Analyzing Automation Technologies and Their Tasks in Patent Texts Using Natural Language Processing

Tomáš Oleš[1]

[1]University of Economics in Bratislava, Faculty of National Economy, Department of Economic Policy, Bratislava, Dolnozemská cesta 1, 852 35, Slovak Republic[1]

tomas.oles@euba.sk

https://doi.org/10.53465/EDAMBA.2023.9788022551274.187-196

Abstract. Technological advances have changed the labor market, and automation technology is a major factor affecting jobs. To study its impact, it's important to accurately measure automation technologies. Patent text is a novel approach to understanding technological progress and identifying the tasks that different automation technologies can perform. The paper uses a dictionary-based approach to identify automation technologies from patent text and categorizes them into three groups: Robots, Software, and Artificial Intelligence. We find that the number of patents related to these groups has grown exponentially since 1980, with AI patents growing the fastest since the 1990s. Most patents are in the Physics (G) patent family. The tasks performed by the different groups of automation technologies vary, with robots focusing on assembly tasks, software on information processing, and AI on higher cognitive tasks. Almost all technologies aim to improve efficiency and reduce costs, according to the patent description.

Keywords: Automation Technologies, Natural Language Processing (NLP), Tasks.

JEL classification: *J23, O33, O34*

1 Introduction

Over the years, technological advancements have significantly changed the job market. Automation technology is one such advancement that has the potential to significantly impact the labor market (Frey and Osborne, 2017). Technology can automate routine and repetitive tasks, leading to increased efficiency and productivity, but it also poses a threat to certain jobs (Acemoglu and Restrepo, 2018). Therefore, it is essential to study the impact of new technologies on the labor market. To assess the impact of these new technologies on the labor market, it is necessary to properly measure the content of technologies that can displace labor. Much of the literature uses different proxies for automation technologies, such as the share of routine tasks in job

descriptions within an industry as a proxy for computerization amenability (Autor et al, 2003; Goos and Manning, 2007), firm-level surveys on the use of computers in the workplace or investments in computer capital (Autor et al., 2003; Beaudry et al., 2010), and finally Graetz and Michaels (2018); Acemoglu and Restrepo (2020); Lábaj and Vitáloš (2021) count the number of robots used in production. The smaller part of the literature uses patent metadata or patent grant text related to automation per se (Mann and Püttmann, 2018) or directly to automation or labor augmentation with reference to the job description (Webb, 2019; Autor et al., 2022). As pointed out by Mann and Püttmann (2018), measuring the adoption and impact of automation technologies is often problematic. However, as Mann and Püttmann (2018) show measuring new technologies directly from patent text provides a novel and somewhat a better approach to understanding technological progress which is less industry biased towards manufacturing sector than the most common approach such as counts of the robots across industries.
In this paper, we follow the pioneering work of Mann and Puttman, (2018), Webb (2020), and Autor et al. (2022) and identify automation technologies using a dictionary-based approach similar to Webb, (2019) and Dechezleprêtre (2021) on almost a universe of granted patents. We differentiate among three broad groups of automation technologies: robotics, software, and artificial intelligence (AI). First we provide a descriptive analysis of patents occurrences labeled by different technology over time and across broad patent families. The main advantage of using patent text as a measure of technological progress embodying automation technologies is that the patent text contains a description of what the technology is capable of. This approach could allow one to calculate the exposure of different tasks performed within different occupations to each particular technology. In the following section, we present the time evolution of patenting activities that embody automation technologies. In addition, we construct, Sankey diagrams of the most common noun-verb pairs (tasks) performed within identified broader groups of automation technologies.

Main findings of this paper could be summarized as follows:
All automation technologies have experienced rapid expansion since 1980, with the appearance of robots technologies in patenting activity dating back to 1940, followed by software technologies two decades later, and the significant development of AI technologies in 1990, which have experienced faster than exponential growth in the last two decades.
The majority of patents were invented within the Physics (G) patent family of the Cooperative Patent Classification (CPC) system. In addition, the tasks performed by different groups of automation technologies vary qualitatively. Robots technology is mainly used to augment or automate assembly tasks, while software technology is used for information processing tasks, and AI technology is capable of performing higher cognitive tasks.
All automation technologies share two common characteristics: improving efficiency and reducing costs. It is important to note that these tasks are capabilities that the technologies can perform and are not a complete representation of their potential and,

more importantly, their implemented functionalities in the production process over time.

First, we extract all patents from the Google Patents Public Dataset including their assigned code, date of first issue, broader patent family classification and title and abstract text. To label our patent text with particular technology (robots, software, and AI) we followed the approach suggested by Webb (2019) and searched for relevant keywords in patent titles and abstracts, which have a high signal-to-noise ratio compared to other text fields. In particular robots technology was identified using the regular expression [robot*|mechatroni(c|cs)|cyber-physical|systems|computer|vision|control systems|sensors], yielding a total of 3,078,364 patents. Software technology was extracted using the regular expression [software|algorithm|computer program|data structure], yielding 2,971,376 patents. Finally, AI technology was identified using the regular expression [artificial intelligence|machine learning|neural network|deep learning], yielding 958,653 patents. We restricted our sample to patents filed between 1940 and 2022, excluding 2023 because a large proportion of patents are currently in the application process and not yet available. Figure 1 shows the total number of patents related to the all groups of technologies grew exponentially over time. In the right panel of Figure 1, the growth rates of robot and software patents appear to resemble linear curves, indicating approximately exponential growth in all time periods. In contrast, the number of patents related to AI technology showed faster than exponential growth since the 1990s.

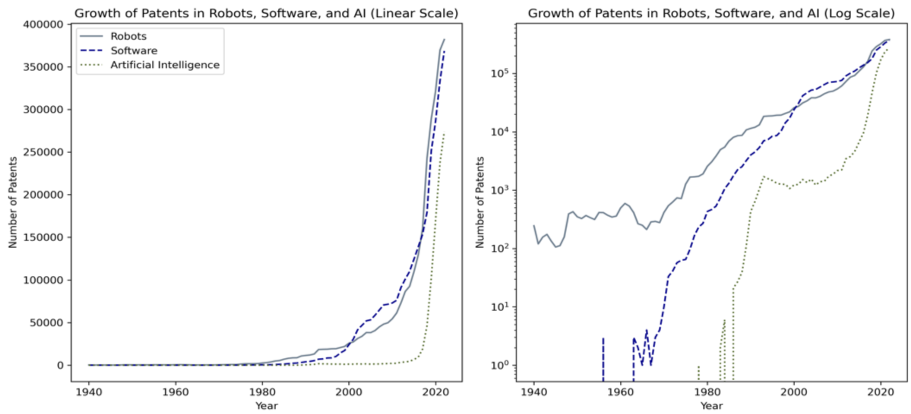


Fig. 1. Total number of patents related to robots, software, AI technology. Source: Autor's elaboration based on Google Patent Public Dataset

Figure 2 shows the relative shares of patents across Cooperative Patent Classification (CPC) invention families for all patents. More detailed decompositions for each technology are shown in Figures A1-A3 in the Appendix. The majority of patents are assigned to the patent family Physics (G), which includes patents from the natural sciences such as computer technologies, optical technologies and biotechnologies. Notably, the second most productive patent family differs across technologies. In robotics, the Performing operation; transporting (B) and Human necessities (A) patent families have experienced significant growth in the relative number of patents granted in recent decades. In software technology, the Electricity (H) patent family has a significant and time-invariant share of all patents. Within AI technology, almost 80% of all patents are concentrated within the Physics (G) family, and the remaining patents fall into the Cross sectional technologies (Y) family, which is a general tagging of CPC for new technological developments that span several sections of the International Patent Classification (IPC).

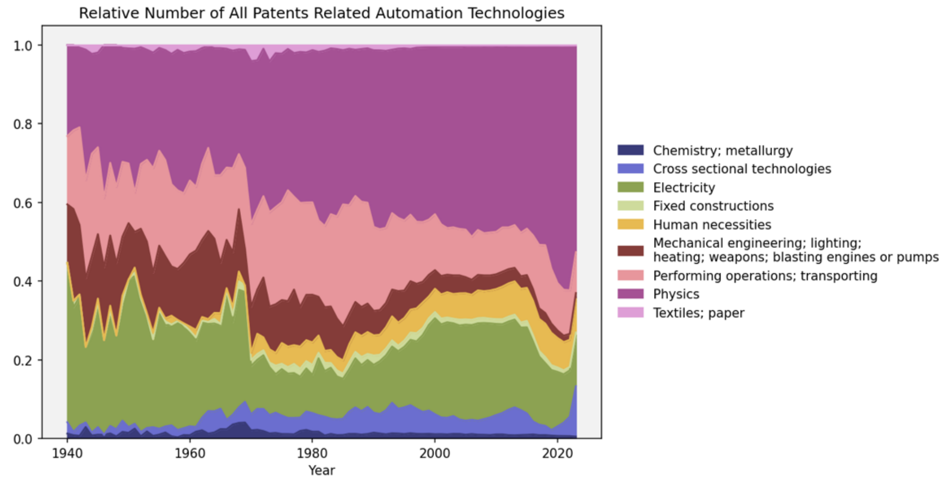
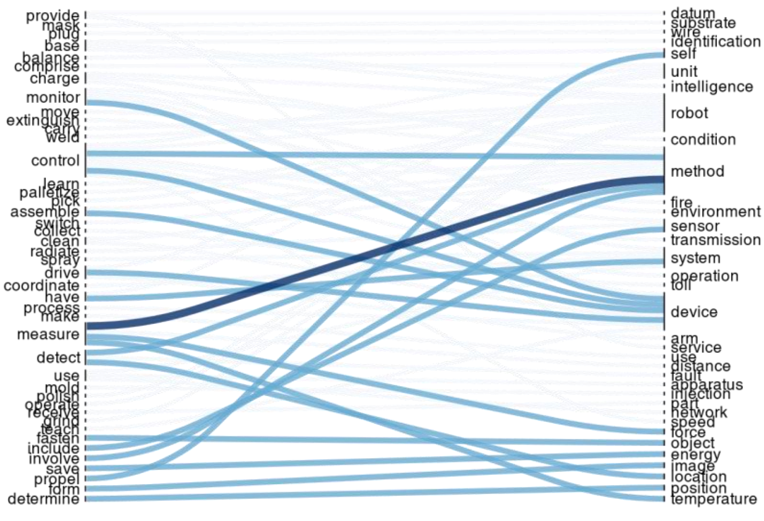


Fig. 2. Flow of all patents over broad CPC patent families in the world. Source: Autor's elaboration based on Google Patent Public Dataset

2 Extracting Capabilities: NLP Analysis of Patents Reveals the Tasks Robots, Software, and AI Can Perform

To illustrate the automation potential of these technologies we have to identify the common tasks performed by these technologies. We employed traditional natural language processing (NLP) techniques on a subsample of patents randomly selected from the entire patent text universe as described in Section 2. Due to computational limitations, we limited our analysis to a random selection of 3,000 patents per robots, software, and AI technology. To test the robustness of our subsamples, we randomly selected 10,000 patents and manually checked whether the most common tasks performed by a particular group of technologies remained largely unchanged. Therefore, we believe that these subsamples of the patent text are representative for identifying patent features of interest that are aggregated tasks performed by the technology, mainly due to the high level of aggregation. Subsequently, we keep only patent titles published in English language, which represents about 80% of all patents published worldwide. As is common practice in NLP literature, we removed all punctuation and stop words from the patent titles. We used dependency parsing algorithm developed by Honnibal and Montanni (2017) to extract all verb-noun pairs from the text of each title, which are commonly considered as proxy for tasks performed by the technological group of patents (see, e.g., Webb, 2019, and Autor et al., 2022). Next, we tokenized and lemmatized[1] all verb-noun pairs to obtain normalized spellings, tenses, and different forms of individual tokens. To construct the most common tasks and make the results of our analysis readable, we selected the 80 most frequent task pairs and removed the first 10 most frequent pairs, which typically contain verb-noun pairs that are semantically related to the patent process rather than the various tasks that patents perform at their core [e.g., disclose ↦ method, invention, system; comprise ↦ method, invention, system; provide ↦ method, invention, system]. Figures 3, 4, and 5 show the Sankey diagrams of 70 combinations of the most common tasks that robots, software, and AI technologies appear to be able to perform. The color intensity shows the relative number of the same tasks across multiple patents, ranging from dark blue for frequently recurring tasks across the group to white blue for tasks that are less common across the analyzed sample, but are still part of the most important tasks that each group of patents can perform. From the same analysis shown in this paper, but on an abstract text, only one fact becomes quite convincing. All automation technologies have two features in common: improving efficiency and saving costs. In fact, the efficiency-improving and cost-saving feature is the most representative characteristic across all patent abstracts. Looking at the figures below from a broad perspective, we can see that robot, software, and AI technologies differ significantly in terms of the tasks they typically perform. Figure 3 shows that tasks that we intuitively ascribe to robots, such as [move, balance, contain, monitor, palletize, assemble, collect, clamp, drive, use, operate ↦ arm, environment, part, object, image, locate, manipulate] are indeed captured in the patent text. We will can call this set of tasks assembling in the production process.



[1] For example, we might use a lemmatization to convert "communicate," "communication," and "communicating" to the base form "communicate."

Fig. 3. The most common combinations of tasks carried out by Robots technology. Source: Autor's elaboration based on Google Patent Public Dataset

In Figure 4 we plot most common tasks performed by Software technology. What is a common denominator of depicted tasks is their relation to the creation, manipulation, transferring, or storing of information. One could immediately create a mental map, that closely resembles the tasks of software technology as we experience it daily, such as [provide, track, use, have, realize, learn, support, select, augment, manage, load, measure, store, route, comprise, bear, face ↦ datum (note: lemmatized expression of data, database, dataset etc.), message, object, parameter, resource, communication, screen, game, function, vision, network, access].

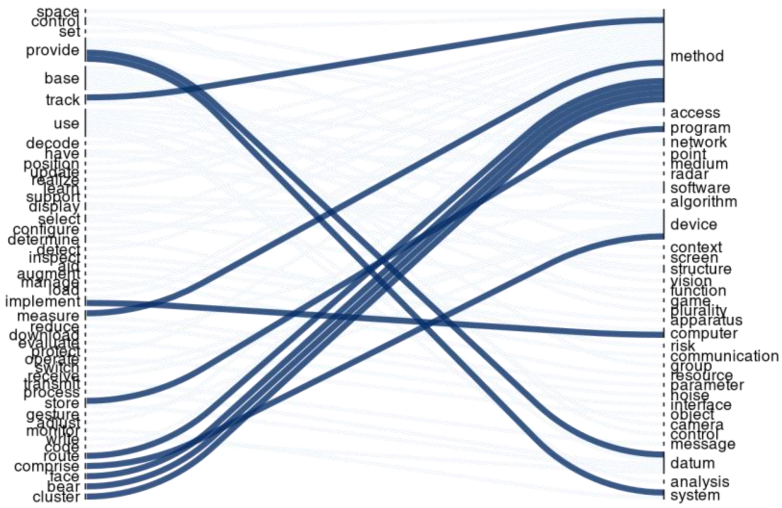


Fig. 4. The most common combinations of tasks carried out by Software technology. Source: Autor's elaboration based on Google Patent Public Dataset

Last figure illustrating tasks that AI technologies perform the most is truly fascinating, cause they are mostly oriented to simulate higher cognitive functions that (to this time only) have humans possessed. As stated by Acemoglu (2021): 'To many commentators, artificial intelligence is the most exciting technology of our age, promising the development of intelligent machines that can surpass humans in various tasks, create new products, services and capabilities, and even build machines that can improve themselves, perhaps eventually beyond all human capabilities.' It is striking to observe that the most common tasks AI is capable of being completely in line with the previous idea [use ↦ intelligence], [learn ↦ system] or [predict ↦ behavior]. Although the vast impact of AI on production and distribution systems is beyond the scope of this paper, we highly recommend further reading on this topic, which can be found in Acemoglu (2021).

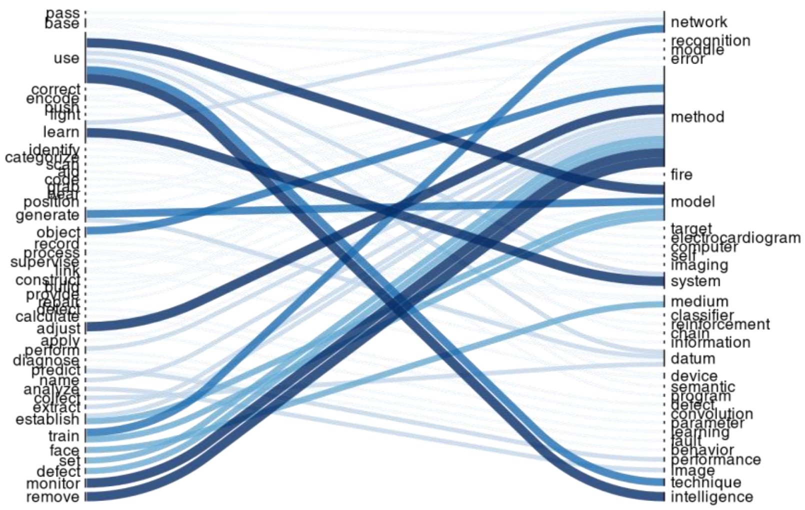


Fig. 5. The most common combinations of tasks carried out by AI technology. Source: Autor's elaboration based on Google Patent Public Dataset

3 Conclusions

In summary, automation technologies, including robots, software, and artificial intelligence, have grown exponentially in recent years and may have the potential to significantly impact the labor market. While these technologies can improve efficiency and productivity, they may also pose a threat to certain jobs. Properly measuring automation technologies is essential to assessing their impact on the labor market, and patent text provides a novel approach to understanding technological progress. Specifically, this paper uses a dictionary-based approach to identify automation technologies from patent text and finds that these technologies vary in the tasks they perform and are mainly developed in the natural sciences, stemming from the broad family of physics patents. Over time, all automation technologies (measured here in three technological groups: robots, software, and AI) experienced rapid expansion since 1980, with the appearance of robot technologies in patenting activity dating back to 1940, followed by software technologies two decades later, and the significant development of AI technologies in 1990, which experienced faster than exponential growth in the last two decades. We focused on identifying the most common tasks that each technology can perform. We documented that robots are mainly used to augment or automate assembly tasks, while software technology is used for information processing tasks, and AI technology is capable of higher cognitive tasks.

4 Acknowledgement

We gratefully acknowledge the financial support from National Bank of Slovakia under the research project 'Enhancing cutting-edge research collaboration between MIT and EUBA in the field of automation', Scientific Grant Agency VEGA under research project 1/0781/21: 'Industrial policy in the context of deindustrialization and automation', and Research Project for Young Teachers, Researchers and Doctoral Candidates in the Internal Form at the University of Economics in Bratislava (I-23-108-00) "Impact of Trade Unions on the Implementation of Automation Technologies Across European Countries".

References

1. Acemoglu, D., Restrepo P.: Robots and jobs: Evidence from US labor markets. *Journal of political economy* 128.6: 2188-2244 (2020).
2. Acemoglu, D., Restrepo P.: The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review* 108.6: 1488-1542 (2018).
3. Acemoglu, D.: Harms of AI. No. w29247. National Bureau of Economic Research, (2021).
4. Autor, D. H., Levy F., Murnane R. J.: The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics* 118.4: 1279-1333 (2003).
5. Autor, D., Chin C., Salomons, A. M., Seegmiller, B.: *New Frontiers: The Origins and Content of New Work, 1940–2018*. No. w30389. National Bureau of Economic Research, (2022).
6. Beaudry, P., Doms M., Lewis E.: Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas. *Journal of political Economy* 118.5: 988-1036 (2010).
7. Dechezleprêtre, A., Hémous, D., Olsen, M., Zanella, C.: Induced automation: evidence from firm-level patent data. University of Zurich, Department of Economics, Working Paper 384 (2021).
8. Frey, C. B., Osborne. A., M.: The future of employment: How susceptible are jobs to computerisation?. *Technological forecasting and social change* 114: 254-280 (2017).
9. Goos, M., Manning, A.: Lousy and lovely jobs: The rising polarization of work in Britain. *The review of economics and statistics* 89.1: 118-133 (2007).
10. Graetz, G., Michaels, G.: Robots at work. *Review of Economics and Statistics* 100.5: 753-768 (2018).
11. Honnibal, M., Montani, I.: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1), 411-420 (2017).
12. Lábaj, M., Vitáloš, M.: Automation and Labor Demand: The Role of Different Types of Robotic Applications. *Finance a Uver: Czech Journal of Economics & Finance* 71.3 (2021).
13. Mann, K., Püttmann, L.: Benign effects of automation: New evidence from patent texts. *Available at SSRN 2959584* (2018).
14. Webb, M. The impact of artificial intelligence on the labor market. *Available at SSRN 3482150* (2019).

Appendix

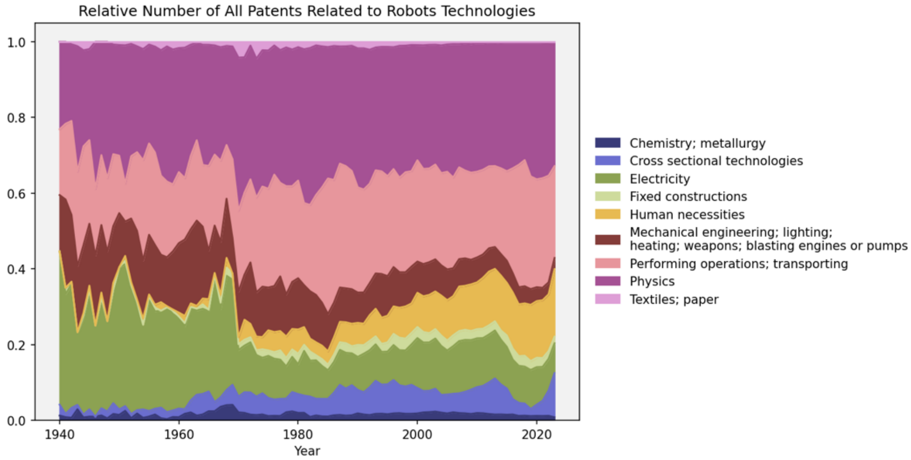
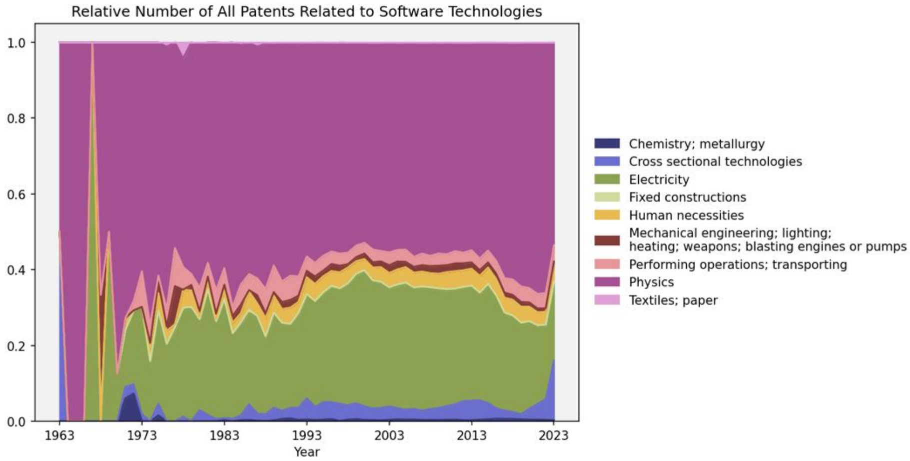


Fig. A1. Flow of patents related to robots technology over broad CPC patent families in the world between 1940 and 2022 period. Source: Autor's elaboration based on Google Patent Public Dataset



Source: Autor's elaboration based on Google Patent Public Dataset

Fig. A2. Flow of patents related to software technology over broad CPC patent families in the world between 1940 and 2022 period. Source: Autor's elaboration based on Google Patent Public Dataset

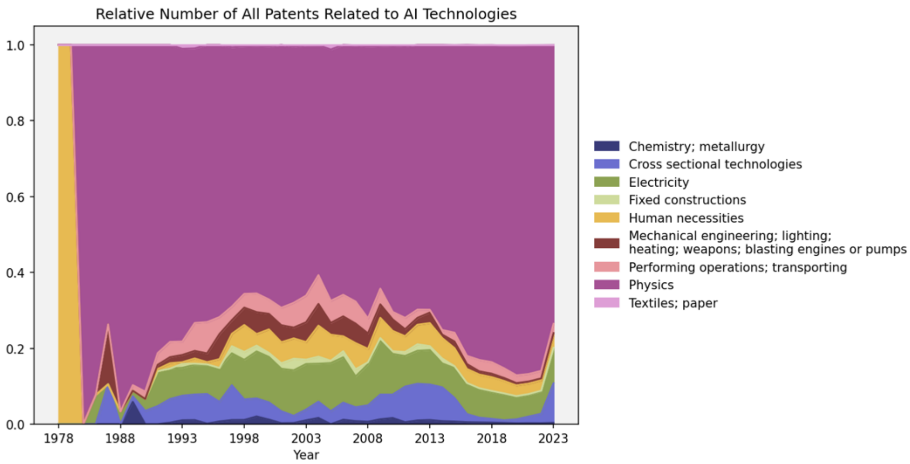


Fig. A3. Flow of patents related to AI technology over broad CPC patent families in the world between 1940 and 2022 period. Source: Autor's elaboration based on Google Patent Public Dataset