Growth Welfare Innovation Productivity

Working Paper

Robots and the origin of their labour-saving impact

Fabio Montobbio
Department of Economic Policy, Università Cattolica del Sacro Cuore; BRICK Collegio Carlo Alberto; ICRIOS Bocconi University

Jacopo Staccioli
Department of Economic Policy, Università Cattolica del Sacro Cuore; Institute of Economics, Scuola Superiore Sant'Anna

Maria Enrica Virgillito
Institute of Economics, Scuola Superiore Sant'Anna; Department of Economic Policy, Università Cattolica del Sacro Cuore

Marco Vivarelli
Institute of Economics, Scuola Superiore Sant'Anna; Maastricht Economic and Social Research Institute on Innovation and Technology; Forschungsinstitut zur Zukunft der Arbeit GmbH

3/2020 February

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

Robots and the origin of their labour-saving impact

Fabio Montobbio[*] [†‡§] Jacopo Staccioli[*] [¶]

Maria Enrica Virgillito[¶*] Marco Vivarelli[*||**]

4th February 2020

Abstract

This paper investigates the presence of explicit labour-saving heuristics within robotic patents. It analyses innovative actors engaged in robotic technology and their economic environment (identity, location, industry), and identifies the technological fields particularly exposed to labour-saving innovations. It exploits advanced natural language processing and probabilistic topic modelling techniques on the universe of patent applications at the USPTO between 2009 and 2018, matched with ORBIS (Bureau van Dijk) firm-level dataset. The results show that labour-saving patent holders comprise not only robots producers, but also adopters. Consequently, labour-saving robotic patents appear along the entire supply chain. The paper shows that labour-saving innovations challenge manual activities (e.g. in the logistics sector), activities entailing social intelligence (e.g. in the healthcare sector) and cognitive skills (e.g. learning and predicting).

JEL classification: O33, J24, C38.

Keywords: Robotic Patents, Labour-Saving Technology, Search Heuristics, Probabilistic Topic Models.

[*]Department of Economic Policy, Università Cattolica del Sacro Cuore, Via Necchi 5 – 20123 Milano, Italy.
[†]BRICK, Collegio Carlo Alberto, Piazza Arbarello 8, 10122 Torino, Italy.
[‡]ICRIOS, Bocconi University, Via Röntgen 1, 20136 Milano, Italy.
[§]To whom correspondence should be addressed: ✉ fabio.montobbio@unicatt.it
[¶]Institute of Economics, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33, 56127 Pisa, Italy.
[||]Maastricht Economic and Social Research Institute on Innovation and Technology, UNU-MERIT, Boschstraat 24, 6211 AX Maastricht, The Netherlands.
[**]Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA), Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany.

1 Introduction

The increasing diffusion of artificial intelligence (hereafter, AI) and robotic technology in the last decade has become a renewed object of analysis in both economics and innovation studies, with contributions reporting a steep adoption of automated processes and industrial robots (e.g. Acemoglu and Restrepo, 2019a; IFR, 2017). The impact of automation and robotics on employment has generated concerns and vibrant debates as well (Autor, 2015; Brynjolfsson and McAfee, 2016; Frank et al., 2019; IFR, 2017). Indeed robots (and intelligent robots more so) are technologies that, within the recent Industry 4.0 wave, are particularly apt to substitute human labour.[1] In fact, on top of standard robotics, AI allows robots to perform an increasing variety of tasks and functions (e.g. Frey and Osborne, 2017; Webb, 2020). Intelligent robots are able to 'sense' and communicate with their environment (e.g. machine-to-machine communication) and operate as mobile, interactive information systems in a wider spectrum of fields, from manufacturing to service sectors (e.g. hospitals, retail outlets).

Pressured by the threat of a new wave of technological unemployment, the extant literature has focussed on both the quantity of jobs potentially displaced by robots, mainly by taking advantage of the International Federation of Robotics (IFR) dataset (e.g. Acemoglu and Restrepo, 2019a,b; Graetz and Michaels, 2018), and on the specific functions and tasks that automation might directly substitute, mainly relying on the U.S. Occupational Information Network (O*NET) (e.g. Autor and Dorn, 2013; Frey and Osborne, 2017). Both streams of literature agree that the most vulnerable occupations and tasks are those performed by low- and medium-skilled workers, mainly executing routinised tasks. More recently, attention has been paid to the impact from AI, which instead appears to be more pervasive for jobs and wage security of high-skilled professionals (Webb, 2020).

Notably, most economic analyses so far have investigated the employment impact of industrial robots on the *adopting* sectors of the economy, chiefly focussing on manufacturing industries (e.g. automotive, electronics, chemicals). Conversely, evidence lacks when it comes to sectors of *origin* of robotic innovations. This constitutes a first gap in the extant literature which the present paper aims at filling. First, since robots are extremely heterogeneous and complex artefacts, it is important to reconstruct their wider technological origin and composition. Second, scale and scope economies, and the position of innovators in the vertical supply chain can influence the nature, rate, and trajectory of innovative activity. For instance, innovations in robot-related technologies can be produced upstream in research intensive labs (e.g. Biomimetics Robotics Lab at MIT) or downstream in large adopters (e.g. Amazon). Clearly, the position of the sector of origin along the vertical supply chain might impact the very nature of a technological artefact. Ultimately, it is important to understand whether innovation in robotics leads to a new product or service (with a potential positive effect on labour demand) and/or to a new labour-saving process.

Indeed, a second major gap in the existing literature lies in the very scant evidence on the origin of innovations which are explicitly meant to be labour-saving. On the one hand,

[1] For a field work analysis on the adoption of Industry 4.0 technological artefacts and their effect on skills and task composition see Cirillo et al. (2018) and Moro et al. (2019).

whether labour-friendly or labour-saving in their use is a question that invests specific technological adopters; on the other hand, a different issue concerns the extent to which technological inventors explicitly manifest labour-saving heuristics when *conceiving* novel robotic technology. Identifying the existence of explicit labour-saving heuristics during the invention process might allow for a fine-grained discovery of functions and tasks which (intelligent) robots are intended to replace.

The present paper addresses these research questions by exploiting all 3,557,435 USPTO patent applications (hereafter, patents) between 1st January 2009 and 31st December 2018, and analysing their full-texts. First, it identifies and describes the wider spectrum of robotics-related technology, it singles out labour-saving (hereafter, LS) innovations, and retraces their origin in terms of specific technological content, geographical location, and positioning of the innovators along the supply chain. Second, it explores the activities in which LS robotic patents are concentrated, linking the technological content to the human tasks intended to be replaced.

We bring novelty both in terms of research questions and methodology. In particular our paper faces three methodological challenges.

(i) What is an innovation in robotics? Robots have many components and interact in a complex way with their environment. It is therefore necessary to understand innovation in robotics-related technologies by defining a broad set of robotic patents that goes beyond the classification codes attributed by patent officers, in order to encompass innovations in complement technologies and process implementations. We tackle this problem by means of an original keyword search in patent texts.

(ii) How can we extract LS innovations and understand their origins? Differently from the extant literature, largely focussing on the economic impact from adopting robots, we look for explicit LS heuristics in the knowledge generation of robotic innovations. In so doing, we rely on textual analysis of patent applications. We perform an in depth semantic study in order to distinguish patents which explicitly claim a direct LS impact. In so doing, we restrict the analysis to a semantic domain which is uncommon to inventors. In particular, we look for a restricted dictionary of words typical of the 'economic slang'. Once LS innovations are identified, we are also able to pinpoint the underlying firms and inventors, their location, and the economic industry of origin.

(iii) We further ask whether LS patents are particularly concentrated in specific domains of robotics-related activities. To answer this question, we estimate a probabilistic topic model, a natural language processing method from the unsupervised machine learning toolbox. In so doing, we are able to measure the frequency of occurrence of semantic topics in LS and overall robotic patents. As a result, our paper generates a human-machine taxonomy which characterises the topics more relevant to LS patents and helps in understanding which specific activities and functions are more exposed to LS innovation.

In a nutshell, our paper shows that the overall number of robotic patents has rapidly increased (3-fold) over the past decade, while LS patents display no specific trend. This supports the idea that the LS property of robotic patents is a rather established heuristics. At the country level, U.S. and Japan appear to largely dominate other countries (although this might be biased by the use of USPTO patents); however, China exhibits a catch-up process. LS robotic patents are largely concentrated in few dominant industries, showing a typical long tail distribution, characterizing cumulative processes (Newman, 2005). Nonetheless, they are quite pervasive in terms of penetration by spanning virtually the entire 2-digit NAICS spectrum. In terms of positioning along the supply chain, LS patent holders are not just constituted by robots producers, but mainly robot adopters. Two archetypical cases are Amazon and UPS. Therefore, LS robotic patents emerge along the entire supply chain, signalling a considerable degree of diffusion.

Moreover, we show that LS patents do not distribute uniformly across all fields as robotic patents. Instead, they cluster in largely human-intensive industries, such as logistics, medical, and health activities. We emphasise in which specific (robotics-related) fields of activity LS patents are relatively more concentrated and position our results within the literature devoted to the analysis of tasks and occupations particularly exposed to automation (which includes, among others, Frey and Osborne, 2017; Webb, 2020).

The remainder of the paper is organised as follows. Section 2 discusses the relevant literature and theoretical framework. Section 3 presents our data and the empirical methodology. Section 4 discusses our results. Finally, Section 5 concludes.

2 Theoretical motivation: search heuristics and labour-saving trajectories

The impact of automation on jobs has (again) become one of the 'trending' topics within both the academic and policy debate. Indeed, fears of technological unemployment have been always accompanying great innovative waves. However, in the history of humanity, periods of intensive LS automation have also coincided with the emergence of new jobs, tasks, activities, and industries.[2] Nevertheless, this time may be different since nowadays the World is on the edge of a new technological revolution (see Bartelsman et al., 2019), dramatically accelerating in the direction of automation driven by pervasive diffusion of robots and AI (see Brynjolfsson and McAfee, 2012, 2016; Frey and Osborne, 2017). Turning the attention to the economic literature, a very recent strand (e.g. Acemoglu and Restrepo, 2018a,b, 2019b; Graetz and Michaels, 2018) accounts for the LS effect of robotisation by looking at indirect measures of penetration (i.e. number of robots acquired within different economic sectors, provided by IFR data). A drawback therein is the likely understatement of the actual role of producers and adopters in assessing the labour impact of robots. Additionally, the use of the number of robots per workers, available at the country/industry level, does not allow to properly dissect the firm-level and local impact of robots. Finally, there is lack of a coherent taxonomy of the knowledge base underlying robotic artefacts and their production/diffusion in a given economy.

[2] In more detail, when a process innovation is introduced, potential market compensation mechanisms are also triggered and these may counterbalance the initial LS impact of innovation (see Dosi and Mohnen, 2019; Freeman and Soete, 1987; Piva and Vivarelli, 2018; Simonetti et al., 2000; Van Roy et al., 2018; Vivarelli, 1995). The present paper focusses on the detection of possible LS heuristics linked to the production and diffusion of robots, while the investigation of the price and income compensation mechanisms lies beyond the scope and aims of the present study (for recent surveys centred on the compensation theory, see Calvino and Virgillito, 2017; Ugur et al., 2018; Vivarelli, 2014).

In the tradition of the economics of innovation, technologies are studied by means of identification of paradigms and trajectories (Dosi, 1982, 1997) underlying the introduction, development, and diffusion of a given artefact. A notable question regards the extent to which the discovery of a given artefact occurs by chance, or it is alternatively driven by some specific *search heuristics* or, put in the words of Rosenberg (1976), *focussing devices*, namely the ensemble of technological bottlenecks, market incentives, and ultimately the cognitive *loci* and the behavioural patterns of who creates those technologies (Dosi and Nelson, 2010, 2013).

Although it is generally hard to identify invariant and ex-ante search heuristics or inducement effects, a specific search heuristic appears to be invariant throughout the history of capitalist societies, namely search efforts aimed at the reduction of human inputs in production.[3] Karl Marx was clear on the point, highlighting how labour resistance, organisation, and claims represent powerful drivers towards mechanisation:

> "In England, strikes have regularly given rise to the invention and application of new machines. Machines were, it may be said, the weapon employed by the capitalists to equal the result of specialised labour. The *self-acting mule*, the greatest invention of modern industry put out of action the spinners who were in revolt. If combinations and strikes had no other effect than of making the efforts of mechanical genius react against them, they would still exercise an immense influence on the development of the industry."
>
> [Marx (1956, p. 161); also cited in Rosenberg (1976, p. 118)]

Granted the pervasiveness of LS heuristics in the space of technological search (Dosi, 1988; Rosenberg, 1976; Tunzelmann, 1995), in the following we aim at understanding whether the current wave of technological innovation is dominated by such heuristics. In our framework, robotics, and indirectly AI, are seen as pervasive *general purpose technologies*, with massive potential in terms of labour substitution across a wide range of skills, occupations, and tasks (see Bresnahan and Trajtenberg, 1995; Cockburn et al., 2018; Trajtenberg, 2018). Patents, as a locus of explicit codified knowledge, represent an appropriate empirical instrument to proxy the rate and direction of innovative activity (Pavitt, 1985). By looking at the textual contents of robotic patents, we aim at isolating the ones explicitly embedding a LS trait. In order to make the identification process as neat as possible, we define a dictionary of words and resort to a semantic analysis. Two excerpts from LS patents follow:

> "Automated systems, such as robotic systems, are used in a variety of industries to **reduce labo[u]r costs and/or increase productivity**. Additionally, the use of human operators can involve increased cost relative to automated systems." [US20170178485A1]

[3] In turns, these efforts are dynamically reinforced by the localised, path-dependent, and irreversible nature of technological progress (see Atkinson and Stiglitz, 1969; Capone et al., 2019; David, 1985; Dosi, 1988).

> "The use of the technology [robots] results in improved management of information, services, and data, increased efficiency, significant reduction of time, **decreased manpower requirements**, and substantial cost savings."
>
> [US20100223134A1]

As we shall reveal in the following, the type of analysis we conduct allows to extend the analysis well beyond the use of the IFR dataset. In fact, leveraging patent data on robotic artefacts, we are able to identify the knowledge generation patterns behind this technology. In so doing, we do not restrict to patents entailing robotic artefacts only as *products*, but also as *processes* (i.e. methods). In this respect, our strategy opens up the possibility of looking at patenting patterns of both robot producers and firms involved in any sort of complementary innovation or developing processes which implement robotic technology. Indeed, the most disruptive impact of robots is plausibly occurring among downstream, non-robotic firms through *embodied* technological change within process innovation (see Barbieri et al., 2018; Dosi et al., 2019; Pellegrino et al., 2019). Clearly, by focussing on patenting firms we inevitably circumscribe our attention to most innovative robots adopters, i.e. large-scale, multi-product firms who have in-house capabilities of integrating processes aimed at cost-cutting and increasing efficiency. In this way, not only we capture those firms who 'know exclusively what they produce', (i.e. robotic firms), but also firms who 'know more than what they produce', namely non-robotic firms holding robotic patents (Dosi et al., 2017; Patel and Pavitt, 1997).

The underlying hypothesis is that robot-as-a-product and robot-as-a-process innovations embodying LS heuristics are patented by different types of firms and present different degrees of pervasiveness in terms of technological diffusion. In fact, LS heuristics might be more deeply rooted in firms located outside robots manufacturing, such as downstream, large-scale adopters.

3 Data and methodology

Our analysis covers the entire set of 3,557,435 patent applications filed at the USPTO between 1st January 2009 and 31st December 2018. Full-texts have been downloaded from the USPTO Bulk Data Storage System[4]. Roughly 350k applications are filed on average each year, showing no clear trend, as depicted in Fig. 1. We match our data to the ORBIS (Bureau van Dijk) database through the relevant patent publication numbers.

Our methodology consists of three steps. First, we single out patents which either directly or indirectly relate to robotics technology. Second, we implement a procedure to detect the underlying LS heuristics and pinpoint the set of explicitly LS patents. Finally, we estimate a probabilistic topic model in order to devise a human-machine taxonomy.

To the best of our knowledge, two extant contributions are methodologically comparable to ours. Dechezleprêtre et al. (2019) devise a multi-step strategy based on both patent classification and multiple keyword search in order to identify automation patents. Mann

[4] Available here: https://bulkdata.uspto.gov/

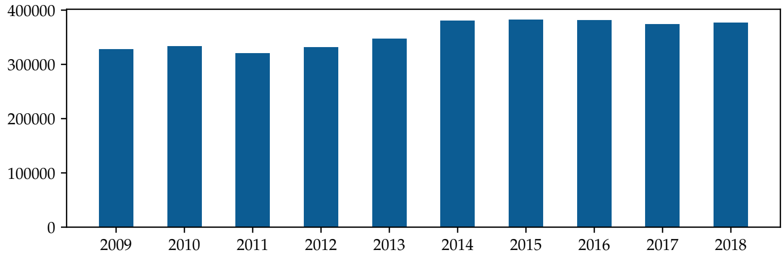


Figure 1: Number of patent applications to USPTO by year.

and Püttmann (2015) manually classify a sample of patents and then use a machine learning algorithm to elicit automation innovations from a larger population.

Our empirical strategy differs from these contributions in many respects. First, we identify a structured dictionary of words entailing a semantic procedure (including a predicate, a direct object, and an object's attribute), rather then a simple word search, which targets LS robotic patents, rather than generic automation ones. In so doing, we overcome the mere search for n-grams and word adjacency, we avoid arbitrary use of the sheer occurrences of 'automation + something', and we control for type I errors. Second, our procedure is general-to-specific, in that we do not restrict our attention to any ex-ante classification by patent examiners; instead, we uncover the entire population of robotic patents identified by *both* patent classification codes and content of patents themselves. Indeed, our search is conducted on the whole population of patents, independently of technological and sectoral classification. Third, in order not to retain false positives, we perform an ex-post validation of *all* patents flagged by the above procedure. Our approach turns out to be more restrictive in terms of requirements, but more comprehensive in terms of technology-industry spectrum. In fact, as we shall see, we are able to span virtually the entire NAICS sectoral classification. In contrast, simply relying on patent officers' classification generally induces a downward bias, since it would only capture those patents mainly associated with robot manufacturers, while patents related to robotics complementary technologies and specific process implementations along assembly lines would be inevitably lost. To overcome this limitation, we have enlarged our scope of analysis beyond the official USPTO classification.

3.1 Robotics patents

In addressing our first methodological challenge, we set up two distinct criteria, one based on the patent classification codes specified within applications, the other based on textual keyword search. A patent is labelled as *robotic* if it obeys at least one of the criteria.

For the first criterion we make use of the official USPTO statistical mapping[5] between former U.S. Patent Classification (USPC) class `901`, entirely dedicated to "Robots" and used by a number of previous contributions to identify robotic patents[6], and the Cooperative

[5] Available here: https://www.uspto.gov/web/patents/classification/cpc/html/us901tocpc.html
[6] Among others, Obama's 2016 "Economic Report of the President": https://www.nber.org/erp/ERP-2016.pdf

Patent Classification (CPC) present in recently published USPTO applications. The concordance table lists 5 distinct CPC codes for each of the 50 subclasses of USPC class `901`, hence a total of 250 target codes, 124 of which are unique when all digits are considered. The CPC system also defines a 'legacy' meta-class `Y10S` which targets "Technical subjects covered by former USPC cross-reference art [...] and digests"; CPC group `Y10S901` then serves as a junction for former art classified as USPC class `901`. There are 50 unique CPC codes in the group, one for each of the original subclasses. Since we wish to embrace the broadest possible definition of robotic technology, the sufficient condition for a patent to be deemed robotic under the first criterion is that the patent exhibits at least one of the 174 mentioned (full-digit) CPC codes (124 from the statistical mapping plus the 50 legacy codes in group `Y10S901`).

Our second criterion looks for the multiple occurrence of the morphological root 'robot' within either section of the full-text of a patent, i.e. including abstract, description, and claims. Looking for this morphological root seems appropriate since it displays a very low degree of ambiguity. Broadly speaking, words that contain the (sub)string `'robot'` are remarkably likely to refer to some sort of robotic technology. However, we initially found that patents with a small number of occurrences therein may well be unrelated to robotics as their core technology; a typical example is innovation in the design of golf balls which makes use of a robotic embodiment merely for final testing, for instance in launching the ball at a certain speed. This prompted a first manual inspection on our behalf. We draw a random sample of a few hundreds patents containing at least one occurrence and order them by the number of total occurrences in order to manually identify an appropriate cut-off. We find that patents with 5 or more occurrences are already quite likely to describe either a core robotic technology, its process implementation, or a close complement technology, with few outliers. We conservatively set the cut-off of our second sufficient condition to 10.

3.2 Labour-saving patents

Our second methodological challenge lies in the discovery of the set of LS patents. From the set of robotic patents identified in the previous section, we now want to single out those which explicitly claim a LS effect of the underlying innovation. We do this by performing a multiple *word* co-occurrence query at the *sentence* level.

To this purpose, we need to preprocess our textual corpus, along the following steps. First, we subdivide, technically *tokenise*, the full-text of each robotic patent (a single `string` concatenating the abstract, the description, and the claims sections) into a `list` of sentences by means of a punctuation `regexp`. Second, we similarly tokenise each sentence into a `list` of words. Third, we filter out a standard set of 182 stop-words, i.e. tokens that are overly common in English (such as `'a'`, `'the'`, `'if'`, ...) and do not convey any useful information to our analysis. Last, we reduce each word in each sentence to its morphological root, by means of a *stemming* algorithm.[7]

[7] In particular, we use the `Porter2` stemmer, an improved version of the original Porter (1980) algorithm, as implemented in the `nltk` Python library.

$$\underbrace{\begin{bmatrix}\texttt{`reduc'}\\ \texttt{`replac'}\\ \texttt{`elimin'}\\ \texttt{`save'}\\ \texttt{`lower'}\\ \texttt{`substitut'}\\ \texttt{`autom'}\end{bmatrix}}_{\text{verbal predicate}} \times \underbrace{\begin{bmatrix}\texttt{`labor'}\\ \texttt{`worker'}\\ \texttt{`human'}\\ \texttt{`employe'}\\ \texttt{`manpow'}\\ \texttt{`job'}\end{bmatrix}}_{\text{direct object}} \times \underbrace{\begin{bmatrix}\texttt{`cost'}\\ \texttt{`expenditure'}\\ \texttt{`expens'}\\ \texttt{`hour'}\\ \texttt{`intens'}\\ \texttt{`task'}\\ \texttt{`time'}\\ \texttt{`skill'}\end{bmatrix}}_{\text{object attribute}}$$

Figure 2: Structure of the *labour-saving* textual query.

At this point we are able to look for the presence of specific words (actually morphological roots, after the aforementioned stemming step) within the whole corpus of robotic patents. We aim at eliciting the heuristic, when present, that the technology described in a patent may somehow reduce human labour requirements if implemented, either in terms of labour cost, worked hours, or the complete substitution of the workers themselves, by automating one or more skills/tasks they previously applied/performed. Accordingly, we develop a methodology by which we scour all the identified sentences and look for the co-occurrence of a certain *verbal predicate*, a *direct object*, and an *attribute*, which jointly convey the desired message, within the same sentence. Fig. 2 shows the selected words we use in our query. In practice, we look for the joint occurrence of a *triplet* of words (which differ from *trigrams*, as we do not require word adjacency), one from each set, within the same sentence, and flag the associated patent as *potentially* LS if at least one sentence contains at least one of the (336, given the Cartesian product of the three sets) triplets.

3.3 Probabilistic topic model and human-machine taxonomy

The selection of LS patents can be used to describe the set of human activities which LS innovations aim at substituting. Thus, our next step entails the technological characterisation of the set of LS patents vis-à-vis the whole class of robotic patents. In principle, this could be done by looking at the CPC codes that each filed application comes with. However, when multiple classification codes are attributed to the same patent, there's no way to assess the actual relevance of each code, either cardinally or ordinally. To overcome this limitation, we estimate the relevance of each CPC code to each patent by leveraging the latent semantic structure of the whole collection of patents' full-text, as identified by a probabilistic topic model.

3.3.1 Probabilistic topic models

A probabilistic topic model (see Blei, 2012; Blei et al., 2003) is a powerful natural language processing tool in the unsupervised machine learning realm which aims at eliciting and quantifying the magnitude of the main subject matters underlying a collection of documents in a fully automated way. Latent Dirichlet Allocation (LDA) is the simplest such model and the one we will use in our analysis. Formally, LDA is a generative probabilistic model of a collection of documents. The underlying assumption is that each document is represented by a random mixture over latent topics and each topic is characterised by a distribution over a fixed vocabulary of words. The intuition is that each document exhibits multiple topics in different proportions; in the generative model, each word in each document is drawn from one of the topics proportionally to their relevance. The generative process for LDA can be represented by the following joint distribution

$$p(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{w}) = \prod_{k=1}^{K} p(\beta_k) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(w_{d,n}|\boldsymbol{\beta}, z_{d,n}) \right) \tag{1}$$

where $\boldsymbol{\beta}$ is the unknown set of K underlying topics β_k, $k = 1, \ldots, K$; $\boldsymbol{\theta}$ is the unknown set of topic proportions $\theta_{d,k}$ for topic k in document d of the collection D; $\boldsymbol{z}$ is the unknown set of topic assignments $z_{d,n}$ for the n-th word in document d; finally, $\boldsymbol{w}$ denotes the observable set of documents, each represented by the underlying sequence of words $w_{d,n}$. LDA is essentially a Bayesian estimator for the *posterior* conditional distribution of the topic structure $p(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{z}|\ \boldsymbol{w})$ given the observed documents $\boldsymbol{w}$. Note that in $\boldsymbol{w}$ the multiplicity of each word is relevant although the specific order in which the words arise is neglected, as per the so-called *bag-of-words* assumption. Crucially, topic model algorithms do not require any prior annotations or labelling of documents, as the topics emerge simply from the analysis of the original texts.

Probabilistic topic models for the analysis of patent data have been occasionally adopted in the past. Venugopalan and Rai (2015) map American solar photovoltaics patents to probability distributions over real world categories and show that linguistic features from topic models can be used to identify the main technology area that a patent's invention applies to in a more effective way compared to traditional classification systems. Lee et al. (2015) attempt to predict the pattern of technology convergence and use a topic model on triadic patents to discover emerging areas of the predicted technology convergence. Chen et al. (2017) and Kim et al. (2015) employ a topic modelling approach for technological trajectories forecasting. Kaplan and Vakili (2015) exploit topic modelling to study the formation of new topics in patent data (e.g. regarding fullerenes and carbon nanotubes) and locate the patents which introduce them.

3.3.2 Topic model estimation

Our analysis proceeds along the following methodological workflow. First, we estimate a topic model on the whole population of robotic patents. This step associates a distribution θ_d of membership over the K-dimensional set $\boldsymbol{\beta}$ of topics to each patent d.

Second, we associate to each topic β_k a distribution of CPC codes, by weighting the original attribution of codes to each patent by the topic proportions θ_d found in the previous step. This informs the labelling of each topic with a quantitative combination of pre-defined technological classes.

Finally, we compare the relevance of each topic for the whole population of robotic patents with that of the subset of LS patents and we draw quantitative conclusions on which technologies are relatively more and less relevant in characterising the latter with respect to the former.

When estimating a topic model, the only relevant parameter the modeller is asked to provide is the number K of topics the model is supposed to identify. There is no general theory in the literature on how to appropriately select the number of topics. There have been a few attempts to address this issue (see e.g. Arun et al., 2010; Cao et al., 2009) but most scholars agree that none is truly universal. The current good practice is to run multiple experiments with different values of K and select the most convincing one. This is less relevant for us however, since we are more interested in characterising the macroscopic technological differences of the two patent corpora, rather than in the specific technological characterisation of each of the two classes. We opt for a relatively low $K = 20$, which allows to maintain overall tractability of exposition.

The estimation of a topic model is an iterative process, as the model incrementally learns the latent semantic structure of the textual corpus and refines the estimate of the underlying topic distribution at every iteration. In our implementation, we feed the whole textual collection to the learning step at each iteration.[8] A typical fitness measure for topic models is the so-called *perplexity*, defined as the inverse of the geometric mean per-word log-likelihood. Perplexity typically decreases at each learning iteration. Rather than imposing ex-ante a fixed number of iterations to the algorithm, we opt for computing the perplexity of the model at each step and terminating the learning process once the perplexity gain runs below a certain threshold, which we conservatively set at 0.1. In our experiment the threshold is reached after 52 iterations and final perplexity equals 667.9838.

The algorithm returns each topic β_k as a list of relevant keywords and a membership value $\theta_{d,k}$ of each patent d to topic k. These membership measures are distributions, in the sense that

$$\theta_{d,k} \geq 0 \qquad \forall\, k = 1,\ldots,K; \quad \forall\, d = 1,\ldots,D \tag{2}$$

$$\sum_{k=1}^{K} \theta_{d,k} = 1 \qquad \forall\, d = 1,\ldots,D \tag{3}$$

Since we are interested in characterising the whole collection of robotic patents, and later the subset of LS patents, we need to construct an aggregate measure of relevance Θ_k of each topic k for an arbitrary collection of D documents. We define this measure as the simple average membership of all documents to each topic, as follows:

$$\Theta_k := \frac{1}{D}\sum_{d=1}^{D} \theta_{d,k} \qquad \forall\, k = 1,\ldots,K \tag{4}$$

From the properties in eqs. (2) and (3), it is straightforward to prove that Θ_k is itself a

[8] Specifically, we use the `LatentDirichletAllocation` module implemented in the `scikit-learn` Python library in `batch` learning mode.

distribution, i.e.

$$\Theta_k \geq 0 \qquad \forall\, k = 1, \ldots, K \tag{5}$$

$$\sum_{k=1}^{K} \Theta_k = 1 \tag{6}$$

3.3.3 Topic labelling with CPC codes

Estimating a topic model returns, for each topic, a list of most important words therein, ranked by their frequency within the collection. The interpretation of each topic is then entirely left to the modeller. Rather than labelling the topics on the basis of obtained keywords alone, we leverage the original attribution of ex-ante equally relevant CPC codes to each patent, together with the membership of each patent to the set of topics. In particular, letting C denote the set of all CPC codes, we define a *cardinal* membership distribution $\Phi_{c,k}$ of each CPC code $c \in C$ to each topic $k = 1, \ldots, K$ as

$$\Phi_{c,k} = \frac{\varphi_{c,k}}{\sum_{k=1}^{K} \varphi_{c,k}} \qquad \forall\, k = 1, \ldots, K; \quad \forall\, c \in C \tag{7}$$

where

$$\varphi_{c,k} = \sum_{d \,\in\, D} \mathbf{1}_{\{c \,\in\, \gamma(d)\}} \cdot \theta_{d,k} \qquad \forall\, k = 1, \ldots, K; \quad \forall\, c \in C \tag{8}$$

$\mathbf{1}_{\{\cdot\}}$ denotes the indicator function and $\gamma(\cdot)$ is a fictitious function which returns the relevant CPC codes originally attributed to the argument patent d by the patent examiner. $\varphi_{c,k} \in [0, +\infty]$ is a (unscaled) measure of membership of CPC code c to topic k, and $\Phi_{c,k} \in [0, 1]$ is the corresponding rescaled version to fit the unit interval.[9] In other words, all the CPC codes of a patent ($\mathbf{1}_{\{c \,\in\, \gamma(d)\}}$) are attributed to the applicable topics proportionately to the relevance of each topic for the patent itself ($\theta_{d,k}$). Together with the *subjective* interpretation of the relevant keywords, the process of finding a suitable label for the topics, carried out in the Section 4.3, can now be informed by (the description) of the most relevant CPC codes therein, as *objectively* (albeit probabilistically) measured by $\Phi_{c,k}$.

4 Results

In the present section we outline the results obtained from the three methodological steps described in the previous section, namely, the definition of robotic and LS patents therein (Section 4.1), their sectoral and geographical characterisation (Section 4.2), and the estimation of the probabilistic topic model (Section 4.3). Finally, we recap our findings on human activities which LS patents intend to replace and discuss their relevance with respect to the literature on technological bottlenecks (Section 4.4).

[9] It is straightforward to prove that the same distribution properties in eqs. (5) and (6) which hold for Θ_k also hold for $\Phi_{c,k}$.

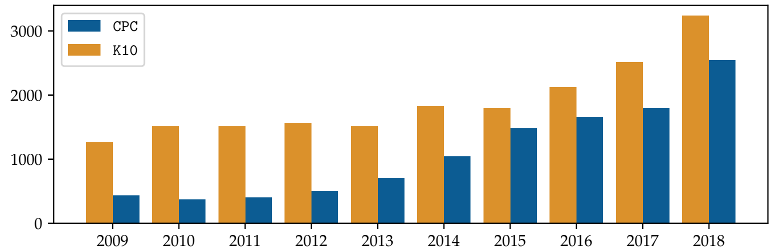


Figure 3: Number of *robotic* patent applications by year.

4.1 Robotic and labour-saving patents

Our first result entails the identification of the set of robotic patents. The CPC-based filter returns 10,929 patents. We label the set of robotic patents according to this criterion as `CPC`. The keyword-based filter returns 18,860 *new* patents, after those already found by the first criterion are discarded. We label the set of new robotic patents according to this criterion as `K10`. The two criteria single out a total of 29,789 unique robotic patents, i.e. approximately 0.84% of the original (universe) population. Fig. 3 shows the yearly evolution in the number of filed robotic patent applications for the two aforementioned subsets. At a first glance, patents under both definitions have almost steadily grown in number over time, with the `CPC` group accounting for most of the relative increase (approximately 6-fold within our reference period). Consistent with discussions in the extant literature, both trends exhibit an acceleration during most recent years (see Brynjolfsson and McAfee, 2012, 2016; Cockburn et al., 2018).

Our second result pertains to the identification of LS patents. The procedure returns 1,666 patents. Since we cannot fully trust the accuracy of the filter with respect to false positives, we proceed with a manual inspection of *all* the potentially LS patents, in order to ensure that the flagged sentence actually conveys the desired message. This conservative manual validation step delivers 1,276 *truly* LS patents (hereafter referred simply as LS patents), i.e. approximately 4.3% of all robotic patents, suggesting our methodology exhibits an accuracy of $\approx 77\%$. Of these, 461 ($\approx 36.1\%$) come from the `CPC` group and 815 ($\approx 63.9\%$) from the `K10` group, indicating that our procedure does not substantially alter the original composition of the whole population of robotic patents.[10] Fig. 4 shows the evolution in the number of LS patents over time, as a fraction of all robotic patents. It is noteworthy that a substantial share of LS patents come from technological fields that do not belong to the standard robot related CPC fields. No clear trend is detectable, suggesting that the underlying LS heuristic has remained quite stable over our reference period. This result is in line with our theoretical assumption that LS heuristics appear to be a robust and invariant driver of technological evolution (c.f. Section 2). Note that our evidence

[10] In order to exclude that our LS patents are the by-product of a specific writing style of a small group of individuals, we exploit the Patent Examination Research Dataset (Public PAIR, available at `https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair`) and find that the number of distinct entities who have been granted power of attorney for the whole set of LS patents exceeds 450, the largest of which administers 37.

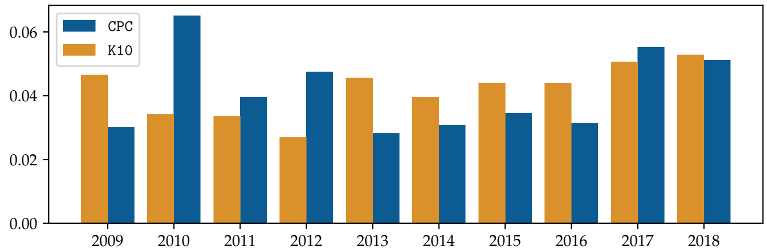


Figure 4: Number of LS patent applications by year, as percentage of robotic applications.

detects both an increasing innovative effort devoted to robotic technology, most of which is not primarily classified as robotics (i.e. the `K10` class, see Fig. 3), and a plateaued search heuristic guided towards labour-displacement, equally distributed on average between the `CPC` and `K10` families (see Fig. 4).

4.2 Firm-level analysis and supply chain

In the present section we characterise LS patents in terms of identity, geographic location, and industrial sector of their current assignee(s). To this purpose, we match our data to the ORBIS (Bureau van Dijk) database through the relevant publication numbers. At the time of writing, the ORBIS database contains information for patent applications published until 31st July 2018; hence, the following analysis is intended over data truncated to that date. 1,136 LS patents ($\approx$ 89% of the original set) find a match, 903 of which ($\approx$ 79%) are assigned to at least one firm, while 233 find no corporate assignment. In total, there are 408 firms which hold at least a LS patent (hereafter, LS firms). Note that patents assigned to more than one firm are deliberately double-counted, since we aim at grasping the actual dispersion of the underlying LS heuristic.

The World map in Fig. 5 gives a glimpse at the geographic distribution of LS patents, given the location of their assignees. The U.S. dominate the picture as the only country with more than 500 LS patents; this is hardly surprising given that all our applications are filed at the USPTO, a primary target for U.S. firms. Japan comes second, as the only other country with more than 100 LS patents, and South Korea, Taiwan, China, and Germany follow suit, holding between 20 and 100 LS patents each. This picture is quite in line with traditional World-class innovation centres when it comes to robotic technology.[11] However, looking at absolute LS patent figures only provides a partial understanding of the associated international patenting activity. Focussing on a relative measure of propensity, i.e. rescaling the number of LS patents by the total number of robotic patents assigned to firms in a given country, allows to infer where LS search efforts are more intensive, compared to the ex-ante capability of producing a robotic patent. This new measure is represented in Fig. 6. While the latter exercise might be biased by the small size of the

[11] An analogous heatmap constructed on the absolute number of robotic patents by country looks strikingly similar and is not included.

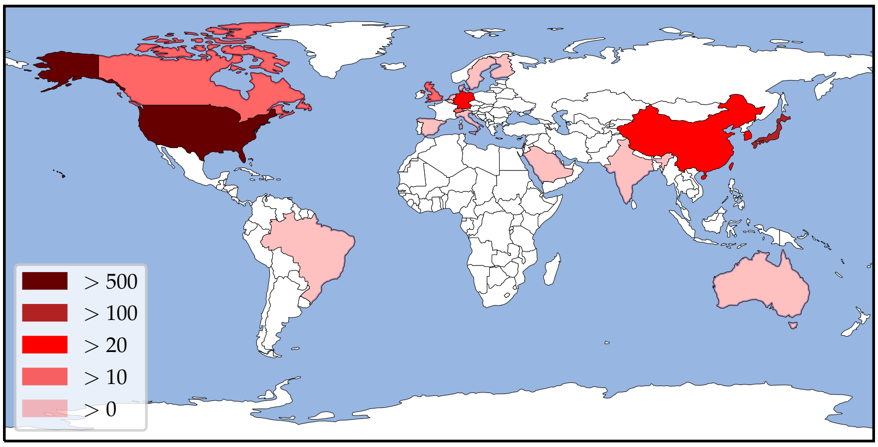


Figure 5: Geographic location of LS patents in absolute terms.

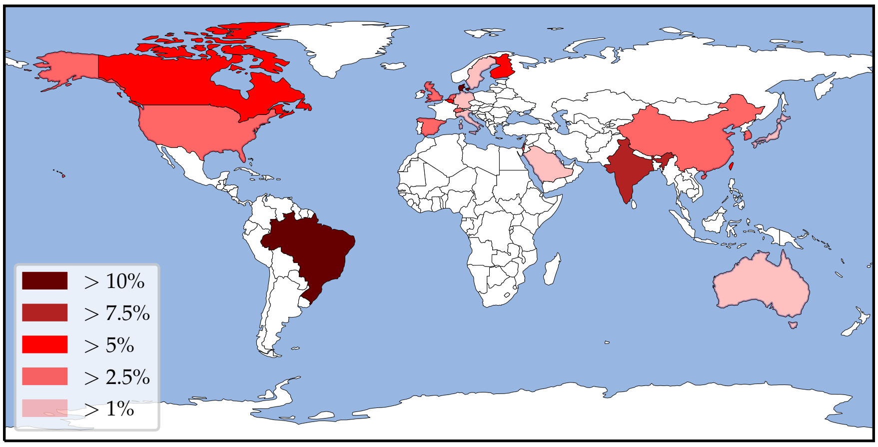


Figure 6: Geographic location of LS patents as percentage of robotic patents.

underlying denominators, it is nonetheless quite informative.[12] Brazil, Hong Kong, and Denmark lead the picture with more than 10% of robotic patents also being LS. Next come, respectively, India and Singapore with at least 7.5%, and Finland, Taiwan, Belgium, and Canada, all beyond the 5% threshold. With the exception of Taiwan, all top robotic patents holders lie below the 5% threshold. In a nutshell, it turns out that countries which hold fewer robotic patents overall, actually hold more LS patents in relative terms.

We now proceed by revealing the identity of top LS patents holding firms and then assessing their sectoral dispersion. Fig. 7 lists the top 15 holders by absolute number of LS patents, while Fig. 8 lists the description of the top 15 primary sectors, identified as 4-digit NAICS codes (2017 revision) assigned to the holders. Both pictures detail the underlying `CPC` and `K10` composition. Boeing, the aircraft manufacturer, is the largest holder of

[12] A strong form of such bias arises for Argentina (excluded from both Figs. 5 and 6) where a single robotic patent has been filed in our focus period, which also happens to be a LS patent.

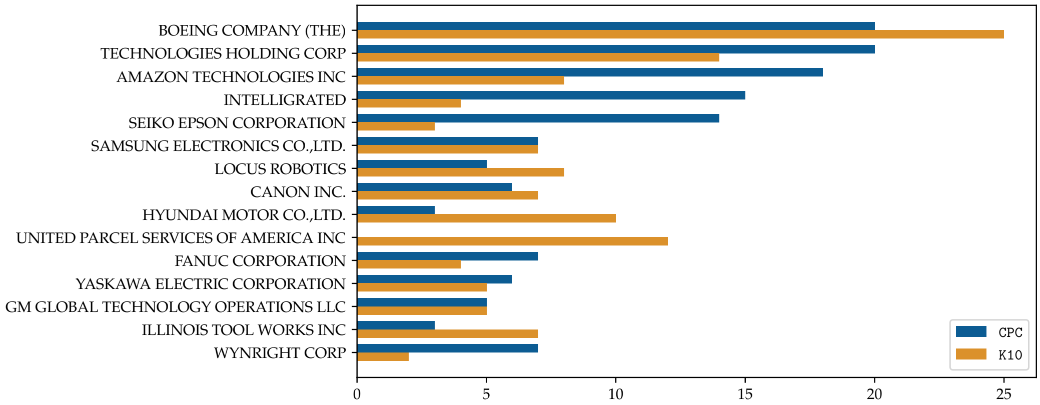


Figure 7: Top 15 firms holding LS patents.

LS patents, with a count of 45.[13] Relatedly, 'Aerospace Product and Parts Manufacturing' is the largest sector within which LS patents reside. Motor vehicles and their parts manufacturing, industries traditionally at the forefront in industrial robots' adoption, also rank very high, as the presence of automotive firms, Hyundai and GM, in the top holders' chart also suggest. Interestingly, retailers (Amazon) and shipping companies (UPS) appear among the top holders, and a deeper inspection of their patents reveals that they are all about fully automatic sorting and routing of packages and drone technology for deliveries. High-tech and R&D intensive firms (e.g. Technologies holding, Intelligrated), robot manufacturers (e.g. Locus, Fanuc), and electronics/software developers, which are the backbones of the robotic value chain (e.g. Seiko-Epson, Samsung), complete the picture. Strikingly, 'Colleges, Universities, and Professional Schools', namely the organisations which are most likely to receive public funding for carrying out research, constitute the 8th largest sector in terms of LS patents holding. Therefore, the industry composition of LS patent holders highlights how robot manufacturers rank lower than robot adopters. The logistics segment, which in our sample emerges from the presence of two international giants, deserves particular scrutiny. The employed workforce in the shipping/delivery industry largely carries out human-intense activities such as conveying, storing, picking, packaging. At this stage of the analysis we cannot conclusively pinpoint the specific human tasks which are more likely to be substituted by LS technology. However, our best guess comprises those phases of the production processes which mainly rely on manpower as their primary input. In the following, we shall attempt to shed some light on the issue. Note that, although we refrain from producing any type of predictive clause, the very fact that LS patents appear to concentrate within large labour-intensive industries is quite interesting.

The frequency distribution of NAICS codes assigned to LS patent holders, pictured in Fig. 9, is also worth noting. In particular, it reveals that, while most of the (408) LS firms are concentrated in a few industries (already shown in Fig. 8), LS patents are overall present in

[13]It is worth noting that these patents relate to the proper aircraft manufacturing process and not to drones or other unmanned aerial vehicles technology.

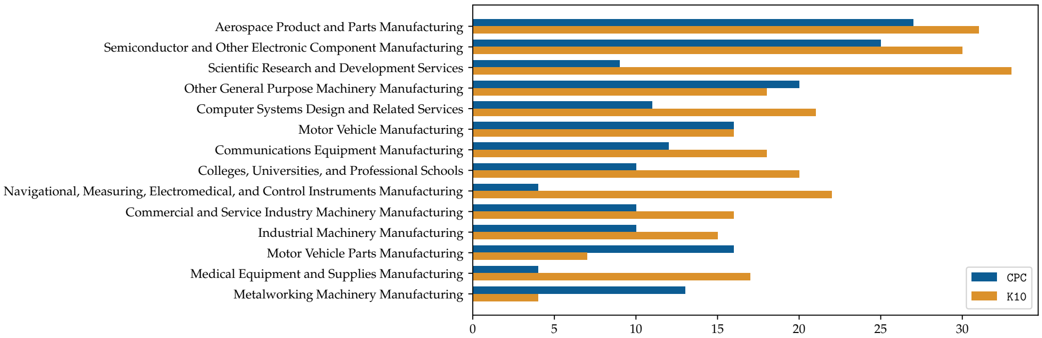


Figure 8: Top 15 industry descriptions of LS patents' holders.

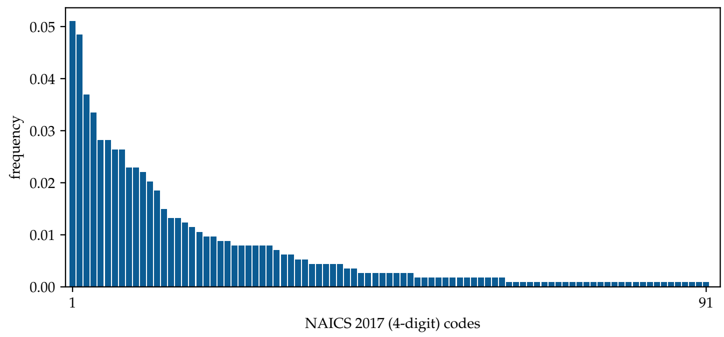


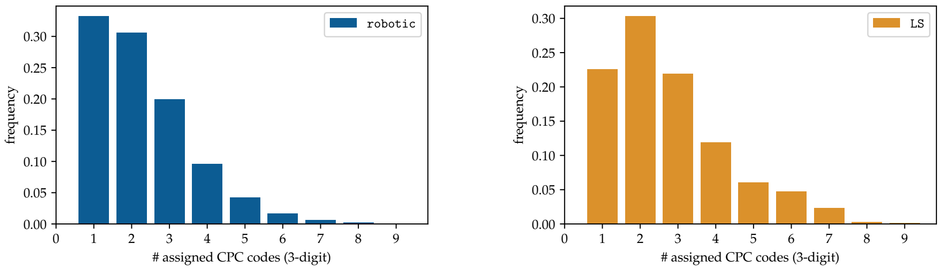
Figure 9: Rank-frequency distribution of NAICS codes across LS firms.

as many as 91 distinct sectors, covering all 2-digit NAICS specifications except *'Agriculture, Forestry, Fishing and Hunting'* (code 11). In other words, the distribution exhibits a 'long tail' across a wide support of NAICS codes. This ultimately suggests that the LS heuristics embedded in robotic technology is quite widespread across the value chain.

4.3 Topic modelling and human-machine taxonomy

At the current stage, we still ignore the human activities LS patents aim at substituting. As already mentioned (in Section 3.3), a simple analysis of CPC codes is not viable because multiple CPC codes are typically attributed to each patent. Fig. 10 shows a histogram with the count of distinct CPC codes at the 3-digit level for robotic (a) and LS patents (b), which reaches up to 9 per patent. While single attributions constitute the modal case for robotic patents, tightly followed by double attributions, the picture is reversed for LS patents, and for these latter triple attributions are almost as widespread as single ones. This suggests that, on average, LS patents are relatively more technologically 'complex' than their robotic superset. To overcome the aforementioned limitation, we estimate the relevance of each CPC code to each patent by leveraging the latent semantic structure of the whole collection of patents' full-text, as identified by a probabilistic topic model.

We now have all the ingredients to assess the technological differences between robotic



(a) All robotic patents.

(b) LS patents only.

Figure 10: Count of 3-digit CPC codes assigned to robotic (a) and LS patents (b).

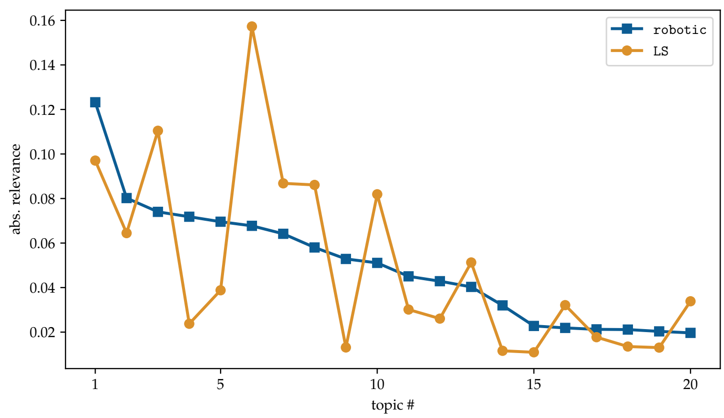


Figure 11: Topic relevance distributions $R(k)^{\mathrm{rob}}$ and $R(k)^{\mathrm{LS}}$ for robotic patents (blue graph) and their LS subset (orange graph).

and LS patents in an informed way, using the CPC-topic matching developed in Section 3.3.3. In particular, we wish to order the identified topics by their relevance to LS patents relative to robotic patents in general. We compute the aggregate relevance distributions Θ_k^{rob} and Θ_k^{LS} defined in Section 3.3.2 for each of the robotic and LS patents populations, respectively. These distributions are pictured in Fig. 11, in which the topics are sorted by decreasing relevance to the robotic patents collection (blue graph). The figure shows sizeable discrepancies between the former and the relevance measure for LS patents (orange graph) for some of the topics. For instance, topic #6 is more than twice as relevant to LS patents than to robotic patents overall. Note however that since the (blue) graph for robotic patents is ranked in decreasing order of relevance, by construction the (orange) graph of LS patents also decreases on average, since the latter are a subset of the former. This implies that the mere difference between Θ_k^{rob} and Θ_k^{LS} (i.e. the vertical distance in Fig. 11) is not fully informative, and an appropriate comparison requires a truly relative measure of relevance for LS patents, which we define as

$$\tilde{\Theta}_k^{\mathsf{LS}} := \frac{\Theta_k^{\mathsf{LS}}}{\Theta_k^{\mathsf{rob}}} \qquad \forall\, k = 1, \dots, K \tag{9}$$

Table 1, which we report in Appendix A for convenience, contains all the relevant information for building the human-machine taxonomy. The first column refers to the topic number, as it appears on the horizontal axis in Fig. 11; the second column, according to which the table is sorted in decreasing order, reports the relative relevance $\tilde{\Theta}_k^{\mathsf{LS}}$ of LS patents to robotic patents, expressed in percentage points; the third column lists the top 10 keywords of the underlying topic; the remaining three columns list the top 5 CPC codes denominations associated to the topic,[14] their weight, and their official description. 3-digit CPC codes are used, except for codes in the `Y10` meta-class reported in full to highlight the original USPC class they point to.

In terms of robotics patents, the five most relevant topics include biochemistry (#1), transmission of digital information (#2), optics (#3), traditional machine tools (#4) and shaping or joining of plastics; additive manufacturing (#5). The latter evidence is comforting in terms of external validity of our topic model. In fact, according to WIPO (2019, p. 17), "among the top 20 technology fields, food chemistry (+13.4%), other special machines (+10.1%), machine tools (+9.2%) and basic materials chemistry (+9.2%) witnessed the fastest average annual growth between 2007 and 2017".

Our relative distance definition allows to single out those topics in which the two populations of patents show strong differences in terms of word occurrence (positive percentage) or similarities (negative percentage). Indeed, LS patents are concentrated in some specific topics. Topic #6 (transport, storage and packaging) displays the highest relative relevance to LS patents (+132.2%), as its relevance more than doubles that of robotic patents overall. This topic can be related to warehouse management and shipping: its most significant CPC code (`B65`) refers to "[c]onveying; packing; storing; [...]" and the top keywords are `'carrier'`, `'conveyor'`, `'item'`, `'gripper'`, `'tape'`. A quick check exposes shipping companies (or companies that have a considerable shipping division) such as Amazon and UPS, as their prime holders (see again Fig. 7). Secondary CPC codes are typically complementary technologies for the robotic embodiment of the underlying artefact; in the present case they mainly relate to "[b]asic electric elements" and "[i]nformation storage".

Other topics with relative relevance above +40% include diagnosis and therapy (#20), transmission of digital information (#10), optics (#3), chemical or physical laboratory apparatus (measuring and testing in chemistry) (#8), and moving parts (either related to prosthetic devices such as exoskeletons or parts of land vehicles) (#16). On the bottom side of Table 1, earth drilling and mining (#9), traditional machine tools (#4), surgery (#14), transmitting and transmission networks (electronic communications) (#15), and shaping

[14]A number of CPC codes widely attributed to robotic patents and whose definition is less informative to the labelling process are discarded. These are: `B25` ("hand tools; portable power-driven tools; manipulators"), `G01` ("measuring; testing"), `G05` ("controlling; regulating"), `G06` ("computing; calculating; counting"), and `Y10S901`, which points to the "Robots" former USPC Class.

or joining of plastics, additive manufacturing (#5), are all topics in which LS patents are less relevant relative to general robotic patents, by at least -40% in our measurement.

4.4 Replaceable human activities and technological bottlenecks

In the final step, we tentatively infer the type of human activities the technology laid out in LS patents is intended to replace. We capture both the formal technological content of inventions using CPC codes definitions and the substantial purpose of single robotic innovations using the vector of words which characterises each topic in the previous analysis. Thanks to this twofold approach, we can describe the fields and activities more exposed to LS innovations.

Notably, topics displaying a larger relative relevance of LS innovations compared to the rest of robotic patents refer to labour-intensive environment, such as the logistics sector, in which workers currently involved in packaging, sorting, and routing items, are particularly threatened. Healthcare constitutes another ground for LS technology, both in the production of medical equipment and in nursing patients or taking care of the elderly. In fact, the second topic in terms of LS relevance relates to medical industry, in line with the development of practices, instruments, and tools possibly able to streamline and reduce the labour intensity required in carrying out medical and healthcare operations and clinical data storage. Notably, medicine appears twice in the table, with both positive and negative relative relevance: while topic #20 (diagnosis and therapy) just discussed displays explicit LS contents, topic #14 mainly relates to collaborative robots for remote surgical activity with a very low relative LS impact. However, it is worth noticing that not all collaborative robots have a labour-friendly impact. Both prostheses and exoskeletons are collaborative in nature and therefore, labour-complementary; yet some features therein (e.g. superhuman force, velocity, speed, reliability) may well save additional human manpower in tasks previously carried out by a team of workers (#16).

Words related to AI emerge in the third most relevant topic (#10, transmission of digital information). `‘learn’`, `‘predict’`, `‘train’`, and `‘evaluate’` all show an high frequency of occurrence. This topic somewhat validates the threat for massive use of ‘intelligent automation’ which might substitute human activities primarily involving qualified professional services (c.f. Section 2). However, manual activities, a common concern for policymakers as they characterise lower-skill jobs, are also frequently mentioned in LS patents (#16, moving parts). High frequency words such as `‘workpiece’`, `‘torque’`, `‘finger’` suggest a particular interest in manual and finger dexterity. Remarkably, this topic spans from medical and veterinary applications to land vehicles CPCs.

Our estimation of the probabilistic topic model allows to pinpoint the technological bottlenecks underlying the search efforts inspiring robotics inventors. Indeed, LS heuristics are concentrated in human activities already identified by other contributions in the literature. Arntz et al. (2016), Frey and Osborne (2017) and Nedelkoska and Quintini (2018) all rely on experts’ judgement (so-called Delphi method) in constructing an automation probability measure of O*NET occupations. For instance, Frey and Osborne (2017) ask technologists to reply to the following question for 70 selected occupations: “Can the tasks of this job be sufficiently specified, conditional on the availability of big data, to be performed by state of the art computer-controlled equipment?" (see Frey and Osborne, 2017, Table 1). They claim that the probability of an occupation being automated is inversely related to: social intelligence, such as the ability to negotiate complex social relationships, including caring for others or evaluate differences; cognitive intelligence, such as the ability of solving complex problems; and finger dexterity and manipulation, such as the ability to carry out precise physical tasks in an unstructured work environment and in awkward positions.

Our work adds evidence in this perspective. The tasks identified by the aforementioned contributions actually map to the semantic domains covered by our LS patents and emerging out of the probabilistic topic model: for example, topics #6 and #16 involve multiple tasks related to perception and manipulation, topic #10 contains tasks related to cognitive intelligence (e.g. *"Systems and methods for consumer-generated media reputation management"*, [US20170286541A1]), and topic #20 involves tasks where social intelligence is important, in particular for assisting and caring for others. Therefore, according to our results, these technological bottlenecks are currently under the spot of cutting-edge research efforts by innovative firms in their knowledge space. This result also aligns with the recent findings of Webb (2020), who shows that the most recent AI driven automation wave targets high-skilled tasks.

5 Concluding remarks

The fast development of robotic(-related) technology, artificial intelligence, and automation has raised concerns about the future of work. Recent literature has focussed, on the one hand, on the analysis of tasks and occupations at risk of automation (e.g. Frey and Osborne, 2017; Webb, 2020). On the other hand, on the labour market impact of industrial robots among adopters (e.g. Acemoglu and Restrepo, 2019b). In this paper, we contribute with an in-depth analysis of the nature of innovations in robotics-related technologies, focussing on their sectors of origin. We study robotic patents explicitly encompassing LS heuristics and trace their distribution in terms of firms, sectors, and geographical location, and we determine whether they differ in terms of technological content with respect to the totality of robotic patents, by emphasising the underlying technological heterogeneity and the vertical supply chain behind. To address these questions we employ advanced textual analysis and machine learning techniques on the universe of USPTO patent applications filed between 2009 and 2018, and we exploit a direct match with ORBIS (Bureau van Dijk) firm-level database.

Our results can be summarised as follows. First, the time evolution of LS patents do not show an explicit trend over time, hinting at a stable and established pattern of LS heuristics. Second, in terms of geographical location, U.S. and Japan still appear to largely dominate other countries. Third, the sectoral distribution of LS robotic patents present a long-tail, signalling a widespread nature of the underlying applications across 4-digit industries. Fourth, patenting firms are not only constituted by robots producers, but mainly adopters, some archetypical cases being Boeing, Hyundai, Amazon, and UPS.

Finally, by means of a probabilistic topic model and a topic-level match with patent classification codes, we construct a human-machine taxonomy highlighting human activities which appear more likely to be displaced. We find that LS patents are particularly concentrated in the following fields: (i) transport, storage and packaging, (ii) diagnosis and therapy, (iii) transmission of digital information, (iv) optical elements, (v) chemical and physical laboratory apparatus (measuring and testing in chemistry), and (vi) moving parts. From our taxonomy it emerges that the typical tasks where LS research effort is focussed include (i) dexterity and manipulations, as in packing, storing, conveying, and handling packages in the logistics industry; (ii) activities entailing social intelligence, such as caretaking patients and the elders; (iii), activities requiring cognitive intelligence and complex reasoning, e.g. the competence in predicting, learning, classifying and evaluating, typical of high-level professional segments. Previous literature has identified the above activities as technological bottlenecks, in the sense of being particularly hard to automate. Our work suggests that search efforts exerted by leading international companies are precisely aimed at defeating these bottlenecks.

The main limitation of our work lies on its excessively conservative measure of LS patents. Indeed, while our procedure completely avoids type I errors (i.e. false positives), the true magnitude of LS innovation is likely to be largely underestimated (type II error). Conceivable extensions of our framework include the analysis of a wider range of patents beyond robotic technology and the potential matching with both firm-level economic indicators (such as sales, wages, and productivity) and dictionaries of occupations and tasks (such as O*NET and PIAAC).

Acknowledgements

The authors wish to thank participants to the 2020 GEOINNO Conference (Stavanger, Norway), the 2019 CONCORDi Conference (Seville, Spain), the 2019 EAEPE Conference (Warsaw, Poland), the 2019 EMAEE Conference (Brighton, UK), the 2019 RENIR Workshop on the impact of automation and artificial intelligence on regional economies (Turin, Italy), and the 2019 Workshop on Innovation, firm dynamics, employment and growth (Greenwich, UK) for helpful comments and insightful suggestions at various stages of this work. Maria Enrica Virgillito acknowledges support from European Union's Horizon 2020 research and innovation programme under grant agreement No. 822781 GROWINPRO – Growth Welfare Innovation Productivity.

References

Acemoglu, Daron and Pascual Restrepo (2018a). 'Artificial Intelligence, Automation and Work'. NBER Working Paper No. 24196.

Acemoglu, Daron and Pascual Restrepo (2018b). 'The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment'. *American Economic Review* 108(6), pp. 1488–1542. DOI: 10.1257/aer.20160696.

Acemoglu, Daron and Pascual Restrepo (2019a). 'Automation and New Tasks: How Technology Displaces and Reinstates Labor'. *Journal of Economic Perspectives* 33(2), pp. 3–30. DOI: 10.1257/jep.33.2.3.

Acemoglu, Daron and Pascual Restrepo (2019b). 'Robots and Jobs: Evidence from US Labor Markets'. *Journal of Political Economy*. Forthcoming. DOI: 10.1086/705716.

Arntz, Melanie, Terry Gregory and Ulrich Zierahn (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis*. OECD Social, Employment and Migration Working Papers No. 189, OECD Publishing, Paris. DOI: 10.1787/5jlz9h56dvq7-en.

Arun, R., V. Suresh, C. E. Veni Madhavan and M. N. Narasimha Murthy (2010). 'On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations'. In: *Advances in Knowledge Discovery and Data Mining*. Ed. by M.J. Zaki, J.X. Yu, B. Ravindran and V. Pudi. Vol. 6118. PAKDD 2010. Lecture Notes in Computer Science. Springer. DOI: 10.1007/978-3-642-13657-3_43.

Atkinson, Anthony B. and Joseph E. Stiglitz (1969). 'A New View of Technological Change'. *Economic Journal* 79(315), pp. 573–578. DOI: 10.2307/2230384.

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

Autor, David H. and David Dorn (2013). 'The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market'. *American Economic Review* 103(5), pp. 1553–1597. DOI: 10.1257/aer.103.5.1553.

Barbieri, Laura, Mariacristina Piva and Marco Vivarelli (2018). 'R&D, embodied technological change, and employment: evidence from Italian microdata'. *Industrial and Corporate Change* 28(1), pp. 203–218. DOI: 10.1093/icc/dty001.

Bartelsman, Eric J., Martin Falk and Eva Hagsten (2019). 'Productivity, technological innovations and broadband connectivity: firm-level evidence for ten European countries'. *Eurasian Business Review* 9(1), pp. 25–48. DOI: 10.1007/s40821-018-0113-0.

Blei, David M. (2012). 'Probabilistic Topic Models'. *Communications of the ACM* 55(4), pp. 77–84. DOI: 10.1145/2133806.2133826.

Blei, David M., Andrew Y. Ng and Michael I. Jordan (2003). 'Latent Dirichlet Allocation'. *Journal of Machine Learning Research* 3(Jan), pp. 993–1022.

Bresnahan, Timothy F. and Manuel Trajtenberg (1995). 'General purpose technologies 'Engines of growth'?' *Journal of Econometrics* 65(1), pp. 83–108. DOI: 10.1016/0304-4076(94)01598-T.

Brynjolfsson, Erik and Andrew McAfee (2012). *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Digital Frontier Press.

Brynjolfsson, Erik and Andrew McAfee (2016). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Calvino, Flavio and M. Enrica Virgillito (2017). 'The innovation employment nexus: A critical survey of theory and empirics'. *Journal of Economic Surveys* 32(1), pp. 83–117. DOI: 10.1111/joes.12190.

Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang and Sheng Tang (2009). 'A density-based method for adaptive LDA model selection'. *Neurocomputing* 72(7–9), pp. 1775–1781. DOI: 10.1016/j.neucom.2008.06.011.

Capone, Gianluca, Franco Malerba, Richard R. Nelson, Luigi Orsenigo and Sidney G. Winter (2019). 'History friendly models: retrospective and future perspectives'. *Eurasian Business Review* 9(1), pp. 1–23. DOI: 10.1007/s40821-019-00121-0.

Chen, Hongshu, Guangquan Zhang, Donghua Zhu and Jie Lu (2017). 'Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014'. *Technological Forecasting and Social Change* 119, pp. 39–52. DOI: 10.1016/j.techfore.2017.03.009.

Cirillo, Valeria, Matteo Rinaldini, Jacopo Staccioli and M. Enrica Virgillito (2018). 'Workers' intervention authority in Italian 4.0 factories: autonomy and discretion'. LEM Working Papers series n. 13/2018. URL: http://www.lem.sssup.it/WPLem/2018-13.html.

Cockburn, Iain M., Rebecca Henderson and Scott Stern (2018). 'The Impact of Artificial Intelligence on Innovation'. NBER Working Paper No. 24449.

David, Paul A. (1985). 'Clio and the Economics of QWERTY'. *American Economic Review* 75(2), pp. 332–337.

Dechezleprêtre, Antoine, David Hémous, Morten Olsen and Carlo Zanella (2019). 'Automating Labor: Evidence from Firm-level Patent Data'. Unpublished. URL: https://www.econ.uzh.ch/dam/jcr:c6844559-38f7-495e-a535-c87e97f16662/Automating%20Labor_Evidence%20from%20Firm-level%20Patent%20Data_201903.pdf.

Dosi, Giovanni (1982). 'Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change'. *Research Policy* 11(3), pp. 147–162. DOI: 10.1016/0048-7333(82)90016-6.

Dosi, Giovanni (1988). 'Sources, Procedures, and Microeconomic Effects of Innovation'. *Journal of Economic Literature* 26(3), pp. 1120–1171.

Dosi, Giovanni (1997). 'Opportunities, Incentives and the Collective Patterns of Technological Change'. *Economic Journal* 107(444), pp. 1530–1547. DOI: 10.1111/j.1468-0297.1997.tb00064.x.

Dosi, Giovanni, Marco Grazzi and Daniele Moschella (2017). 'What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification'. *Small Business Economics* 48(2), pp. 413–429. DOI: 10.1007/s11187-016-9783-0.

Dosi, Giovanni and Pierre Mohnen (2019). 'Innovation and employment: an introduction'. *Industrial and Corporate Change* 28(1), pp. 45–49. DOI: 10.1093/icc/dty064.

Dosi, Giovanni and Richard R. Nelson (2010). 'Technical Change and Industrial Dynamics as Evolutionary Processes'. In: *Handbook of the Economics of Innovation*. Ed. by Bronwyn H. Hall and Nathan Rosenberg. Vol. 1. North Holland. Chap. 3. DOI: 10.1016/S0169-7218(10)01003-8.

Dosi, Giovanni and Richard R. Nelson (2013). 'The Evolution of Technologies: An Assessment of the State-of-the-Art'. *Eurasian Business Review* 3(1), pp. 3–46.

Dosi, Giovanni, Mariacristina Piva, M. Enrica Virgillito and Marco Vivarelli (2019). 'Embodied and Disembodied Technological Change: The Sectoral Patterns of Job-Creation and Job-Destruction'. IZA Discussion Paper No. 12408.

Frank, Morgan R., David H. Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, Dashun Wang, Hyejin Youn and Iyad Rahwan (2019). 'Toward understanding the impact of artificial intelligence on labor'. *Proceedings of the National Academy of Science* 116(14), pp. 6531–6539. DOI: 10.1073/pnas.1900949116.

Freeman, Christopher and Luc Soete (1987). *Technical change and full employment*. B. Blackwell.

Frey, Carl Benedikt and Michael A. Osborne (2017). 'The future of employment: How susceptible are jobs to computerisation?' *Technological Forecasting and Social Change* 114, pp. 254–280. DOI: 10.1016/j.techfore.2016.08.019.

Graetz, Georg and Guy Michaels (2018). 'Robots at Work'. *Review of Economics and Statistics* 100(5), pp. 753–768. DOI: 10.1162/rest_a_00754.

IFR (2017). 'The Impact of Robots on Productivity, Employment and Jobs'. Positioning paper. Updated April 2018. URL: https://ifr.org/downloads/papers/IFR_The_Impact_of_Robots_on_Employment_Positioning_Paper_updated_version_2018.pdf.

Kaplan, Sarah and Keyvan Vakili (2015). 'The double-edged sword of recombination in breakthrough innovation'. *Strategic Management Journal* 36(10), pp. 1435–1457. DOI: 10.1002/smj.2294.

Kim, Gab Jo, Sang Sung Park and Dong Sik Jang (2015). 'Technology Forecasting using Topic-Based Patent Analysis'. *Journal of Scientific and Industrial Research* 74, pp. 265–270.

Lee, Won Sang, Eun Jin Han and So Young Sohn (2015). 'Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents'. *Technological Forecasting and Social Change* 100, pp. 317–329. DOI: 10.1016/j.techfore.2015.07.022.

Mann, Katja and Lukas Püttmann (2015). 'Benign Effects of Automation: New Evidence from Patent Texts'. Available at SSRN. URL: https://ssrn.com/abstract=2959584.

Marx, Karl (1956). *The Poverty of Philosophy*. Lawrence and Wishart (for the Foreign Languages Publishing House, Moscow).

Moro, Angelo, Matteo Rinaldini, Jacopo Staccioli and M. Enrica Virgillito (2019). 'Control in the era of surveillance capitalism: an empirical investigation of Italian Industry 4.0 factories'. *Journal of Industrial and Business Economics* 46(3), pp. 347–360. DOI: 10.1007/s40812-019-00120-2.

Nedelkoska, Ljubica and Glenda Quintini (2018). *Automation, skills use and training*. OECD Social, Employment and Migration Working Papers, No. 202, OECD Publishing, Paris. DOI: 10.1787/2e2f4eea-en.

Newman, MEJ (2005). 'Power laws, Pareto distributions and Zipf's law'. *Contemporary Physics* 46(5), pp. 323–351. DOI: 10.1080/00107510500052444.

Patel, Pari and Keith Pavitt (1997). 'The technological competencies of the world's largest firms: Complex and path-dependent, but not much variety'. *Research Policy* 26(2), pp. 141–156. DOI: 10.1016/S0048-7333(97)00005-X.

Pavitt, Keith (1985). 'Patent statistics as indicators of innovative activities: Possibilities and problems'. *Scientometrics* 7(1–2), pp. 77–99. DOI: 10.1007/bf02020142.

Pellegrino, Gabriele, Mariacristina Piva and Marco Vivarelli (2019). 'Beyond R&D: the role of embodied technological change in affecting employment'. *Journal of Evolutionary Economics* 29(4), pp. 1151–1171. DOI: 10.1007/s00191-019-00635-w.

Piva, Mariacristina and Marco Vivarelli (2018). 'Technological change and employment: is Europe ready for the challenge?' *Eurasian Business Review* 8(1), pp. 13–32. DOI: 10.1007/s40821-017-0100-x.

Porter, Martin F. (1980). 'An algorithm for suffix stripping'. *Program: electronic library and information systems* 14(3), pp. 130–137. DOI: 10.1108/eb046814.

Rosenberg, Nathan (1976). *Perspectives on Technology*. 1st ed. Cambridge University Press. DOI: 10.1017/cbo9780511561313.

Simonetti, Roberto, Karl Taylor and Marco Vivarelli (2000). 'Modelling the employment impact of innovation'. In: *The Employment Impact of Innovation*. Ed. by Mario Pianta and Marco Vivarelli. Routledge. Chap. 3, pp. 26–43.

Trajtenberg, Manuel (2018). 'AI as the next GPT: a Political-Economy Perspective'. NBER Working Paper No. 24245.

Tunzelmann, G. Nick VON (1995). 'Time-Saving Technical Change: The Cotton Industry in the English Industrial Revolution'. *Explorations in Economic History* 32(1), pp. 1–27. DOI: 10.1006/exeh.1995.1001.

Ugur, Mehmet, Sefa Awaworyi Churchill and Edna Solomon (2018). 'Technological innovation and employment in derived labour demand models: A hierarchical meta-regression analysis'. *Journal of Economic Surveys* 32(1), pp. 50–82. DOI: 10.1111/joes.12187.

Van Roy, Vincent, Dániel Vértesy and Marco Vivarelli (2018). 'Technology and employment: Mass unemployment or job creation? Empirical evidence from European patenting firms'. *Research Policy* 47(9), pp. 1762–1776. DOI: 10.1016/j.respol.2018.06.008.

Venugopalan, Subhashini and Varun Rai (2015). 'Topic based classification and pattern identification in patents'. *Technological Forecasting and Social Change* 94, pp. 236–250. DOI: 10.1016/j.techfore.2014.10.006.

Vivarelli, Marco (1995). *The Economics of Technology and Employment: Theory and Empirical Evidence*. Edward Elgar Pub.

Vivarelli, Marco (2014). 'Innovation, Employment and Skills in Advanced and Developing Countries: A Survey of Economic Literature'. *Journal of Economic Issues* 48(1), pp. 123–154. DOI: 10.2753/JEI0021-3624480106.

Webb, Michael (2020). 'The Impact of Artificial Intelligence on the Labor Market'. Available at SSRN. URL: https://ssrn.com/abstract=3482150.

WIPO (2019). 'World Intellectual Property Indicators 2019'. Geneva: World Intellectual Property Organization. URL: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2019.pdf.

Appendix A

Table 1

Topic #	$\tilde{\Theta}_k^{LS}$	Words	CPC	Weight	Description
6	+132.2%	carrier	B65	24.4%	Conveying; packing; storing; handling thin or filamentary material
		conveyor			
		item	H01	6.8%	Basic electric elements
		gripper	G11	6.0%	Information storage
		tape	Y02	4.6%	Technologies or applications for mitigation or adaptation against climate change
		articl			
		convey	B23	4.3%	Machine tools; metal-working not otherwise provided for
		tray			
		packag			
		door			

Continues on the next page

Table 1 – continued from the previous page

Topic #	$\tilde{\Theta}_k^{LS}$	Words	CPC	Weight	Description
20	+72.2%	weld	A61	47.8%	Medical or veterinary science; hygiene
		patient	B23	17.7%	Machine tools; metal-working not otherwise provided for
		medic			
		cathet	G16	4.1%	Information and communication technology [ict] specially adapted for specific application fields
		treatment			
		tissu	H04	3.1%	Electric communication technique
		lumen	G09	2.0%	Education; cryptography; display; advertising; seals
		electrod			
		needl			
		transduc			
10	+60.3%	node	H04	16.3%	Electric communication technique
		learn	A61	14.4%	Medical or veterinary science; hygiene
		predict	Y02	7.8%	Technologies or applications for mitigation or adaptation against climate change
		train			
		evalu	G08	4.2%	Signalling
		estim	H01	3.6%	Basic electric elements
		score			
		neural			
		behavior			
		sampl			
3	+49.2%	beam	H04	17.7%	Electric communication technique
		ray	A61	17.6%	Medical or veterinary science; hygiene
		eye	G02	12.1%	Optics
		scan	H01	6.4%	Basic electric elements
		len	G09	6.3%	Education; cryptography; display; advertising; seals
		pixel			
		fiber			
		detector			
		radiat			
		fluoresc			
8	+48.4%	sampl	B01	23.5%	Physical or chemical processes or apparatus in general
		assay			
		pipett	C12	13.5%	Biochemistry; beer; spirits; wine; vinegar; microbiology; enzymology; mutation or genetic engineering
		reagent			
		vessel			
		dispens	Y10T436	13.5%	Chemistry: analytical and immunological testing
		reaction			
		specimen	A61	6.7%	Medical or veterinary science; hygiene
		cartridg	B65	5.4%	Conveying; packing; storing; handling thin or filamentary material
		analyt			

Continues on the next page

Table 1 – continued from the previous page

Topic #	$\tilde{\Theta}_k^{LS}$	Words	CPC	Weight	Description
16	+46.7%	workpiec	A61	23.2%	Medical or veterinary science; hygiene
		torqu	B62	12.2%	Land vehicles for travelling otherwise than on rails
		leg			
		finger	A63	7.9%	Sports; games; amusements
		trajectori	Y10T74	6.3%	Machine element or mechanism
		pose	B23	6.1%	Machine tools; metal-working not otherwise provided for
		ball			
		foot			
		veloc			
		walk			
7	+35.3%	substrat	H01	45.0%	Basic electric elements
		chamber	C23	9.1%	Coating metallic material; coating material with metallic material; chemical surface treatment; diffusion treatment of metallic material; coating by vacuum evaporation, by sputtering, by ion implantation or by chemical vapour deposition, in general; inhibiting corrosion of metallic material or incrustation in general
		wafer			
		gas			
		film			
		semiconductor			
		deposit			
		polish			
		chuck	G03	4.0%	Photography; cinematography; analogous techniques using waves other than optical waves; electrography; holography
		holder			
			B08	3.2%	Cleaning
			B24	3.0%	Grinding; polishing
13	+27.2%	surgic	A61	71.3%	Medical or veterinary science; hygiene
		patient	Y10T74	4.8%	Machine element or mechanism
		implant	G16	3.0%	Information and communication technology [ict] specially adapted for specific application fields
		surgeon			
		marker	H04	2.0%	Electric communication technique
		bone	G09	1.9%	Education; cryptography; display; advertising; seals
		surgeri			
		endoscop			
		master			
		tissu			
17	−16.5%	suction	A01	15.6%	Agriculture; forestry; animal husbandry; hunting; trapping; fishing
		cleaner			
		nozzl	A47	15.5%	Furniture; domestic articles or appliances; coffee mills; spice mills; suction cleaners in general
		milk			
		teat	H04	5.7%	Electric communication technique
		valv	H01	4.8%	Basic electric elements
		cup	E04	4.1%	Building
		discharg			
		dairi			
		pool			

Continues on the next page

Table 1 – continued from the previous page

Topic #	$\tilde{\Theta}_k^{LS}$	Words	CPC	Weight	Description
2	–19.6%	server	H04	26.1%	Electric communication technique
		request	A61	8.1%	Medical or veterinary science; hygiene
		video	A63	7.0%	Sports; games; amusements
		messag	G10	6.8%	Musical instruments; acoustics
		databas	G08	5.3%	Signalling
		audio			
		voic			
		search			
		game			
		client			
1	–21.3%	cell	C12	23.1%	Biochemistry; beer; spirits; wine; vinegar; microbiology; enzymology; mutation or genetic engineering
		acid			
		probe			
		protein	B01	11.0%	Physical or chemical processes or apparatus in general
		nucleic			
		polypeptid	A61	7.1%	Medical or veterinary science; hygiene
		hybrid	Y10T436	6.1%	Chemistry: analytical and immunological testing
		dna			
		molecul	C07	5.0%	Organic chemistry
		plant			
11	–33.1%	vehicl	B60	14.3%	Vehicles in general
		autonom	B62	7.5%	Land vehicles for travelling otherwise than on rails
		obstacl			
		navig	H04	7.2%	Electric communication technique
		charg	Y02	7.0%	Technologies or applications for mitigation or adaptation against climate change
		uav			
		rout	A47	6.7%	Furniture; domestic articles or appliances; coffee mills; spice mills; suction cleaners in general
		batteri			
		self			
		dock			
19	–35.9%	clamp	A61	12.2%	Medical or veterinary science; hygiene
		roller	B23	7.4%	Machine tools; metal-working not otherwise provided for
		seal			
		ring	Y10T74	6.8%	Machine element or mechanism
		rod	F16	6.4%	Engineering elements and units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general
		connector			
		fasten			
		axial			
		cylind	H01	6.3%	Basic electric elements
		bend			

Continues on the next page

Table 1 – continued from the previous page

Topic #	$\tilde{\Theta}_k^{LS}$	Words	CPC	Weight	Description
18	–36.0%	gear	A61	15.9%	Medical or veterinary science; hygiene
		smart	Y10T74	12.5%	Machine element or mechanism
		pulley	F16	10.7%	Engineering elements and units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation in general
		home			
		rotari			
		occup			
		detector	H04	7.2%	Electric communication technique
		hazard	G08	5.8%	Signalling
		mesh			
		nut			
12	–39.2%	reson	H02	16.9%	Generation; conversion or distribution of electric power
		voltag			
		charg	H01	11.8%	Basic electric elements
		batteri	B60	11.6%	Vehicles in general
		conductor	Y02	9.3%	Technologies or applications for mitigation or adaptation against climate change
		induct			
		coil	H03	9.1%	Basic electronic circuitry
		imped			
		circuitri			
		trace			
5	–44.2%	electrod	B29	9.2%	Working of plastics; working of substances in a plastic state in general
		coat			
		print	H01	8.9%	Basic electric elements
		mold	B05	6.2%	Spraying or atomising in general; applying liquids or other fluent materials to surfaces, in general
		build			
		panel			
		fabric	B23	4.9%	Machine tools; metal-working not otherwise provided for
		adhes			
		sheet	Y02	4.8%	Technologies or applications for mitigation or adaptation against climate change
		paint			
15	–52.0%	transmitt	H04	28.7%	Electric communication technique
		inspect	A61	7.2%	Medical or veterinary science; hygiene
		sound	H01	6.0%	Basic electric elements
		lumin	G08	6.0%	Signalling
		recept	A47	4.3%	Furniture; domestic articles or appliances; coffee mills; spice mills; suction cleaners in general
		exposur			
		bright			
		ultrason			
		antenna			
		server			

Continues on the next page

Table 1 – continued from the previous page

Topic #	$\tilde{\Theta}_k^{LS}$	Words	CPC	Weight	Description
14	–63.7%	effector	A61	55.6%	Medical or veterinary science; hygiene
		surgic	Y10T29	4.9%	Metal working
		stapl	H01	4.8%	Basic electric elements
		articul	Y10T74	4.2%	Machine element or mechanism
		closur	B23	3.5%	Machine tools; metal-working not otherwise provided for
		cartridg			
		elong			
		jaw			
		anvil			
		fire			
4	–66.9%	teach	A61	10.7%	Medical or veterinary science; hygiene
		vibrat	H01	9.3%	Basic electric elements
		calibr	B23	8.5%	Machine tools; metal-working not otherwise provided for
		postur			
		veloc	H04	7.8%	Electric communication technique
		board	Y02	6.0%	Technologies or applications for mitigation or adaptation against climate change
		mark			
		piezoelectr			
		angular			
		acceler			
9	–75.2%	heater	H01	8.6%	Basic electric elements
		hydrocarbon	E21	6.6%	Earth drilling; mining
		conductor	B23	5.5%	Machine tools; metal-working not otherwise provided for
		conduit			
		pipe	Y10T29	4.4%	Metal working
		treatment	Y02	4.4%	Technologies or applications for mitigation or adaptation against climate change
		drill			
		gas			
		cool			
		insul			