slight decrease for connectivity and security intensity as well. The AI intensity of twin technologies has also been very stable over time, it is only in the last two years that we see a substantial increase from 0.8% to 1.6%.

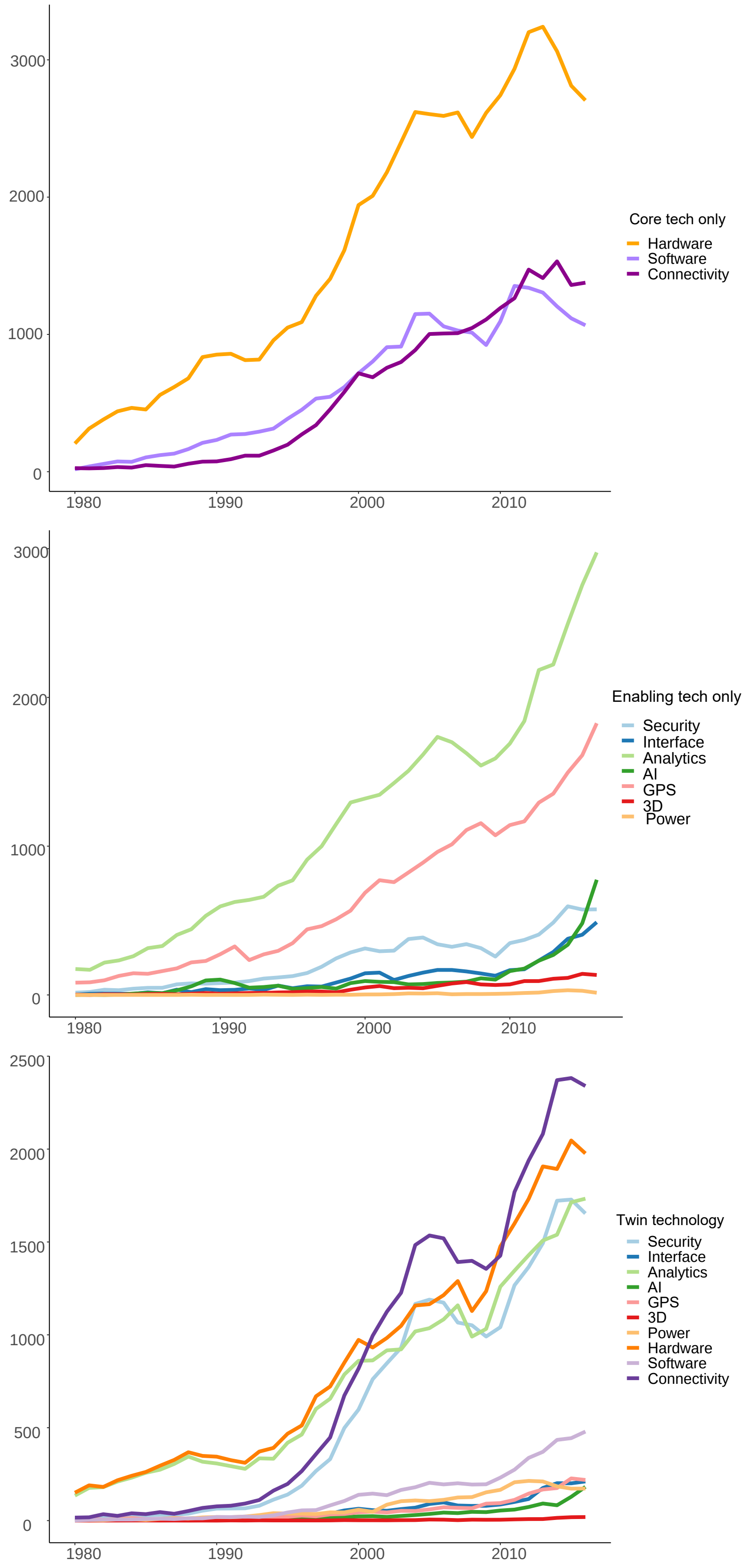
Accounting for the digital intensity of different technologies, Figure 7 shows the number

Figure 6: Technology Intensity over Time, by Technological Areas



Notes: Depicted is the annual mean technology intensity for each technology. For example, the hardware intensity is the average share of CPC classes related to hardware technology features over all CPC classes. Source: PATSTAT 2020 - spring edition, own calculation.

Figure 7: Number of 4.0 Patents (Fractional Counts) over Time, by Technology within Technological Areas

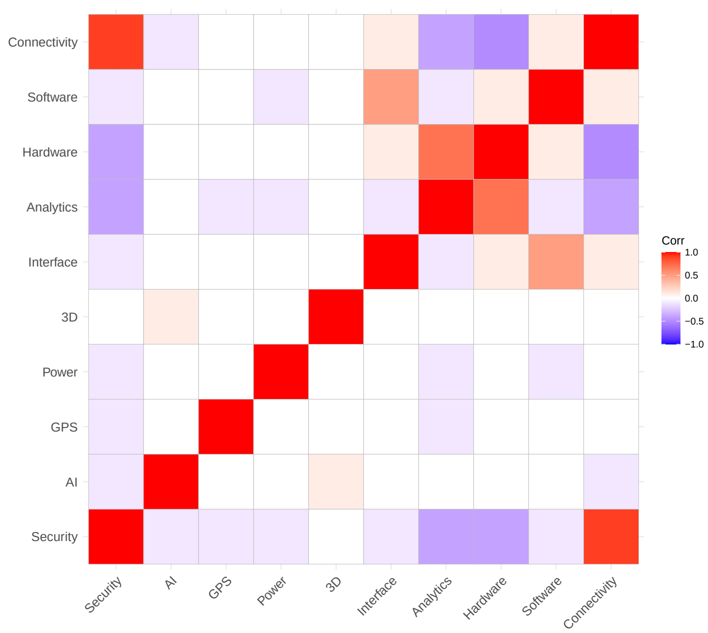


Source: PATSTAT 2020 - spring edition, own calculation.

of patents using fractional counts.

Finally, for twin technology patents, the peculiarities of 4.0 technological recombinations are analyzed by means of a correlation analysis. Figure 8 depicts the correlation matrix of technology intensities at the patent level. On the colour scale, red indicates a positive and blue a negative correlation between technologies. The figure shows that different types of core technologies are more strongly correlated with each other than enabling technologies. In more detail, connectivity technologies are strongly positively correlated with security technologies and to a lesser extent also with interface technologies for data transmission and software technologies. Within twin technologies, however, connectivity is less frequently combined with hardware and data analytics. Security technologies stand out as they are disproportionately combined with all other technologies except connectivity. The correlation pattern for software and interface technology is very similar. They are positively correlated with each other and show a weak positive correlation with hardware and connectivity but a negative one with data analytical purposes and security applications. 3D, GPS and AI are technologies that are rather independent of other technologies. 3D is only weakly combined with AI, and AI additionally shows a weak negative correlation with security and connectivity.

Figure 8: Correlation of 4.0 Technology Intensities for Twin Technology Patents

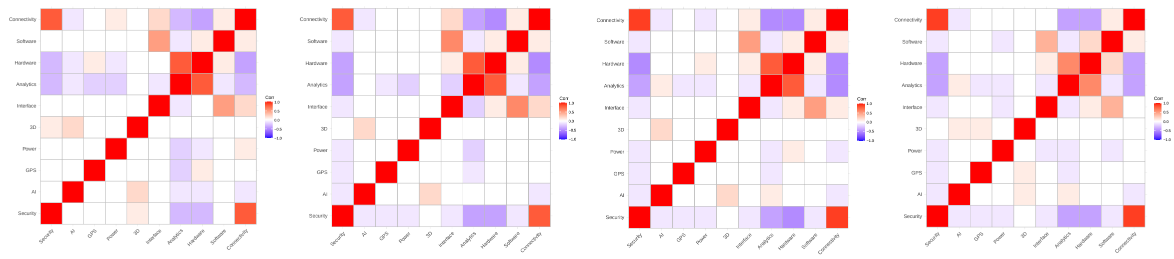


Source: PATSTAT 2020 - spring edition, own calculation.

Figure 9 shows that this technology recombination pattern of twin technology patents has been fairly stable over the four decades, but with some interesting exceptions. For example, the positive correlation between connectivity and security has become stronger in the last two periods, while the positive correlation between connectivity and interface technologies has become weaker, even reaching zero in the most recent period. With regard to hardware technology, it can be observed that it has been combined more frequently

Figure 9: Correlation of 4.0 Technology Intensities for Twin Technology Patents, by Time Period

(a) 1980-1989 (b) 1990-1999 (c) 2000-2009 (d) 2010-2016



Source: PATSTAT 2020 - spring edition, own calculation.

with software infrastructure like cloud computing, mobile operating systems or blockchain technologies and has become more neutral towards enabling technologies. Also interesting is the correlation between data analytics and AI, which has changed from negative in the first decade, to zero in the second decade and to positive in the last two periods.

3.3 Development of 4.0 Patenting across Industries over Time

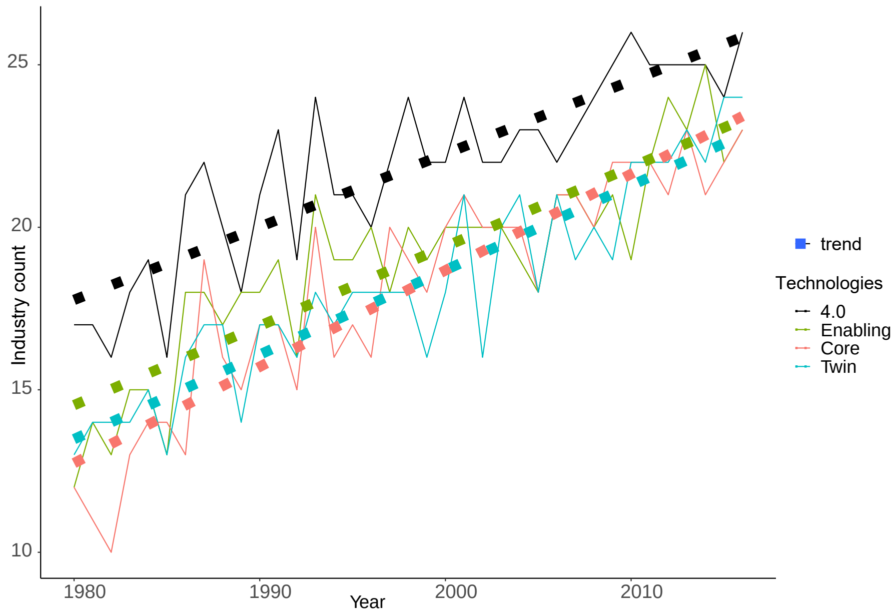
This section studies the development of 4.0 patenting over time at a sectoral level. In order to perform a sector level analysis, we have to classify each patent to an industry.[2] We use the IPC classes of each patent and the 2-digit NACE Rev.2 concordance table from Patstat, developed by Van Looy et al. (2015). The concordance table assigns 640 IPC codes to 26 industries.

Figure 10 shows the evolution of the number of industries that have a least one 4.0 patent application in a given year overall and by technological area. Two findings stand out. First, the number of industries fluctuates from year to year, but shows a clear upward trend. Starting with 17 industries in 1980, over 22 industries in 2000, we observe that all 26 Nace 2-digit industries have patented 4.0 technologies in 2016. Thus, the diffusion of the generation of 4.0 technologies has strongly increased over time. Second, the rate of penetration across industries has increased faster among purely enabling technologies than twin and core technologies only. However, the trends for the three technological areas seem converging towards the overall trend of 4.0 technologies over the most recent period since 2011, leading to 24 industries patenting twin technologies and 23 industries active in inventing core and enabling technologies in 2016.

Of course, not each industries contributes the same amount to 4.0 patenting. We therefore calculate the proportion each industry contributes to the number of 4.0 patents (fractional counts) in a given year to see whether the importance of different industries has changed over time. The 26 2-digit Nace industries consist of 23 manufacturing industries which we summarize based on their average R&D intensity into high, medium-high,

[2] We use the term industry and sector interchangeably.

Figure 10: Number of Industries Patenting in 4.0 Technologies



Notes: The dotted lines approximate the underpinning trends with a linear regression.
Source: PATSTAT 2020 - spring edition, own calculation.

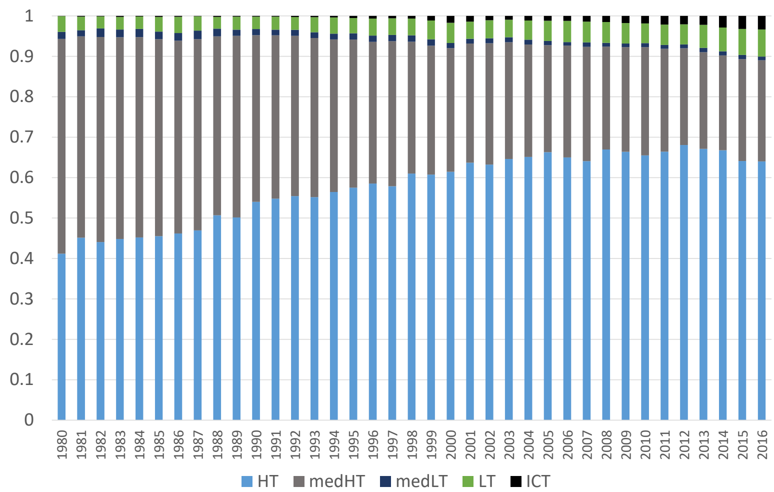
medium-low and low technology using the Eurostat classification scheme.[3] The only Nace 2-digit industry in services produced by the concordance table is Nace 62. The latter captures computer programming, consultancy and related activities and we shortly call it ICT in the following.[4] While the absolute numbers show a substantial increase in 4.0 patenting for all five industries over time, this increase has not been proportional. This has led to a clear shift in the contribution of different industries to the overall number of 4.0 patents as Figure 11 shows. The share of 4.0 patents invented by the high tech industries has steadily increased from about 45% in the early years to about 65%. This has mainly come at the cost of medium high technology whose share has almost halved from about 50% in 1981 to 25% in 2016. The share of ICT is still comparably small but steadily increasing. Between 2000 and 2016 it has doubled from about 1.7% to 3.4% in line with the nature of core technologies (e.g. software).

Finally, Figure 12 splits the industry contribution by technological area. The industry composition and its evolution over time is different in all three areas. In core and twin technologies high tech and low tech together have the lion share of the patenting activity with more than 90%. For both technological areas we observe an increasing concentration of high tech industries. Starting from an already high level of about 60%, high tech industries, mainly the computer and electronics industry, have further increased their share to about 80% in later periods. In contrast, the proportion of high tech was rather small in inventing

[3]High tech (HT) consists of Nace2 2-digit 21 (manufacture of basic pharmaceutical products and pharmaceutical preparations) and 26 (manufacture of computer, electronic and optical products), medium high tech (medHT) of 20 and 27-30, medium low tech (medLT) of 19 and 22-25 and low tech (LT) of 10-18, 31 and 32.

[4]In the following we leave out Nace 2-digit 42 and 43 in construction.

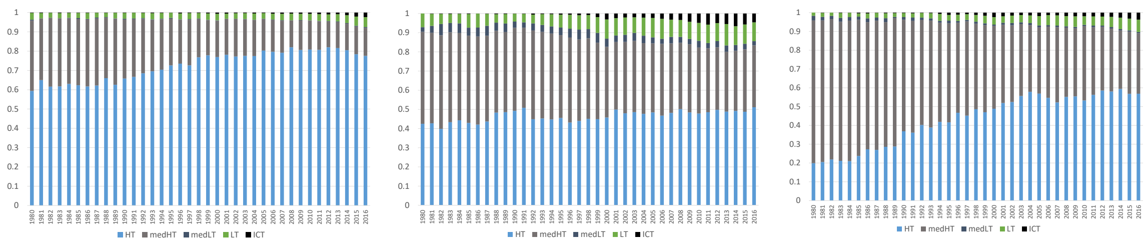
Figure 11: Contribution to 4.0 Patenting (Fractional Counts) by Industry over Time



Source: PATSTAT 2020 - spring edition, own calculation.

Figure 12: Contribution to 4.0 Patenting (Fractional Counts) by Industry and Technological Area

(a) Core tech only (b) Enabling tech only (c) Twin tech



Source: PATSTAT 2020 - spring edition, own calculation.

twin technologies in the first decade at about 20%, but is has almost tripled to about 60% in recent years. In both areas medium high tech industries sharply lost importance to high tech, low tech and ICT industries. In contrast, the evolution of industrial composition looks much more stable for enabling technologies, where high tech industries gain only slightly and medium high tech industries lose only moderately. The ICT and low technology industries mainly contribute to inventing enabling technologies.

4 Diffusion of 4.0 Patents over Time

Stoneman and Battisti (2010) define technological diffusion as “the process by which the market for a new technology changes over time and from which production and usage patterns of new products and production processes result”. The cumulative adoption of innovation over time follows an S-shaped curve, forming a technology adoption life cycle (Rogers 2010). This general pattern has been established in the literature even if specific industry or technology studies show different time-spans to reach this S-shaped curve (see

Stoneman and Battisti 2010, for an overview). Numerous factors can influence the time span needed to adopt new technologies, among those can be regulations, standards, technological trajectories, technological opportunities, market competition, network externalities, skills, and technological and organizational complementarities (Silverberg 1991). Besides these macro and meso determinants, specific technological are likely to influence the scope and pace of technological diffusion across sectors. Some technologies are more pervasive than others and are likely to encompass a larger set of applications across sectors. These technologies, known as General-Purpose Technologies (GPT), tend to follow two main diffusion phases: first a phase of slow diffusion, followed by a phase of strong acceleration (Bresnahan 2010). Within this process, market conditions have been brought to the fore as a crucial argument to explain the diffusion across sectors (Griliches 1957). Following previous studies for information and communication technologies (Hall and Trajtenberg 2004), we investigate the pattern of diffusion of industry 4.0 technologies across sectors and to what extent their use alters inter-sectoral technological linkages.

This section studies the diffusion of 4.0 technologies from the supply-side using citation analyses. Forward citations are the most common proxy to measure the diffusion of technologies. They count the number of citations that a given patent receives in future patent documents, and in our case they therefore measure technological developments that build upon 4.0 technologies. Forward citations are furthermore a proxy for the importance and economic value of a patent (Trajtenberg, 1990; Hall, et al., 2005; Harhoff et al., 2003). This section provides an overview of the forward citation pattern of 4.0 patents compared to non-4.0 patents. By looking at how often they are cited and by whom, we get first-hand information about interesting differences in the diffusion process depending on the 4.0 technology subgroup, the country of origin of the citations, the NACE sector and how long it takes for this knowledge to spread. Subsection 4.4 then looks at the top most cited 4.0 patents as indicator for breakthrough technologies and their diffusion process as well as at the top 4.0 filing companies over time and lists the individual top 30 most cited 4.0 patents and what technology they relate to.

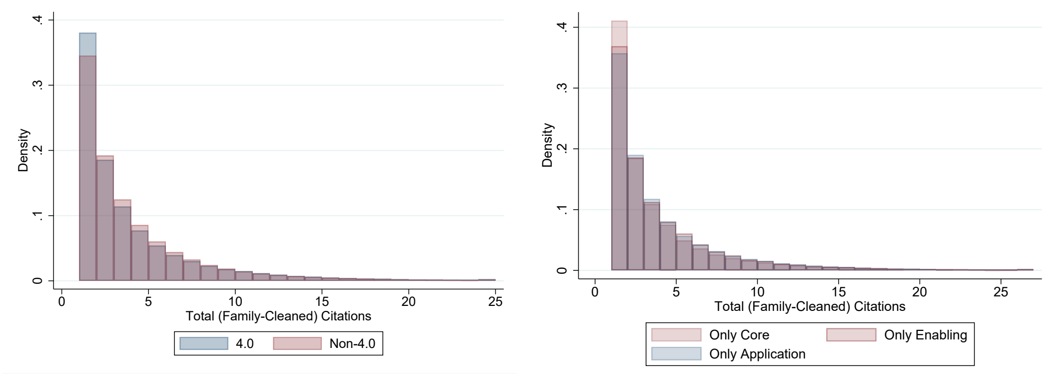
As starting point, we use the same sample of 513,880 distinct patent families earlier identified as 4.0 patents and collect all forward citations for these patents. Since older patents have had more time to receive citations which lead to a bias when comparing it with more recent patents, we follow if necessary the common approach and use a five-year time window after filing date, in which citations are counted. In this case as a result, only patents filed until the end of 2011 (around 2.73 million applications, of which 385,826 are 4.0 patents) are considered in order to account for the full five-year window. Furthermore, when differentiating between citations from 4.0 and non-4.0 patents, we can only use the citations made by other EP patents, as only these can be classified into 4.0 technologies. In each graph or table, the notes below indicate which data have been considered. A patent or a citing patent is identified as being 4.0 if at least one of its CPC technology classes falls within the EPO-defined list of CPC codes relevant to Industry 4.0.

4.1 Overall Diffusion: Number of Citations

We measure the overall diffusion of a patent using its total number of citations. 59.4% (52.4%) of all EP patents filed between 1980-2016 (1980-2011) have received at least one citations to date (spring 2020). On average, 4.0 patents of that period have received less citations: 1.3 compared to 1.7 citations for non-4.0 patents. The left graph of Figure 13 shows the distribution of citations of all EP patents conditional on being cited. In addition, it shows the difference in the distribution of 4.0 and non-4.0 patents. For example, the blue part of the bar for one citation indicates that the density for 4.0 is higher than that for non-4.0 patents by that amount. Overall, we can conclude that the citation patterns are pretty similar for 4.0 vs non-4.0 patents, although the distribution is slightly more skewed to the right for non-4.0 patents. This means that also conditional on being cited, non-4.0 patents are cited slightly more than 4.0 patents. The slower rate of diffusion is in line with the argument of a slow diffusion of GPTs in the first phase (Bresnahan 2010). Even when broken down into the three main 4.0 categories - core, enabling and smart applications - as can be seen on the right graph of Figure 13, citation patters do not deviate much. Only slight deviation can be observed indicating that core technology patents receive comparatively less citations while application patents are slightly cited more frequently.

Figure 13: Distribution of All Citations

(a) 4.0 vs Non-4.0 (b) 4.0 by Technological Area

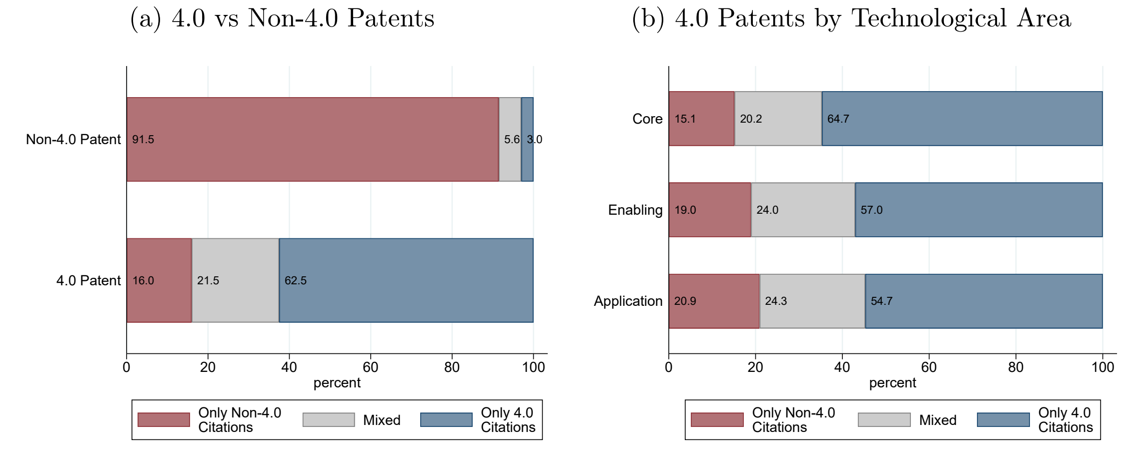


Notes: Data used all EP applications filed between 1980-2016 and counts all family-cleaned citations. No restrictions are made wrt. a citation time-window. For the figure to the right, only patents with mutually exclusive categories were used. Outliers (top 1%) were dropped.
Source: PATSTAT 2020 - spring edition, own calculation.

4.2 Diffusion Towards 4.0 and Non-4.0 Technologies

In this subsection, we investigate to which extent 4.0 patents follow the same pattern of diffusion across other 4.0 technologies and non-4.0 technologies. We therefore classify all citations into 4.0 (non-4.0) citations in a first step if the citing EP patent is a 4.0 (non-4.0) patent, and then classify each patent in a second step according to whether it receives only 4.0 citations, only non-4.0 citations or mixed citations, indicating that these inventions have

Figure 14: Diffusion Towards 4.0 and Non-4.0 Technologies Using Citation Segregation



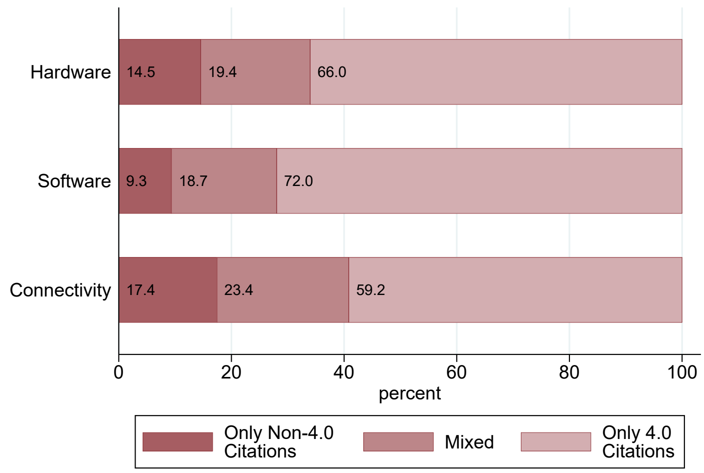
Notes: Data used are all EP applications filed between 1980-2016 that received citations by other EP patents. No restrictions have been made regarding a citation time-window, but only citations by other EP applications can be accounted for.
Source: PATSTAT 2020 - spring edition, own calculation.

diffused to both 4.0 and non-4.0 technological areas. The left graph of Figure 14 shows that 62.5% of 4.0 patents receive all their citations only from other 4.0 patents. This is a rather substantial share, considering that only a comparatively low share (16.2%) of EP patents protect 4.0 technology inventions. Conversely, 91.5% of non-4.0 patents are cited only by other non-4.0 patents. This indicates a rather strong segregation. From another perspective, only 8.6% (5.6+3) of non-4.0 patents serve as a technological foundation for 4.0 inventions. For 4.0 patents, the corresponding share is significantly higher; 37.5% of 4.0 patents serve as a technological foundation for non-4.0 inventions, among them are 16% of 4.0 patents that have only diffused to non-4.0 technologies and 21.5% that have diffused to both types of technologies. Overall, these findings are in line with the expectation that 4.0 patents have a broader industrial applicability. The right graph of Figure 14 shows the origin of citations received by 4.0 patents split by technological area. Not only do core technology patents receive slightly fewer citations, they also diffuse more towards other 4.0 technologies (64.7%) than enabling technology patents (57%) and smart application patents (54.7%). Or to put it differently, 45.3% of smart application patents and 43% of enabling technology patents diffuse toward non-4.0 technologies and serve as a technological foundation for non-4.0 inventions.

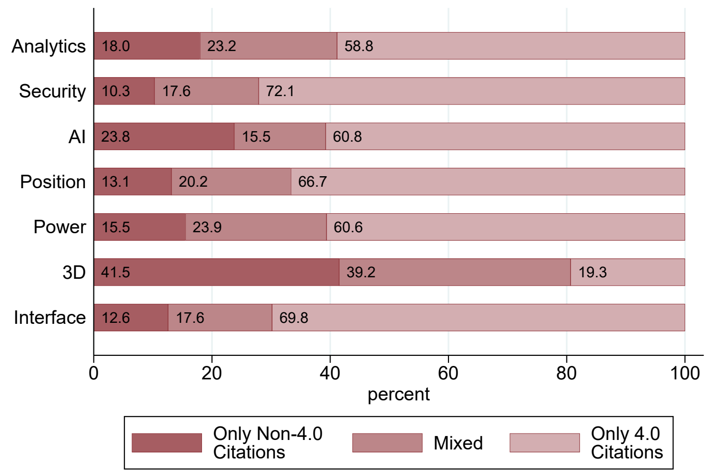
Some 4.0 technologies diffuse into non-4.0 inventions more than others. Figure 15 shows the diffusion pattern of cited 4.0 by splitting them into their respective 16 technology subgroups. Here it becomes clear that among core technologies software infrastructure patents diffuse much less to non-4.0 technologies (28%) than hardware or connectivity patents. For example, more than 40% of patents related to connectivity have served as technological foundation for non 4.0 inventions. Out of the enabling technologies, it is 3D-support systems that diffuses the most to Non-4.0 patents: 80.7% of 3D patents diffuse into non-4.0 technologies, these are almost evenly split between those that diffuse exclusively into non-4.0 technologies and those that are picked up by both 4.0 and non-4.0 technology

Figure 15: Diffusion Towards 4.0 and Non-4.0 Technologies Using Citation Segregation, by 4.0 Technology Subgroup

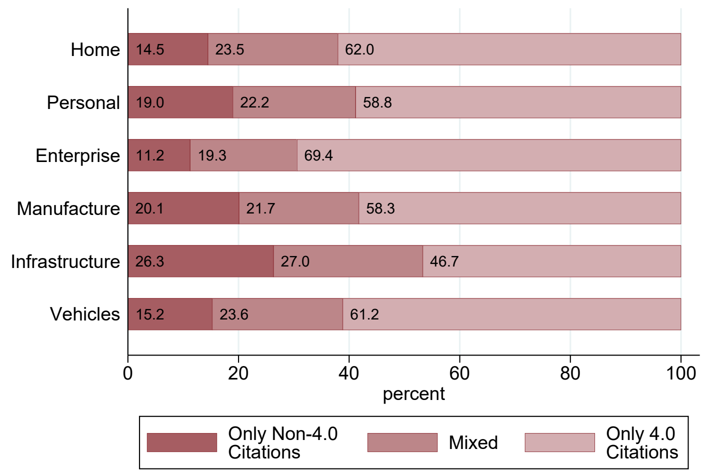
(a) Core Technologies



(b) Enabling Technologies



(c) Application Domains



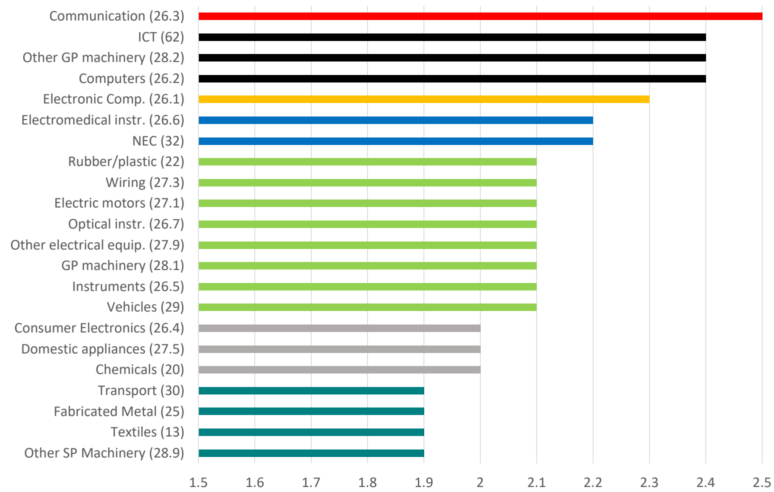
Notes: Data used all EP 4.0 patent applications filed between 1980-2016 that received citations by other EP patents. Even though the 16 different technology subgroups of the 4.0 cartography are not mutually exclusive, we only used those cited EP 4.0 patents for this graph that can be identified in a mutually exclusive manner. No restrictions have been made regarding a citation time-window.
Source: PATSTAT 2020 - spring edition, own calculation.

inventions. When including 'mixed' citations, then data analytics (41.2%), intelligent power handling (39.4%) and AI (39.2%) also play a comparatively larger role in diffusion to non-4.0 technologies.

4.3 Diffusion of 4.0 patents by NACE Industry

Industries differ greatly in their role in the value-added process, some are primarily producers of components and inputs for other industries while other industries produce primarily for private end consumers. As a result, the diffusion of new 4.0 technologies can also vary greatly between industries. Figure 16 shows the mean number of citations for all industries with 1000 or more 4.0 patents in the period 1980 to 2011.[5] The communication industry (Nace 26.3) has by far the highest number of 4.0 patents. Almost one third of all 4.0 patents belong to this industry, followed by the computer equipment industry with another 17.4%. Not only do these industries produce the largest amount of 4.0 patents, they also get cited the most on average. 4.0 patents from communication equipment producers receive on average the highest number of citation (2.5 per patent), closely followed by the ICT, other general purpose machinery and computer equipment producers (all 2.4) and electronic component producers (2.3). Comparatively less diffused are 4.0 patents from industries like chemicals, transport, fabricated metal, textiles or other special purpose machinery.

Figure 16: Mean Number of Citations Received by 4.0 Patents, by NACE Industry

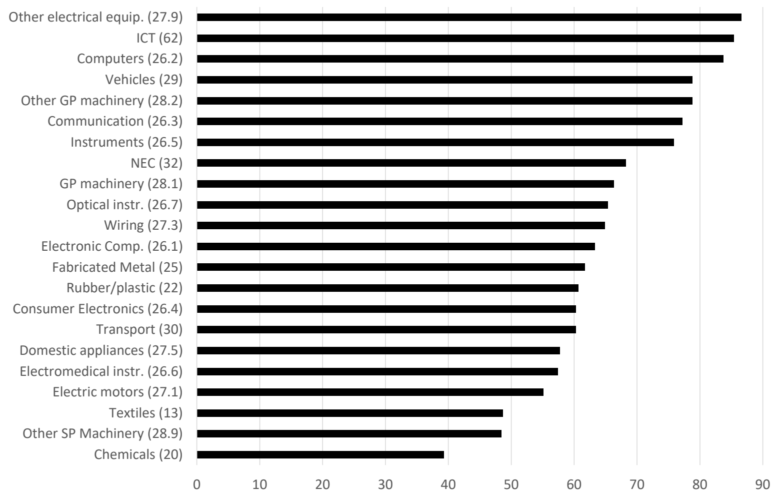


Notes: Mean number of forward citations of 4.0 patents in industries with 1000 or more 4.0 patents filed between 1980-2011 that received citations by other EP patents. Industry classification based on 2-/3-digit NACE Rev.2 concordance table from Patstat (Van Looy et al. 2015). For a full industry list, see Table 8. Source: PATSTAT 2020 - spring edition, own calculation.

Figure 17 shows for the same set of industries to what extent these 4.0 patents diffuse to other 4.0 technologies. Especially in industries related to the manufacturing of electrical equipment, ICT and computer equipment, 4.0 patents primarily serve as a technological

[5]We use this threshold in order to prevent that outliers in industries with a very small number of 4.0 patents have strong impact on the results. Table 8 in the Appendix shows the total number of 4.0 patents and different citation indicators for all industries using the 2-/3-digit NACE Rev.2 concordance table from Patstat (Van Looy et al. 2015).

Figure 17: Share of 4.0 Citations of 4.0 Patents, by NACE Industry

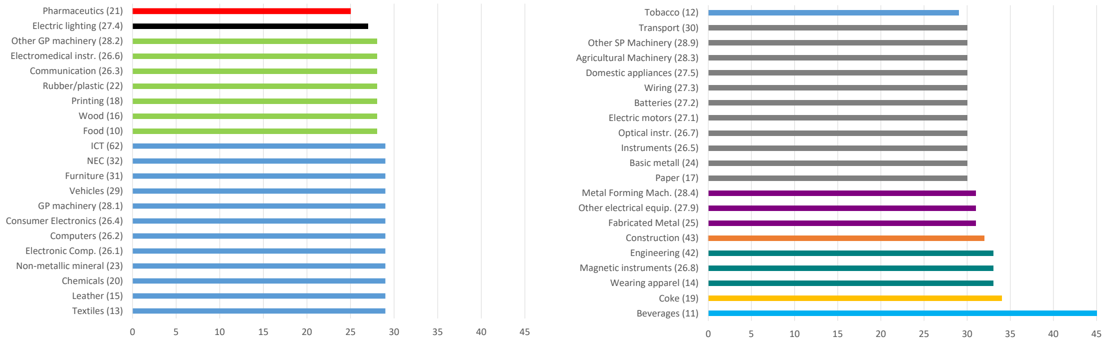


Notes: Share of forward citations that stem from 4.0 citing patents. Basis are all 4.0 patents in industries with 1000 or more 4.0 patents filed between 1980-2011 that received citations by other EP patents. Industry classification based on 2-/3-digit NACE Rev.2 concordance table from Patstat (Van Looy et al. 2015). For a full industry list, see Table 8.
Source: PATSTAT 2020 - spring edition, own calculation.

foundation for subsequent follow-on 4.0 inventions, since more than 83% of their citations are 4.0 citations. The second group of industries consists of vehicles, other general purpose machinery, communication and instruments with a 4.0 citation share of 76 to 80%. For all other industries this share is below 70%. At the lower end of the distribution, we find industries like electric motors, textiles, other special purpose machinery or chemicals whose 4.0 patents subsequently stimulate relatively more non 4-0 inventions.

Citations can also be used to examine the temporal pattern in the diffusion process. A well-known indicator is the time span how long it takes for a patent to receive its first citation. Figure 18 shows for all 4.0 patents in a given industry the number of months to receive the first citation. While the total number of 4.0 patents in the pharmaceutical industry is rather low (488 4.0 patents in total), these patents are disseminated very strongly (mean citation rate of 2.5) and very quickly. With 25 months to get the first citation, pharmaceutical is in the top position, followed by electrical lightening (27 months) and a number of industries consisting of food, wood, printing, rubber and plastic, communication, electromedical instruments and other general purpose machinery, all with 28 months. Interestingly, the industries with the highest number of 4.0 patents and the highest citation rates (communication and computer) are not in the top two positions in terms of time span to first citation. But overall, Figure 18 also shows that the time to first citation is fairly evenly distributed across industries. In more than 3 out of 4 industries, the time span is only between 25 and 30 months.

Figure 18: Number of Months to first citation of 4.0 patents by NACE Industry



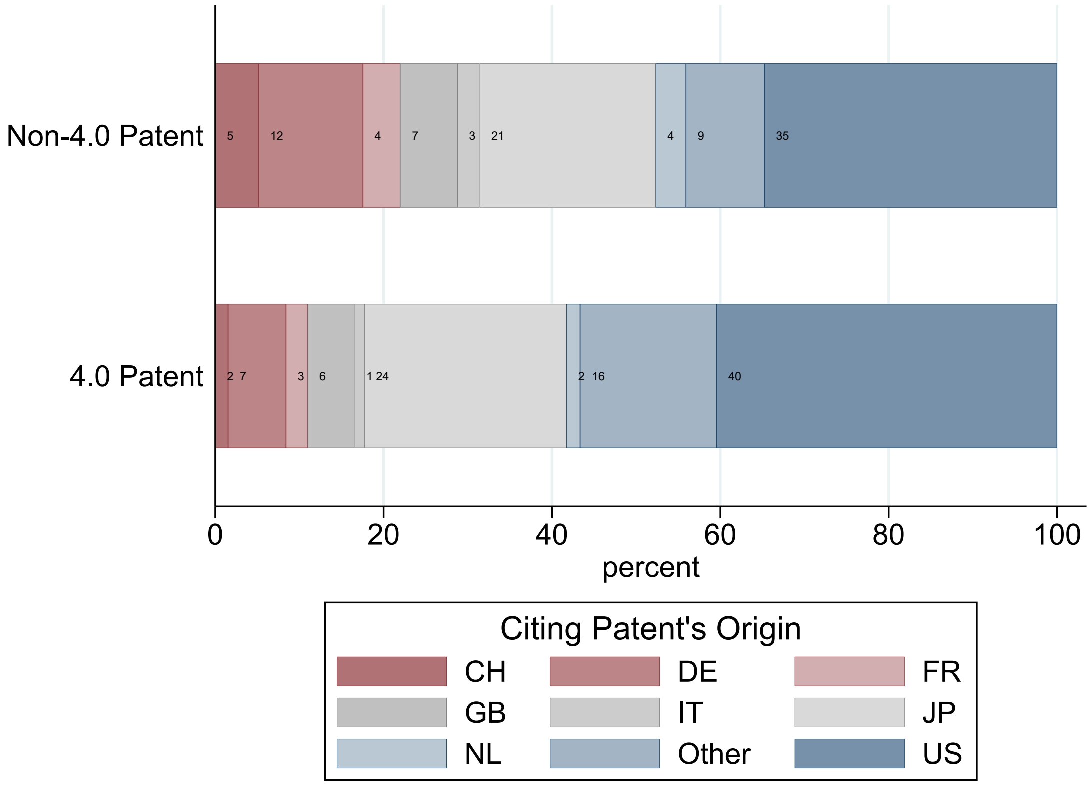
Notes: Data used all EP 4.0 patent applications filed between 1980-2011 that received citations by other EP patents. Industry classification based on 2- and 3-digit Nace Rev2.
Source: PATSTAT 2020 - spring edition, own calculation.

4.4 Diffusion Pattern of Top Cited 4.0 Patents and Top 4.0 Patent Applicants

The top cited patents are considered to be breakthrough innovations because they have the largest impact on society and technological advancement. In order to get a better understanding of the type of technologies that are most diffused, Table 9 in the Appendix lists the individual top most cited 4.0 patents. Most of the patent owners are well-established multinationals residing in Japan and the US. In fact, 14 and 9 of the top 30 most cited 4.0 patents are filed by these countries respectively. Almost half (14) and thus a disproportionately high share were patents whose main NACE Sector was communications equipment (code 26.3). Texas Instruments' electronic television program guide system received the most citations within 5 years after application.

For the subsequent analysis, we follow Ahuja and Lampert's (2001) approach to identify breakthrough innovations and focuse on the top 1% most cited 4.0 patents and their diffusion pattern. Figure 19 gives us an idea of the geographic location that top cited 4.0 patents diffuse to by presenting the origin of the citing patent. The country of origin is defined by the firm's assignee location. For comparison, we also show the origin of forward citations for the top 1% most cited non-4.0 patents. For both 4.0 and non-4.0 patents the United States account for the largest share of forward citations. Firms located in the US represent 40% of the citations that the top 1% most cited 4.0 EP patents received. This is relatively larger than the US's share for Non-4.0 EP patents (35%), indicating that 4.0 technologies diffuse more-so to the US than non-4.0 technologies do. Japan accounts for the second largest share of citations; 24% and 21%, respectively. Among European countries, Germany make up the largest share with 7% of the citations of top 4.0 patents. In contrast to the US and Japan, however, they represent a relatively larger share of citations in non-4.0 patents. A similar pattern is found for France, Great Britain or Italy. The patterns described for the top 1% most cited 4.0 patents are also in line with the geographical diffusion of patents citing 4.0 patents using the total sample of 4.0 patents, as it is shown in Figure 20).

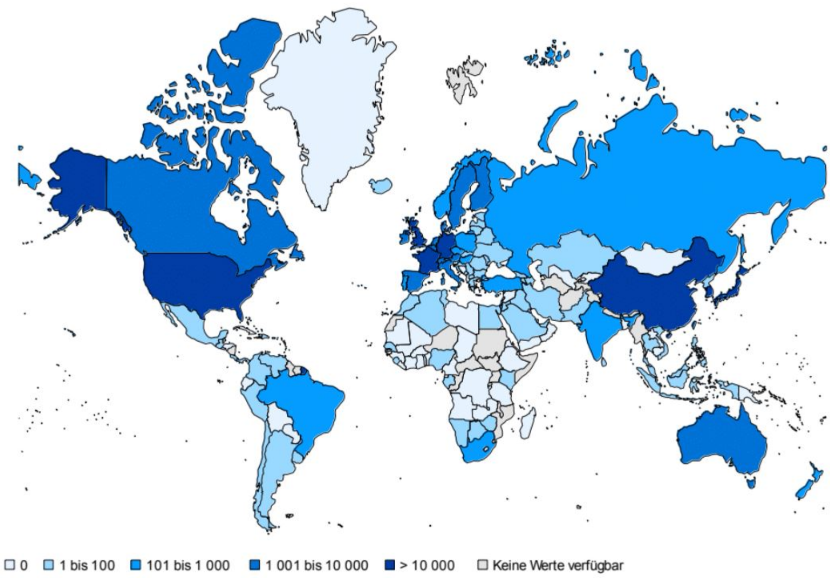
Figure 19: Geographical Diffusion of Patents Citing Top Cited 1% 4.0 and Non-4.0 Patents



Notes: Data used are all forward citations from the top 1% most cited 4.0 and Non-4.0 EP applications filed between 1980-2016 that received citations by other EP patents. No restrictons are made wrt a citation time-window. The origin of the citation is defined as the country that the citing firm (assignee location) is residing in.
Source: PATSTAT 2020 - spring edition, own calculation.

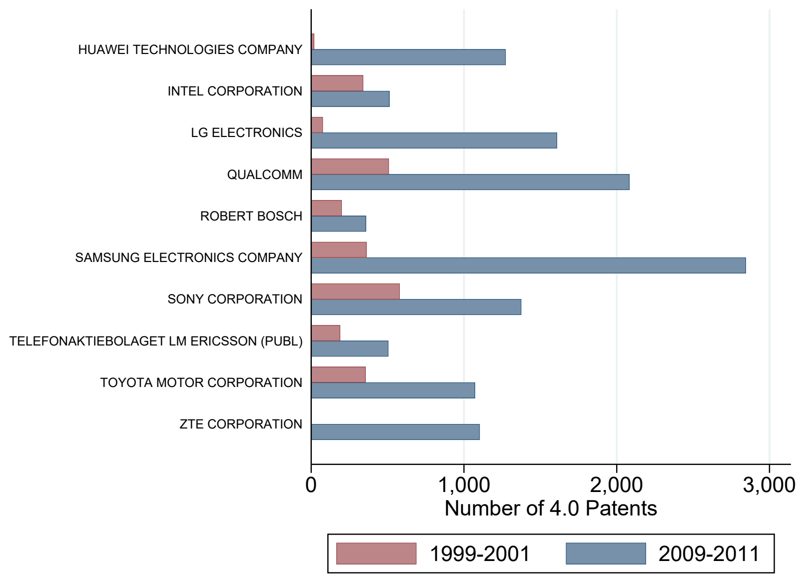
Finally, we study the top 10 firms with the most 4.0 patent applications filed in 2009-2011 and how their patents diffuse. Together they accounted for a substantial share of 18.9% of all 4.0 patents filed during that period. Taking non-4.0 patents additionally into account, these 10 firms alone applied for 7.6% of the EP applications filed in 2009-2011. Figure 21 shows that Samsung takes first place among 4.0 inventors, followed by Qualcomm and LG Electronics. Strikingly, all of the top 10 firms have seen a noteworthy increase in the absolute number of 4.0 patents filed compared to 1999-2001. For many but not for all (Intel, Bosch) top 10 firms, this increase has been stronger than the growth for the total sample (see Figure 2). When weighting their applications by citations, however, the ranking changes. From Figure 22 it becomes apparent that now LG Electronics is in first place, followed by Sony and Samsung. Furthermore, the older 4.0 patents (filed in 1999-2001) are far more valuable than the more recently filed 4.0 patents, despite having restricted both to a 5-year citation window.

Figure 20: Geographical Diffusion of Patents Citing 4.0 Patents (Absolute Number)



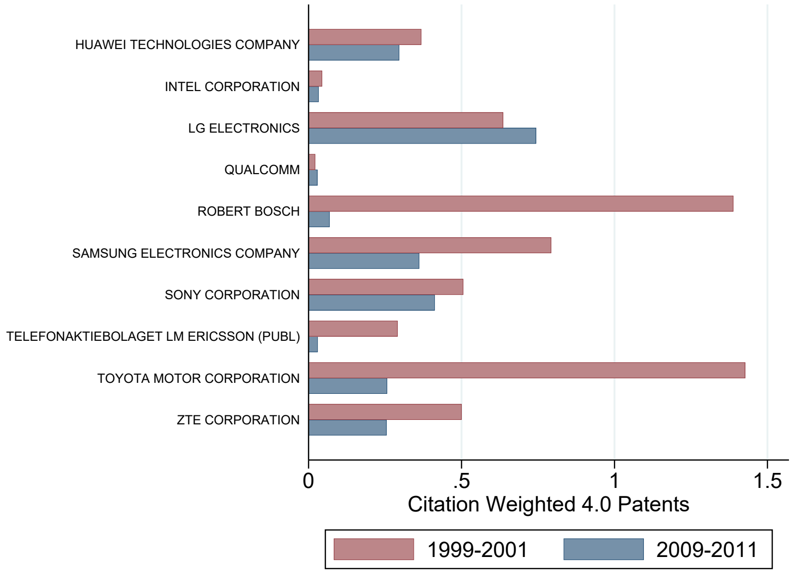
Notes: Data used all citations of EP 4.0 patents filed between 1980-2016 that received citations by other EP patents. No restrictons are made wrt a citation time-window. The origin of the citation is defined as the country that the citing firm (assignee location) is residing in.
Source: PATSTAT 2020 - spring edition, own calculation.

Figure 21: Top 10 4.0 Applicants Over Time



Source: PATSTAT 2020 - spring edition, own calculation.

Figure 22: Top 10 4.0 Applicants and Their Citation Weighted Count



Notes: The graph weighs the applicant's 4.0 patents by dividing the number of family-cleaned citations the 4.0 patents received within five years after application by the applicant's absolute number of 4.0 patents. Source: PATSTAT 2020 - spring edition, own calculation.

5 Impact of Adopting AI on Productivity

Up to now, we have studied the generation and diffusion of new inventions related to the Fourth Industrial Revolution. A particular focus was on distinguishing the underpinning technological fields in terms of their functional role in digitizing economics activities, for example translating activities into information, analysis of this information, and interconnection between machines generating and treating information. This section complements the analysis on the generation and diffusion of 4.0 technologies conducted in the previous two sections in an important way: It examines firm-level impacts of using 4.0 technologies. More specifically, we analyse the productivity effects of developing and adopting artificial intelligence (AI) technologies. We would have liked to explore this question for 4.0 technologies in general, but the available data limit our analysis to AI. On the other hand, AI is supposed to be the core technology of the the Fourth Industrial Revolution that has shown the strongest growth since 2010 and especially since 2015.

The diffusion lag between the date of the initial invention and the moment when firms realize the economic benefits associated with the use of this technology has been widely studied (see Hall and Trajtenberg 2004, for an illustration). Several determinants have been empirically well-established to influence firms' technology adoption decisions (Stoneman and Battisti 2010). R&D capabilities, the availability of skills, and compatible organizational routines are among the most important determinants (Cohen and Levinthal 1989; Silverberg 1991; David 2000). A.I. technologies are no exception to this rule (Brynjolfsson et al. 2017; Haskel and Westlake 2018). The interesting point related to A.I. technologies lies in its complementarity with data. In contrast to other generic technologies, the use of A.I. does not depend too much on technological complementarity, but the availability of "good data". Firms must collect large and harmonised datasets to be able to grasp the full benefits from A.I. in easing decision-making and analysis. Developing in-house capacities to harmonise efforts in collecting relevant data that reflects the actual firms' problems and environment is hence key to get relevant predictions. This decision to jointly adopt A.I. with complementary investment in data infrastructure and how this affects productivity is studied in this section. We set the stage in subsection 5.1 with a short analysis of the diffusion of AI in Germany using patent data, followed by descriptive evidence on the adoption of AI and data infrastructure investments using the Mannheim Innovation Panel (MIP). The MIP is a German firm-level data set that we use to analyse the productivity effects of AI adoption. In the cross-section 2018 it included a specific and detailed set of questions of the adoption of AI methods that we exploit. Productivity estimates are presented in subsection 5.2.

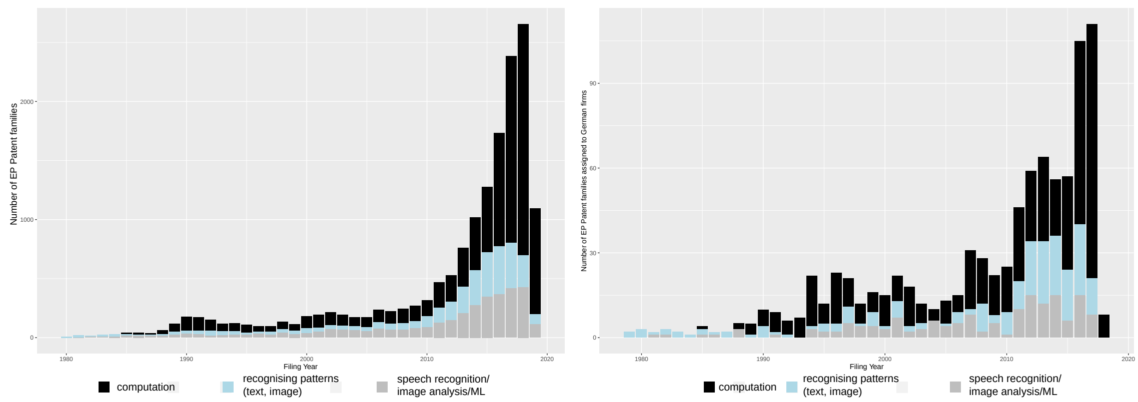
5.1 Patenting and Adoption of AI in Germany

5.1.1 AI Patenting in Germany

We start with a short overview of patenting activities in AI related fields in Germany before we exploit a unique dataset that documents the adoption trend of such technologies at the firm level in Germany. Figure 23 depicts the patenting activities at EPO over time and across the main AI applications areas for Europe compares it with Germany. The figures rely on the AI delineation developed by WIPO and suggested as the best practice in a recent comparative analysis on the topic. That is, we distinguish computation, pattern recognition, and a cluster of connected technologies in speech recognition, image analysis and machine learning (Baruffaldi et al. 2020). Both graphs exhibit a structural uptake of all three areas since 2010, but especially in computation AI applications. The latter field has more than two and a half times as many patents as those in pattern recognition and speech recognition. The German case shows a similar pattern in the increasing patenting activity, but differ in terms of composition over time. German companies seem to specialize their A.I. patenting activities in patterns recognition and lately in computing even stronger than other countries since the 2010s.

Figure 23: AI Patents in Europe and Germany over Time, by AI Areas

(a) Europe (b) Germany



Notes: Patents (patent-family cleaned patent applications) with at least one CPC class that has been identified as being relevant for AI at the EPO.
Source: PATSTAT 2020 - spring edition, own calculation.

5.1.2 Firm-Level Evidence on the Adoption of AI in Germany

Up to now we have measured the development and diffusion/adoption of 4.0 technologies and AI using patent data and related citation data. However, not all AI inventions may be patented and the citations to AI patents only cover a specific type of knowledge diffusion which leads to a patented follow-on invention. For example, the pure acquisition of an AI-based production technology would not be captured as adoption activity. Hence, we additionally use the Mannheim Innovation Panel (MIP). The MIP is the German contribution to the European-wide Community Innovation Surveys. The CIS is a representative survey among all German firms with more than 5 employees in manufacturing (Nace 2.0 classes 10-33), mining (Nace 2.0 classes 05-09), energy and water supply and recycling (Nace 2.0 classes 35-39) and core service sectors (Nace 2.0 classes 46 wholesale, 49-53 transport, 58-63 information and communication services, 64-66 financial services 69-74 technical services and 78-82 other business related services). Our main estimation sample consists of 4360 firms with non-missing information on all relevant variables, which we define below. As the sample is representative for the German population, the vast majority of sampled firms is rather small. The median firm size is 27 employees, while the average firm size is about 192.3 employees, indicating that firm size is highly skewed (see Table 5).

Most importantly, the German CIS2018 included a specific set of questions on artificial intelligence which better captures the use of AI methods in firms since it is not restricted to AI patenting activities. *AI* is a binary variable that equals one if the firm uses artificial intelligence. In the survey artificial intelligence is defined as a method of information processing that allows computers to autonomously solve problems and examples like language understanding, image recognition, machine learning and knowledge-based systems are given. The survey additionally asks each firm who developed the AI method(s) employed, and we correspondingly define three indicators: *AI in-house* equals 1 if the firm mainly develops the AI methods within the firm, *AI joint* equals 1 if they develop it in cooperation with external partners and *AI outsourced* is 1 if the AI methods used were mainly developed by external parties. The latter can include the use of standard AI solutions as well as firm-specific special solutions developed by external parties within the framework of an external research contract. *AI joint/out* is a combined dummy that indicators that external parties were involved in developing the AI method employed, either partly or fully.

Our analysis aims to study the impact of the use of AI technologies, which is supposed to be complementary to the availability of "good" data. Data can either be data collected in-house or external data bases. The MIP survey also asks firms whether and how much they have invested in setting up new or maintaining existing internal data bases (including internal expenditures for software infrastructure programming) and the same for purchasing external data bases collected by others. *Data in-house* and *Data external* are dummy variables indicating positive expenditures for internal and external data bases respectively, while *ln(Data in-house)* and *ln(Data external)* measure the corresponding log expenditures.

Table 4 shows the overall AI adoption rate and its pattern across sectors and firm size. In 2018, 7.9 of German firms used AI methods in their company.[6]. The use of AI is still very heterogeneously distributed across industries and firm size. Among small and medium-sized enterprises with less than 250 employees, about 6% of the firms employ AI-based technologies and solutions, while in the group of large companies with more than 1000 employees, it is more than 4 out of 10. In terms of sectors, knowledge-based service firms employ AI by far the most (15.5%), followed by high-tech firms (10.9%). Higher

[6] This is very close to the extrapolated figure of 5.8% for Germany as a whole (Rammer 2020), indicating that our estimation sample does not give rise to selectivity concerns.

adoption rates in these sectors are not surprising as they both offer more technological and market opportunities to exploit AI technologies. In all other industries, it is so far only 3.5 to 4.3% of companies that use AI.

We furthermore evidence that, on average, firms tend to adopt AI via acquiring these technologies and not developing them in-house. Only around 1.5% of companies have developed them entirely in-house. As many as 2.3% of the firms have developed them in cooperation with external partners, while 4.0% of the companies have completely outsourced the development of their AI technologies. Interestingly in most sectors and across size categories, the proportion of firms that fully outsource the development of AI among the AI using firms is similar at about 50%. Finally, the last two columns of Table 4 show the rate of firms that invest in data bases which can be potentially exploited using AI technologies. In total, 37% of the firms invest in setting up new or maintaining existing internal data bases while 40.8% of firms acquire external data bases. Large firms with more than 1000 employees invest more frequently in own than external data bases, while smaller firms rely more often on the acquisition of external data bases. It is interesting to note that high-tech companies tend to focus relatively more on internal databases, while it is the opposite for knowledge-intensive service companies.

Table 4: AI Adoption Rate and Data Expenditure Rate

	AI	AI in-house	AI joint	AI outsourced	Data in-house[1]	Data external[1]
High Tech	0.109	0.024	0.039	0.043	0.461	0.449
Low Tech	0.035	0.003	0.006	0.025	0.292	0.337
KIS	0.155	0.033	0.046	0.074	0.463	0.484
LKIS	0.037	0.004	0.007	0.027	0.313	0.354
Energy	0.043	0.005	0.014	0.021	0.316	0.421
5-99	0.062	0.013	0.016	0.032	0.338	0.365
100-249	0.060	0.008	0.017	0.035	0.477	0.564
250-999	0.160	0.017	0.067	0.076	0.506	0.579
1000+	0.402	0.0667	0.125	0.200	0.614	0.551
Total	0.079	0.015	0.023	0.040	0.370	0.408

Notes: [1] Data in-house is a binary indicator that equals 1 if the firm has invested in building up internal data based (including internal software infrastructure programming). Data external is a binary indicator that equals 1 if the firm has acquired external data bases (including software). In the estimation the corresponding expenditures values are used. Sector classification is based on a Eurostat definition using 2-digit Nace Rev2 information. KIS: Knowledge-intensive services, LKIS: less knowledge-intensive services.
Source: Source: ZEW - Mannheim Innovation Panel CIS2018.

This uneven adoption pattern suggests that the nature of activities and presumably the size of investments matter to grasp the benefits of AI.

5.2 Econometric Evidence on the Productivity Impact of AI

5.2.1 Econometric Framework

We measure the benefits of AI adoption using a standard Cobb Douglas production framework as theoretical framework (see, for example, Mairesse and Sassenou 1991). We estimate in the first stage firm-level *Total Factor Productivity (TFP)* using the Levinson-Petrin (LP) approach and in the second stage the contribution of AI and investment in data on TFP.

The output of firm i at time t can be described by the Cobb-Douglas production function in log linear version as:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \omega_{it} + \eta_{it}, \tag{1}$$

where y_{it}, m_{it}, and k_{it} denote firm $i's$ value of log output, log material, and log physical capital in year t, and l_{it} denotes its log labour input, measured as the log number of employees. In addition to the observed inputs labour, capital and material, the CD production function contains a term A_{it} which is a measure of firm $i's$ level of efficiency, commonly referred to as Total Factor Productivity (TFP).[7] In our log linear specification $ln(A_{it})$ is decomposed into three elements, β_0, ω_{it}, and η_{it}. The first element, β_0, represents mean efficiency across all firms, and ω_{it} denotes the time- and firm-specific deviation from that mean, while η_{it} is a true error term that contains unobserved shocks and measurement errors. ϵ_{it} is i.i.d normally distributed while we follow the usual assumption that ω_{it} evolves according to a first-order Markov process.

We assume that the firm-specific deviation from the mean efficiency is observable by the firm when it makes its investment decision, but not by the econometrician. This leads to an endogeneity problem since firm's input choices will likely be correlated with its productivity and thus with the error term of the productivity equation (Marschak and Andrews 1944). In this case, OLS leads to inconsistent estimates. The variable inputs like labour and materials are expected to have an upward bias and the coefficients associated with quasi-fixed inputs like capital are expected to be biased downwards in OLS (Olley and Pakes 1996). In order to estimate the production function, we use the non-parametric econometric estimation method by Levinson and Petrin. This approach solves the endogeneity problem using a control function approach. The main idea is to use observable material input to proxy for unobserved productivity shocks. The basic assumption is that $m_{it} = f_t(\omega_{it}, k_{it})$ and f_t is invertible. Therefore $\omega_{it} = f_t^{-1}(m_{it}, k_{it})$. In a first step, we therefore regress y_{it} on labour l_{it} and a non-parametric function $\phi(m_{it}, k_{it})$ that we approximate with a second order polynomial to get the coefficient β_l and $\widehat{\phi_{it}}$. In a second step, we exploit the moment conditions that capital k_{it} and lagged material m_{it-1} are uncorrelated with the error term to estimate the remaining production function parameters.

[7] The effect of A_{it} on Y_{it} is assumed to be Hicks-neutral, so TFP is additively separable from the other production factors.

In a second stage, we estimate TFP using the coefficient estimates from the first stage and use $\widehat{TFP}_{it}$ as dependent variable to estimate the impact of AI adoption and investment in data bases on productivity:

$$\widehat{TFP}_{it} = \gamma_1 AI_{it} + \gamma_2 ln(DI) + \gamma_3 AI_{it} \times ln(DI) + \gamma_4 ln(DE) + \gamma_5 AI_{it} \times ln(DE) + X_{it} + \varepsilon_{it} \quad (2)$$

where AI denotes AI adoption, and $ln(DE)$ and $ln(DI)$ measure the log expenditure of firms for setting up internal data bases and acquiring external data bases, respectively. In order to test for for complementarity, our specification additionally accounts for the interaction of AI with both investment in internal and external data investment. In an alternative specification we differentiate between inhouse AI development, joint AI development and outsourced AI and adjust the interaction terms accordingly.

In addition to AI and investment in internal and external data bases, a set of control variables X_{it} is included. The latter includes industry dummies, a location dummy for firms from East Germany and a dummy for firms belonging to a company group.

Table 5: Summary Statistics (overall sample)

	N	Mean	S.D.	p10	p50	p90
Labour Prod.[1]	4360	0.184	0.326	0.045	0.114	0.333
ln(Labour Prod.)	4360	-2.141	0.863	-3.095	-2.171	-1.099
TFP	4360	0.001	0.491	-0.461	-0.008	0.460
Labour	4360	192.269	1164.842	7	27	266.5
ln(Labour)	4360	3.570	1.464	1.946	3.296	5.585
ln(Capital)	4360	-2.777	2.398	-5.943	-3.013	0
ln(Material)	4360	-3.516	1.550	-5.545	-3.401	-1.691
AI	4360	0.079	0.270	0	0	0
AI in-house	4354	0.015	0.120	0	0	0
AI joint	4354	0.023	0.150	0	0	0
AI outsourced	4354	0.0402	0.196	0	0	0
Data in-house[1,2]	1600	0.221	1.861	0.001	0.015	0.200
ln(Data in-house)	4140	-1.327	2.093	-4.605	0	0
Data ext[1,2]	1833	0.226	1.303	0.002	0.022	0.250
ln(Data ext)	4177	-1.444	2.062	-4.605	0	0
East	4360	0.415	0.493	0	0	1
Group	4360	0.315	0.465	0	0	1

Notes: [1] For illustration purposes, the variable is shown in million Euro while the log value is used in the estimation. [2] Descriptive statistics for firms with positive expenditure only. In the estimation, the transformed variable log(1+expenditure) is used.
Source: Source: ZEW - Mannheim Innovation Panel CIS2018.

Table 5 shows the summary statistics for all variables used in the econometric analysis. The average labour productivity is about 184 thousand Euro sales per employee. If firms invest in setting up or maintaining internal data bases, they spend on average 221 thousand Euro. However, the distribution on data base investments is highly skewed. The median is much smaller with about 15 thousand Euro. The expenditures for acquiring external data bases is slightly higher with mean of 226 thousand Euro and a median of 22 thousand Euro.

Table 6 presents results on simple differences in means of productivity and investment in databases by AI adoption. Both average labour productivity and TFP is higher among AI adopters than among non-AI adopters. We also confirm that investment in both in-house and external databases is significantly larger among AI adopters. However, these differences could also be driven by differences other observed and unobserved characteristics like industry or firm size. We therefore perform an econometric analysis.

Table 6: Productivity and Investment in Databases by AI adoption

	AI adoption		
	No AI	AI	p-value
Labour Prod	0.181	0.218	0.0757*
ln(Labour Prod)	-2.152	-2.004	0.0021***
TFP	-0.002	0.044	0.0627*
Data in-house	0.154	0.737	0.0289**
Data external	0.176	0.620	0.0075***
Obs.	4,015	345	

Notes: N=4,360. Stars indicate significance of t-test of mean difference between firms with and without AI adoption, allowing for unequal variances. ***(,**,*): p<0.01(,0.05,0.10)

5.2.2 Econometric Results

Table 7 presents the results of the second stage estimates linking AI adoption with labour productivity. Dependent variable is TFP that we estimated using the Levinsohn-Petrin approach. For comparison, OLS estimates for labour productivity are presented in Table 10 in the Appendix. Overall, they show very similar results, so we focus on the LP results.

Column (1) shows the estimation results for equation (2), assuming that γ_2 to γ_5 are zero. We therefore focus only on the impact of AI adoption. The productivity impact of 4.7% is rather large but slightly fail to be significant at conventional levels. The productivity difference is very close to the one reported in the simple mean difference test, however using cluster-robust standard errors we do not find this difference to be significant anymore. Thus, we cannot confirm that adopting AI technologies has on average a significant impact on TFP.

While adopting AI in general does not guarantee to go with productivity gains, we check in columns (2) and (3) whether the adoption process (i.e. in-house development, joint development, and outsourced) has an impact on the productivity gains that firms achieve as a result of using AI. Model (2) confirms significant differences. Firms that purely develop their AI technologies with firms' internal R&D capacities yield a significant productivity boost of about 7.1%. In contrast, productivity gains are smaller and show a much higher variance when firms develop them in cooperation with external partners or purely outsource them. The results in column (3) supports the previous findings by pooling together joint and outsourced development of the adopted AI technologies.

While Models (1)-(3) focus solely on the role of AI, we extend the analysis in models (4) to (6) to study the role of AI adoption, investment in data infrastructure and its interplay. Column (4) adds investments for internal and external data infrastructure to the specification. Both turn out to be insignificant showing that investing in data infrastructure as such is also not associated with higher productivity. However, the picture changes when we explicitly allow for the interaction between AI adoption and in-house and external data infrastructure investments. The results suggest a high productivity gain of adopting AI, when we simultaneously control for the related data investment strategies. Firms on average boost productivity by 9.6% when they adopt AI technologies. This effect is highly significant at the 1% level. Regarding investments in internal and external data infrastructures, we corroborate the earlier finding that they do not significantly increase productivity per se. However, we do find a complementary and enhancing effect of investing in internal data infrastructure for firms which simultaneously adopt AI technologies. On average, an increase in internal data infrastructure investment by 1% leads to an increase in productivity by 0.038% when combined with AI adoption. In contrast, there is no significant interaction effect of external data infrastructure investment and AI adoption. The complementarity found for internal data infrastructure investment and AI is in line with previous studies showing the importance of building a data infrastructure in the ICT diffusion to fully grasp the benefits from using such technologies (David 2000). However, our results add to this literature by showing that the source of data matters for productivity effects of AI adoption.

In model (6) we simultaneously investigate the interplay between the underlying source of AI development and firms' data strategies. That is, we further extend our analysis by distinguishing pure in-house versus joint outsourced AI and its interplay with internal and external data infrastructure investments. Regarding our previous finding of complementarity between AI and internal data infrastructure investments, model (6) details that this complementary relationship exists between both in-house AI as well as joint outsourced AI and internal data infrastructure investments. However, firms with in-house AI still benefit from a higher complementarity effect from in-house data infrastructure (0.057) than firms outsourcing AI or developing it with external partners (0.034). Surprisingly, in model (6) we do not find any statistical significance in adopting AI which has been developed in-house. This in combination with the strong interaction effect with internal and external data confirms that the initial finding on in-house AI adoption was driven by the large amount of firms who jointly develop an in-house and external data infrastructure. In contrast, firms that develop their AI solutions in-house AI and simultaneously focus on external database investment achieve significantly lower (short-run) productivity gains (-0.031). This detrimental effect on productivity highlights the importance of investing in-house to create a data-oriented organizational culture and the capabilities to provide large standardized datasets to address specific firms' problems and to analyze various A.I.

Table 7: Impact of AI Adoption on TFP Across Firms' Strategies

	(1)	(2)	(3)	(4)	(5)	(6)
AI	0.047 (0.032)			0.032 (0.033)	0.096*** (0.035)	
AI in-house		0.071** (0.035)	0.071** (0.035)			0.107 (0.078)
AI joint		0.025 (0.071)				
AI outsourced		0.048 (0.031)				
AI joint/outsourced			0.039 (0.046)			0.093** (0.037)
Ln(data in-house)				-0.009 (0.007)	-0.010 (0.007)	-0.011 (0.007)
AI × Ln(data in-house)					0.038*** (0.010)	
Ln(data ext.)				0.004 (0.006)	0.003 (0.006)	0.003 (0.006)
AI × Ln(data ext.)					0.005 (0.016)	
AI in-house × Ln(data in-house)						0.057** (0.023)
AI in-house × Ln(data ext.)						-0.031* (0.017)
AI joint/outsourced × Ln(data in-house)						0.034*** (0.010)
AI joint/outsourced × Ln(data ext.)						0.014 (0.018)
Constant	-0.002 (0.016)	-0.002 (0.016)	-0.002 (0.016)	-0.018 (0.017)	-0.021 (0.017)	-0.021 (0.017)
Observations	4360	4354	4354	4055	4055	4050

Dependent variable: TFP, estimated using the Levinsohn-Petrin approach.
Clustered standard errors in parentheses (by 3-digit industry level).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

scenarios (Brynjolfsson et al. 2017; Haskel and Westlake 2018). While model (2) initially suggests that sourcing A.I. inside the company is key to increase productivity, model (6) rather shows that the underpinning data strategies explain why the in-house adoption seems more efficient.

In summary, our results show firstly that the adoption of AI does not automatically lead to productivity gains and that only firms that invest in internal complementary intangible assets (data) do so. Firms that invest in building in-house data capabilities enhance the effect of AI adoption on productivity, regardless of whether AI technologies are developed internally or external AI solutions are used. Second, firms that have adopted AI from an external source increase their productivity by 10 percent on average in comparison to firms that have not adopted AI. Third, if the firms with external AI also invest in an in-house data infrastructure, they benefit from an additional increase in productivity. Productivity increases by about 0.037% when the investment for in-house data infrastructure increases by 1%. Fourth, firms that choose the strategy of adopting AI through in-house development increase their productivity, when they simultaneously invest in an internal data infrastructure, but not when they combine their strategy of in-house AI development with external data. Firms that developed their AI technologies internally double the gains from complementarity with in-house data infrastructure investments compared to external data infrastructure.

6 Conclusion

This paper studies the role of 4.0 technologies. In a first part, based on a new classification for 4.0 patents developed by the EPO, we examine the patterns and trends in the emergence of 4.0 technologies in their entirety in Europe, as well as differentiated by technology groups (core, enabling and twin technologies) and sectors. In contrast to previous literature, we focus on the digital intensity of such 4.0 patents, as 4.0 patents show strong heterogeneity in terms of how large the share of 4.0-related technical features of the protected technology is. Furthermore, we investigate the diffusion of 4.0 technologies for the first time. We measure diffusion by looking at European 4.0 patents and their respective pattern of forward citations. In doing so, we differentiate whether 4.0 technologies diffuse into other 4.0 or non-4.0 technology fields, to which countries and sectors knowledge diffuses primarily.

In the second part, we examine the impact of the introduction of AI technologies on firm performance. The novelty of our study is to examine the complementarity between the adoption of AI and investments in complementary intangible assets, more specifically data infrastructure. To what extent complementary intangible assets explain the productivity paradox observed with A.I. technologies? This question has been mostly answered at the macro level (Brynjolfsson et al. 2017), or at the industry level for a broader set of digital technologies (Gal et al. 2019). In this paper, we rely on an unique dataset to empirically assess at the micro level the importance of investing in internal intangible assets (i.e. data and software) in translating the A.I. adoption into productivity gains.

In summary, as suggested by Brynjolfsson et al. (2017), our results show firstly that the adoption of A.I. does not automatically translate to productivity gains. We evidence a positive effect of AI adoption on TFP if firms jointly invest in internal complementary intangible assets. Firms that invest in building internal data capabilities significantly increase the effect of AI adoption on productivity, regardless of the source of AI developments (in-house versus external development). Second, firms that have adopted AI from an external source increase their productivity by 10 per cent on average compared to firms that have not adopted AI. Third, if the firms with external AI also invest in an internal data infrastructure, they benefit from an additional increase in productivity. Productivity increases by about 0.037% when the investment for internal data infrastructure increases by 1%. Fourth, firms that choose the strategy of adopting AI through in-house development increase productivity when they simultaneously invest in an internal data infrastructure, but not when they combine their strategy of in-house development with external data infrastructure. Firms that developed their AI technologies in-house double the gains from complementarity with internal data infrastructure investments compared to external data infrastructure.

We do not document a substitution effect coming from external investments in acquiring external data and software, suggesting that complementarity goes further than generating data inputs. This result echoes the literature on technological change that stresses the role of organizational culture and practices in grasping the benefits from ICT technologies (David 2000; Brynjolfsson and Hitt 2000; Aral et al. 2012; Brynjolfsson et al. 2021). Our findings show that only firms willing to bear the costs of adapting their internal organizational culture and skill-sets to A.I. predictions can fully grasp the related productivity gains.

Our study contributes to a larger set of studies stressing the increasing importance of intangible assets as a form of capital that deepens inequalities across firms in polarizing productivity gains among the leader firms (Andrews et al. 2016; Haskel and Westlake 2018; Gal et al. 2019). However, our study is not without limitations. We cannot establish a causal link between the decision of adopting AI and its impact on productivity. Our results must be then seen as correlation and descriptive evidence of the complementarity between internal intangible assets and AI Further studies may address this issue with external variations in the price of complementary assets to better establish the causal effect of AI adoption on productivity. Moreover, our data allows us to estimate the short-run productivity gains, ignoring potential effect in the long-run. Despite those limitations, the large and significant effect of AI and internal intangible assets suggest a few policy and managerial implications. More attention should be given to developing a "data culture" within the organization and providing appropriate training to adapt the firm's skill-set. Providing the right building blocks in machine learning (e.g. statistics and computer sciences) in general postgraduate education would help to avoid the polarization of the productivity gains in the hands of a few (large) companies. Building skills and an "AI friendly culture" are two key ingredients to develop complementary assets to those into productivity gains in the entire economy.

References

Andrews, D., C. Criscuolo, and P. N. Gal (2016): "The best versus the rest: the global productivity slowdown, divergence across firms and the role of public policy," .

Aral, S., E. Brynjolfsson, and L. Wu (2012): "Three-way complementarities: Performance pay, human resource analytics, and information technology," *Management Science*, 58, 913–931.

Baruffaldi, S., B. van Beuzekom, H. Dernis, D. Harhoff, N. Rao, D. Rosenfeld, and M. Squicciarini (2020): "Identifying and measuring developments in artificial intelligence," .

Bresnahan, T. (2010): "General purpose technologies," *Handbook of the Economics of Innovation*, 2, 761–791.

Brynjolfsson, E. and L. M. Hitt (2000): "Beyond computation: Information technology, organizational transformation and business performance," *Journal of Economic perspectives*, 14, 23–48.

Brynjolfsson, E. and A. McAfee (2011): *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, Digital Frontier Press.

——— (2014): *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W. W. Norton.

Brynjolfsson, E., D. Rock, and C. Syverson (2017): "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics," Tech. rep., National Bureau of Economic Research.

——— (2021): "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, 13, 333–72.

Calvino, F., C. Criscuolo, L. Marcolin, and M. Squicciarini (2018): "A taxonomy of digital intensive sectors," .

Cohen, W. M. and D. A. Levinthal (1989): "Innovation and learning: the two faces of R & D," *The economic journal*, 99, 569–596.

David, P. A. (2000): "Understanding digital technology's evolution and the path of measured productivity growth: present and future in the mirror of the past," *Understanding the digital economy: data, tools, and research. MIT, Cambridge, MA.*

Fuller, A., Z. Fan, C. Day, and C. Barlow (2020): "Digital Twin: Enabling Technologies, Challenges and Open Research," *IEEE Access*, 8, 108952–108971.

Gal, P., G. Nicoletti, T. Renault, S. Sorbe, and C. Timiliotis (2019): "Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries," .

GORDON, R. J. (2012): "Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds," Working Paper 18315, National Bureau of Economic Research.

GRILICHES, Z. (1957): "Hybrid corn: An exploration in the economics of technological change," *Econometrica, Journal of the Econometric Society*, 501–522.

HALL, B. H. AND M. TRAJTENBERG (2004): "Uncovering GPTs with patent data," Tech. rep., National Bureau of Economic Research.

HASKEL, J. AND S. WESTLAKE (2018): *Capitalism without capital: The rise of the intangible economy*, Princeton University Press.

MAIRESSE, J. AND M. SASSENOU (1991): "R&D Productivity: A Survey of Econometric Studies at the Firm Level," Working Papers 3666, National Bureau of Economic Research, Inc, NBER.

MARSCHAK, J. AND W. H. ANDREWS (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12, 143–205.

MARTINELLI, A., A. MINA, AND M. MOGGI (2020): "The enabling technologies of industry 4.0: Examining the seeds of the fourth industrial revolution," *Industrial and Corporate Change, Forthcoming*.

MÉNIÈRE, Y., I. RUDYK, AND J. VALDÉS (2017): "Patents and the Fourth Industrial Revolution: The inventions behind digital transformation," Tech. rep.

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

QI, Q., F. TAO, T. HU, N. ANWER, A. LIU, Y. WEI, L. WANG, AND A. NEE (2019): "Enabling technologies and tools for digital twin," *Journal of Manufacturing Systems*.

RAMMER, C. (2020): "Dokumentation zur Innovationserhebung 2019," Tech. rep.

ROGERS, E. M. (2010): *Diffusion of innovations*, Simon and Schuster.

SILVERBERG, G. (1991): "Adoption and diffusion of technology as a collective evolutionary process," in *Diffusion of technologies and social behavior*, Springer, 209–229.

STONEMAN, P. AND G. BATTISTI (2010): "The diffusion of new technology," in *Handbook of the Economics of Innovation*, Elsevier, vol. 2, 733–760.

TEECE, D. J. (2018): "Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world," *Research Policy*, 47, 1367–1387.

VAN LOOY, B., C. VEREYEN, AND U. SCHMOCH (2015): "Patent Statistics: Concordance IPC V8 – NACE REV.2," Tech. rep.

7 Appendix

Table 8: Forward Citations of 4.0 Patents by Sector

NACE	Sector Description	Patent Count^w	Share Receiving Citations^w	Share Receiving Citations	Median Citations	Mean Citations	75^{th} Percentile Citations	95^{th} Percentile Citations	Maximum Number of Citations	4.0 Citation Count^w	Non-4.0 Citation Count^w	4.0 Citation Share^w	Months to First Citation
10	Food Products	142	5.9	16.0	2	3.3	3.0	11.4	23	28	49	36.4	28
11	Beverages	2	2.1	13.3	2	1.5	1.8	2.0	2	0	0	0.0	45
12	Tobacco Products	35	8.8	20.0	1	5.9	9.3	21.0	24	12	43	21.8	29
13	Textiles	3,127	12.5	22.9	1	1.9	2.0	5.0	14	675	710	48.7	29
14	Wearing Apparel	103	4.4	17.8	1	1.5	2.0	3.0	9	10	16	38.5	33
15	Leather and Related Products	24	3.2	7.2	1	1.3	1.8	2.0	2	0	2	0.0	29
16	Wood and of Products of Wood and Cork, Except Furniture	6	5.6	26.9	2	2.4	2.5	5.8	7	3	1	75	28
17	Paper and Paper Products	144	4.7	22.8	1	2.0	3.0	5.0	11	21	35	37.5	30
18	Printing and Reproduction of Recorded Media	583	7.8	29.7	2	2.6	3.0	7.0	23	281	128	68.7	28
19	Coke and Refined Petroleum Products	51	4.5	15.2	1	1.7	1.0	5.6	7	3	8	27.3	34
20	Chemicals and Chemical Products	2,632	7.4	22.5	1	2.0	2.0	5.0	20	446	689	39.3	29
21	Basic Pharmaceutical Products and Pharmaceutical Preparations	488	3.2	11.3	2	2.5	3.0	7.0	18	40	57	41.2	25
22	Rubber and Plastic Products	1,116	10.6	24.2	1	2.1	3.0	5.0	34	342	221	60.7	28
23	Other Non-Metallic Mineral Products	572	6.0	21.5	1	2.2	3.0	6.0	32	102	137	42.7	29
24	Basic Metals	79	6.8	23.4	1	2.1	3.0	6.0	7	18	16	52.9	30
25	Fabricated Metal Products, Except Machinery and Equipment	2,789	13.4	24.3	1	1.9	2.0	5.0	13	769	478	61.7	31
26.1	Electronic Components and Boards	3,568	9.4	25.8	2	2.3	3.0	6.0	33	1,289	748	63.3	29
26.2	computers and peripheral equipment	67,121	12.1	20.5	1	2.4	3.0	6.0	66	23,764	4,629	83.7	29
26.3	Communication Equipment	124,384	15.4	20.7	2	2.5	3.0	7.0	66	47,883	14,149	77.2	28
26.4	Consumer Electronics	1,369	10.0	21.5	1	2.0	2.0	5.0	16	348	229	60.3	29
26.5	Instruments and Appliances for Measuring, Testing and Navigation	22,720	11.4	22.2	1	2.1	2.0	5.0	43	7,129	2,269	75.9	30
26.6	irradiation, electromedical and electrotherapeutic equipment	1,541	6.1	10.7	1	2.2	3.0	6.0	15	178	132	57.4	28
26.7	Optical Instruments and Photographic Equipment	2,322	10.0	24.4	1	2.1	2.0	6.0	19	718	380	65.4	30
26.8	Magnetic and Optical Media	20	4.0	23.5	1	2.0	2.0	5.8	9	2	5	28.6	33
27.1	Electric Motors, Generators, Transformers and Electricity Distribution	2,084	11.0	26.5	1	2.1	2.0	5.0	24	545	445	55.1	30
27.2	Batteries and Accumulators	264	9.6	29.9	1	2.1	2.0	6.0	15	56	92	37.8	30
27.3	Wiring and Wiring Devices	1,263	10.8	27.1	1	2.1	2.0	6.0	17	437	236	64.9	30
27.4	Electric Lighting Equipment	487	14.1	28.0	2	2.6	3.0	7.7	15	206	205	50.1	27
27.5	Domestic Appliances	2,479	11.9	24.9	1	2.0	2.0	5.0	27	653	479	57.7	30
27.9	Other Electrical Equipment	4,748	11.0	27.9	1	2.1	2.0	6.0	31	2,139	332	86.6	31
28.1	General-Purpose Machinery	13,185	14.7	28.1	1	2.1	2.0	6.0	34	4,836	2,457	66.3	29
28.2	Other General-Purpose Machinery	50,107	14.9	25.8	1	2.4	3.0	7.0	63	23,398	6,279	78.8	28
28.3	Agricultural and Forestry Machinery	681	15.9	26.5	2	2.3	3.0	6.0	16	234	209	52.8	30
28.4	Metal Forming Machinery and Machine Tools	610	6.9	20.5	1	2.1	2.0	5.0	34	136	104	56.7	31
28.9	Other Special-Purpose Machinery	3,632	8.9	21.0	1	1.9	2.0	5.0	12	649	692	48.4	30
29	Motor Vehicles, Trailers and Semi-Trailers	32,098	18.2	28.7	1	2.1	2.0	6.0	34	14,330	3,860	78.8	29
30	Other Transport Equipment	1,918	9.5	21.7	1	1.9	2.0	5.0	14	417	275	60.3	30
31	Furniture	308	6.5	18.8	1	1.8	2.0	4.8	8	64	29	68.8	29
32	Other Manufacturing	20,654	9.2	17.8	1	2.2	3.0	6.0	43	4,692	2,192	68.2	29
42	Civil Engineering	46	5.8	20.6	1	1.6	2.0	3.7	4	6	7	46.2	33
43	Specialised Construction Activities	185	4.9	17.1	1	1.6	2.0	3.0	8	18	28	39.1	32
62	Computer Programming, Consultancy and Related Activities	16,139	6.1	12.4	1	2.4	3.0	7.0	39	2,971	508	85.4	29
Total (*Mean)		385,826	8.9*	21.7*	1*	2.2*	2.5*	5.9*		139,848	43,560	0.5*	30

Notes: This table counts all EP 4.0 patents filed between 1980-2011. Citations are measured as the number of citations a given patent has received by another EP patent within 5 years after filing. Since patents may belong to multiple NACE sectors, columns indicated with w have been weighted by the percentage of the patent belonging to the respective sector. Industry classification is based on 2-/3-digit NACE Rev.2 concordance table from Patstat (Van Looy et al. 2015). NACE codes 10 to 31 are the manufacturers of the listed products. Source: PATSTAT 2020 - spring edition, own calculation.

Table 9: Top 30 Most Cited 4.0 Patents - EP Citations within 5 Years of Application

Company Name	Filing Year	Share 4.0 CPC	Country of Origin	Main Sector	Sector Weight	Forward Citations	Patent Title
TEXAS INSTRUMENTS	1998	0.68	US	26.3	0.94	66	Electronic television program guide system and method
TOSHIBA CORPORATION	1996	0.44	JP	28.2	0.61	63	A recording medium on which a data containing navigation data is recorded, a method and apparatus for reproducing a data according to navigation data, a method and apparatus for recording a data containing navigation data on a recording medium, and a system for transferring data via a communication route on the basis of navigation data
MOTOROLA MOBILITY	1998	1.00	US	26.3	0.87	55	Prediction and coding of bi-directionally predicted video object planes for interlaced digital video
NTT MOBILE COMMUNICATIONS NETWORK	1996	0.09	JP	26.3	1.00	54	Access method, mobile station and base station for CDMA mobile communication system
NOKIA NETWORKS	2007	0.22	FI	26.3	1.00	52	Communication network element and method of switching activity states
SONY CORPORATION	1995	0.07	JP	26.3	0.58	44	Method and device for recording data, data recording medium, and method and device for reproducing data
CANON	1997	0.17	JP	26.5	0.50	43	Remote maintenance system
INVENSYS SYSTEMS	1997	0.20	US	26.3	0.67	41	Methods and apparatus for remote process control
HP (HEWLETT-PACKARD COMPANY)	1997	0.29	US	28.2	0.87	40	Replaceable part with integral memory for usage, calibration and other data
POINTCAST	1996	0.78	US	26.2	0.40	39	Information and advertising distribution system and method
APPLE	2007	0.67	US	26.2	1.00	39	Multimedia communication device with touch screen responsive to gestures for controlling, manipulating and editing of media files
HP (HEWLETT-PACKARD COMPANY)	1997	0.67	US	26.2	0.56	39	Web interfacing device
SONY CORPORATION	1992	0.52	JP	28.2	0.50	38	Disk recording medium and reproducing device therefor
TOSHIBA CORPORATION	1994	0.48	JP	26.3	0.65	38	Video signal compression/decompression device for video disk recording/reproducing apparatus
HITACHI	1996	0.62	JP	26.3	1.00	37	Decoder for compressed and multiplexed video and audio data
NOKIA MOBILE PHONES	1995	0.38	FI	26.3	1.00	37	System for transmitting packet data in digital cellular time division multiple access (TDMA) air interface
TELEFONAKTIEBOLAGET LM ERICSSON	2000	0.29	SE	26.3	1.00	36	A mobile communication network
SIEMENS INFO. & COMM. NETWORKS	1997	0.03	US	26.3	1.00	36	Method and system for increasing quality of service at or below a treshold cost
SWISSCOM	1999	0.86	CH	26.3	0.64	36	Method for purchasing goods or services with a mobile telephone
TOSHIBA CORPORATION	1992	0.81	JP	26.3	1.00	35	Band-compressed signal processing apparatus and VTR
LG ELECTRONICS	1995	0.17	KR	26.3	0.67	35	Illegal view and copy protection method in digital video system and controlling method thereof
TOSHIBA CORPORATION	1996	0.29	JP	26.3	0.62	35	Recording medium, recording apparatus and recording method for recording data into recording medium, and reproducing apparatus and reproducing method for reproducing data from recording medium
TOYOTA MOTOR CORPORATION	1996	0.21	JP	28.1	0.67	34	A method for purifying exhaust gas of a diesel engine
NTT DOCOMO	1997	0.17	JP	26.3	1.00	34	CDMA communication method and group spreading modulator
CUBITAL	1987	0.25	IL	26.2	0.43	34	Three-dimensional modelling apparatus.
FUJITSU	1996	0.78	JP	28.2	0.88	33	Plasma display panel, method of driving the same performing interlaced scanning, and plasma display apparatus
SUN MICROSYSTEMS	1997	0.83	US	26.3	0.75	33	System for context-dependent name resolution
HITACHI	1994	0.32	JP	26.3	0.67	32	Digital video recording device
LIGHTPAT	2003	0.04	DE	23	0.33	32	Glazing comprising a luminous element
NEC CORPORATION	1996	0.04	JP	26.3	0.54	32	Method and system for inserting a spread spectrum watermark into multimedia data

Table 10: Impact of AI Adoption on Labour Productivity Across Firms' Strategies (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
AI	0.062		0.096*	0.109***		
	(0.038)		(0.050)	(0.040)		
AI in-house		0.097*			0.129*	0.145*
		(0.050)			(0.077)	(0.077)
AI joint		0.056				
		(0.084)				
AI outsourced		0.050				
		(0.034)				
AI joint/out			-0.044			0.103**
			(0.052)			(0.043)
Ln(data in-house)				-0.018**	-0.016**	-0.018**
				(0.008)	(0.008)	(0.008)
AI × Ln(data in-house)				0.044***		
				(0.012)		
Ln(data ext)				0.011	0.013*	0.012
				(0.007)	(0.007)	(0.007)
AI in-house × Ln(data in-house)					0.067***	0.071***
					(0.021)	(0.021)
AI in-house × Ln(data ext)					-0.035	-0.032
					(0.022)	(0.022)
AI joint/out × Ln(data in-house)						0.039***
						(0.013)
AI joint/out × Ln(data ext)						0.008
						(0.024)
Ln(labour)	-0.003	-0.003	-0.003	-0.006	-0.004	-0.006
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Ln(intermediate)	0.365***	0.365***	0.365***	0.364***	0.365***	0.364***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Ln(capital)	0.054***	0.054***	0.054***	0.056***	0.056***	0.056***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Dummy capital missing	-0.289***	-0.289***	-0.289***	-0.294***	-0.294***	-0.294***
	(0.041)	(0.041)	(0.041)	(0.043)	(0.043)	(0.043)
Constant	-0.786***	-0.786***	-0.786***	-0.769***	-0.774***	-0.768***
	(0.163)	(0.163)	(0.163)	(0.159)	(0.159)	(0.159)
Observations	4360	4354	4354	4055	4050	4050

Clustered standard errors in parentheses (by 3-digit industry level).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls encompass: location, group, missing answers to data infrastructure dummies