Pillars – Pathways to Inclusive Labour Markets

Report

Taking Stock of the Effects of Technological Change on Labour Markets:

A Systematic Literature Review of the Design of Digital Automation Technologies in the Technical and Engineering Literature

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

PILLARS – Pathways to Inclusive Labour Markets: D1.1

Taking Stock of the Effects of Technological Change on Labour Markets:

A Systematic Literature Review of the Design of Digital Automation Technologies in the Technical and Engineering Literature [1]

Maria Savona[2], Tommaso Ciarli[3], Tommaso Portaluri[4], Ed Steinmueller[5], and Simone Vannuccini[6],

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

[1] We are grateful to Lorraine Mackenzie for the excellent research assistance on the empirical analysis, and to Tomiwa Akin-Onigbinde, Megha Bhattacharyya, Saradha Krishnamoorthy, Jonathan Puerta Pereira, and Alisha Shaik for the excellent research assistance on the coding of the technical and engineering papers. All errors and omissions are responsibility of the authors. We gratefully acknowledge remarks from the participants to the II Biannual Workshop on Technological Change, health, inequality and data for policy evaluation LABOR, Laboratorio Revelli; GEOINNO Conference. Part of this deliverable has been published as a CESifo contribution to the PILLARS Conference Special Issue: The Design of Digital Automation Technologies: Implications for the Future of Work by Savona, Maria; Ciarli, Tommaso; Steinmueller, Ed; Vannuccini, Simone. CESifo Forum; München Vol. 23, Iss. 5, (Sep 2022): 4-10.

[2] SPRU, University of Sussex, (UK) and Dpt. Economics and Finance, Luiss University, (IT)

[3] UNU MERIT (NL) and SPRU, University of Sussex, (UK)

[4] SPRU, University of Sussex, (UK)

[5] SPRU, University of Sussex, (UK)

[6] SPRU, University of Sussex, (UK)

Table of Content

List of Figures

1 Introduction

This paper addresses the core **objectives of WP1 within PILLARS**, that is taking stock of the extant literature on the potential effects of technological change on labour outcomes and represents Deliverable 1.1 (D1.1). We focus on the link between technological change, jobs and tasks.

The paper makes a crucial contribution to existing reviews of the technology-employment nexus, by focusing on the technical and engineering literature, that describes the design of new technologies and how these execute tasks and jobs across industries.

Earlier contributions provide excellent evidence on broadly-defined technologies, such as ICT or robots, or more in general innovation. For example, Léon Ledesma et al. (2010) reviews works that have estimated the **labour elasticity of substitution of technology**. The coefficient of the elasticity of substitution of labour to technology ($\boldsymbol{\sigma}$) is identified for different time periods by relevant scholars in the field. Different assumptions are formulated on the nature of technical change (Hicks Neutral or Factor augmenting) and several range of variation of $\boldsymbol{\sigma}$ are provided. As such, "technological change" is a broad umbrella that encompasses very different technologies.

Building on a wealth of evidence from several OECD countries, Acemoglu and Autor (2011) provide a more nuanced understanding of how (automation) technologies change the composition of labour, either by changing the demand for different skills, or by codifying the knowledge required to perform working tasks, which become routinised, and therefore easy to perform for a machine with little manual intervention. This line of research has introduced the routine-biased approach to the study of the technology-employment nexus (Autor, 2019), which has become the prevalent one in the field.

Taking a broader perspective, Calvino and Virgillito (2018) summarise the literature on the impact of firm innovation, both as automation (process) and as new products, on employment. They outline different compensation mechanisms through which employment adjusts after an innovation is introduced, and illustrate recent empirical evidence on such mechanisms. This exercise is extended in Montobbio et al. (2022), where the authors take stock of (and organise) the last three decades of literature on technology and work distinguishing studies by technology, level of analysis, and empirical methods used.

Unlike this, and the burgeoning number of contributions within the field of labour economics, our objective is to "reverse engineer" the labour elasticity to technology and the impacts summarised in the economic literature by unpacking technology. We aim at identifying the extent to which technologies that are usually combined in a single group, were designed to undergo specific tasks in specific sectors, codifying the knowledge to undergo those tasks. We also assess the extent to which such technologies were designed to substitute or complement specific tasks.

To do so, we start by providing a reasoned and **fine-grained classification** of the emerging technologies included in the current black box of "Automation". This allows us to dig into the idiosyncratic link between **sectoral exposure to** and **adoption of** each of these automation technologies (or a combination thereof) and labour outcomes in the exposed and adopting sectors.

The strategy of **increasing the granularity** of the investigation of both the types of automated devices and processes and the types of tasks (and capabilities) that automation **replaces** on the one hand, and the **reconfiguration** of the tasks affected on the other hand, is the one that we consider the most appropriate and novel in the context of PILLARS. In addition, the adoption of a finer level of granularity in this analysis will also allow us a higher degree of precision in further, complementary analyses of how the labour market outcomes of automation **may be mediated by international trade and structural changes of economies**.

The present effort of systematic review allows us to unpack the relevant factors that contribute to automating jobs and awareness of what specific technologies are affecting labour markets in the context of the latest wave of automation.

Table 1: Empirical Studies of Aggregate Elasticity of Substitution and Technological Change in the US.

				Estimated annual rate of efficiency change		
Study	Sample	Assumption on technological change	Estimated elasticity of substitution $\hat{\sigma}$	Hicks neutral: $\hat{\gamma}_N = \hat{\gamma}_K$	Labor augmenting: $\hat{\gamma}_N$	Capital augmenting: $\hat{\gamma}_K$
Arrow et al. (1961)[a]	1909–1949	Hicks neutral	0.57	1.8	—	—
Kendrick and Sato (1963)[b]	1919–1960	Hicks neutral	0.58	2.1	—	—
Brown and De Cani (1963)[a]	1890–1918	Factor augmenting	0.35	Labor saving ($\gamma_N - \gamma_K = 0.48$)		
	1919–1937		0.08	Labor saving ($\gamma_N - \gamma_K = 0.62$)		
	1938–1958		0.11	Labor saving ($\gamma_N - \gamma_K = 0.36$)		
	1890–1958		0.44	?		
David and van de Klundert (1965)[a]	1899–1960	Factor augmenting	0.32	—	2.2	1.5
Bodkin and Klein (1967)[c]	1909–1949	Hicks neutral	0.5–0.7	1.4–1.5		
Wilkinson (1968)[d]	1899–1953	Factor augmenting	0.5	Labor saving ($\gamma_N - \gamma_K = 0.51$)		
Sato (1970)[a]	1909–1960	Factor augmenting	0.5–0.7	—	2.0	1.0
Panik (1976)[a]	1929–1966	Factor augmenting	0.76	Labor saving ($\gamma_N - \gamma_K = 0.27$)		
Berndt (1976)[e]	1929–1968	Hicks neutral	0.96–1.25	?	—	—
Kalt (1978)[e]	1929–1967	Factor augmenting	0.76	—	2.2	0.01
Antràs (2004)[e]	1948–1998	Hicks neutral	0.94–1.02	1.14	—	—
		Factor augmenting	0.80	Labor saving ($\gamma_N - \gamma_K = 3.15$)		
Klump et al. (2007a)[f]	1953–1998	Factor augmenting	0.51	—	1.7	0.4

Notes: Key to estimation method: a = linear single equation; b = implicit; c = linear and nonlinear single equation; d = linear system; e = linear single equation and linear system, f = nonlinear system (Klump et al. 2007a report both constant and time-varying factor-augmenting growth cases; the results reported in this table are from the former case.) This table is updated from Kalt (1978) and Klump et al. (2007b). A full source of references to the above studies may be found therein.

Source: Leon-Ledesma et al., 2010

In sum, the main aim of this exercise is to perform a literature review that addresses the following research questions:

- Can emerging automation technologies (as identified in this study) potentially substitute, complement, and/or reconfigure specific technical tasks within occupations within sectors, that are executed by humans?

- What are these specific tasks across sectors? Can we derive insights on the extent to which specific industries are exposed to specific automation technologies at higher level of granularity compared to the received literature in economics?
- Can we infer trends and dynamics related to the future of work in relation to the contemporary wave of emerging automation technologies, characterised by complex interdependency and the pervasive presence of artificial intelligence (AI)?

The paper contributes as follows.

First, we **offer a novel classification** of automation technologies;[7]

Second, we explore **the technical and engineering literature** that focuses on each of the classified technologies, describing – among other things – prototypes and mature automation devices; technical characteristics and conditions of adoption; the extent of automation of the production process and success and failure of prototypes, all of which affects occupational tasks and their associated levels of skills, knowledge codification, levels of routinisation, degree of substitution or complementarity with respect to humans, to various extent.

For instance, within the main class of Robots are included machines that contribute to real time monitoring, autonomous driving, process automation, automated platforms; it features semi-autonomous robots, service robots, co-bots, swarm robots. In sum, for the family of robots, the technologies we explore in the technical literature are both what Sheridan (2016, p.525) labels *telerobots*, namely machines "capable of carrying out a limited series of actions automatically, based on a computer program, and capable of sensing its environment and its own joint positions and communicating such information back to a human operator who updates its computer instructions as required", and *teleoperators*, that is machines that "perform manipulation and mobility tasks in the remote physical environment in correspondence to continuous control movements by the remote human." To provide a few examples of the variety of technologies in development or use, in the analysis we reviewed studies on robotic arms, or welding robots, but also robot vehicles for the exploration of underwater mines as well as dismissed nuclear power plants. All of these robots might be adopted and used in applications that vary from bar tendering to very complex surgical operations, within which they replace or complement tasks that might be more or less routinised and involve the use of higher or lower skills or complex knowledge or very tacit knowledge.

Third, the **methodology** of this review allows organising the collected information into an **interpretative grid** that serves as a general scheme to assess the employment impacts of fine-grained-defined technologies, and helps to compare whether this most recent wave of automation

[7] The **classification** produced in this paper will also be used to identify technologies in several other Work Packages in the whole of the project.

is a qualitatively different phenomenon compared to previous waves. For this reason, we frame our analysis into a historical background accounting for the evolution of automation technologies.

The **results** of the systematic technical literature review will guide the interpretation of the economic results throughout PILLARS.

The paper is structured as follows:

- We start by offering a brief historical review of previous waves of technological change based on mechanisation and electrification **(Section 2)**. This allows to contextualise our contribution, which illustrates the processes of automation linked to digitisation in comparison to previous waves. This historical contextualisation underpins the rationale of our proposed classification.
- Next, we lead the reader through the journey that led us to design our original methodological and empirical approaches to address the objectives above **(Section 3 and in much greater detail in the Appendix – Section 8)**. Our methodology represents a new benchmark for a multidisciplinary approach to look at the sectoral exposure and adoption of automation technologies and their impact on the future of work.
- Finally, we proceed with a quantitative and qualitative analysis of the content of the selected technical papers, along a standardised set of dimensions that are relevant to our purposes (Section 4): we focus on which tasks the technologies are designed to perform, and in which sectors; the extent to which the technologies codify knowledge and routinise tasks; the extent to which they substitute or complement existing jobs; and the skills that are needed to deal with them **(Section 5)**. We then summarise the key messages **(Section 6)**.

2 Technical change, automation and employment: A brief historical overview

Contemporary discussions about automation and employment echo a long history of development of labour-saving innovation. This history has unfolded in waves of disruption to existing labour practices and routines and has been accompanied by anxieties and protests (Mokyr et al., 2015).

In previous eras, anxieties and protests declined as new occupations were created and economic growth continued to raise the demand for labour albeit in very different jobs than those that were lost. In every one of these historical occasions of automation anxiety, the claim was made that epochal changes were occurring and that if past incidences of concern were raised, the additional claim was '**this time will be different**.' This history resembles the classical fable of the boy who cried wolf. At its heart, therefore, discussion of modern forms of automation in the form of robots, AI, and other manifestations of information and communication technologies (ICT) pose the question – is this time *really* different, is the wolf going to arrive?

Answering this question requires several empirical assessments. The most important of these is the assessment of **emerging capabilities for labour saving devices** associated with employment. The phrase 'associated with employment' may seem odd, but is an acknowledgement that some labour saving devices are related to activities that are generally not paid – e.g. household appliances. Emerging capabilities are those that extend and deepen the capacity of labour saving devices to substitute for *employed* labour.[8]

A considerable amount of attention has been devoted to **robotics and AI** and this seems to be the result of the proclivity of humans to anthropomorphise such devices – so a robot arm or a decision-making AI algorithm receives greater attention than an automated measuring system for filling containers or a logistical system that monitors the location of parcels in transit. In all these cases, of course, there are implications for human labour, and it is important to consider capabilities broadly to make an accurate assessment of emergent labour-saving solutions.

A second consideration is **the nature of jobs**. Job classifications and designations often reduce the complexity of the tasks that actual workers do. Consequently, the mapping of capabilities to employment may over-estimate the extent that labour-saving technologies may mean job losses. In many cases, deploying labour-saving devices results in the **reconfiguration rather than elimination of jobs** – with some net loss of employment and often an increase in output or productivity (which can then be reflected in lower prices that increase the size of a market or market share). In other words, the net impact on jobs of the introduction of labour-saving devices is

[8] Nonetheless, since almost all household services have paid equivalents (e.g. commercial laundry service is a substitute for home laundry), there is a two-way interaction between innovations in the paid and unpaid version of the services (e.g. home sewing machines are adaptations of commercial sewing machines).

complex and difficult to ascertain *a priori* – it often requires ***ex post* assessment** after allowing some time for adjustments to occur.[9] The complexity of this second assessment is why **greater precision in assessing emerging capabilities** is the most promising empirical strategy.

A third type of empirical assessment that takes a national or regional unit of analysis involves taking account of labour-saving technologies on the **international division of labour** and hence on a country or region's pattern of international trade. In the latter two decades of the 20th century and well into the present century, companies often resolved make or buy decisions by outsourcing production to lower wage countries, off-shoring, that could meet product or service quality standards and an immense infrastructure of international logistics was constructed that facilitated the movement of parts and finished goods, and in some cases, provided the means to deliver services across national or regional boundaries. For example, containerised shipping is not generally considered part of automation, but if one thinks about past arrangements for merchandise trade, there were many labour saving implications of the growth of this transport method. For a variety of reasons including rising wages, the Covid-19 crisis, and perceived inequities to domestic workers in consuming countries, **automation is taken as an opportunity to 're-shore' production**. It is, however, also a means for off-shore producers to improve the productivity of workers and hence to compete with re-shoring trends. Furthermore, emerging automation technologies such as AI have induced new waves of offshoring decisions, this time aimed at externalising to the "global south" data work, such as labelling and moderation (Casilli, 2021). Like the second consideration, this is also a very complex process whose outcome is very difficult to predict. Similarly, however, the examination of labour-saving technologies in terms of **capabilities** is an important first step for an empirical analysis of impact.

To assess the potential for studying improving capabilities, it is helpful to look a bit more closely at the **history of labour-saving innovation**. Along with the periodic eruption of concerns about the automation of jobs, there is a continuing development of the waves of the past (von Tunzelmann, 1995). Thus, **mechanisation** continues to develop and indeed is influenced by **electrification** and what we have come to call **digitisation** influences both the previous waves of development. The interactive and cumulative effects are important. For example, early mechanisation was driven (literally) by **steam power** which dictated that plants would be vertically organised due to the constraints in distributing motive power horizontally. With electrification, the organisation of factories, and the nature of jobs could be transformed, first by extension and development of horizontal '**assembly lines**' and more recently a plethora of different '**work station**' and '**machine cluster**' groupings in which partially finished goods continue to be brought to workers and then transferred for additional work or final packaging.

The basic principles of mass production involve the creation of interchangeable parts that can be assembled rapidly by the movement of these parts to the assembly worker (through an assembly

[9] Part of the problem here is that task efficiency is often easier to estimate than production system efficiency because the latter involves multiple changes that have synergistic effects.

line or other means). Interchangeability requires manufacturing to precise tolerances and rapid movement and assembly operations will be facilitated by flexible arrangements for delivering mechanical power to the worker (e.g. pneumatic tools and motorised conveyors). The technologies for achieving all these functions were developed early in the 20th century and accompanied the surge in development following electrification.

The 20th century is an extended history of **industrial mass production displacing craft production** in which the individual worker does a variety of operations in fabricating and finishing a product. The basic logic of mass production has been extended to operations in **service industries** such as processing payments in banks, the use of a battery of power tools and other equipment in dentistry, or the surgical operating theatre. In many services, there are residual craft elements that continue to rely on the skill of the 'operator.' In some of these cases, the logic of mass production has been reversed so that the customer becomes the operator as is the case with automated teller machines (ATMs) which largely displace the role of bank clerks who previously received and dispensed cash (Savona and Steinmueller, 2013). The ATM operates under the control of the customer rather than the bank employee, substituting customer 'labour' for the labour of the bank clerk.

Mechanisation using electrical motors (hence, **electrification**) was also progressively refined throughout the 20th century. In many cases this involved the redesign of products to make them more compatible with mechanised technique – milk bottles became tetra-packs, butchers parcelled out meat in shrink wrap packaging, and stockings were knit and finished with very little human intervention. The idea of **co-opting customer labour** was further refined by shipping products with 'some assembly required' and Ikea's refinement of 'flat pack' furniture that could be transported by the customer from large retail showroom/warehouses and assembled by the customer at home.

It is important to note that parts assembly-based mass production is not the only means by which industrial progress has been achieved. The evolution of technique stemming from the exploitation of petroleum as a source of fuel and as a feedstock for chemical manufacture involved several innovations that, together, created the petroleum refinery. Petroleum refineries are examples of continuous flow manufacturing in which human intervention is only required to monitor and control the flows of raw petroleum to finished products (e.g. petrol, diesel fuel, asphalt base, fuel oils, heating oil, paraffin and liquefied petroleum gas). In chemical production, a mixture of continuous flow and 'unit process' technologies are employed with the transfer of materials undertaken by pumps or, in some cases, conveyors. These basic operations have been extended to a variety of other industries such as food processing, ore refining, and pharmaceutical production.

What is important to observe about this brief account of technological history is that processes of automation involving **mechanisation** and **electrification** have been underway for an extended time. In many cases, labour saving innovations have already greatly improved individual worker productivity and most of the job losses in global North manufacturing have already occurred, not so much as the result of outsourcing of production, but by the combined influences of mechanisation and electrification.

Digitisation combined with international logistics and transport have continued this process. The ability to codify designs, arrange supply contracts, communicate about production issues, trace and monitor transport of parts and partially finished goods, and efficiently manage inventories in relation to the flow of production and consumption have all been greatly improved by digitisation. This is the result of digital technologies layering up on existing techno-economic systems and lowering a series of costs: search, tracking, replication, verification, and experimentation among others (Goldfarb and Tucker, 2019). These, in turn, lead to reshaping economic actors' incentives and, therefore, to reallocation of working activities.

The uncertainties that now confront economies throughout the world arise from the technological potentials of newer generations of **cyber-physical systems** that have the potential to transform the mass production paradigm, as they are technology systems integrating sophisticated hardware (i.e. sensors) and software (i.e. prediction or scheduling algorithms) to executing functions in a flexible manner. These new potentials involve new physical technologies such as additive manufacturing which is evolving from a substitute for hand machining and other forming processes for the tools and dies involved in mass production toward efficient direct production of final parts and products. If this trajectory were to continue, it is possible to imagine city or even neighbourhood-based '**manufactories'** capable of producing a very wide range of products for business and household use. Competing with this possibility is an adaptation of the '**fabless integrated circuit company**' which produces designs that can be used to create physical products in larger-scale flexible manufacturing facilities. The location of these larger-scale facilities are one of the greatest uncertainties concerning the future of international division of labour, as well as a policy theme. For example, it is possible to imagine an initial period of development in which currently labour-intensive factories in middle income countries are automated with this new generation of technologies. Alternatively, since the new technologies will be **far less labour intensive**, it is possible to imagine **a major trend to reshoring of production**.[10]

In the countries that are currently most wealthy, the absolute dominance of service sector employment raises questions about the role of **new technologies in replacing these service sector jobs**. For example, the historical occupation of data entry operator has experienced dramatic reductions as new technologies for data acquisition including user data entry displace centralised facilities for data entry and filing, creating a 'data cloud' which itself offers a myriad of opportunities for the application of machine-learning AI to create predictive models and, increasingly, to manage data-intensive service provision. As we mentioned, last-century's data entry occupations have been replaced with data labelling and moderation tasks executed thorough microwork platforms, often resorting to cheap labour in quasi-monopsonist settings. In any case, a key challenge is to improve the **human-computer interface** so that opportunities, choices and services can be presented in ways that are customised to the users' needs. The flexibility and scalability of robotic equipment

[10] The recent disturbances in international supply chains associated with the Covid-19 pandemic have provided a motivation for reconsidering the international division of labour and rebalancing off-shoring with new domestic capabilities.

has major implications for employment in ‘customer facing’ jobs, if what is facing the customer is a cyber-physical system rather than a human being. As in past waves of automation, it seems likely that **the initial applications will be specialised to tasks that can be clearly specified** (e.g. picking items from warehouse shelves to fulfil an online order). However, impacts can cascade – in the same example, if packing and delivery also are performed by autonomous systems (e.g. robotic stocking and packing and autonomous vehicle delivery), employment implications become much more significant. Assessing the potential for the emerging new wave of automation, and whether this time it will be different, begins with a **careful assessment of the emergence of new capabilities in the cyber-physical systems that are the current subjects of research, development and initial deployment**.

3 Methodology

This section and the related Appendix (section 8) outline the original methodology we designed for the study. The presentation is framed as a chronological journey that the research team has endeavoured in to identify and refine in-the-making the literature search and selection protocol. This allows us to articulate in detail the rationale underpinning our methodological choices and the solid grounding of the selected technical literature that supports our findings. We trust this to become a methodological template for any in-depth investigation of multidisciplinary literature aimed at shedding light on such a complex phenomenon.

In particular, as we detail in the Appendix, we reconstruct the step-by-step decisions we took to deal with the criteria of classification of digital automation technologies, from the initial scoping to the devising of a final classification that has served as a standard for the whole PILLARS project, particularly WP3 which deals with the new and emerging technologies. We then detail the articulated search strategy of the relevant papers, from how the classification devised informed keywords selection to the scrupulous manual screening of the relevant papers, to the identification of the further relevant keywords that served as controlled terms to expand the initial selection of core papers.

We organize the literature review in **six main steps**, which are summarised in Figure 1 and further detailed in the Appendix (Section 8).

First, given differences among technologies and their applications, we have run a separate review for each different family of emerging automation technologies. We identify eight families (in parenthesis their acronym):

A. Robots (R) **–** *technologies that sense and (autonomously) act based on data*
B. Physical data acquisition technologies (DA) – *technologies that harvest and record information*
C. Software based data management (DM) **–** *technologies for storing, protecting, managing/handling and acquiring data*
D. Computing (C) – *technologies used to compute/calculate*
E. AI & Intelligent Information System (AI) – *technologies using algorithms and advanced methods to make sense out of the data*
F. Additive manufacturing (AM) – *technologies that produce bottom-up based on digital models*
G. Networking (N) – *technologies for communicating between machines (data transmission) or connecting machines*
H. User interface (UI) – *technologies for human interaction with machines or data*

Second, for each of the technology families we identify relevant records in the Scopus database (Figure 1). We use literature and our own expertise to build a seed query to search titles, abstracts and keywords of publications. As we are not interested in technological development per se, but in what technology can do, and in particular what tasks it can perform, we build a three-part query that combines keywords identifying (i) the *technology* (e.g. robot OR human worker), (ii) *its functions and applications* (e.g. process OR routine), and (iii) *the tasks that it can perform* (e.g. interact OR recognize OR weld).[11] We selected only documents published after the year 2000, in the form of original articles, reviews, or conferences papers, and in the top percentiles by citations (by year). The number of percentiles varies across technologies and queries in order to maintain a comparable and manageable number of documents to screen manually (approximately over 500 documents per technology).

Third, to maximise precision in our selection of relevant papers, we manually screen a sample of papers to decide on inclusion in the sample (Figure 1). Each document's title and abstract is screened by two independent reviewers, and conflicting cases decided by a third reviewer. Detailed screening rules are reported in the Appendix, but mainly consist in filtering out papers that are not about the technology of interest, nor about production of goods and services (e.g. house appliances), which are only conceptual, or do not clearly indicate what task the technology performs, even if they clearly describe the skills needed. For example, a paper on robots that can move in a small space, avoiding obstacles, but with no mention of their application to specific tasks (e.g. picking objects from shelves and bringing them to a different point of a warehouse) is excluded.

Fourth, to increase recall, we expand our initial query to documents in Scopus that, although not including in their title and abstract terms considered in our initial query (for example because they refer to tasks that we did not consider initially) may be relevant to the development of the technologies in the family under consideration. We take advantage of the *Engineering Index Thesaurus*, a thesaurus of controlled engineering terms that are manually allocated to documents in Scopus. Using the relevant documents selected after manual filtering (see previous step), we use text mining metrics (TF-IDF) to select from the thesaurus the engineering terms that better identify these relevant papers (and the technologies they illustrate) (Figure 1). We combine engineering terms from the thesaurus that are ubiquitous, i.e. appear in documents related to different technologies (e.g. automation OR intelligent robots), and that relate to distinctive technologies (e.g. crops OR architectural design). We then use these engineering terms to run a second search in Scopus to identify papers that we did not find with our initial query, but which are assigned relevant engineering terms.

Fifth, to maximise precision in our selection of relevant papers, we manually screen this sample of papers – the expansion set – as described in the third step (Figure 1).

Sixth, and finally, as a result of manual screening of both the first (third step above) and expanded (fifth step above) searches, we are left with a number of selected documents per technology family (Figure 1), each referring to one or more technology. In our final step, we read and code the full text

[11] A full list of the detailed queries is available in the Appendix, Table .

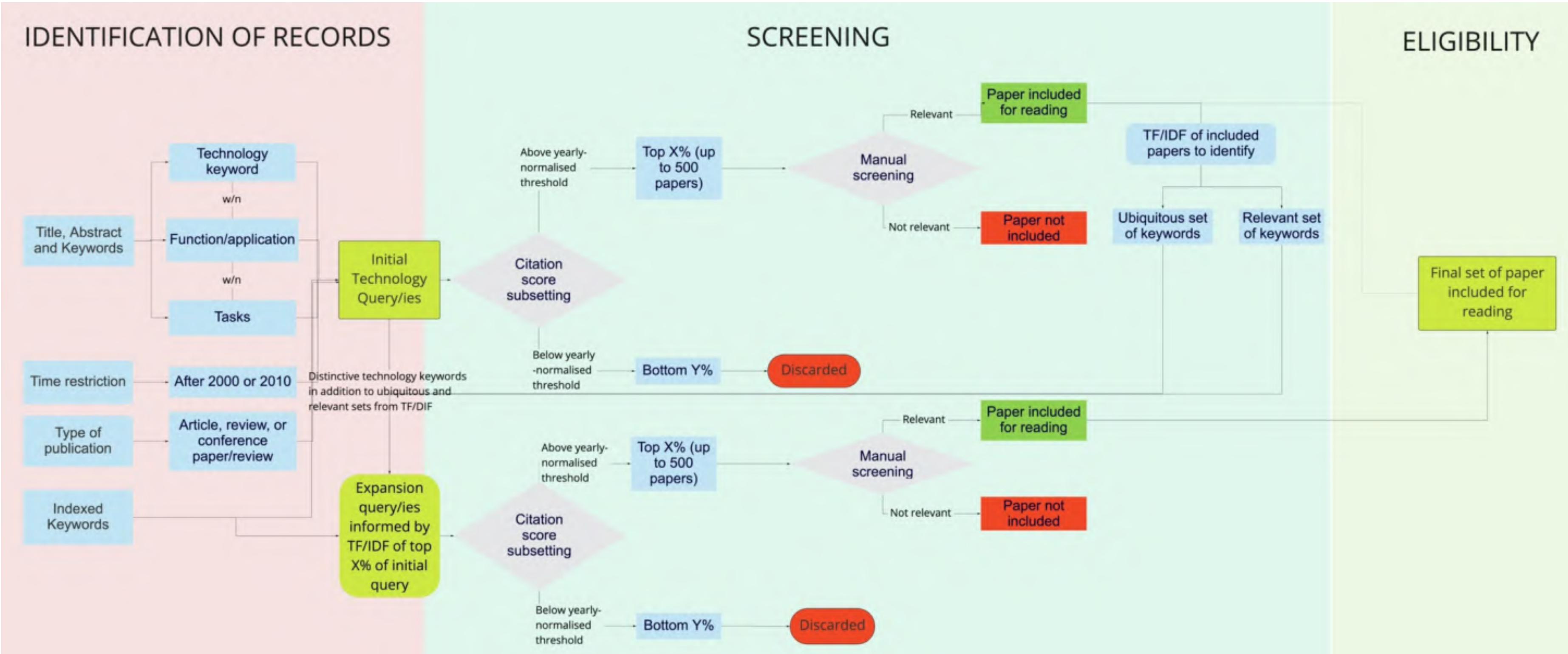
of the selected publications, extracting the following information about the technology, where available:

1. Level of adoption of the technology
2. Development stage of the technology
3. Routinisation: the ability to perform a task without any human intervention
4. Knowledge codification: the ability to make all instruction explicit (i.e. codified)
5. Whether the technology works with people, symbols or objects (based on Reich, 1991)
6. Level of skills required to use the technology
7. Whether the technology substitutes or complements humans
8. Whether the technology improves the product/service, or the process to produce products/services
9. Intended sectors of application
10. Intended tasks that the technology performs
11. Geographical area of development/use
12. Type of organization that is likely to use the technology
13. Size of the organization that is likely to use the technology

We describe all these steps in detail in the Appendix (Section 8).

The reading and coding allow to describe, for each technology family, whether the technologies identified in the literature potentially substitute, complement humans, and/or reconfigure specific technical tasks within occupations and sectors; and what are these specific tasks across sectors, and which are the specific industries are exposed to specific automation technologies and routinisation of tasks. The codification allows us to quantify our discussion of the literature, which provides some qualitative understanding of tasks, sectors, and type of impact that these different technologies may have on employment. When relevant, we split results into two sub-periods, according to the year of publication, before and after 2010. This is expected to give some indications on whether there has been any changes in the intended use and potential impacts of the technologies under analysis

Figure 1: Flow-chart of the literature review process



4 A quantitative and qualitative analysis of technical papers on emerging technologies: the interpretative grid

This section outlines the variables we collected when evaluating and classifying the final sample of records. The analysis organises the collected information into an interpretative grid, with the aim of producing a general scheme to assess impacts of fine-grained-defined technologies. Hence, the results come from the construction of a qualitative typology derived from reading the papers, which can be measured by the relative frequency of each observation across typologies. More precisely we codify the records according to the characteristics detailed below. We then look at frequencies of papers focused on the specific technology family across the relevant *job-related* characteristics.

This type of analysis (qualitative reading and coding in a pre-structured grid) allows us to conduct a quantitative analysis (frequencies of occurrence; cross-tabulations) of the number of records that can be classified along each sub-category, both in terms of row-share of number of records and along the time dimension. The Appendix reports all the categories of classification for each of these characteristics (see tables Table -A and Table -A).

- ***Exposure and adoption:***

1. **Level of Adoption of the technology.** Technologies experience different levels and patterns of adoption, which depend on an array of variables ranging from actors' features, propensities and thresholds, to the structures of interconnections in adopters' networks up to the very technical features of the technologies (Rogers, 2010). In particular emerging technologies (Rotolo et al., 2015), especially when articulated and complex, can show different adoption levels for different sub-technologies. In our work, we classify this variable in low, medium and high, depending on whether the technology is a prototype or displays a level of maturity that is ready to be adopted.
2. **Development (function design) stage.** This can be at the stage of invention; conceptual definition; or, more downstream, at the experimental level (prototype); definition of product; at the stage of ready to deploy; or mature. We use this variable to capture the level of maturity of the technologies (Albert, 2016) and to map in a stylised manner the stages of new product (or process) development (Takeuchi and Nonaka, 1986).

- ***Tasks reconfiguration:***

3. **Task/routinisation.** Technologies execute functions at the level of single operations or activities, with different degrees of autonomy. This ability impacts the very structuring of how

products and services are made, i.e. the 'factory physics' of production (Hopp and Spearman, 2011); the reach of automation and its particular deployment depends on how much given tasks can be routinised (and the relevant knowledge for the activities to be executed codified – see below for another variable capturing this), which in turns also depends on how single operations can be separated or consolidated (Combemale et al., 2021). For our purposes, we classify the technology on the basis of the ability to perform a task without any human intervention, including the possibility that a task is further decomposed into an automated and humanly supervised segments.

4. **Task/knowledge codification.** This variable classifies the technology on the basis of the ability to make all instruction explicit (i.e. codified) without the use of any tacit knowledge. The capability to develop and deal with tacit know-how is a cornerstone of the economics of knowledge (Cohendet and Steinmueller, 2000; Cowan et al., 2000; Foray and Steinmueller, 2003) and so far usually considered a monopoly of human action.
5. **Works with People/Symbols/Objects.** This characteristics is borrowed by the taxonomy suggested by Robert Reich (1991) and considers whether a technology deals mainly with people, integrating the use of objects or targeting the use of symbols. This simple categorisation has a very interesting (potential) explanatory power, as it allows to assess whether recent emerging technologies such as artificial intelligence algorithms, that seem to be lowering prediction costs in tasks usually involved in symbol manipulation, produce a similar impact compared to, for instance, robots.

- ***Compensation mechanisms:***

6. **Skills.** The idea that technologies display different degrees of complementarity with humans along the skills distribution has been widely explored by the literature on skill-biased technological change, and refined by the approach of routine-biased technical change (Autor 2015; 2019; Ciarli et al., 2021; Goldin et al., 2020). With this variable we measure the extent to which the technology may need low, medium or high skills to operate, as described by the authors.
7. **Complement vs. substitute.** Technologies might (traditionally) complement or replace the use of human labour, producing what have been labelled respectively a productivity or substitution effect (Acemoglu and Restrepo, 2019). Recent work has been identifying the conditions and the mechanisms for the decrease of the labour share (hence, the substitution of human labour by capital) both at the aggregated and firm level (Ray and Mookherjee, 2021; Acemoglu et al., 2020). We record papers with a dummy indicator complement or substitute; in addition, we consider the possibility to identify segments of tasks (or sub-tasks) that are replaced and others within the same task that are complemented. In these cases records are identified as both substitute *and* complement.

8. **Time saving or product/process innovation.** The many sorts of automation technologies we cover in the analysis intervene in different areas of firms' activities; some are process improvement aiming at restructuring production, provide efficiency gains such as cost reduction or time saving, others are embedded into the design of new product offerings. We record this information to understand the direction of change within economic actors, loosely referring to Savona and Steinmueller (2013) for services.

- ***Sector, tasks and geographical areas of application:***

9. **Geographical area of provenience.** This codes the technology described in the academic publication on the basis of where the technology has been deployed.[12] Table -A in the Appendix includes all values.
10. **Sector of application.** This variable organises technologies according to the main sector of use based on 3-digit ISIC classification for manufacturing and 2-digit for other sectors; it is usefully cross-referenced with the tasks of application (point 11 below).
11. **Task of application.** This variable classifies technologies on the basis of what work task(s) it is likely to replace or be integrated into. We code these using O*NET aggregate work activities list. Table -A in the Appendix describes all values the variable can take.

The database of technical documents constructed within this Deliverable is very rich in information that can be selectively focused on. However, for the purpose of a readable, digestible and focused discussion, the analysis reported in the remainder of the paper will focus on a subset of the characteristics above, namely **3 and 11 (Task of application and routinization)**[13], **7 and 8 (employment compensation and time saving/process innovation)** and **10 (sector of application)**. We use evidence from the other variables to complement the main discussion on tasks and sectors.

For what concerns the other dimensions, and particularly the **maturity and development stage of the technologies**, the large majority of papers describes robots at the experimental stage (almost two thirds), with only a minority at the mature stage. The prototypes and the conceptual stages have similar shares, interestingly regardless of the time span considered (e.g. before or after 2010).

[12] This dimension is hard to codify as often the place/country of first development or application of the technology is not explicitly mentioned or deducible. When this is the case, the geographical area is attributed through the affiliation of the author. Because of this, we have not taken into consideration this category when looking at the results.

[13] There is a high correlation between routinising the activity and codifying the knowledge to perform a task, so results are similar for variable 4.

This confirms the relevance of our final selection, which privileges technologies which are currently emerging and have not yet been fully deployed. This reflects also the fact that engineering and technical publications over the last two decades are concentrated in novel applications of robotisation, mainly at the experimental stage, which most likely requires the academic community recognition to move forward towards downstream stages of development.

Interestingly, the numbers and shares of papers describing data driven technologies is slightly less concentrated in the experimental category, with a higher presence in the mature stage of application, despite the large majority of papers have been published after 2010. This might be linked to the fact that the adoption and use of data acquisition, management techniques, particularly those that are software-based, have a higher pervasiveness in terms of range and sectors of applications than the robots' applications. This might imply that the degree of novelty of publications is not necessarily correlated to the stage of technology development and the technical features of the robot-related automation, but in the domain of (service) use, rather than the technical feature of the software application used for data management (Saviotti and Metcalfe, 1984).

The large concentration of papers in the experimental stage of technological development might imply an idiosyncratic boundary of the tasks and labour functions actually affected by these technologies, due to their degree of diffusion and sectoral exposure being by definition still embryonal. We will focus on this in the next section.

Some of the technologies will be displaying a low number of contributions, others are more fully explored by engineering and technical scholarly work, and this affects the number of papers that figures in the frequency analysis. This emerges from the tables, nonetheless we provide a meaningful comparison of the above dimensions across all the technologies.

5 Results: Tasks, complementarity and sectoral exposure

5.1 Tasks within work activities and complementarity with humans

We first discuss the focus of the sampled technical literature on specific work activities (tasks) by technology family, and the degree to which the technologies in each family is designed to complement workers. We report results for the variables related to tasks of application (11 as per above) and complementarity (7). We are assessing how technical scholarly work has described how each technology is designed to execute tasks that are categorised according to the O*NET classification (Table 2), and whether such design is aimed at complementing or substituting the participation of humans (Table 3). In particular, Table 2 shows the shares of papers that refer to each of the work activity, for each of the technology families. Work activities are sorted according to the geometric mean over all technology families, reported in the last column. A colour coding helps visualising where the higher shares are across technologies and tasks. The last row indicates in both tables the number of papers coded and the number of observations. The latter might be higher than the number of papers as each paper might refer to more than one task and/or more than one frequency of complementarity or substitution.

The eight technologies differ substantially in the tasks they are designed to execute, particularly between robots and the data technologies (DM, DA, AI, C) – what we label the 'data value chain'. Around 50% of the papers mention that robots carry out tasks related to "Handling and Moving Objects" or "Identifying Objects, Actions, and Events". Technologies pertaining to the data value chain of DA, DM, AI and Computing are similar to each other with respect to the work activities they carry out. However, there are important differences also among those.

Amongst the selected papers, most of the applications of robots relate to production processes, welding, surgery, supporting human mobility (i.e. passive robotics for the elderly), inspecting objects in human non-friendly environments (i.e. underwater ship hulls, mines), moving vehicles (especially in agriculture-related activities) and, indeed, moving objects (from large – i.e. with robotic cranes – to nano – i.e. robotic tweezers to move cells). Instead, DM technologies are developed to organise and process information in different fields, such as scheduling in warehouses and manufacturing pipelines, purely intangible information systems to manage personnel, but also systems to organise workload in scientific laboratories. DM technologies papers also refer to digital assistants as front-end interfaces to provide humans access to information. Papers on DA technologies often display a more engineering flavour, especially as DA technologies are more embedded in devices (e.g. sensors), with several studies exploring new materials to build sensors or fields of application, in particular monitoring at different levels, from infrastructure to floods and spillage down to healthcare applications.

A substantial number of papers in AI refer to "Identifying Objects, Actions, and Events", and "Estimating the Quantifiable Characteristics of Products, Events, or Information", tasks hardly mentioned by papers in DA and DM. However, only a few discuss technologies for "Getting information", which are instead widespread among papers discussing DA and DM technologies and Networking. The highest share of sampled papers describes how these technologies carry our tasks of "Processing Information" for DM, 'Analysing data' for AI and a less complex or analytic task such as "Monitoring Processes and Materials" for DA. The other technologies are related to a more distributed number and type of tasks. This evidence reinforces the suggestion that some of these technology families act in coordination, within more complex technology systems. In particular, it appears clear a "division of labour" between AI technologies, executing prediction tasks, and data acquisition and management technologies, preparing the ground for AI to perform its functions by respectively collecting and organising data.

Networking is very similar to DM, and other data intensive technologies, while Additive Manufacturing shows a hybrid profile, displaying tasks that are common to other data technologies, such as "Getting information" and "Analysing data and information" but it results designed also to execute more material tasks, such as “Controlling Machines and Processes” and “Controlling Machines and Processes”.

None of the automation technologies we study seem (yet) to be executing tasks implying interactions with people, from “Coaching and Developing others” to the management of human resources. However, this could be precisely the result of division of labour in an activity in which humans and technology are complementary. In fact, papers in DM include some which are discussing solutions for information systems in the domain of personnel management – in our analysis, these studies are assigned to tasks of information processing as they aid humans to organise information. “Algorithmic HR” might be captured by AI technologies, but it appear not to be a major emerging technology in our sample.

Overall, in terms of tasks executed, Robots technologies seem to show a quite different profile compared to other technology families, except for the tasks of “Identifying objects, actions and events” which is highly represented also in AI papers. Despite some similarities, unpacking the idiosyncratic characteristics of the technical and design functions underpinning each of the technology family allows understanding how they singularly or jointly execute specific tasks. A cluster of tasks related to *physical* activities are expectedly exclusively linked to robots, such as “Identifying” and “Moving” objects, actions or “performing general physical activities”. Nonetheless, as mentioned, the identification part is also common to AI technologies. Interesting is also what emerges in the data value chain technologies, whose scholarly work describes execution of tasks in sequences, such as “Getting”, “Processing” and “Analysing” information.

In a nutshell, in terms of tasks it is possible to trace differences down to the very technological nature (and trajectories) of software and hardware technologies, with the former progressively executing tasks related to data, prediction, and decision, and the latter evolved in the direction to provide increasing malleability and degrees of freedom in the execution of physical work. Technology families standing in-between this continuum, such as networking and additive manufacturing, tend to be less polarised in the type of tasks executed. This point leads us to a further insight: the distribution of tasks performed tends to be less concentrated the more hybrid the technologies are. Tasks that are overlapping across technologies are often obtained from papers that describe technologies that integrate a core family with features from other families – for example, consider robots augmented by computer vision sensors and algorithms, information systems and databases accessible by handheld devices (which belong to the human interfaces class), digital assistants that merge AI capabilities with data management technologies, or data acquisition-based sensors that expand the skills of surgery robots. All these examples suggest that further advances in the convergence of our technologies into more sophisticated cyber-physical technology systems might extend the task coverage for each family.

In what follows, we go beyond idiosyncrasies and commonalities across the eight technologies in the specific tasks they execute, and look at the complementarities and substitution with human activity (Table 3).

A low share of robot technologies mentioned in at least 5% of the coded papers are designed to complement humans. Around 50% of the coded papers mention that robots that carry out tasks related to “Handling and Moving Objects” or “Identifying Objects, Actions, and Events” will **complement** workers.

Most of the remaining papers discuss activities designed to **replace** workers, with a small share (approx. 15%) designed to **both complement and substitute workers**, for instance in the case of automation that requires human supervision. Only a limited share (20-30%) of papers on robots unveils a design that complements human workers for tasks such as “Controlling Machines and Processes” or “Performing General Physical Activities”. In the case of “Controlling Machines and Processes”, though, 50% of the papers mention both complementing and substituting, while in the case of “Performing General Physical Activities” the combination of complementing and substituting is mentioned in only approximately 15% of the papers. This suggests that different cohorts of robots with different degrees of capabilities co-exist, with some only improving efficiency and facilitating workers’ operation, while others fully automate processes, for example by opening the way for flexible factory-floors with reconfigurable assembly systems (Kousy et al, 2018).

Unlike robots, more than 80 of the papers that discuss the data value chains – DM, AI and DA and Computing – suggest that these data-intensive technologies **complement** human workers in all main work activities that they carry out, such as ‘Analysing Data or Information’, ‘Processing Information’ and ‘Getting Information’. There are some exceptions, such as DA technologies related to ‘Inspecting Equipment, Structures, or Material’: in this case, 60% of the papers suggest that DA complement workers, with the remaining share of the papers suggesting substitution, as in the data filler operator example mentioned earlier on.

In the case of Additive Manufacturing, Networking and User Interface, a significant share of papers show that they are complementing humans for tasks such as to “Controlling Machine and Processes” and “Monitoring processes”, revealing that their design is still very much functional (and most likely dependant on) the intervention of humans. “Identifying object” too seems to be a task that is executed by these technology and highly complementing the human factor. [14]

In sum, the data value chain technologies (Acquisition, Management, AI and Computing) share a high degree of complementarity with humans, which are the repository of the tacit knowledge needed to complement automated and routinised data acquisition and processing. Human knowledge and activities in these tasks acts as an enabler or, better, an essential factor. Even tasks

[14] The highest shares of 100% usually reveal that the technology described in the paper describes a process that is complementing a single task.

usually considered as conquest ground for automated systems, such as scheduling work and activities, seem to be grounded by the importance of non-codified knowledge, which acts as a 'reverse salient' – a front of activity that generates bottlenecks.

The technologies that, even visually with a higher number of cells in green, show a low-medium shares of paper describing complementarity with humans are certainly Robots, though Networking and User Interface complement humans in a lower number of cases related to the "Monitoring processes" and "Judging the value of things".

Following Agrawal et al. (2019), labour is replaced when capital substitutes prediction and decision tasks. When automating prediction-related activities raises instead the returns of employing humans in decision tasks, we should expect complementary relationship between technology and labour. This is precisely what we find in the case of the data value chain technologies, that lowering the cost of collecting, processing, and using data to predict (or, more generally, to offer insights and analytics), increase the value of human labour, still relevant to take decisions. In the case of robots (and at a limited extent also additive manufacturing), automating physical tasks might spill-over to decision making tasks as well – as in the case of autonomous vehicles in unfriendly environments – reinforcing the substitution effect.

Table 2 – Share of papers describing tasks executed within work activities by technology family.

	Share of Papers By Work Activity								
TASKS	**Robots**	**Software Data Management**	**Data Aquisiton**	**Computing**	**AI**	**Additive Manufacturign**	**Networking Table**	**User Interface**	**GeoMean**
Processing Information	2.0%	29.7%	21.9%	19.7%	15.0%	7.2%	10.6%	15.7%	12.1%
Monitor Processes, Materials, or Surroundings	3.8%	4.1%	28.3%	8.9%	6.5%	16.7%	22.9%	13.4%	10.3%
Analyzing Data or Information	3.1%	19.7%	17.0%	15.4%	30.9%	0.5%	16.3%	14.9%	9.3%
Identifying Objects, Actions, and Events	18.4%	1.6%	10.3%	5.5%	17.9%	5.4%	5.8%	11.9%	7.6%
Getting Information	7.5%	15.9%	13.8%	4.9%	2.1%	1.4%	16.1%	10.9%	6.7%
Controlling Machines and Processes	3.8%	0.3%	1.3%	2.8%		16.7%	4.0%	4.8%	2.8%
Estimating the Quantifiable Character of Products, Events or Information	0.3%	0.3%	7.1%	5.5%	12.9%	11.3%	1.3%	3.1%	2.6%
Making Decisions and Solving Probelems	1.7%	2.5%	1.9%	7.1%	2.5%		5.5%	0.6%	2.4%
Assisting and Caring for Others	6.5%							0.8%	2.3%
Monitoring and Controlling Resources	0.3%	1.9%	3.5%	11.4%	0.8%	5.4%	3.0%	0.8%	2.0%
Perfroming General Physcial Activties	9.9%	1.3%			0.6%	9.5%	0.3%	1.9%	1.8%
Scheduling Work and Activities	1.0%	8.1%	0.3%	6.5%	0.4%		4.0%		1.7%
Handling and Moving Objects	21.5%			0.3%		2.3%	0.3%	2.3%	1.5%
Judging the Qualities of Things, Services, or People	1.0%	1.6%	0.6%	1.5%	5.6%	2.7%	0.5%	1.7%	1.5%
Inspecting Equipment, Structures, or Material	6.1%	0.9%	5.1%	0.3%	1.5%	7.7%	0.3%	0.4%	1.4%
Interacting with Computers	0.7%	1.3%		3.7%	0.2%	1.4%	1.5%	6.7%	1.4%
Operating Vehicles, Mechanized Devices, or Equipment	3.1%			0.3%		9.0%	0.5%	1.0%	1.4%
Documenting/Recording Information	1.7%	5.6%	2.3%	0.9%	0.2%		2.5%	0.8%	1.3%
Updating and Using Relevant Knowledge			1.0%	0.3%	0.4%	1.8%	2.5%	2.5%	1.0%
Developing Objectives and Strategies	1.4%			1.8%		0.5%	1.0%	1.0%	1.0%
Evaluating Information to Determine Compliance with Standards	0.3%	0.6%	1.9%	1.2%	1.2%	1.4%	0.8%	1.0%	0.9%
Training and Teaching Others	1.0%				0.4%			2.1%	0.9%
Organizing, Planning, and Prioritizing	1.0%	1.9%		1.5%				0.2%	0.9%
Interpreting the Meaning of Information	1.4%	0.9%		0.3%	0.6%			0.2%	0.5%
Communicating with Supervisors, Peers, or Subordinates	0.7%	0.3%							0.5%
Establishing Information to Determine…	0.3%								0.3%
Performing for or Working Directly with the Public	0.3%		0.3%						0.3%
Performing Administative Activites	0.3%	0.6%			0.2%		0.3%		0.3%
Staffing Organizational Units		0.3%							0.3%
Communicating with Persons Outside the Organisation				0.3%					0.3%
Coaching and Developing Others	0.3%				0.2%			0.2%	0.2%
Repairing and Maintaining Mechanical						0.5%	0.3%	0.2%	0.3%
Provide Consultation and Advice to Others	0.3%	0.3%						0.2%	0.3%
Selling Or Influencing Others		0.3%			0.2%			0.2%	0.2%
Total Observations	293	320	311	325	521	221	398	477	
Total Papers	**154**	**122**	**177**	**162**	**251**	**134**	**156**	**184**	

Notes: the table reports the share of papers that were coded as describing tasks related to each of the O*NET broad work activities under column (1), for each of the following families of technologies: robots (2), Software based data management (3), Physical data acquisition technologies (4), Computing (5), AI & Intelligent Information System (6), Additive Manufacturing (7), Networking (8) and User Interface (9). The final row reports the total number of papers that were coded in relation to each work activity: one paper can refer to more than one work activity, therefore the number of work activities is larger than the number of papers.

Table 3 – Share of papers describing the degree of complementarity of tasks with human workers by technology family.

	Complementing Workers								
TASKS	Robots	Software Data Management	Data Aquisiton	Computing	AI	Additive Manufacturign	Networking Table	User Interface	Mean
Updating and Using Relevant Knowledge							90%	100%	95%
Estimating the Quantifiable Character of Products, Events or Inform			86%	89%	97%	96%	100%	100%	95%
Evaluating Information to Determine Compliance with Standards			100%		100%			80%	93%
Scheduling Work and Activities		69%		100%			94%		87%
Identifying Objects, Actions, and Events	61%	100%	81%	72%	92%	92%	87%	95%	84%
Interacting with Computers				58%			100%	100%	84%
Analyzing Data or Information	44%	89%	96%	88%	94%		92%	86%	82%
Getting Information	59%	90%	91%	63%	91%		91%	98%	82%
Processing Information	33%	88%	87%	84%	90%	94%	98%	100%	81%
Training and Teaching Others								80%	80%
Monitor Processes, Materials, or Surroundings	36%	69%	80%	86%	100%	78%	95%	95%	77%
Operating Vehicles, Mechanized Devices, or Equipment	67%					85%		80%	77%
Handling and Moving Objects	49%					100%		91%	76%
Developing Objectives and Strategies				50%				100%	71%
Monitoring and Controlling Resources		33%	64%	81%		100%	100%		70%
Documenting/Recording Information	40%	89%	100%				60%		68%
Controlling Machines and Processes	36%			56%		89%	81%	96%	67%
Inspecting Equipment, Structures, or Material	28%		63%		100%	94%			64%
Organizing, Planning, and Prioritizing		50%		80%					63%
Perfroming General Physcial Activties	24%					95%		100%	61%
Assisting and Caring for Others	58%								58%
Making Decisions and Solving Probelems	40%	50%	33%	74%	100%		45%		53%
Judging the Qualities of Things, Services, or People		100%		20%	79%	100%		25%	52%
Coaching and Developing Others									
Communicating with Persons Outside the Organisation									
Communicating with Supervisors, Peers, or Subordinates									
Establishing Information to Determine...									
Interpreting the Meaning of Information									
Performing Administative Activites									
Performing for or Working Directly with the Public									
Provide Consultation and Advice to Others									
Repairing and Maintaining Mechanical									
Selling Or Influencing Others									
Staffing Organizational Units									
Total Observations	293	320	363	325	521	221	398	477	
Total Papers	**154**	**122**	**177**	**162**	**251**	**134**	**156**	**184**	

Notes: the table reports the share of papers that suggest that the technology complements workers, related to each of the O*NET broad work activities under column (1), for each of the following families of technologies: robots (2), Software based data management (3), Physical data acquisition technologies (4), Computing (5), AI & Intelligent Information System (6), Additive Manufacturing (7), Networking (8) and User Interface (9). The final row reports the total number of papers that were coded in relation to each work activity: one paper can refer to more than one work activity, therefore the number of work activities is larger than the number of papers.

5.2 Sectoral exposure, process innovation and routinisation

Table 4 shows the share of papers that discuss technologies related to specific sectors for each technology family. Table 5 reports the share of papers that describes, within each sector, whether and how technologies improve efficiency, as opposed to those that improve the quality of the good/service. Table 6 shows the share of papers that describes, within each sector, whether and how technologies routinise activities.

Considering the sectors mentioned in at least 5% of the publications, the academic literature focuses on a few, recurrent sectors across technologies. While different technology families apply to several work activities, they are all relevant only for a small subset of sectors. The most common across technologies is "Professional, scientific and technical activities (M)", particularly DA, DM, and User Interface. This feeds R&D activities that allow prototypes, technical design, and subsequent deployment. Beyond "Professional, scientific and technical activities (M)" (21% of the papers across all technologies), there are important differences across technology families. While Robots and Additive Manufacturing focus on Manufacturing (C), tasks related to "Analysing Data or Information", executed by DM, AI, DA and Computing are discussed in relation to "Information and communication (J)", while 'Human health and social work activities (Q)' show a high concentration of papers focusing on AI (39%), Additive Manufacturing (37%) and User Interface (33%), and on average 16% of all papers across technologies. The number of papers concentrating on applications of these technologies in the realm of Public Health (rather than social work activities) is large, fortunately. The sector focus is similar also for tasks related to "Processing Information", carried out by DM, AI, Computing, DA and Networking. Highly human interaction-intensive and creative services as in Arts, entertainment and recreation, do not seem to be the focus of papers concentrating on any of the digital automation and data driven technologies.

As we could identify the sectors that are most spread across technologies, we can also explore the spread of each technology across sectors. A simple way to do that is to calculate a concentration metric for each column of Table 4. We compute the Hirschman-Herfindal Index (HHI) and find that the most concentrated technology (HHI=0.34) is Computing, largely specific to the information and communication sector, followed by additive manufacturing (HHI=0.31), which concentrates in manufacturing and human health and social work activities. This evidence needs to be interpreted with caution: for example, it is well known that computing technologies are ubiquitous and the engine of the current techno-economic paradigm (Lombardi and Vannuccini, 2022). What we identify here is the field of application of ongoing technological developments, which likely contribute to intra-sector advances before spurring economy-wide consequences. In the case of additive manufacturing, instead, the evidence point at an interesting trend – the use of this technology family to print devices used in the medical field (i.e. "smart" skin patches).

Each sector shows a varying degree of task concentration. The uniqueness of tasks within a sector can be considered as a rough indicator of the *degree of exposure of a sector* to fully-fledged automation across the task-spectrum, which should occur in particular when the tasks are routinised. Conversely, a low concentration of tasks within each sector means that a higher number of tasks remain executed by humans, either because their specific mode of execution in a particular sector is not prone to automation or because routinisation and knowledge codification has not yet progressed enough in the context of that sector-task tandem. In sectors with lower robot-powered task concentration, existing operational practices, barriers or bottlenecks might provide protection from the risk of being fully exposed to automation-related substitution.

In general, there is a large gap in terms of task concentration (uniqueness in specific sectors) between physical and non-physical tasks, most likely concentrated in several manufacturing sectors rather than in a few services. The non-physical ones – such as administration, information processing, scheduling activities - are comparatively less diffused.

Table 5 and 6 show large differences in how the technologies concentrated in different sectors serve the purpose of increasing efficiency via process innovation or higher routinisation of tasks: we find a larger focus on improving efficiency in AI, DA and Networking papers, whilst Robots, DM and Computing place a comparatively stronger focus on improving the product or service. Some of the technologies, as referring to specific sectors, did not show relevant information on process innovation.

Interestingly, Table 6 shows that a considerable share of papers describes AI and DA as routinising tasks in most personal services (Accommodation and Food (I) Administrative support (N), Real Estate (N) and Finance (K)), which are the most pervasively exposed to data-intensive technologies. Despite the differences in relation to complementing labour, it is interesting to note that robots and DM technologies have a lower tendency to mention the routinisation of activities than AI and DA. This suggests that, although they do not substitute workers, these technologies are able to make these tasks highly replicable.

The share of papers that indicate a substantial routinisation role are concentrated in DA, DM, AI and Computing, the most data driven technologies, and mainly in the Information and Communication sector.

In relation to routinisation and knowledge codification, we also look at a further dimension that qualifies the above dimensions. We classify tasks within occupations on the basis of how the use of **technologies involves interactions mainly with people, things or symbols**. This characteristics is borrowed from an interesting - and fairly under-estimated – contribution by Robert Reich in his book “The Work of Nations. Preparing Ourselves for 21st Century Capitalism” (Reich, 1991). Quite evocatively, Reich considers the North American labour markets at the dawn of the 1990s, and divides jobs into three categories. The “symbolic analytic” services, the “routine production” services, and the “in person” services.

According to Reich (1991), the Symbolic Analysists are workers that mainly interact with technologies through symbols, and hence include, among other, what has been alternatively classified as “Knowledge Intensive Business Services”: engineers, lawyers, scientists, academics, consultants and other intellectual activities. The routine and the in-person services respectively interact with technologies through Things –performing what has later been considered routinised tasks (as assembly workers, data processors, machine supervisors) – and through People, as in personal services, such as care workers, essential services, janitors and so on.

We predict that this classification will offer a good explanatory power in making sense of how the automation technologies considered here will reconfigure the tasks in terms of human-machine composition.

Currently, technical papers show that the non-automated, non-routinised and complex tasks are less likely to interact mainly with things or people, unlike routinised ones, which tend instead to involve interactions with things.

For instance, in the case of robots, technical papers published between 2000 and 2020 describe prevalently robotisation processes that affect a very few number of tasks, which tend to mainly target interactions with things and people. Robots interacting with symbols belong to the future and, likely, to the interaction of symbol manipulation with more complex technology systems integrating both physical and data processing and analysis capabilities.

Data driven technologies are described in the literature as having a high number of interactions with symbols, rather than people or things, or a combination of symbols and things. A joint interactions with people things and symbols is infrequent across all technologies.

Table 4 – Share of papers describing the sector of adoption by technology family.

		Shares of papers by sector								
NACE	**Labels**	**Robots**	**Software Data Management**	**Data Aquisiton**	**Computing**	**AI**	**Additive Manufacturign**	**Networking Table**	**User Interface**	**Mean**
A	Agriculture, forestry and fishing	7%	3%	10%	1%	4%		12%		5%
B	Mining and quarrying	2%	1%	0%		1%				1%
C	Manufacturing	26%	13%	12%	8%	6%	37%	9%	12%	13%
D	Electricity, gas, steam and air conditioning supply	1%	3%	1%	1%	3%		3%		2%
E	Water supply; sewerage, waste management and remedia	2%	1%	2%	0%	5%		6%	0%	2%
F	Construction	3%	5%	4%	1%	4%	6%	3%	2%	3%
G	Wholesale and retail trade; repair of motor vehicles and m		1%					0%	0%	1%
H	Transportation and storage	5%	3%	3%	2%	3%		3%	0%	2%
I	Accommodation and food service activities	3%		1%		0%	1%	0%	0%	1%
J	Information and communication	1%	25%	8%	49%	6%	1%	18%	16%	8%
K	Financial and insurance activities		2%		0%	3%				1%
L	Real estate activities		1%			1%		2%		1%
M	Professional, scientific and technical activities	25%	29%	40%	16%	17%	16%	14%	26%	21%
N	Administrative and support service activities	6%	3%	1%	1%	1%		3%	2%	2%
O	Public administration and defence; compulsory social secu		1%		3%	1%		3%	0%	1%
P	Education					1%			2%	1%
Q	Human health and social work activities	10%	5%	11%	13%	39%	37%	16%	33%	16%
R	Arts, entertainment and recreation	3%				0%		0%	5%	1%
S	Other service activities	5%	1%			0%		0%		1%
T	Activities of households as employers; undifferentiated go	1%								1%
U	Activities of extraterritorial organisations and bodies									
NA			5%	5%	4%	3%	3%	6%	0%	3%
Total Papers		**154**	**122**	**177**	**162**	**251**	**134**	**156**	**184**	
Total Observations		187	153	270	233	313	174	210	237	

Notes: The table reports the share of papers that were coded as describing industries related to each NACE sector (column 1), for each of the following families of technologies: robots (2), Software based data management (3), Physical data acquisition technologies (4), Computing (5), AI & Intelligent Information System (6), Additive Manufacturing (7), Networking (8) and User Interface (9). The final row reports the total number of papers that were coded in relation to each sector: one paper can refer to more than one sector, therefore the number of observations is larger than the number of papers.

Table 5 – Share of papers describing the process versus quality improvement by sector and technology family.

		Process Improvement Shares								
NACE	Labels	Robots	Software Data Management	Data Aquisiton	Computing	AI	Additive Manufacturign	Networking Table	User Interface	Mean
A	Agriculture, forestry and fishing	86%		114%		100%		100%		99%
B	Mining and quarrying									#NUM!
C	Manufacturing	86%	80%	124%	132%	156%	106%	117%	96%	110%
D	Electricity, gas, steam and air conditioning supply					100%		100%		100%
E	Water supply; sewerage, waste management and remediation activit			133%		112%		108%		117%
F	Construction	67%	29%	109%		164%	140%	100%		88%
G	Wholesale and retail trade; repair of motor vehicles and motorcycles									
H	Transportation and storage	67%		100%		90%		83%		84%
I	Accommodation and food service activities	80%								80%
J	Information and communication		76%	95%	123%	110%		116%	95%	101%
K	Financial and insurance activities					100%				100%
L	Real estate activities									
M	Professional, scientific and technical activities	66%	89%	105%	38%	98%	70%	100%	79%	77%
N	Administrative and support service activities	82%						86%	100%	89%
O	Public administration and defence; compulsory social security				67%			83%		75%
P	Education									
Q	Human health and social work activities	78%	71%	84%	90%	98%	72%	91%	71%	81%
R	Arts, entertainment and recreation	60%							100%	77%
S	Other service activities	60%								60%
T	Activities of households as employers; undifferentiated goods- and se									
U	Activities of extraterritorial organisations and bodies									
NA			114%	107%	78%	100%	83%	92%		95%

Notes: The table reports the share of papers, for each technology family and sector, which suggests that the technology improves the efficiency in producing the good/service, as opposed to improving their quality. The final row reports the total number of papers that were coded in relation to each sector: one paper can refer to more than one sector, therefore the number of observations is larger than the number of papers

Table 6 – Share of papers describing the routinization of tasks by sector and technology family.

		Routinization Shares								
NACE	Labels	Robots	Software Data Management	Data Aquisiton	Computing	AI	Additive Manufacturign	Networking Table	User Interface	Mean
A	Agriculture, forestry and fishing	100%	25%	111%	100%	86%		100%		79%
B	Mining and quarrying	33%		100%		125%				75%
C	Manufacturing	69%	40%	70%	111%	156%	100%	117%	93%	88%
D	Electricity, gas, steam and air co		75%	100%	100%	90%		100%		92%
E	Water supply; sewerage, waste	75%		133%	100%	94%		108%	100%	100%
F	Construction	50%	43%	109%	50%	157%	100%	100%	125%	83%
G	Wholesale and retail trade; repa		100%					100%	100%	100%
H	Transportation and storage	67%	75%	86%	75%	90%		100%	100%	84%
I	Accommodation and food servic	100%		75%		100%	100%	100%	100%	95%
J	Information and communication	50%	79%	95%	115%	115%		118%	113%	94%
K	Financial and insurance activitie		33%		100%	78%				64%
L	Real estate activities		100%			100%		100%		100%
M	Professional, scientific and techn	34%	43%	106%	76%	83%	41%	97%	95%	66%
N	Administrative and support servi	64%	75%	50%	67%	100%		100%	80%	74%
O	Public administration and defen		50%		83%	100%		100%	100%	84%
P	Education					75%			75%	75%
Q	Human health and social work a	83%	71%	81%	81%	82%	62%	103%	84%	80%
R	Arts, entertainment and recreati	20%				100%		100%	100%	67%
S	Other service activities	60%	100%			100%		100%		88%
T	Activities of households as empl	50%								50%
U	Activities of extraterritorial orga									
NA			86%	100%	67%	100%	83%	92%	100%	89%

Notes: The table reports the share of papers, for each technology family and sector, which suggest that the technology allows to routinise the task on which they focus. The final row reports the total number of papers that were coded in relation to each sector: one paper can refer to more than one sector, therefore the number of observations is larger than the number of papers.

6 Key messages and final remarks

This paper identifies, selects, and reviews a large sample of academic papers from engineering and technology-related disciplines, and presents and discusses eight different families of automation technologies that execute or complement tasks across different sectors. Our proposed classification of the technologies allows to offer evidence on a much more granular representation of digital automation than received studies in economics and social sciences.

Therefore, this study is able to provide an articulated understanding of how the technical design of automation technologies that have been emerging since the early 2000s may affect different aspects of employment, exploiting first-hand information from scholarly work by technology designers and developers.

One of the main contributions of this paper is a methodological benchmark to approach from a multidisciplinary angle a literature review of technical papers that focus on the conception, experimentation and development of emerging technologies. We offer a detailed reconstruction of the methodological journey that has led to a novel classification of emerging automation technologies, and a grounded multi-steps literature review of technical records that are most relevant for our purpose of fine graining technologies and tasks affected. This is included in Section 3 and in a detailed Appendix in Section 8.

We then provide a quantitative descriptive analysis of the qualitative codification of the papers extracted, focusing on the relevant employment-related variables illustrated at length in Section 4.

Overall, the analysis of the technical literature provides a rich zoomed-in picture of the economic evidence, which usually clusters automation in one single technology with either a substitution effect on routinised tasks requiring codified knowledge or a productivity effect resulting from labour compensation mechanisms.

We summarise the key messages below.

First, automation technologies, including within the same family, are fundamentally different in their design and the tasks they can execute. These tasks tend to be specific to one sector, but often extend to several sectors (such as analysing data or information) and might have different impacts. As highlighted in the introduction to this paper, the objective of unpacking different technologies is important and is so also in view of predicting the next trends of the future of work when the range of automation technologies changes shape, composition and effects over time.

Second, the number of sectors that attract the development of most automation technologies is still relatively limited, but expanding. From this type of work, policymakers can form expectations about what occupations and industries are more likely to be affected by digital automation technologies in the future. For instance, Software-based data management but more in general data-intensive technologies, hence also AI and Data acquisition technologies, are more pervasive in services than in manufacturing sectors, though we expect new applications to emerge, particularly as a result of the novel regulatory framework emerging at the EU levels, such as the Digital Market Act, the Digital Service Act and the Data Act. This calls for policy to extend its focus from robots to other, more pervasive, forms of automation.

Third, automation related to robotisation is likely to become more and more substitutive of tasks performed by humans, notwithstanding some of these technologies are at the very experimental stage, as the literature on robots published post 2010 shows that they tend to become more substitutive than they were at the beginning of the years 2000s. In contrast, so far data-intensive technologies are consistently more complementary to tasks performed by humans. As it turns out, this is driven by the type of service produced, which is an input to other activities, rather than by the inability of routinising tasks, which is also higher for data-intensive technologies.

Fourth, the use of codified or tacit knowledge is fairly associated with routinisation, whereby the most routinised tasks performed by these technologies seem to make use of codified more than tacit knowledge. Also, data intensive technologies interact largely more with symbols, rather than with things or people. In fact, the technologies included in this category emerge as covering tasks of processing and analysing information, unlike robotisation-based automation is described as mainly interacting with things or people, in this latter case when employed to automate processes that are supervised or managed by humans.

Finally, the future of work depends on technologies' evolution, their idiosyncrasies, their stage of development and adoption, the specific sectors that are mostly exposed to each of them, the specific tasks they complement or replace, and finally their convergence towards complex technology systems integrating both software and hardware components. Labour market policies should rely on evidence on digital automation at a greater level of granularity to be properly informed about their heterogeneous effects on tasks reconfiguration within sectors.

7 References

Acemoglu, D. & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. in Handbook of Labor Economics (eds. Card, D. & Ashenfelter, O.) 4, 1043–1171 (Elsevier, 2011).

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives, 33(2), 3-30.

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020). Competing with robots: Firm-level evidence from france. In *AEA Papers and Proceedings* (Vol. 110, pp. 383-88).

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, *33*(2), 31-50.

Albert, T. (2016). Measuring Technology Maturity: Theoretical Aspects. In Measuring Technology Maturity (pp. 9-113). Springer Gabler, Wiesbaden.

Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of economic perspectives, 29(3), 3-30.

Autor, D. (2019). Work of the Past, Work of the Future (No. w25588). National Bureau of Economic Research.

Bakhshi, H., Downing, J. M., Osborne, M. A., & Schneider, P. (2017). *The future of skills: Employment in 2030*. Pearson.

Balland, P. A., & Boschma, R. (2021). Mapping the potentials of regions in Europe to contribute to new knowledge production in Industry 4.0 technologies. *Regional Studies*, 1-15.

Benmelech, E., & Zator, M. (2022). *Robots and Firm Investment* (No. w29676). National Bureau of Economic Research.

Bresnahan, T., & Yin, P. L. (2017). Adoption of new information and communications technologies in the workplace today. *Innovation policy and the economy*, *17*(1), 95-124.

Casilli, A. A. (2021). Waiting for robots: the ever-elusive myth of automation and the global exploitation of digital labor. *Sociologias*, *23*, 112-133.

Calvino, F. & Virgillito, M. E. (2018). The innovation-employment nexus: A critical survey of theory and empirics. J. Econ. Surv. 32, 83–117.

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

Cohendet, P., & Edward Steinmueller, W. (2000). The codification of knowledge: a conceptual and empirical exploration. Industrial and corporate change, 9(2), 195-209.

Combemale, C., Whitefoot, K. S., Ales, L., & Fuchs, E. R. (2021). Not all technological change is equal: how the separability of tasks mediates the effect of technology change on skill demand. Industrial and Corporate Change, 30(6), 1361-1387.

Cowan, R., David, P. A., & Foray, D. (2000). The explicit economics of knowledge codification and tacitness. Industrial and corporate change, 9(2), 211-253.

Foray, D., & Steinmueller, W. E. (2003). The economics of knowledge reproduction by inscription. Industrial and corporate change, 12(2), 299-319.

Goldin, C., & Katz, L. F. (2020). Extending the race between education and technology. In AEA Papers and Proceedings (Vol. 110, pp. 347-51).

Hopp, W. J., & Spearman, M. L. (2011). Factory physics. Waveland Press.

IBM Classification of 4.0 technologies, accessible at https://www.ibm.com/topics/industry-4-0 (last access: March 14, 2022)

Inaba, T. and M. Squicciarini (2017), "ICT: A new taxonomy based on the international patent classification", OECD Science, Technology and Industry Working Papers, No. 2017/01, OECD Publishing, Paris, https://doi.org/10.1787/ab16c396-en.

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3-43.

Keisner, C.A. Raffo, J., & Wunsch-Vincent, S. (2015), *Breakthrough technologies – Robotics, innovation and intellectual property*, Economic Research Working Paper No. 30. WIPO

Kousi, N., Michalos, G., Aivaliotis, S., & Makris, S. (2018). *An outlook on future assembly systems introducing robotic mobile dual arm workers*. Procedia CIRP, 72, 33-38.

Kousi, N., Michalos, G., Aivaliotis, S., & Makris, S. (2018). An outlook on future assembly systems introducing robotic mobile dual arm workers. *Procedia CIRP*, *72*, 33-38.

León-Ledesma, M. A., McAdam, P., & Willman, A. (2010). *Identifying the elasticity of substitution with biased technical change*. American Economic Review, 100(4), 1330-57.

Lombardi, M., & Vannuccini, S. (2022). Understanding emerging patterns and dynamics through the lenses of the cyber-physical universe. *Patterns*, *3*(11), 100601.

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

Ménière, Y., Rudyk, I., and Valdes, J. (2017). *Patents and the Fourth Industrial Revolution: The inventions behind digital transformation*. European Patent Office.

Mokyr, J., Vickers, C., and Ziebarth, N. L. (2015). The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different? Journal of Economic Perspectives, 29(3):31–50.

Montobbio, F Staccioli, J. Virgillito, M. and Vivarelli, M. (2022). The Empirics of Technology, Employment and Occupations: Lessons Learned and Challenges Ahead. IZA Discussion Paper No. 15731.

OECD (2017), *OECD Science, Technology and Industry Scoreboard 2017: The digital transformation*, OECD Publishing, Paris http://dx.doi.org/10.1787/9789264268821-en

OECD (2017), *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264271036-enOECD.

OECD. (2018). Transformative Technologies and Jobs of the Future.

Penny, Simon (2013) 'Art and Robotics: Sixty Years of Situated Machines. [REVIEW],' *AI and Society* 28 (2):147-156.

Ray, D., & Mookherjee, D. (2021). Growth, automation, and the long-run share of labor. Review of Economic Dynamics.

Rogers, E. M. (2010). Diffusion of innovations. Simon and Schuster.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology?. Research policy, 44(10), 1827-1843.

Saviotti, P. P., & Metcalfe, J. S. (1984). A theoretical approach to the construction of technological output indicators. *Research policy*, *13*(3), 141-151.

Savona, M. and W.E. Steinmueller (2013). Service output, innovation and productivity: A time-based conceptual framework. Structural Change and Economic Dynamics. Elsevier, ISSN 0954-349X. - Vol. 27.2013, C, p. 118-132

Servoz, M. (2019), *AI, the Future of work? Work of the Future! On How Artificial Intelligence, Robotics and Automation are Transforming Jobs and the Economy in Europe*. Brussels:

European Political Strategy Centre (European Commission), Brussels

Sheridan, T. B. (2016). Human–robot interaction: status and challenges. *Human factors*, *58*(4), 525-532.

Takeuchi, H., & Nonaka, I. (1986). The new new product development game. Harvard business review, 64(1), 137-146.

Von Tunzelmann, N. (1995). Time-saving technical change: The cotton industry in the English industrial revolution. Exploration in Economic History, 32:1–27

Wirkierman, AL. (2022). Structural dynamics in the era of smart technologies, in Kurz, H. D., Schütz, M., Strohmaier, R., & Zilian, S., *The Routledge Handbook of Smart Technologies* (pp. 1-14). Routledge.

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D. N., Buffington, C., ... & Dinlersoz, E. (2021). *Advanced Technologies Adoption and Use by US Firms: Evidence from the Annual Business Survey* (No. w28290). National Bureau of Economic Research

8 Methodological Appendix

This section details the final literature search protocol, which is the outcome of several iterations, both to define technology families, as well as to select queries and their expansions. This allows us to articulate in detail the rationale underpinning our methodological choices and the solid grounding of the selected technical literature that supports our findings. We propose this as a methodological template for any in-depth investigation of multidisciplinary literature that seeks to shed light on such a complex phenomenon.

The remainder of the Appendix is organised as follows. The first part (Section 8.1) describes the rationales behind the classification into families of automation technologies that will be used across several PILLARS Work Packages, particularly WP3 which deals with the phase of emerging technologies. The second part (Section 8.2) details the complex and rigorous search strategy for relevant papers, which attempts to maximise both precision (selection of relevant papers) and recall (covering as many technology as are available in each family). We describe the basic query to search for documents in Scopus, the manual screening of the sampled papers, the identification of the relevant engineering controlled terms to expand the initial selection of core papers to a more comprehensive set.

The search protocol described in this Appendix has been exemplified with reference to two technology families: Robots and Data Management. The protocol has been fully consistency applied to all the technologies, though not reported here (except the original and expansion queries reported in Table 11-A) for reasons of space.

8.1 Identification of Automation Technologies

To identify the families of digital technologies that contribute to the automation of production processes in all sectors of the economy that involve workers (e.g. excluding households' digital appliances). We followed four steps:

a) First, team members listed and classified technologies based on the literature in economics and management and on their expert judgement. This was followed by two validation exercises.
b) We used reports and academic publications from academia, industry, government and the public on Industry 4.0, future of work, and automation, to build a concordance table between the classification developed in step (a) and classifications used by experts across sectors.
c) We constructed a control classification, developed by experts in a different research team in the PILLARS project, to build another concordance table with the classification developed in step (a)
d) We studied the concordances between our first classification, the classifications used by experts across sectors (b) and the second PILLARS expert classification in point (c) and we

produced a final classification. We discussed this final classification with the whole PILLARS consortium and made minor refinements.

For replicability, we describe each step in detail below.

8.1.1 First classification

The initial classification into families of digital automation technologies was performed based on the literature and the expertise of the research team, especially by Professor Ed Steinmueller. This classification was guided by industry observation and writing in the areas of industrial economics, innovation studies and technology history. It included eight classes, listed in Table -A, each with a list of prominent technologies. To some extent, the list of prominent technologies is arbitrary and reflects an early 21st century understanding of uses and purposes of these technologies. As we move through this century, there will certainly be additional subclassifications (e.g. 'robots' introduced into human bodies to diagnose or repair). The reported subclassifications reflect areas where substantial investments have occurred and that have a potential or demonstrable effect on human skills or employment.

Table 1-A: Classification of digital automation technology families (first classification)

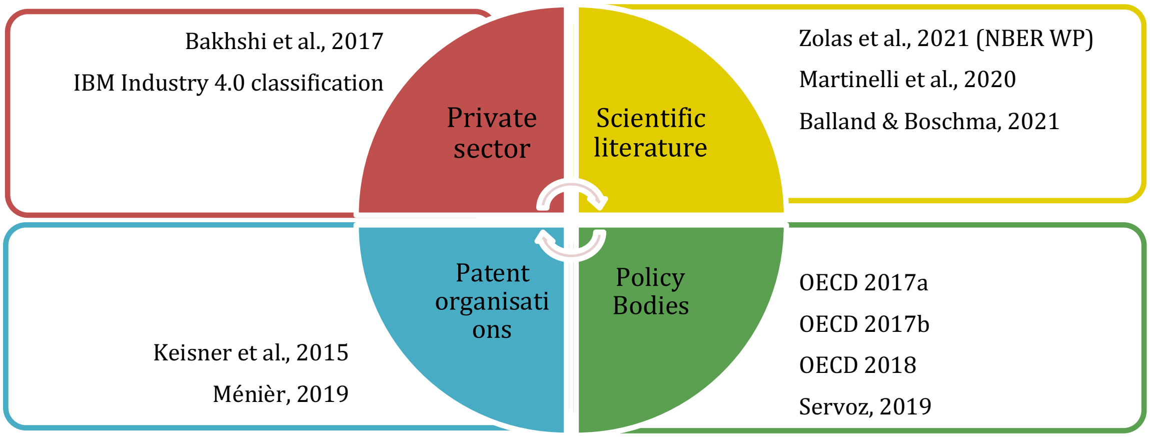
Codes	Automation technology families and prominent technologies
A	**Robots (Primarily stationary)**
1	Machine vision and real-time monitoring
2	Selective Compliance Assembly Robots (SCARA)
3	Articulated
4	Cartesian
5	Dual Arm
6	Co-bots (non-mobile)
7	Swarm robotics
8	Cylindrical
9	Service robotics
B	**Robots (Primarily mobile)**
1	Co-bots (mobile)
2	Semi-autonomous (e.g. bricklaying)
3	Automated platforms/vehicles
4	Tunnel boring and mining robots
5	Submersible robots
6	Drones
7	Space vehicles and rovers
8	Service robotics
C	**Data Acquisition Technologies**
1	IoT (including Radio-frequency identification (RFID) systems)
2	Scanners
3	Sensors
4	Remote Sensing
5	GPS
6	CCTV
7	Scientific and engineering instruments
8	Healthcare instruments (including Personal health instrumentation)

D	**Cloud, PC and Smartphone Computing and Services**
1	Big data analytics
2	Gaming
3	Streaming services
4	Automated storage and retrieval systems
5	Database systems
6	Relational databases
7	Computer architectures
8	Cryptography and security
9	5G
E	**AI (not directly as a cloud service)**
1	Simulation
2	Machine learning (predictive systems)/ Deep Learning
3	NLP
4	Machine vision (Image recognition)
5	Expert systems
6	Speech recognition and production
7	Text recognition and production
8	Machine Translation
F	**Additive manufacturing (using any material – e.g. powder metallurgy as well as bioplastic filament)**
1	Prototyping (including CAD)
2	Tools production
3	Production at scale
G	**User interface**
1	Conventional input devices (e.g. keyboard, mice, pens, webcams)
2	Display devices (conventional)
3	Augmented reality
4	Haptics and Tele-haptics
5	Virtual Reality (including 3D Visualisation)
6	Touchscreens/kiosks for customer interface
7	Sound technologies (e.g. noise cancellation)
8	Neuro-scanning
H	**Other**
1	Machine Tools
2	Wireless Identification Tags/Beacons

8.1.2 Concordance with classifications in the literature

A first validation of the classification presented in Section 8.1.1 (and Table -A) was done by reviewing existing classifications of automation technologies published by different organisations across different sectors of the economy: private sector, patent authorities, policy organisations, and academia (Figure -A). The literature refers to automation technologies under several frameworks, the most common one being Industry 4.0. We focused on documents especially focused in understanding the adoption of automation technologies and their impact on labour, specifically the automation of tasks. The selected documents include three OECD reports (OECD, 2017a; OECD, 2017b; OECD, 2018), one report each from the EC (Servoz, 2019), WIPO (Keisner et al., 2015), and EPO (Ménière et al., 2017)), a report compiled by Nesta and Pearson (Bakhshi, 2017), the IBM classification (https://www.ibm.com/topics/industry-4-0), and three academic papers (Balland & Boschma, 2021; Martinelli et al., 2021; Zolas et al., 2021)).

Figure 1-A: Reference documents from four sectors of the economy



From each of the documents selected we extracted the technology classification proposed. Whenever available, we also collected a list of prominent technologies within each family. Based on this information, we created a concordance table to compare our first classification with the ones received from the literature. The rationale for this exercise was to identify potential gaps, missing technologies or entire families of technologies, or mis-classifications of prominent technologies into the technological families proposed in our first classification. The concordance table is included below in Table 2-A.

Table 2-A: Concordance table of automation technologies classification (first classification and literature)

Literature			First classification	
Report	**Families**	**Sub-categories**	**Family**	**Technology**
WIPO 2015	Robotics - Remoted controlled	Telepresence robots	B	
		Remoted controlled humanoid robots		
		Robotic assisted surgical devices		
		Exoskeletons		
		Drones		B.6
	Robotics - Semi autonomous		B	B.2
	Robotics - Fully autonomous		B	B.3
	Artificial Intelligence		E	E
Pearson-Nesta, 2017 | The Future of Skills	Robots		A/B	
	Artificial Intelligence		E	E
	Big data		D	D.1
	Internet of Things		C	C.1
OECD STI 2017	Mobility		B	

	Cloud computing		D	
	Internet of Things		C	C.1
	Artificial Intelligence		E	
	Big data analytics		D	D.1
AI Future of Work | European Commission, 2019	Artificial Intelligence	General AI: broad cognitive abilities	E	
		Narrow AI: human-level intelligence		
	Robotics		A/B	
	Internet of Things		C	C.1
	Blockchain technology			
	New and advanced materials		F	N/A
	Autonomous devices		B	
Balland & Boschma, 2021	Additive manufacturing		F	
	Artificial Intelligence		E	E
	Augmented reality		G	G.3
	Autonomous robots		B	
	Autonomous vehicles		B	
	Cloud computing		D	
	Cybersecurity			
	Machine tools		D	N/A
	Quantum computers		D	N/A
	System integration		D	D.6
OECD Next Production Revolution 2017	Simulations		E	E.1
	Artificial Intelligence		E	E
	System integration		D	D.6
	Big data		D	D.1
	Cloud computing		D	
	Internet of Things		C	C.1
OECD 2018 | Transformative technologies & jobs of the future	Artificial Intelligence		E	E
	Internet of Things		C	C.1
	Blockchain technology			
IBM IR 4.0 classification	Internet of Things		C	C.1
	Cloud computing		D	
	AI and machine learning		E	E / E.2
	Edge computing		D	N/A
	Cybersecurity			
	Digital twin		D	N/A
Zolas et al., 2021 (NBER)	Augmented reality		G	G.3
	Automated guided vehicles (AGV)		B	B.3
	Automated storage and retrieval systems		D	D.4
	Machine learning		E	E.2
	Machine vision		E	E.4
	Natural language processing		E	E.6

	Radio-frequency identification (RFID) system		C	C.1
	Robotics		A/B	
	Touchscreens/kiosks for customer interface		G	G.7
	Voice recognition software		E	E.6
Martinelli et al., 2020	Internet of Things		C	C.1
	Big data / Industrial analytics		D	D.1
	Cloud manufacturing		D	
	Robotics		A/B	
	Artificial Intelligence		E	E
	Additive manufacturing		F	F
European Patent Office	Core technologies	Hardware	D/G	
		Software	C/D/E	
		Connectivity	D/G	
	Enabling technologies	Analytics	D	D.1
		Security	D	
		AI	E	
		Position determination	D	
		Power supply	N/A?	
		3D systems	G	
		User interfaces	G	
	Application domains	Home, Personal, Enterprise, Manufacturing, Infrastructure, Vehicles	N/A	

There were few discrepancies between our first classification and list of technologies and the various classifications produced in the literature by different groups of experts.

Terms in yellow cells identify prominent technologies which were not explicitly mentioned in our first classification, but which were covered in existing families either as synonyms or as subcategories of a wider technological group. These were mainly technologies part of Cloud, PC and Smartphone Computing and Services (D), such as Quantum computers, Edge computing, Digital twins, System integration and position determination (part of big data analytics). New and advanced materials was classified under additive manufacturing (F). And Position determination was classified as one of the enabling technologies part of the broader big data analytics, also in family D.

Terms in red cells identify prominent technologies which were not explicitly mentioned in our first classification, and which are not easily classified under our first classification. These are Blockchain technologies and Cybersecurity technologies. As we discuss below, following our second comparison, we define a new technology family called 'software based data management', which also includes Blockchain technologies and Cybersecurity technologies.

8.1.3 Control classification

The control classification into families of digital automation technologies was performed based on the literature and the expertise of the research team preparing a Delphi survey on the emerging automation technologies. This classification was also guided by industry observation, the literature, and the OECD classification of ICT (Inaba and Squicciarini, 2017). It included only seven classes, listed and described in Table -A, each with a list of prominent technologies.

Table 3-A: Classification of digital automation technology families (control classification)

Code	**Automation technology families and prominent technologies**
A	**Robotics**
1	Programmable Robots
2	(Semi) Autonomous Robots
3	Service Robotics
4	Exoskeletons
5	Autonomous Vehicles
6	Drones
7	Swarm Robotics
8	Remote sensing
9	Smart Equipment
B	**Networking**
1	Streaming Services
2	5G
3	Internet of Things
4	Wireless Identification Tags/Beacons
C	**Data Management**
1	(Relational) Databases
2	Big Data
3	Cryptography
4	Cyber Security
5	Blockchain
6	Cloud Storage
7	Data Mining/Scraping
D	**Computing**
1	Edge Computing
2	Cloud Computing
3	Quantum Computing
4	High Performance Computing
5	Grid Computing
6	Simulation/Digital Twins
E	**Artificial Intelligence**
1	Machine/Deep Learning
2	Predictive Maintenance
3	Natural Language Processing
4	Machine Vision
5	Text Recognition and Production
6	Speech Recognition and Production
7	Machine Translation
F	**Additive Manufacturing**
1	3D printing
2	Computer Aided Design
G	**Human Machine Interaction**
1	Virtual Reality (VR)

Code	**Automation technology families and prominent technologies**
2	Augmented Reality (AR)
3	(Tele) Haptics
4	Neuro-control (Brain Controlled)
5	Holograms
6	Telepresence
7	Voice control
8	Visual Interfaces (Displays, monitors, touchscreens)
9	Biometrics
10	Apps
11	Gamification (Serious games)

8.1.4 Final classification

After building a concordance table between the first (step a), Section 8.1.1) and the control classifications (step c), Section 8.1.3), experts who developed the two classifications met to discuss results from the two concordance table and converge on a final, integrated, classification. As a final robustness check, this final table was discussed with the entire consortium. The final classification of automation technologies contains nine families and is reported in Table -A.

Table 4-A: PILLARS classification of emerging automation technology families

A	**Robots (Articulated, Cylindrical, Cartesian, Dual Arm, SCARA)**
1	Machine vision and real-time monitoring
2	Co-bots
3	Swarm robotics
4	Service robotics
5	Semi-autonomous (e.g. bricklaying)
6	Automated platforms/vehicles
7	Tunnel boring and mining robots
8	Drones
9	Robotic vehicles (including Space vehicles and rovers, autonomous vehicles, submersible robots)
10	Exoskeletons
11	Robotic Process Automation (RPA) (including software robots)
B	**Physical Data Acquisition Technologies**
1	Scanners
2	Sensors
3	Remote Sensing
4	GPS
5	CCTV
6	Scientific and engineering instruments
7	Healthcare instruments (including personal health instrumentation)
8	Data scraping
C	**Software-based Data management**
1	Database systems
2	Relational databases (including API)
3	Cryptography, security, and blockchain mining)
4	Big data analytics
D	**Computing**
1	Automated storage and retrieval systems (including cloud storage)
2	Computer architectures (including edge computing, cloud computing, HPC, grid computing)
E	**AI (not directly as a cloud service) & Intelligent Information System**
1	Simulation

2	Machine learning (predictive systems)/ Deep Learning
3	NLP
4	Machine vision (image recognition)
5	Expert systems
6	Speech recognition and production
7	Text recognition and production (including machine translation)
F	**Additive manufacturing (using any material e.g. powder metallurgy and bioplastic filament)**
1	Prototyping (including tools production, production at scale)
2	3D printing
3	CAD/CAM (prototype and/or production)
G	**Networking**
1	IoT (including Radio-frequency identification (RFID) systems)
2	Wireless communication (including 5G)
H	**User interface**
1	Conventional input devices (e.g. keyboard, mice, pens, webcams)
2	Display devices (conventional)
3	Augmented reality (including holograms)
4	Haptics and Tele-haptics (including all tele-operations of physical machinery by human operator requiring feedback)
5	Virtual Reality (including 3D Visualisation)
6	Touchscreens/kiosks for customer interface
7	Sound technologies (e.g. noise cancellation)
8	Neuroscanning
9	Gamification
I	**Other**
1	Machine Tools
2	Factory control system

As any other classification, the one presented in Table -A is based on a specific interpretation of these digital automation technologies. Our classification serves the purposes of the analysis of the literature in this paper, and the analysis of emerging technologies both using documents metadata (publications and patents) and using expert's view in the Delphi survey. To provide more clarity on our classification, below we provide our own definition and description of these technologies by family. It should also be noted that several of the prominent technologies listed in Table -A may be allocated in more than one family. As any classification, the boundaries between the technology families proposed here overlap, and technologies may fall under more than one family.

Technology family A – Robots (Articulated, Cylindrical, Cartesian, Dual Arm, SCARA)

Definition: Technologies that sense and (autonomously) act based on data

Although robots were initially imagined as autonomous and human-like in science fiction, their actual implementation reflects a series of incremental steps to integrate specific functionalities, e.g. initial efforts to develop software that could recognise and differentiate objects (e.g. wooden alphabet blocks led to the development of robot 'hands' and arms capable of arranging blocks or other objects). This basic capability was then further developed through the form of the robot during which the issue of robots and humans occupying the same space needed to be addressed (a robot that is sturdy enough to move large objects at speed is a potential hazard to human co-workers). Robot mobility is almost inevitably linked with human presence (e.g. exceptions did exist such as robots used to move into hazardous spaces such as nuclear accident sites) and also led to robots navigating different terrains and performing a few specialised functions (e.g. tunnel boring).

Technology family B – Physical Data Acquisition Technologies

Definition: technologies that harvest and record information

Computer systems originated as engines for data processing. The data that they employed needed to be fed carefully prepared (in order to be machine readable), a process that began as translation of paper records and proceeded to direct entry by data terminals. More recently, a collection of technologies has been developed for data acquisition that do not require this careful preparation. These new technologies stem from the use of digital measurement devices in laboratories and the use of locational beacons (e.g. GPS). The sensory capabilities of these devise were extended outside of the laboratory in the world, paralleling the rise of computer vision applications in robotics.

Technology family C – Software-based data management

Definition: Technologies for storing, protecting, managing/handling and acquiring data

The stores of data themselves became a site for innovation. What were once paper records generated by human activities became electronic records, audiovisual content became data streams and means of securing data became security and cryptographic innovations. The software used for these purposes is also included in this family.

Technology family D – Computing

Definition: Technologies used to compute/calculate

Computers were initially employed to automate the work of ‘human computers’ and quickly exceeded human capabilities in computational tasks. As a consequence, several activities that were previously performed by humans such as compiling tables of the values of mathematical functions for different parameters (e.g. values of cosine by angle) are so well established that they are excluded here. Innovation in the architectures for data storage and for performing computations are still occurring and constitute this family.

Technology family E – AI (not directly as a cloud service) & Intelligent Information System

Definition: Technologies using algorithms and advanced methods to make sense out of the data

The existence of large data stores (B) has opened a new range of computational capabilities commonly referred to as Artificial Intelligence (AI) which (so far) involves a collection of algorithms and some related hardware to implement machine learning. Machine learning has proven a powerful means of addressing the older issue of machine vision and the previously unreachable goal of Natural Language Processing (NLP) as well as more limited goals such as speech recognition or production.

Technology family F – Additive manufacturing (using any material)

Definition: Technologies that produce bottom-up based on digital models

A by-product of the development of mechanical and materials control for printing was the recognition that material objects could be 'printed' by precisely positioned deposits of material, what has come to be known as additive manufacturing.

Technology family G – Networking

Definition: Technologies for communicating between machines (data transmission) or connecting machines

The distribution of data entry and computation created a demand for rapid and reliable data communication between computer systems and the technologies for performing this function are networking technologies. As new technologies from family B were developed, they came to employ networking technologies as well. Networking technologies employ both 'wired' (physical connection) and 'wireless' (connection using the radiofrequency spectrum) methods. The extension of these networks through 'inter-networking' led to the Internet and with the Internet previous technologies such as telephony and broadcast media were increasingly accomplished with data communications (as these previous technologies were already automated, their applications are omitted. The areas of active innovation in networking are the integration of technologies from family B into the Internet (hence IoT) and the use of mobile phones as both receivers of broadcasts and as interactive devices (e.g. for voice and text communication).

It shall be noted that there are still areas where innovation is occurring that are not covered by IoT and 5G – e.g. edge networks and variants of the Internet protocol (e.g. IPv6). However, innovations in these areas are variants in the form of technology in already automated technologies and are therefore not directly implicated in employment (except to the extent that they extend or modify the skills needed to work in the data communications areas)

Technology family H – User interface

Definition: Technologies for human interaction with machines or data

The interaction between humans and computational capabilities is achieved through human computer interfaces, an area of continuing innovation. The technologies under this heading reflect those interactions that involve the presence of a human being exchanging information with a computer system through a variety of methods.

This classification is used to divide the systematic literature review into different blocks of coherent literature. To study the properties of these technologies in relation to working tasks, we conduct a separate literature review for each of the nine technology families, following the same procedure described below. The prominent technologies within each family are used to construct detailed search queries to identify relevant literature at a rather fine-grained level of technological detail.

8.2 Systematic literature review of automation technologies: the protocol

The novelty of the rationale proposed here with respect to existing reviews of the literature on the impact of selected technologies on employment is the focus on the technical literature rather than the economic literature, to go beyond the coarse view that is available in current measures of, e.g., robots and AI. The focus on technical features of automation technologies, as described by academics working on them, has the power of offering a better understanding of the nature of human-machines interdependence, depending on the nature of the technology, the tasks they can perform, and on the nature of human-machine interactions.

To this end, we followed a six-steps protocol, each including several sub-steps. It was first validated on robots, on which we have performed further checks (explained below). Figure -A shows the final sequence of steps, whereas Figure -A shows the sequence including the validation (on the literature on robots).

This protocol was applied to and shown in this Appendix for robots and software-based data management (see Table -A). The protocol has been fully consistently applied to all other technology families though not reported here for reasons of space.

8.2.1 Step 1: Identification of records: query

Because we are not interested in the technological development per se, but in what the technology can do, and in particular what tasks that it can perform, we first build a query that is composed in the following way. We define three different sets of keywords, each identifying

- the technologies in a given family (and synonymies),
- their functions or applications, and
- the tasks that they can perform.

To increase the precision of the documents retrieved by the query, the three sets of terms were connected using the proximity operator (W/n) which requires terms in the three sets to be no more than *n* terms away. For instance, if n=2, the search would retrieve only documents in which the technology term is followed (or preceded) by that term identifying the function at most two terms away, and the term identifying the task at most two terms away from the term identifying the function. In practice, this means that those three terms, on average, would appear in the same sentence of the abstract.

The identification of relevant keywords for the three sets of terms is based on our technology classification and on additional keywords identified via relevant papers from the Scopus and/or from core papers in the literature which study these technologies, especially in relation to labour. Although we present here our protocol as a linear process, the refinement of queries was an iterative process that necessarily required expert assessment, also based on the documents retrieved by several queries. We also experimented with different values of n, ranging between one and three.

8.2.2 Step 2: Identification of records: extraction of documents

We applied the query in Scopus and downloaded all documents which

- Were retrieved by our query (Step 1) in title, abstract, or keywords (TITLE-ABS-KEY)
- Were Published after 2000 or 2010 (depending on the technology) ((PUBYEAR)).
- Are among the following type of publication (DOCTYPE): article (ar) OR Conference Proceedings (cp), OR Conference Review (cr) OR review (re). We decided to include conference proceedings and reviews, despite the lack peer review quality check, because of the key goal of the analysis to capture emerging applications of technologies and the tasks they may be able to perform.
- Include one or more keyword(s) among from the indexed terms (INDEXTERM), to restrict to specific technologies, when available in the engineering thesaurus used by the Ei Compendex.[15]

We select the top X% of cited papers (normalized by year). The exact percentile is determined based on the size of the corpus of papers downloaded with the query; we aim at more or less 500 papers to be screened for each technology family. See Figure -A and Figure -A for details on the number of documents included and excluded in each step for robots and software based data management.

8.2.3 Step 3: Screening: documents selection

To create a sample of relevant documents to read and code for our literature review, we first screened the title and abstract of all the ~500 documents sampled in Step 2. Each paper was screened by two random reviewers independently.

[15] https://www.elsevier.com/about/press-releases/archive/science-and-technology/5th-edition-of-the-ei-thesaurus-now-available

The five reviewers agreed on the following rule to include or exclude documents from the sample of relevant documents. They excluded documents that were conceptually describing a technology, documents which were describing abstracts proofs of concept, and documents which did not explicitly refer to the performance of specific tasks. For instance, in the case of robot technologies, in relation to tasks, we included documents that explicitly referred to the automation of tasks, or robotic control in specific domains, both supervised and unsupervised, and the interactions between humans and robots. We instead excluded document referring to basic research on robotics, basic research on the capacity to move, or sense, but without a specific explanation about the task, basic research on the performance interaction between robots, again without a specific description of the tasks performed.

Figure 2-A: Flow-chart of the literature review process (including robot validation)

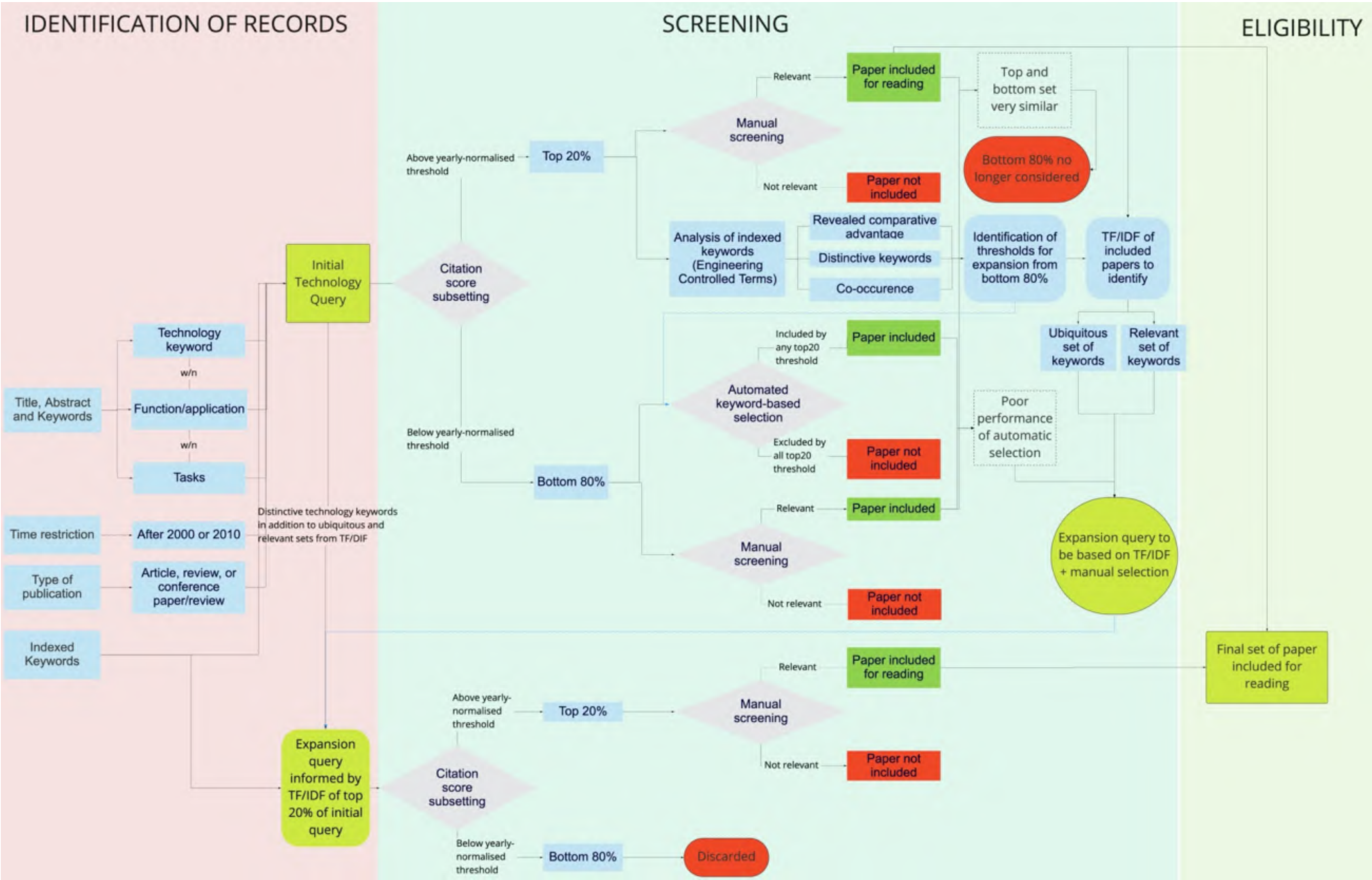
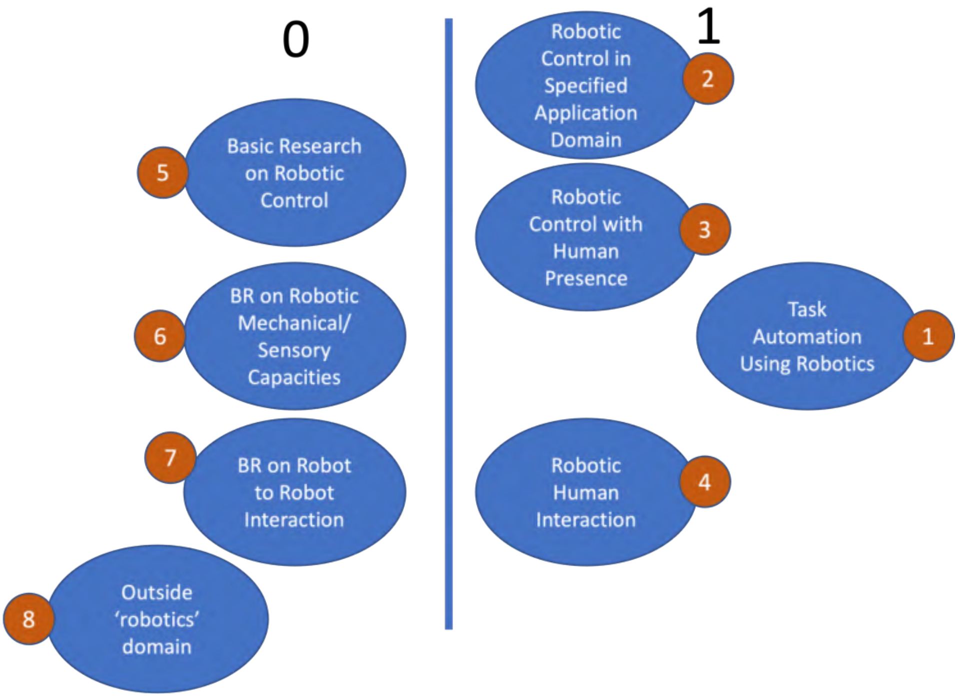


Figure -A illustrates the above distinction, were the section labelled with 0 refers to documents that were excluded from the selected sample and 1 refers to documents that were included in the selected sample.

Figure 3-A: Task criteria for excluding/including documents

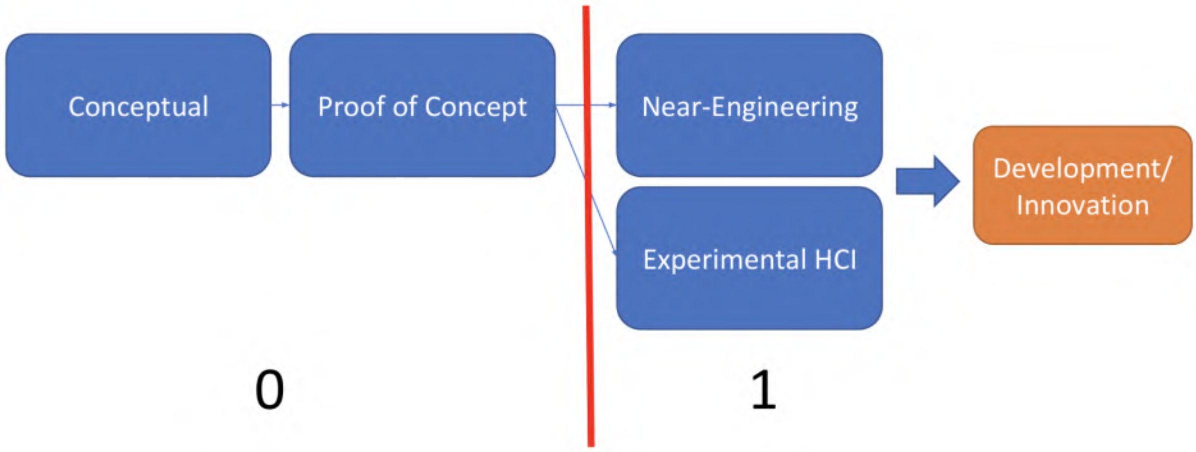


Instead, we included documents that describe the development of specific technologies, including proof of concepts that lead to such developments, documents that describe experiments of the technologies, and which refer to tasks. Figure -A provides a visual description of the criteria used to select screened documents based on their level of applicability. Below we define the four levels of applicability:

- Conceptual – A new framework or approach is described and argued to be valuable
- Proof-of-Concept – A particular approach is tested in a limited number of trials, sometimes with reference to existing practice but more commonly in relation to the claim made by the concept
- Near-engineering – Proof of concept is implicit but further work is done to characterise and test the robustness of the solution in ways that would be useful in the development process
- Experimental HCI – Explicit tests and co-design with users of new HCI approach. In one sense this is proof of concept but because it involves 'live subjects' it is likely to inform development efforts.

In case of disagreements between the two reviewers a third reviewer made the decision on whether to include or not the document in the selected sample. Extracted data included: title, abstract, keywords, authors, affiliation, journal, publication year, and type of study (article or research letter/comment/editorial). See Figure and Figure for details on the number of documents included and excluded in each step for robots and software based data management.

Figure 4-A: Application criteria for excluding/including documents



8.2.4 Step 4: Identification of records: query expansion

Any strategy to select the literature on the basis of a string of keywords may identify documents that are not relevant to the topic (low precision) and not identify documents that are relevant to the topic (low recall). Low precision occurs because the terms used are too broad. For instance, "robots" may retrieve documents from science fiction. We controlled for precisions using a relatively complex query with proximity operators, and by manually screening titles and abstracts as described in the steps above. To control for recall (i.e. retrieve documents that were not found by our query, but which are relevant for a given technology), we exploited the *engineering controlled terms*, a set of keywords in the Scopus documents metadata assigned by librarians manually. *engineering controlled terms* provide a list of subject terms for the content of a document in a specific and consistent way, because they for a curated list of terms reviewed and standardised. According to Elsevier, "controlled terms are assigned by professional indexers to capture the concepts a document deals with in a standard manner".[16]

Although we did not have access to the "ei Thesaurus", which is integrated in Scopus, for each document we retrieved the engineering controlled terms as part of the metadata. Only for the case

[16] https://www.sciencedirect.com/topics/computer-science/controlled-vocabulary

of robots, we examined different ways in which we could use this controlled vocabulary to identify an automatic way to select distinctive keywords for papers in the selected sample (after screening) (Figure -A). And then use these distinctive controlled terms to search for more paper in Scopus, which were assigned the same key terms, but which were not retrieved by our query.

In the robot pilot test case, out of the 498 papers screened, 87 were included in the selected sample. These 87 papers have been labelled with a total number of 511 distinctive engineering controlled terms. The top ten most frequent terms are presented in Table -A as an example.

Table 5-A: Top 10 most frequent engineering controlled terms for robot technologies

Keywords	**n**
Robotics	29
Robots	17
Robot programming	9
Industrial robots	8
Intelligent robots	8
Man machine systems	7
Mobile robots	7
Artificial intelligence	6
Motion planning	6
Agricultural robots	5
Automation	5

To identify engineering controlled terms that are specific to the 87 documents in the selected sample, and are not relevant for documents in the residual non-selected sample (411 documents) we used four different methods:

- **Balassa index (RCA):** considering *s* a sample of documents s and k the engineering controlled terms, we define the RCA of a given engineering controlled term in a given sample of documents as the ratio between the normalized frequency of an engineering controlled term[17] *k* in the sample *s* over the normalized total number of engineering controlled terms *K* in that set and the normalized frequency of the same engineering controlled term *k* in the whole sample of screened publications *S* over

[17] Because documents differ in relation to the number of key terms they are labelled with, and documents with more terms are more likely to be labelled against one term, we normalised the frequency (F) of engineering controlled terms by the number of engineering controlled term of each document. That is, if a document had 10 engineering controlled term, each term was assigned 1/10 as frequency; the same term was assigned 1/6 a document with 6 keywords.

the normalized total number of engineering controlled term in the whole sample of screened publications. Formally:

$$RCA_{sk} = \frac{F_{sk} / \sum_{k' \in K} F_{sk'}}{\sum_{s' \in S} F_{s'k} / \sum_{s' \in S, k' \in K} F_{s'k'}}$$

where F_{sk} is the normalized frequency of engineering controlled term sets k and k' in the samples s and s'

- We used two samples of documents, based on the screening (Step 3): the sample selected for final analysis and the sample of all documents excluded by at both reviewers.[18]
- We consider the following set of engineering controlled terms that are specific to the selected sample: all terms in the sample of selected publications with RCA > 1.
- Note that these indicators include engineering controlled terms that may appear both in the selected sample and in the excluded sample, but which are more common to appear in the selected sample, and therefore may characterize the topic of the sample that we manually selected as relevant.

- **Co-occurrence**
  - We compute all combinations of two or more engineering controlled term that appear in any one document in the selected sample.
  - We consider the following two sets of engineering controlled terms that are specific to the selected sample: we restrict our analysis to the documents including at least 3 or 4 terms out of the overall set of engineering controlled terms in the selected sample (excluding robots and robotics because they are too common also in the non-selected sample of documents, by construction).
- **Distinctive keywords:**
  - We consider the set of engineering controlled terms that appear only in documents in the selected sample and not in documents in the non-selected sample.
- **Single keywords selected**
  - We rank engineering controlled terms by their frequency in the selected sample of documents. We consider the set of engineering controlled terms with an absolute frequency higher or equal than 3 (top 6%).

We then assess if any, or a combination, of these four methods and sets of engineering controlled terms distinguishes relevant documents in a larger corpus. If this were true, we could then use the selected set of engineering controlled terms to search in Scopus to expand our query to other

[18] We did not include documents excluded by one reviewer, because their non-relevance is more ambiguous.

relevant documents and increase recall. To do this assessment, we first manually screened the 865 documents retrieved by our query (Step 1), not included among the 20% most cited (normalised by year) (Step 2),[19] and retrieved by at least one of the four methods explained above to select engineering controlled terms. For the screening we followed the same procedure described in Step 3.

As a result of the manual screening, 124 documents were included in the selected sample by two reviewers (14%), 592 were excluded (68%), and 149 were included by only one reviewer (17%). The controversial papers were analysed by a third reviewer, who included 76 of those in the selected sample. In total, 200 papers out of the 865 were included in this selected sample (for the 80% least cited documents) by means of manual screening.

We compared these documents allocated to the selected sample manually with the documents selected by each of the four methods (i.e. documents that contain engineering controlled terms in the sets selected by each of the methods). Results were not encouraging.

Most documents (815) were retrieved by at least one of the methods – the most inclusive being the single keyword selection (79%; 682 documents) and the most selective being co-occurrence (33% for three keywords; 287 documents). The overlap between the sample of documents selected manually, and the sample of documents selected by the four methods is rather low. Only between 32% (co-occurrence) and 78% (single keyword selected) of documents selected manually were also selected by any of the four methods. That is between 22%-68% of the documents selected manually, were not identified by any of our methods. Combining the four methods did not improve much the performance of the automatic methods to select relevant document in a larger corpus[20].

This pilot test on robots led to two important findings for our protocol. First, in relation to extending our initial query to other relevant papers in Scopus, the sparsity of engineering controlled terms (in the corpus of papers related to robots) is too high and their frequency too low to determine an automated method to select relevant documents.[21]

Second, in relation to the choice of focusing on the top X% of top cited documents (normalized by years) the documents in the bottom 80% cover very similar topics to those covered by documents in the top 20%. There does not seem to be a substantial added value/information in considering all documents for screening.

To extend our initial query, for all technologies, we thus followed a different strategy, which addresses the issue of the sparsity of the engineering controlled terms in the corpus of documents downloaded from Scopus. We identify relevant engineering controlled terms using Term Frequency

[19] That is, the bottom 80% least cited documents (normalised by year).

[20] Both pair-wise combinations and a higher number of concurrent methods were analysed.

[21] We also made an attempt using machine learning methods (Word2vec), but the size of the corpus is too small for the algorithm to recognise any pattern in the selected sample.

Inverse Document Frequency (TF-IDF). TF-IDF is a text mining technique used to identify relevant words in a corpus of text. It combines the frequency of a term in the corpus (TF),[22] with their inverse document frequency (IDF), which measures the extent to which words are used frequently within a given document but not in the entire corpus of documents. By multiplying TF and IDF, we obtain the frequency of a term adjusted for how rarely it they are used. Where IDF is mathematically defined as

$$IDF(term) \ = \ ln(\frac{n_{docs}}{n_{docs\ contaning\ term}})$$

We first build the matrix of co-occurrences of all engineering controlled terms across the documents in our selected sample (in the top X%). Columns and rows are the engineering controlled terms, and the cells measure the number of documents in which each pair of engineering controlled terms appears together in the same document.

We then compute the TF-IDF using the co-occurrence matrix in the following way. We consider the weighted combination of each engineering controlled terms with all the engineering controlled terms with which it appears as the record in which a term can appear – that is, all columns of the co-occurrence matrix. In other words, we consider all the terms that appear together with each of the other engineering controlled terms in the same document. Table -A provides an example extracted from the case of software based data management. "State estimation" and "internet of things" in the rows both appear together with four other terms, but "internet of things" appear more frequently with some terms than with others, which means they will have a higher IDF.

Table 6-A Extract of co-occurrence matrix of engineering controlled terms – Software-based data management

	5g mobile communication systems	internet of things	state estimation	wireless sensor networks	wireless telecommunication systems
5g mobile communication systems	0	2	1	2	1
internet of things	2	0	1	3	2
state estimation	1	1	0	1	1
wireless sensor networks	2	3	1	0	2

[22] Measuring only how frequently a term occurs in documents would privilege terms that may not be as relevant to identify a specific corpus. For instance stop words, or in the case of the robot technology "robots".

wireless telecommunication systems	1	2	1	2	0

Another way to look at this is to consider each engineering controlled term as a record, whose text is the sequence of all engineering controlled term with which it appears together in the same documents, repeated for the number of times they co-occur (as per the co-occurrence matrix). Table -A provides an example of these records for three engineering controlled terms included in Table -A.

Table 7-A From the co-occurrence matrix to TF-IDF records and terms

Doc	Text
5g mobile communication systems	internet of things, internet of things, state estimation, wireless sensor networks, wireless sensor networks, wireless telecommunication systems, quality control, global positioning system, optimization, quality of service, accident prevention, automotive industry, military applications, surveys, wireless local area networks (wlan)
internet of things	5g mobile communication systems, 5g mobile communication systems, state estimation, wireless sensor networks, wireless sensor networks, wireless sensor networks, wireless telecommunication systems, wireless telecommunication systems, automation, network architecture, network architecture, network security, sensor nodes, sensor nodes, sensor nodes, zigbee, process control, semantics, diseases, data privacy, distributed database systems, fog, fog, fog, fog, fog computing, fog computing, health care, health care, middleware, middleware, global positioning system, cloud computing, cloud computing, cloud computing, digital storage, digital storage, distributed computer systems, distributed computer systems, manufacture, data acquisition, data acquisition, optimization, security of data, quality of service, accident prevention, accident prevention, automotive industry, automotive industry, military applications, surveys, wireless local area networks (wlan), budget control, decision making, decision making, edge computing, edge computing, electric batteries, energy efficiency, green computing, job analysis, monitoring, parallel algorithms, solar cells, agriculture, heart, data handling, intelligent buildings, internet, web services, matrix algebra, embedded systems, agricultural robots, chemical sensors, gas detectors, knowledge based systems, gas sensing electrodes, risk assessment, data transfer, safety engineering, conducting polymers, design, integration, security systems, medical applications, remote patient monitoring, planning, urban growth, complex networks, intelligent systems, blood pressure, cardiology, patient treatment, infrastructure as a service (iaas), knowledge representation, multimedia services, multimedia systems
state estimation	5g mobile communication systems, internet of things, wireless sensor networks, wireless telecommunication systems, trees (mathematics), global positioning system, kalman filters, digital storage, vehicles, online systems, uncertainty analysis, cellular telephones, location based services, telecommunication services, mobile phones, telephone sets, probes, highway traffic control, hybrid systems,

	lagrange multipliers, street traffic control, telecommunication equipment, traffic control

We then compute TF as the number of times a term appears in the column (record), and IDF as the log of the share between the total number of columns (records) and the number of columns (columns) in which it appears (records). In this way, we consider both the frequency of a given engineering controlled term (TF) and a measure of their relevance in combination with other terms, that is, with how many other engineering controlled terms they appear together. For example, in Table -A the TF of the term "human" in the record (engineering controlled term) "article" is the ratio between n (its frequency) and the total number of terms in that record (total). The IDF instead is computed as the log of the ratio between the total number of records (2060) over the total number of records in which "human" appears (256), that it the total number of engineering controlled term with which "human" co-occur.

Table 8-A: TF-IDF for extension query (example)

Record	Term	n	total	tf	idf	tf_idf
article	human	14	331	0,042296	2,043213	0,08642
human	article	14	379	0,036939	2,158191	0,079722
chemical sensors	gas detectors	13	206	0,063107	2,967022	0,187239
gas detectors	chemical sensors	13	158	0,082278	2,667617	0,219487
process control	sensors	12	256	0,046875	2,352347	0,110266
sensors	process control	12	254	0,047244	2,372966	0,112109
article	priority journal	11	331	0,033233	2,245966	0,074639
priority journal	article	11	278	0,039568	2,158191	0,085396
human	priority journal	10	379	0,026385	2,245966	0,05926
priority journal	human	10	278	0,035971	2,043213	0,073497

Note: **Record** is the name of the columns in the engineering controlled term co-occurrence matrix, that is all engineering controlled terms; **Term** is the list of engineering controlled term in the column of the engineering controlled term co-occurrence matrix (also all engineering controlled terms); **n** is the total number of times an engineering controlled term appears together with other engineering controlled terms; **total** is the number of times that all engineering controlled terms co-occur with an engineering controlled term; **tf** is term frequency as defined above; **idf** is Inverse Document Frequency as defined above; **tf-idf** is the product between tf and idf as defined above.

We use both TF and IDF to construct the expansion query, which is meant to identify additional areas of application of the technology that were not included in the string of keywords in our initial query. We build the query using two sets of engineering controlled terms. A set of ubiquitous terms, i.e., engineering controlled terms that are very frequent and appear in most of the documents in the selected sample, which we assume characterise the technology family of interest. And a set of relevant specific terms, i.e. engineering controlled terms that are specific to sub-technologies or sub-applications, which are frequent only in homogeneous subsets of the documents in the selected sample.

The first set is identified ranking engineering controlled terms by term frequency (TF); the second set is identified ranking engineering controlled terms by inverse frequency (IDF). The selection is always manual, although informed by this methodology, as the expansion is meant to cover additional areas which might be missed by the initial query, hence human judgment is always needed.

The full expansion query based on the engineering controlled terms is reported in Table -A. See Figure -A and Figure -A for details on the number of documents included and excluded in each step for robots and software based data management.

8.2.5 Step 5: Identification and screening of expansion documents

We repeat Steps 2 and 3 to download from Scopus the documents retrieved with the expansion query, select the top X% most cited (by year)[23], and screen them to build a second sample of selected documents to be included in the literature review.

8.2.6 Step 6: Reading and coding selected documents

We finally read the text of all documents in the selected sample. A small share of these documents were discarded from the review because considered not relevant in relation to the criteria discussed under Step 4 when reading the full text, or because we could not find in electronic form to be downloaded. See Figure -A and Figure -A for details on the number of documents included and excluded in each step for robots and software based data management as examples of how the procedure was completed. The remaining papers were coded to capture the following features, when available. These are described at length in the text and recalled here for convenience:

Emerging technologies, exposure and adoption:

[23] With X depending on the number of documents retrieved, aiming to screen around 500 documents

1. **Level of Adoption of the technology:** Depending on whether the technology is already used or not in industry.
   Values: low, medium or high.
2. **Development stage:** Depending on the level of maturity of the technology described
   Values: conceptual; experimental; prototype; ready to deploy; mature

Task Routinisation:

3. **Routinisation.** This classifies the technology on the basis of the ability to perform a task without any human intervention, including the possibility that a task is further decomposed into an automated and humanly supervised segments
   Values: Yes; No
4. **Knowledge codification.** This classifies the technology on the basis of the ability to make all instruction explicit (i.e. codified) without the use of any tacit knowledge (so far assumed as being a human-specific characteristic)
   Values: Yes; No
5. **Works with People/Symbols/Objects.** This characterize the type of tasks that can be performed in terms of what the technology works with
   Values: People ("in-person" services (requires high soft skills)); Things (routine production services (does not require soft skills)); Symbols ("symbolic- analytic" services (does not require soft skills))

Technologies in relation with occupations and skills:

6. **Skills.** Skills needed to use the technology
   Values: Low; Medium; High
7. **Substitute or complement.** Whether the technology is meant to complement or substitute the human workers. In the documents reviewed, it is also possible to identify segments of tasks (or sub-tasks) that are replaced and others within the same task that are complemented. In these cases records are identified as substitute and complements.
   Values: Complement ; Substitute
8. **Time saving or product/process innovation.** Whether the technology lead to mainly improvements on the quality of the product or service produced and/or on the production process (quantity, i.e. time saving)
   Values: Process; Product

Sectors, firms, and geography categories

9. **Sector of application.** This classifies technologies according to the main intended sector of use according to the 3-digit ISIC classification for manufacturing and 2-digit for other sectors

Values: 3-digit ISIC classification for manufacturing and 2-digit for other sectors.

10. **Task of application.** This classifies technologies on the basis of the tasks they are intended to perform, classified based on the ONET work activities
Values: See Table -A

11. **Geographical area of provenience:** We record where the technology has been developed and mainly deployed, based on the UN M49 classification[24]
Values: See Table -A

12. **Type of organizations:** This question as left open to identify the different type of organizations for which the technology is intended

13. **Size of organisations:** the typical size of the organisations for which the technology is intended
Values: Micro<10 employees; Small<50 employees; Medium<500 employees; Large>550 employees

Table 9-A: Geographical area of the technology

Geographical area
015 Northern Africa
202 Sub-Saharan Africa
419 Latin America and the Caribbean
021 Northern America
010 Antarctica
143 Central Asia
030 Eastern Asia
035 South-eastern Asia
034 Southern Asia
145 Western Asia
151 Eastern Europe
154 Northern Europe
039 Southern Europe
155 Western Europe
009 Oceania

[24] This dimension is hard to codify as often the place/country of first development or application of the technology is not explicitly mentioned or deducible. When this is the case, the geographical area is attributed through the affiliation of the author. Because of this, we have not taken into consideration this category when looking at the results.

Table 10-A: ONET work activities

ONET work activities
Getting Information
Monitor Processes, Materials, or Surroundings
Identifying Objects, Actions, and Events
Inspecting Equipment, Structures, or Material
Estimating the Quantifiable Characteristics of Products, Events, or Information
Judging the Qualities of Things, Services, or People
Processing Information
Evaluating Information to Determine Compliance with Standards
Analyzing Data or Information
Making Decisions and Solving Problems
Thinking Creatively
Updating and Using Relevant Knowledge
Developing Objectives and Strategies
Scheduling Work and Activities
Organizing, Planning, and Prioritizing Work
Performing General Physical Activities
Handling and Moving Objects
Controlling Machines and Processes
Operating Vehicles, Mechanized Devices, or Equipment
Interacting With Computers
Repairing and Maintaining Mechanical Equipment
Documenting/Recording Information
Interpreting the Meaning of Information for Others
Communicating with Supervisors, Peers, or Subordinates
Communicating with Persons Outside Organization
Establishing and Maintaining Interpersonal Relationships
Assisting and Caring for Others
Selling or Influencing Others
Resolving Conflicts and Negotiating with Others
Performing for or Working Directly with the Public
Training and Teaching Others
Guiding, Directing, and Motivating Subordinates
Coaching and Developing Others
Provide Consultation and Advice to Others
Performing Administrative Activities
Staffing Organizational Units
Monitoring and Controlling Resources

Table 11-A: Search queries by automation technology family

<table>
<tr><th>Code</th><th>Technology</th><th colspan="2">First Query</th><th>Expansion</th><th>Notes</th></tr>
<tr><td>A</td><td>Robots</td><td colspan="2">TITLE-ABS-KEY ((robot* OR "human worker") W/2 (process* OR routine OR task OR service) W/2 (automat* OR repetitive OR autonomous* OR smart OR intelligen* OR self-learn* OR interact OR recogn* OR weld OR control OR move OR clean OR walk OR carry OR detect OR drive OR predict OR detect OR identify OR determine OR control OR generate OR classify) OR cobot* OR "co-bot*" OR "collaborative robot*") AND DOCTYPE (ar OR cp OR cr OR re) AND INDEXTERMS (robot*) AND PUBYEAR > 2000</td><td>INDEXTERMS ((robotics AND robots) AND (automation OR "intelligent robots" OR "service robots" OR "mobile robots" OR "multi agent systems" OR "process control") AND ("service industry" OR crops OR "information management" OR "risk management" OR "architectural design" OR personnel OR "large scale systems" OR aircraft OR welding OR navigation OR surgery OR assembly)) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2000</td><td>• After 2000
• function include routine, task, process, service</td></tr>
<tr><td>B</td><td>Physical Data Acquisition Technologies</td><td>Top 10%
TITLE-ABS (("Data Acquisition Tech*" OR das OR daq OR dau OR "Scanner*" OR converter OR sensor OR "Remote Sens*" OR gps OR cctv OR spectrometer* OR</td><td>Top 50%
TITLE-ABS (("Data Acquisition Tech*" OR das OR daq OR dau OR "Scanner*" OR converter OR sensor OR "Remote Sens*" OR gps OR cctv OR spectrometer* OR "Polymerase chain</td><td>Top 4%
INDEXTERMS (("data") AND ("in-situ measurement" OR "mass spectrometry" OR "biosensors" OR "online monitoring" OR "data processing" OR "data correlations" OR microelectronics OR "spatial data infrastructure" OR "x ray diffraction" OR “Data handling” OR “smart process”) AND ("drug delivery" OR "industrial sensor" OR "weather forecasting" OR "microelectronics" OR “process control” OR “Remote sensing” OR “Quality</td><td>• After 2010</td></tr>
</table>

		"Polymerase chain reaction" OR "x-ray" OR "scanning electron microscope" OR "Atomic force microscopy" OR "scanning force microscopy" OR "blood analyser" OR microfluidics OR tomography OR "blood oxygen monitor" OR ekg OR mri OR neuroimaging OR dialysis OR "insulin pumps" OR "Cyber Physical Systems" OR "Data Acquisition Systems" OR "Sensor node*" OR "Satellite imagery" OR radar OR imaging OR "Computer Assisted Tomography" OR "Laser scann*" OR "Handheld scann*" OR	reaction" OR "x-ray" OR "scanning electron microscope" OR "Atomic force microscopy" OR "scanning force microscopy" OR "blood analyser" OR microfluidics OR tomography OR "blood oxygen monitor" OR ekg OR mri OR neuroimaging OR dialysis OR "insulin pumps" OR "Cyber Physical Systems" OR "Data Acquisition Systems" OR "Sensor node*" OR "Satellite imagery" OR radar OR imaging OR "Computer Assisted Tomography" OR "Laser scann*" OR "Handheld scann*" OR "Navigation System" OR "Satellite Navigation Aid" OR photogrammetry OR "Automatic Identification	control” OR “environmental sensor” OR “automated visual inspection” OR “satellite imagery” OR “satellite sensor” OR “medical imaging” OR “body sensor networks” OR “hydraulic model” OR “weather forecast”)) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010	

		"Navigation System" OR "Satellite Navigation Aid" OR photogrammetry OR "Automatic Identification System" OR "Surveillance Systems" OR "Medical Imaging" OR "Laboratory Automation" OR radar OR "microwave radar" OR "millimiter wave radar" OR "image sensor" OR "radiometer" OR "radio wave") (measurement OR supervising OR observing OR monitor* OR acqui* OR autom* OR scan* OR classif* OR extract* OR estimat* OR forecast* OR diagnos* OR	System" OR "Surveillance Systems" OR "Medical Imaging" OR "Laboratory Automation" OR radar OR "microwave radar" OR "millimiter wave radar" OR "image sensor" OR "radiometer" OR "radio wave") (measurement OR supervising OR observing OR monitor* OR acqui* OR autom* OR scan* OR classif* OR extract* OR estimat* OR forecast* OR diagnos* OR segment OR track OR calibrat* OR sequenc* OR conver* OR process OR analyse OR reconstruct OR scrap*) W/2 (service OR inform OR "quality control" OR "process control") AND (task)) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010 AND		

		segment OR track OR calibrat* OR sequenc* OR conver* OR process OR analyse OR reconstruct OR scrap*) W/2 (service OR inform OR "quality control" OR "process control")) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010 AND INDEXTERMS (data* OR sens*)	INDEXTERMS (data* OR sens*)			
C	Software-based Data management	TITLE-ABS ((("Database system*" OR "Information Management" OR "Query process*" OR "Information retrieval" OR "Search engine*" OR "Digital storage" OR "Relational Database" OR "Application Programming Interface" OR "Graph Database" OR "Cryptograph*" OR "Data Security" OR "Blockchain" OR "Data encrypt*" OR "Data privacy" OR "Network security" OR "Embedded system*" OR "Map-reduce" OR "Mapreduce") W/2 (automat* OR autonomous* OR smart OR intelligen* OR "self-learn*" OR interact OR recogn* OR clean OR detect OR predict OR		**No automation, top 2%** INDEXTERMS (("data collection" OR "relational database" OR "digital storage" OR "distributed computer systems" OR "data acquisition" OR "data processing" OR "information management" OR blockchain OR "network	**Automation as separate, 50%** INDEXTERMS (automation AND ("data collection" OR "relational database" OR "digital storage" OR "distributed computer systems" OR "data acquisition" OR "data processing" OR	• After 2010

		identify OR generate OR classify OR acqui* OR stor* OR organi* OR access* OR retriev* OR extract* OR maintain* OR convert* OR encod* OR decod* OR encrypt* OR decrypt*) W/2 (manag* OR "problem solving" OR backoffice OR "back office" OR organis*)) OR ((data) W/1 (min* OR reduc* OR handl* OR integr* OR entr* OR enter* OR report* OR clean*) W/2 (manag* OR "problem solving" OR backoffice OR "back office" OR organis*)) OR (("big data") W/1 (analy*) W/2 (manag* OR "problem solving" OR backoffice OR "back office" OR organis*))) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2000 AND INDEXTERMS (data*)	architecture" OR "real time systems") AND ("laboratory information management system" OR "image watermarks" OR "information hiding" OR "enterprise resource planning" OR "program processors" OR "copyright protection" OR "watermarking relational databases" OR "electronic document identification systems" OR "scheduling" OR "pervasive monitoring" OR "interactive querying")) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010	"information management" OR blockchain OR "network architecture" OR "real time systems") AND ("laboratory information management system" OR "image watermarks" OR "information hiding" OR "enterprise resource planning" OR "program processors" OR "copyright protection" OR "watermarking relational databases" OR "electronic document identification systems" OR "scheduling" OR "pervasive monitoring" OR "interactive querying")) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010	

D	Computing	**Top 10%** TITLE-ABS-KEY (("retrieval system" OR "cloud computing" OR "quantum computing" OR "Edge computing" OR “fog computing” OR neuromorphic OR “ubiquitous computing” OR hpc OR “cluster comput*” OR "grid computing" OR "distributed computer system*" OR supercomput* OR super-comput* OR “super comput*” OR “high performance comput*”) W/4 (allocate* OR schedul* OR automat* OR efficien* OR autonomous* OR intelligen* OR recogn* OR detect OR predict OR identify OR retriev* OR extract* OR select OR estimat* OR decide OR solv* OR forecast* OR simul* OR optimis* OR stor* OR collect* OR calculate OR compu*) W/3 (“problem solving” OR “information management” OR “cognitive data management” OR “comput* capability” OR “comput* infrastructure” OR “resource schedul*” OR “flexible electronic*” OR “Large-scale Distributed System” OR “data visualization” OR “parallel application” OR “resource allocation” OR scheduling OR “iterative methods” OR “data storage” OR “drug discovery”)) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010 AND INDEXTERMS (cloud OR comput*)	**Top 7%** INDEXTERMS ((cloud OR comput*) AND ("cloud computing" OR "digital storage" OR fog OR "distributed computer systems" OR "internet of things" OR "data storage" OR "computing capacity" OR "real-time application" OR "real time system" OR "smart cities" OR "evolution algorithms" OR "data centre" OR "scale modeling" OR "nonlinear programming" OR "heuristic algorithm" OR "scheduling algorithm" OR "service provider" OR "apache hadoop") AND ("access control" OR "content distribution" OR "content management" OR "information modeling" OR "network-intensive applications" OR "resource management" OR "collusion attack" OR "fault detection" OR "video streaming" OR "video recording" OR "software as a service" OR "parallel processing" OR "statistical learning" OR "image analysis" OR "performance anomaly" OR mhealth OR multitasking OR profitability)) AND DOCTYPE (ar OR cp OR cr) AND PUBYEAR > 2010	

E	AI (not directly as a cloud service) & Intelligent Information System	TITLE-ABS-KEY (("Artificial Intelligence" OR "Machine learning" OR "deep learning" OR "neural networks" OR "support vector machine" OR "supervised learning" OR "unsupervised learning" OR “Reinforcement learning” OR “Foundation model*” OR “GAN” OR “Generative adversarial network” OR “large language model*” OR LLM) OR ((deep OR depth OR migration OR machine OR supervised OR unsupervised OR reinforcement OR network OR model OR transfer OR classifier OR algorithm) PRE/0 (learn* OR train*)) OR (((r?cnn) OR (cnn AND convolut*) OR dbn OR dnn OR ann OR lstm OR rnn OR "u net" OR gan OR gru OR crf OR ssd OR svm OR capsule) AND (network OR model OR algorithm)) OR (((recurrent OR recursive OR spike OR wavelet OR adversarial OR anti* OR deep OR belief OR capsule OR confrontation OR resistance OR countermeasure OR neural OR generative OR artificial OR convolut* OR elman OR rbf OR "feed forward") PRE/0 (network OR classifier OR classification)) AND (train* OR learn*)) OR ((incremental OR supervis* OR unsupervis* OR semi*supervis* OR machine OR deep OR depth OR statistical OR reinforcement OR ensemble) AND (learn*) AND (data*) AND (train*)) OR ((knn OR "nearest neighbor" OR "restrict* boltzmann" OR	**Top 10%** INDEXTERMS (("artificial intelligence" OR AI) AND ("machine learning" OR "feature select*" OR "learning system" OR "neural network" OR "decision tree" OR "detection method" OR "natural language processing" OR nlp OR "learning algorithm" OR "hidden markov models") OR ("adaptive boosting" OR "random forest" OR "markov models" OR "discriminant analysis" OR "support vector machine" OR svm OR "frequency domain analysis" OR "factorization model") AND ("discriminative feature" OR "nearest neighbor" OR "decision making" OR "numerical model" OR "medical decision making" OR forecasting OR optimization OR prediction OR "predictive analytics" OR mapping OR "performance assessment" OR "damage detection" OR decision theory OR sustainability OR "dynamical systems" OR "navigation system" OR "gene encoding" OR "gene expression" OR scheduling OR "classification accuracy" OR "pattern recognition" OR "pattern recognition problems" OR "real world environment" OR "urban planning" OR "behavioral research" OR "water management" OR "electricity market" OR "environmental management" OR "industry 4.0" OR "construction" OR "human resource management" OR "e-learning" OR "activity analysis" OR	

		"random forest" OR "decision tree" OR "naive bayes" OR svm OR "support vector machine" OR "bayesian network" OR “gradient descent”) AND (learn* OR train*)) OR ((learn* OR train* OR predict*) AND ((ml OR svm OR ai OR elm) W/1 (model OR algorithm))) OR ((adaboost OR xgboost) AND (train* OR learn*)) OR (((machine OR computer) PRE/1 vision)) OR (("self adapt*" OR adaptive OR intelligent OR "self learning" OR fuzzy OR pid OR plc OR "programmable logic") PRE/1 controller) W/3 (automat* OR autonomous* OR intelligen* OR “self-learn*” OR recogn* OR detect OR predict OR identify OR classify OR retriev* OR extract* OR select OR estimat* OR decide OR solv* OR synthe* OR discriminat* OR forecast* OR segment) W/3 (process OR “problem solving” OR decision OR feature OR “Natural Language Processing” OR “Machine Vision” OR “Image recognition” OR “Speech recogn*” OR “Text recogn*” OR “Pattern recogn*” OR “Pattern analysis” OR “Object detection” OR "machine translation" OR "grammar parser" OR "speech synthesis" OR "sentiment analysis" OR "sentiment score" OR "voice recogn*" OR "voiceprint recogn*" OR "feature extract*" OR classification OR "knowledge graph" OR "vector represent*" OR summar* OR "semantic	"automatic detection" OR "medical image segmentation" OR "process control" OR "process monitoring" OR "supply chains" OR "precision agriculture" OR "electronic trading" OR transportation OR "object detection" OR health)) AND DOCTYPE (ar OR cp OR cr) AND PUBYEAR > 2010	

		analysis" OR "semantic representation" OR "word segmentation" OR "text segmentation" OR "word embed*" OR "word vector" OR "key*word extract*" OR "key*phrase extract" OR "n*gram extract" OR "key*word search" OR "key*phrase search" OR OCR OR "optical recognition" OR "character recognition" OR "entity recognition" OR "object recognition" OR "expression recognition" OR "pattern recognition" OR "voiceprint recognition" OR "face recognition" OR "facial recognition" OR "automatic recognition" OR "speech recognition" OR "asr recognition") PRE/1 recognition))) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010 AND INDEXTERMS ("artificial intelligence")		
F	Additive manufacturing	**Top 7%** TITLE-ABS-KEY (("Additive manufacturing" OR "Additive layer manufacturing" OR "layered manufacturing" OR "three dimensional print*" OR "3d print*" OR "3-d print*" OR "3d-printing" OR "print* press*" OR "4d print*" OR bioprint* OR "bio-print*" OR "acoustic levitation" OR "deposition model*" OR "Computer Aided Design" OR "Rapid prototyping") W/3 (automat* OR autonomous* OR	INDEXTERMS (("Additive Manufacturing" OR "3d print*" OR "three dim* print*" OR "direct metal laser sintering" OR dmls OR "selective laser sintering" OR sls OR "high pressure die casting" OR dmld OR "Direct Metal Laser Deposition" OR "fused deposition model*" OR stereolithography OR "layered manufactur*") AND (textiles OR "methacrylic acid" OR eudragit OR "transition metals" OR nanofiber OR "silk fibroin" OR	

<table>
<tr><td></td><td></td><td>intelligen* OR simul* OR design OR interact OR extract* OR select OR print* OR synthe* OR model* OR craft* OR press* OR fabricat* OR "product design" OR "tissue engineering" OR "material testing" OR construction OR microfluidics OR aerospace OR integrat*) W/3 (process OR workflow OR design OR prototype OR streamline OR "shape optimisation" OR "laser melting")) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010 AND INDEXTERMS ("Additive manufacturing" OR "3d print*")</td><td colspan="2">"batch production" OR "drug delivery" OR biology OR "particle size analysis" OR "digital storage" OR "food processing" OR photopolymerization OR "tissue regeneration" OR "construction industry" OR bioink OR lysine OR "food processing" OR electrochemistry OR "surface treatment" OR polymers OR "ductile fracture" OR "shape memory" OR construction OR concretes OR "concrete construction" OR actuators OR "intelligent materials" OR carbides OR "hybrid materials" OR "synthetic fibers" OR "textile fibers" OR "titanium compounds" OR "methacrylic acid derivative" OR "silk fibroin" OR "vinyl derivative" OR "aerospace industry" OR "jet engines" OR "artificial organs" OR "dental equipment" OR "polymeric implants" OR ophthalmology OR optics OR spectacles OR "aeronautical components")) AND DOCTYPE (ar OR cp OR cr) AND PUBYEAR > 2010</td><td></td></tr>
<tr><td>G</td><td>Networking</td><td>Top 15%
TITLE-ABS-KEY ((“Internet of Things” OR iot OR rfid OR “radio frequency identification” OR “wireless telecommunication” OR “mobile communication” OR 5G OR “wireless network*” OR “network</td><td>INDEXTERMS (("wireless sensor
network" OR wsns OR "po wer grids" OR "smart grid" OR "wireless</td><td></td><td></td></tr>
</table>

		architecture” OR “wireless sensor networks” OR “telecommunication system*” OR zigbee OR “virtual private connection” OR “near-field comm*” OR NFC) W/4 (automat* OR autonomous* OR intelligen* OR recogn* OR detect OR identify OR retriev* OR extract* OR select OR decide OR optimi* OR communicat* OR schedule OR comm* OR “self-div*” OR agriculture OR warehouses OR “resource allocation” OR “information management” OR “network protocols” OR “radio broadcasting” OR “remote monitoring”) W/4 (process OR “problem solving” OR interoperability OR monitor* OR control* OR broadcast OR connect* OR trasmission)) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2000 AND INDEXTERMS (network* OR iot OR “internet of things” OR wireless OR comm* OR VPN OR NFC)	sensor" OR "internet of things" OR iot OR "routing protocols" OR "distributed computer systems" OR "radio frequency identification" OR rfid OR " mobile telecommunication systems" OR "mimo systems" OR "sensor communication" OR "vehic ular sensor network" OR "vehicular wireless" OR internet) AND (“scheduling” OR "signal processing" OR irrigation O R middleware OR "smart city" OR "industrial management" OR "precisio n agriculture" OR "financial transactions" OR "industria l automation" OR parking O R "micro-climate		

			monitoring" OR "urban traffic" OR "message transmissions" OR "water management" OR "sensing data" OR “patient monitoring” OR interoperability)) AND DOCTYPE (ar OR cp OR cr) AND PUBYEAR > 2010		
H	User interface	**Top 10%** TITLE-ABS-KEY (("user interface" OR "interactive computer system" OR "human engineering" OR "computer interface" OR "augmented reality" OR "mixed reality" OR "haptic interfaces" OR haptics OR "computer aided interaction" OR haptic* OR telehaptic* OR input OR "computer keyboards" OR "virtual keyboards" OR "touch screens" OR mouse OR "pen input" OR "pointing device*" OR "display device*" OR touchscreen* OR "touch screen*" OR "display devices" OR "multi-touch" OR "sound synthesis" OR "noise cancellation" OR neuroscan* OR neuro-scan* OR neuroimaging OR “motion	INDEXTERMS (("user interface" OR "human computer interaction" OR "Human Robot Interaction" OR "computer interface" OR "brain computer interface" OR "man machine system" OR "human engineering" OR "human machine interaction" OR "hci system" OR "human computer interfaces") AND ("augmented reality" OR "virtual reality" OR ergonomics OR "mixed reality" OR "mover's distance" OR "interface state" OR "helmet mounted displays" OR "three dimensional computer graphics" OR smartphones OR "depth camera" OR "graphical user interfaces" OR "interfaces computer") AND("hand shape" OR "hand detection" OR "eye tracking"		

		sensing" OR kiosk*) W/3 (automat* OR autonomous* OR intelligen* OR interact OR recogn* OR detect OR process OR identify OR assist OR display OR simul* OR synth* OR "E-learning" OR augment OR visual* OR scan OR "data acquisition" OR "computer music" OR display) W/3 (process OR "problem solving" OR interface OR engagement OR "decision making" OR interaction)) AND DOCTYPE (ar OR cp OR cr OR re) AND PUBYEAR > 2010 AND INDEXTERMS (interface OR interact* OR reality)	OR "motion tracking" OR "palmprint recognition" OR "hand gesture" OR "hand-gesture recognition" OR "industrial research" OR "industry 4.0" OR "integration testing" OR "feature extractor" OR "temporal pooling" OR "visual surveillance" OR "image segmentation" OR "stereo image processing" OR "stereo vision" OR "spatial temporal" OR Kinect OR biomechanics OR "patient monitoring" OR prosthetics OR neurofeedback OR rehabilitation OR "computer game" OR "distance metrics" OR "shape matching" OR "emotion recognition" OR visualization OR diagnosis OR "person re identifications" OR ergonomics OR diseases OR "patient treatment" OR "roads and streets" OR vehicles OR "attention deficit disorder")) AND DOCTYPE (ar OR cp OR cr) AND PUBYEAR > 2010	

Figure 5-A: Selection of documents for literature review: robots

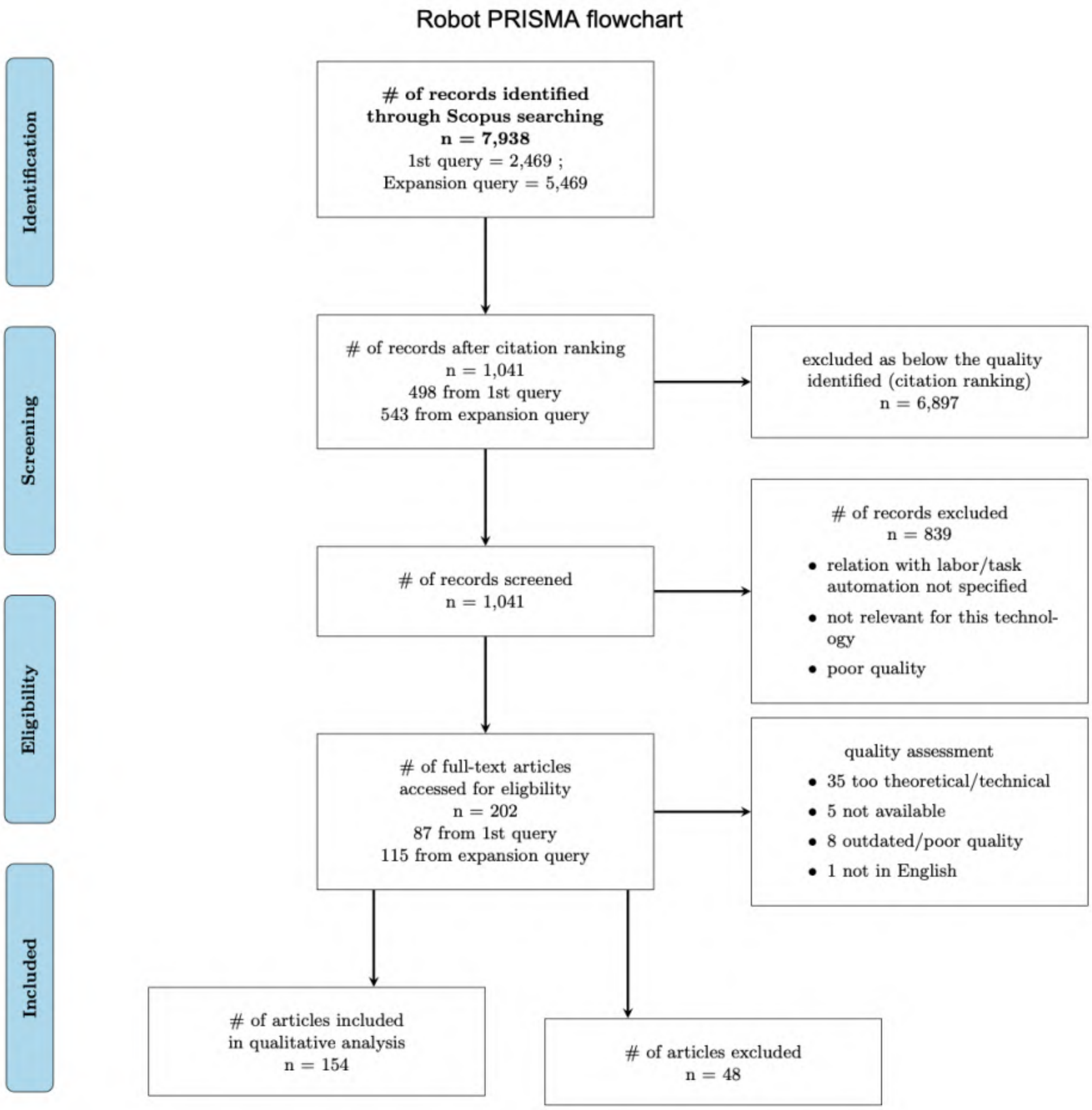


Figure 6-A: Selection of documents for literature review: software based data management