Patterns of patenting activity in AI and robotics. A time series analysis

G. Guidetti, R. Leoncini, M. Macaluso

Abstract

The last two decades have been characterized by major advances in Artificial Intelligence (AI) and robotics in the wake of the so-called "Fourth Industrial Revolution" (4IR). These advances brought about renewed concerns among both academics and policy makers with reference to their impacts on, for example, performance processes, inequality, occupational structure, the nature of labour, the demand for skills, competitiveness of the economic actors (Brynjolfsson and McAfee, 2011; 2014; Arntz et al., 2016; Frey and Osborne, 2017; OECD, 2019; Benassi, Grinza and Rentocchini, 2020; Leoncini, Guidetti and Cattani, 2021). Frey and Osborne (2013) calculated that 47% of US jobs are at risk of automation, while according to the World Bank (2016) the same share accounts up to 57% in the OECD countries. OECD (2019) estimates that 14% of existing jobs could disappear because of automation in the next 15-20 years, but another 32% are likely to be radically transformed as individual tasks are automated. On the other hand, automation and new technologies are creating new types of jobs, where new tasks and activities can productively combine labour with technologies (Acemoglu and Restrepo, 2019; World Economic Forum, 2018; International Federation of Robotics, 2019). However, the boundaries between AI and other innovations are frequently blurred and constantly evolving. In addition, definitions have changed over time and are in some cases contentious (see e.g. ISO 8373; OECD, 2019, 2020; European Commission, 2018; HLEG, 2019; WIPO, 2019; Samoili et al., 2020). This paper relies on the suggestion provided by Acemoglu and Restrepo (2019), according to whom on the one side, AI should be thought of as a technology platform, which can automate tasks previously performed by labour or create new tasks and activities in which humans can be productively employed. On the other side, robotics is thought of as a new technology that "often makes use of AI and other digital technologies for processing data but is distinguished from other digital technologies by its focus on interacting with the physical world". This paper mainly aims to answer to the questions: "Is it possible to discriminate between AI and robotics applications based on patent information?" and "If so, are there any differences between the two in terms of diffusion and evolution over time?". This paper examines the evolution of 4IR patent applications over the reference period 1980 - 2019, by identifying the peculiarities of the trends of AI- and robotics- related patents. By combining different sources of information (e.g. EPO, 2017; European Commission, 2020 on AI taxonomy, IPO, WIPO), we want to provide a novel comprehensive analysis of the technology trends in 4IR-related patents.

Our identification strategy relies on a novel approach by using a hierarchical analysis (also called multilevel analysis). This approach is often used in the social sciences in order to control for the effect of unobserved background variables and is used to take into account the social contexts as well as the individual respondents or subjects. The importance of multilevel modelling is due mainly to the determination of research constructs that consider the existence of nested data structures, in which certain variables show variation between distinct units that represent groups. We first focus on recognizing the data structure. Specifically, our final dataset involves a 2-level nested data structure: level 1 involves robotics (intelligent vs others), level 2 includes the broader category of AI (core and transversal technologies as suggested by JRC, 2020). In this way, we want to identify a hierarchical structure, where we include patents that refer to AI and patents that refers to intelligent robots.

Our analysis relies on PATSTAT (Spring 2021 version) data to gather information on patents filed at the European Patent Office (EPO) in the period 1980-2019: namely, patent families, technological classification codes with relative abstracts and titles, citations (forward and backward) and other patent applicants' and inventors' variables, including countries of origin. Following standard

practice in the literature, we exploit PATSTAT to identify AI-related patents, based on a technology classification search. To do so, we use different sources of information (EPO, IPO, WIPO), which provide lists of technological classification codes associated to selected AI-related technologies. Specifically, for each patent, we obtained the list of its assigned CPC (Cooperative Patent Classification) codes. Following the literature (see e.g. Hall and Helmers, 2013; Barbieri, Marzucchi and Rizzo, 2020), we use patent family as the unit of analysis to avoid double counting of the same invention, which can occur when patents are issued in more than one country. To identify patent family, we primarily take the maximum value of forward citation within the patent family or 2) if this information is missing, we take the maximum value of backward citations (see Verhoeven et al. (2016) and Barbieri, Marzucchi and Rizzo, 2020).

To mitigate against the risk of omitting patents if those areas are relied on completely, patents related to robotics were retrieved that include certain keywords in the titles and abstracts (which summarises the technical content of the patent). To do so, we use the keyword list provided by the JRC (European Commission), which proposes a list of keywords to determine the boundaries of the collection of areas that represents AI (core and transversal technologies, domains and subdomains). In order to avoid the occurrence of false positives as much as possible, JRC proposes to use some terms in combination with intrinsic-AI terms (keywords used to identify the relevant active agents in AI). To identify intelligent robots, we use terms such as “industrial robot”, “social robot”, “service robot” and “robot system” after conditioning its occurrence with other terms, as “cognitive systems”, “control theory”. In addition, we use a text-mining techniques by exploiting a Python algorithm to do a search keyword inside robots-related patents. To this end, the text contained in patent abstracts or titles was parsed using a series of keywords from different sources (JRC, IPO, OECD) to search and extract the first occurrence information (word frequencies, either for single words or combinations of words) in AI-related robots. Finally, we create a dataset that distinguishes between intelligent robots and other robots.