OECD*publishing*

LABOUR-SAVING TECHNOLOGIES AND EMPLOYMENT LEVELS

ARE ROBOTS REALLY MAKING WORKERS REDUNDANT?

OECD SCIENCE, TECHNOLOGY AND INDUSTRY
POLICY PAPERS
January 2022 **No. 124**

Federal Ministry
of Labour and Social Affairs

OECD
BETTER POLICIES FOR BETTER LIVES

OECD Science, Technology and Industry Policy Papers

This paper was approved and declassified by written procedure by the Committee for Industry, Innovation and Entrepreneurship (CIIE) on 29 June 2021 and prepared for publication by the OECD Secretariat.

This publication contributes to the OECD's Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) programme, which provides policymakers with new evidence and analysis to keep abreast of the fast-evolving changes in AI capabilities and diffusion and their implications for the world of work. The programme aims to help ensure that adoption of AI in the world of work is effective, beneficial to all, people-centred and accepted by the population at large. AI-WIPS is supported by the German Federal Ministry of Labour and Social Affairs (BMAS) and will complement the work of the German AI Observatory in the Ministry's Policy Lab Digital, Work & Society. For more information, visit *https://oecd.ai/work-innovation-productivity-skills* and *https://denkfabrik-bmas.de/.*

Note to Delegations:

This document is also available on O.N.E. under the reference code

DSTI/CIIE/WPIA(2021)4/FINAL

Labour-Saving Technologies and Employment Levels: are robots really making workers redundant?

Mariagrazia Squicciarini (UNESCO)*,
Jacopo Staccioli (Catholic University of the Sacred Heart, Milan)

This paper exploits natural language processing techniques to detect the presence of explicit labour-saving goals in inventive efforts in robotics and assess their relevance for different occupational profiles and the impact on employment levels. The analysis relies on the universe of patents published by the European Patent Office (EPO) between 1978 and 2019, and on firm-level data from ORBIS® IP. It investigates innovative actors engaged in labour-saving technologies and their economic environment (identity, location, industry), and identifies the technological fields and associated occupations which are particularly exposed to them. Labour-saving patents are especially concentrated in Japan, the United States, and Italy, and seem to affect both low-skilled and blue-collar jobs, along with highly cognitive and specialised professions. While labour displacement of these occupations may occur in the future, a preliminary analysis does not find an appreciable negative effect on employment shares in OECD countries over the past decade. Further research to econometrically investigate the relationship between labour-saving technological developments and employment would be helpful.

JEL classification: O33, J24, C38.

Keywords: Robotics Patents, Labour-Saving Technologies, Technological Unemployment, Search Heuristics

*Mariagrazia Squicciarini led this work while serving as a Senior Economist and Head of Unit at the OECD Directorate for Science Technology and Innovation. She is now UNESCO's Chief of the Executive Office of the Social Human Sciences Sector.

Synthèse

Ce document exploite les techniques de traitement du langage naturel pour détecter la présence d'objectifs explicites d'économie de main-d'œuvre dans les efforts d'inventions en robotique et évaluer leur pertinence pour différents profils professionnels ainsi que leur impact sur les niveaux d'emploi. L'analyse s'appuie sur l'univers des brevets publiés par l'Office européen des brevets (OEB) entre 1978 et 2019, et sur les données au niveau des entreprises d'ORBIS® IP. Il étudie les acteurs innovants engagés dans les technologies économes en main-d'œuvre et leur environnement économique (identité, localisation, industrie), et identifie les domaines technologiques et les métiers associés qui y sont particulièrement exposés. Les brevets permettant d'économiser du travail sont particulièrement concentrés au Japon, aux États-Unis et en Italie et semblent affecter à la fois les emplois peu qualifiés et les cols bleus, ainsi que les professions hautement cognitives et spécialisées. Bien que le déplacement de la main-d'œuvre employée dans ces professions puisse se produire à l'avenir, une analyse préliminaire ne révèle pas d'effet négatif significatif sur les parts d'emploi dans les pays de l'OCDE, au cours de la dernière décennie. Il serait utile de poursuivre les recherches par une analyse économétrique de la relation entre le développement des technologies peu intensives en main-d'œuvre et l'emploi.

Kurzfassung

Diese Studie nutzt Techniken der natürlichen Sprachverarbeitung (Natural Language Processing, NLP), um das Vorhandensein expliziter arbeitssparender Maßnahmen bei Erfindungen in der Robotik zu erkennen und ihre Relevanz für verschiedene Berufsprofile und ihre Auswirkungen auf das Beschäftigungsniveau zu bewerten. Die Analyse basiert auf der Gesamtheit der vom Europäischen Patentamt (EPA) zwischen 1978 und 2019 veröffentlichten Patente sowie auf Daten auf Unternehmensebene aus ORBIS® IP. Sie untersucht die innovativen Akteure, die sich mit arbeitssparenden Technologien befassen, sowie deren wirtschaftliches Umfeld (Identität, Standort, Branche). Ferner identifiziert sie die Technologiefelder und die damit verbundenen Berufe, die besonders stark von ihnen betroffen sind. Arbeitssparende Patente sind besonders in Japan, den USA und Italien zu finden und scheinen sowohl gering qualifizierte und gewerbliche, als auch hoch kognitive und spezialisierte Berufe zu betreffen. Auch wenn es in Zukunft zu einer Verdrängung dieser Berufe kommen kann, zeigt eine vorläufige Analyse zumindest im letzten Jahrzehnt keinen nennenswerten negativen Effekt auf die Beschäftigung dieser Gruppen in OECD-Ländern. Weitere Forschung zur ökonometrischen Untersuchung der Beziehung zwischen arbeitssparenden technologischen Entwicklungen und Beschäftigung wären hilfreich.

Executive Summary

The question of technological unemployment – namely whether technological change has the potential to cause massive job losses with robots making workers redundant – has been under intense scrutiny during recent years. Automation, which is an essential part of the wider "Industry 4.0" concept and the Next Production Revolution, has the aim of enabling production processes and procedures requiring minimal human intervention through use of technologies such as robotics. In turn, the field of robotics has in recent decades leveraged on a number of breakthroughs in related technological fields, such as Artificial Intelligence (AI) and the so-called Internet of Things (IoT), increasingly acquiring the connotation of a pervasive technology.

This work aims to inform the discussion about developments in robotics substituting workers or otherwise complementing human activities. To this end, we implement a natural language approach to detect discussions of labour-saving (LS) intentions, effects or outcomes of specific inventive efforts, with the latter proxied by patented technologies. To this end, we exploit the full text of patents published by the European Patent Office between 1978 and 2019, and we link patent data to firm-level data from ORBIS®, to characterise the leading actors in the robotics space and their LS innovations. Again, using a natural language processing approach, we estimate a similarity measure associating different LS patents to one or more occupations, based on the description contained in the 2008 International Standard Classification of Occupations (ISCO-08). This allows us to identify the occupations that are more likely to be affected by LS developments. We further contrast this with data about employment levels by occupation over time, for a set of 31 OECD countries over the period 2011-19.

The main findings of the analysis and their implications for policymaking are:

- Despite the number of robotics patents steadily increasing since 1978, and at an especially fast pace in the last decade, the share of labour-saving patents has been quite stable over time, confirming that labour-saving goals behind technological innovation are not a new phenomenon, but rather a quite established one.
- Only about a quarter of all labour-saving patents published by the EPO are filed by firms located in European countries or in other members of the European Patent Convention (EPC). The remaining three quarters are evenly distributed between Japan, the United States, and other non-EPC countries, respectively.
- Overall, traditional robotics competence centres, such as Japan, the United States, and Italy, dominate the picture, although a few emerging economies are also present, such as the People's Republic of China (hereafter "China"), India, Turkey, and Brazil.
- Robot manufacturers, especially from Japan, account for most labour-saving product innovations. However, labour-saving goals are found to emerge in many industries, including aerospace, mining, and retail. Also, food processing patents incorporate labour-saving robotics technology (i.e. a process innovation), which could lead to employment disruption once implemented.
- To the extent that labour-saving innovations are implemented in firms' production processes, the occupations most exposed to them would include not only low-skilled and blue-collar jobs (e.g. vehicle drivers and cleaners), but also highly cognitive and specialised professions (e.g. system analysts and application

programmers). While labour displacement of these occupations may occur in the future, a preliminary analysis does not find any evidence of such displacements when looking at employment shares in selected OECD countries over the past decade. Further research to econometrically investigate the relationship between labour-saving technological developments and employment would be helpful.

Overall, our findings point to the need to have a balanced policy discussion about the possible threat posed by automation to labour. The analysis for the first time offers evidence relying on real technological developments as compared to discussions based on the hypothetical automatability of different tasks and occupation. Results suggest that labour saving developments happen in a context of stable employment levels of the same occupations, which these labour-saving technologies have the potential to offset. This points to the existence of mechanisms, such as increased demand, triggered by lower prices as automation unfolds, or supply diversification strategies leading to more and diverse goods and services being produced, that may compensate for the jobs possibly lost due to development and adoption of these labour-saving technologies.

Résumé

La question du chômage technologique – autrement dit si le changement technologique a le potentiel de provoquer des pertes d'emplois massives au moyen de robots rendant les travailleurs redondants - a fait l'objet d'un examen minutieux ces dernières années. L'automatisation, qui est une partie essentielle du concept plus large d' «Industrie 4.0» et de la prochaine révolution de la production, a pour objectif de favoriser des processus et des procédures de production nécessitant une intervention humaine minimale grâce à l'utilisation de technologies telles que la robotique. À son tour, le domaine de la robotique a exploité au cours des dernières décennies un certain nombre de percées dans des domaines technologiques connexes, tels que l'intelligence artificielle (IA) et le soi-disant Internet des objets (IoT), gagnant de plus en plus la réputation d'une technologie omniprésente.

Ce travail vise à éclairer la discussion sur les développements de la robotique remplaçant les travailleurs ou complétant autrement les activités humaines. À cette fin, nous mettons en œuvre une approche en langage naturel pour détecter les discussions sur les intentions, les effets ou les résultats d'efforts inventifs spécifiques en matière d'économie de main-d'œuvre (EMO), ces derniers étant représentés par des technologies brevetées. Pour cela, nous exploitons le texte intégral des brevets publiés par l'Office Européen des Brevets entre 1978 et 2019, et nous relions les données de brevets aux données de niveau entreprise d'ORBIS®, pour caractériser les principaux acteurs de l'espace robotique et leurs innovations en EMO. Encore une fois, en utilisant une approche de traitement du langage naturel, nous estimons une mesure de similitude associant différents brevets EMO à une ou plusieurs professions, sur la base de la description contenue dans la Classification internationale type des professions de 2008 (CITP08). Cela nous permet d'identifier les professions les plus susceptibles d'être affectées par les développements de l'EMO. Nous comparons encore cela avec les données sur les niveaux d'emploi par profession au fil du temps, pour un ensemble de 31 pays de l'OCDE sur la période 2011-19.

Les principales conclusions de l'analyse et leurs implications pour l'élaboration des politiques sont les suivantes:

- Malgré un nombre de brevets de robotique en constante augmentation depuis 1978, et cela à un rythme particulièrement rapide au cours de la dernière décennie, la part des brevets permettant d'économiser de la main-d'œuvre est restée assez stable dans le temps, confirmant que les objectifs d'économie de main-d'œuvre derrière l'innovation technologique ne sont pas un phénomène nouveau mais plutôt bien établi.
- Seul un quart environ de tous les brevets permettant d'économiser du travail publiés par l'OEB sont déposés par des entreprises situées dans des pays européens ou dans d'autres pays membres de la Convention sur le brevet européen (CBE). Les trois quarts restants sont répartis également entre le Japon, les États-Unis et d'autres pays non-CBE, respectivement.
- Dans l'ensemble, les centres traditionnels de compétences en robotique, tels que le Japon, les États Unis et l'Italie, dominent le tableau, bien que quelques économies émergentes soient également présentes, comme la République populaire de Chine (ci-après dénommée “Chine”), l'Inde, la Turquie et le Brésil.
- Les fabricants de robots, en particulier au Japon, sont à l'origine de la plupart des innovations de produits permettant d'économiser du travail. Cependant, des objectifs d'économie de main-d'œuvre sont susceptibles d'apparaitre dans de

nombreuses industries, notamment l'aérospatiale, l'extraction minière et la vente au détail. En outre, des brevets de transformation des aliments intègrent une technologie robotique permettant d'économiser du travail (c'est-à-dire une innovation de procédé), ce qui pourrait entraîner une rupture de l'emploi une fois mise en œuvre.

- Dans la mesure où les innovations en EMO sont intégrées dans les processus de production des entreprises, les professions les plus exposées couvrent non seulement les emplois peu qualifiés et les cols bleus (par exemple les conducteurs de véhicules et les aides de ménage), mais également les professions hautement cognitives et spécialisées (par exemple les analystes de systèmes et les programmeurs d'applications). Bien que le déplacement de la main-d'œuvre employée dans ces professions puisse se produire à l'avenir, une analyse préliminaire des parts d'emploi ne révèle pas de tels déplacements au cours de la dernière décennie pour les pays de l'OCDE étudiés. Il serait utile de poursuivre les recherches par une analyse économétrique de la relation entre le développement des technologies peu intensives en main-d'œuvre et l'emploi.

Dans l'ensemble, nos résultats soulignent la nécessité d'avoir une discussion politique équilibrée sur la menace potentielle posée pour l'emploi par l'automatisation du travail. L'analyse offre pour la première fois des preuves reposant sur des développements technologiques réels par rapport à des débats basés sur une hypothétique automatisation de différentes tâches et professions. Les résultats suggèrent que l'évolution des économies de main-d'œuvre se produit dans un contexte de niveaux d'emploi stables pour les mêmes professions que ces technologies liées à l'économie de main-d'œuvre pourraient compenser. Cela indique l'existence de mécanismes, tels qu'une demande accrue, déclenchée par des prix plus bas à mesure que l'automatisation se développe, ou des stratégies de diversification de l'offre conduisant à la production de biens et de services plus nombreux et diversifiés, qui peuvent compenser les emplois éventuellement perdus en raison du développement et de l'adoption de ces technologies permettant des économies de main-d'œuvre.

Zusammenfassung

Die Frage der technologischen Arbeitslosigkeit - nämlich ob der technologische Wandel das Potenzial hat, massive Arbeitsplatzverluste durch Roboter zu verursachen, die Arbeitnehmer überflüssig machen - wurde in den letzten Jahren intensiv untersucht. Die Automatisierung ist ein wesentlicher Bestandteil des umfassenderen Konzepts "Industrie 4.0" und der nächsten Produktionsrevolution. Sie versucht, durch den Einsatz von Technologien wie der Robotik Produktionsprozesse und -verfahren zu ermöglichen, die nur minimale menschliche Eingriffe erfordern. Eine Reihe von Durchbrüchen in verwandten Technologiebereichen wie der künstlichen Intelligenz (KI) und dem sogenannten Internet der Dinge (Internet of Things, IoT) in den letzten Jahrzehnten haben dem Bereich der Robotik genutzt, sodass er zunehmend den Ruf einer allgegenwärtigen Technologie erlangt hat.

Diese Arbeit zielt darauf ab, die Diskussion über Entwicklungen in der Robotik, die Arbeiter ersetzen oder anderweitig menschliche Tätigkeiten ergänzen, anzuregen. Zu diesem Zweck implementieren wir einen natürlichsprachlichen Ansatz, um Diskussionen über arbeitssparende (AS) Absichten, Auswirkungen oder Ergebnisse spezifischer Erfindungen, repräsentiert durch patentierte Technologien, zu erkennen. Dazu werden Daten aus Originaltexten von Patenten, die von der Europäischen Patentorganisation (EPO) zwischen 1978 und 2019 veröffentlicht wurden, mit Daten auf Firmenebene aus ORBIS® verknüpft, um die führenden Akteure im Bereich Robotik und ihre AS Innovationen zu charakterisieren. Unter erneuter Verwendung natürlicher Sprachverarbeitung wird ein Schätzwert für ein Ähnlichkeitsmaß errechnet, das verschiedene AS Patente einem oder mehreren Berufen laut Beschreibung der Internationalen Standardklassifikation der Berufe 2008 (ISCO-08) zuordnet. Diese Resultate ermöglichen es uns, die Berufe zu identifizieren, die am ehesten von AS Entwicklungen betroffen sein werden. Ferner vergleichen wir sie mit Daten über die Beschäftigungsniveaus in den einzelnen Berufsgruppen für 31 OECD-Länder im Zeitraum 2011-19.

Die wichtigsten Ergebnisse der Analyse und ihre Implikationen für die Politikgestaltung sind:

- Obwohl die Zahl der Robotikpatente seit 1978 stetig und im letzten Jahrzehnt besonders schnell zugenommen hat, ist der Anteil der AS Patente im Laufe der Zeit recht stabil geblieben. Diese Tatsache bestätigt, dass AS Ziele hinter technologischen Innovationen kein neues Phänomen sind, sondern ein recht etabliertes.
- Nur etwa ein Viertel aller von der EPO veröffentlichten AS Patente werden von Unternehmen mit Sitz in europäischen oder anderen Ländern des Europäischen Patentübereinkommens (EPÜ) angemeldet. Die restlichen drei Viertel verteilen sich jeweils zu gleichen Teilen auf Japan, die USA und andere Nicht-EPÜ-Länder.
- Insgesamt dominieren die traditionellen Robotik-Kompetenzzentren wie Japan, die USA und Italien das Bild, obwohl auch einige Schwellenländer wie die Volksrepublik China (nachstehend “China”), Indien, die Türkei und Brasilien vertreten sind.
- Die meisten AS Produktinnovationen kommen von Roboterherstellern, insbesondere aus Japan. Allerdings werden AS Ziele in vielen Branchen beobachtet, darunter Luft- und Raumfahrt, Bergbau und Einzelhandel. Auch Patente in der Lebensmittelverarbeitung beinhalten entsprechende

Robotertechnologie (i.S. einer Prozessinnovation). Diese könnte nach ihrer Einführung zu einer Unterbrechung der Beschäftigung führen.

- In dem Maße, in dem AS Innovationen in die Produktionssprozesse der Unternehmen integriert werden, umfassen die am stärksten davon betroffenen Berufe nicht nur gering qualifizierte und gewerbliche Tätigkeiten (z. B. Fahrzeugführer und Reinigungskräfte), sondern auch hoch kognitive und spezialisierte Berufe (z. B. Systemanalytiker und Anwendungsprogrammierer). Obwohl es in Zukunft zu einer Verdrängung von Arbeitskräften in diesen Berufen kommen kann, zeigt eine vorläufige Analyse unter Berücksichtigung der Beschäftigungsanteile in den ausgewählten OECD-Ländern im letzten Jahrzehnt bisher keine Anzeichen dafür. Weitere Forschung zur ökonometrischen Untersuchung der Beziehung zwischen arbeitssparenden technologischen Entwicklungen und Beschäftigung wären hilfreich.

Insgesamt weisen unsere Ergebnisse auf die Notwendigkeit einer ausgewogenen politischen Diskussion über die mögliche Bedrohung der Beschäftigung durch die Arbeitsautomatisierung hin. Die Analyse bietet zum ersten Mal Fakten, die auf tatsächlichen technologischen Entwicklungen beruhen, im Gegensatz zu Debatten, die auf der hypothetischen Automatisierbarkeit verschiedener Aufgaben und Berufe basieren. Die Ergebnisse deuten darauf hin, dass AS Entwicklungen in einem Kontext stabiler Beschäftigungsniveaus der Berufe stattfinden, die AS Technologien ausgleichen können. Es scheint Mechanismen zu geben, die möglicherweise verlorene Arbeitsplätze durch die Entwicklung und Einführung dieser AS Technologien kompensieren können. Dazu gehören z.B. eine erhöhte Nachfrage ausgelöst durch niedrigere Preise im Zuge der Automatisierung oder Strategien für Angebotdiversifizierung, die zur Produktion von mehr und vielfältigeren Gütern und Dienstleistungen führen.

Table of contents

Tables

Figures

1. Introduction

The question of technological unemployment – namely whether technological change has the potential to cause massive job losses with robots making workers redundant – has been under intense scrutiny during recent years (see e.g. Acemoglu and Restrepo, 2019, 2020; Manyika, Chui and Miremadi, 2017).

> *"While the foundation of our economic system presumes a strong link between value creation and job creation, the Great Recession reveals the weakening or breakage of that link. This is not merely an artifact of the business cycle but rather a symptom of a deeper structural change in the nature of production." [Brynjolfsson and McAfee, 2011]*

> *"Industrial robots are on the verge of revolutionizing manufacturing. As they become smarter, faster and cheaper, they're being called upon to do more—well beyond traditional repetitive, onerous or even dangerous tasks such as welding and materials handling. They're taking on more "human" capabilities and traits such as sensing, dexterity, memory, trainability, and object recognition. As a result, they're taking on more jobs." [PwC, 2014]*

Automation, which is an essential part of the wider "Industry 4.0" concept[1] and of the Next Production Revolution (OECD, 2017), has the aim of enabling production processes and procedures requiring minimal human intervention. The field of robotics, whose initial seeds date back to the ancient world, has recently leveraged on a number of breakthroughs in related technological fields, such as Artificial Intelligence (AI) and the so-called Internet of Things (IoT), increasingly acquiring the connotation of a pervasive technology.

This work aims to inform the discussion about developments in robotics substituting workers or otherwise complementing human activities. To this end, we implement a natural language approach to detect the presence of explicit labour-saving (hereafter, LS) heuristics[2] in inventive efforts, as proxied by patented technologies. We exploit the full text of patents published by the European Patent Office (EPO) between 1978 and 2019, and we link patent data to firm-level data from ORBIS® IP (Bureau van Dijk) to identify and characterise the leading actors in the robotics space and their LS innovations.

Again, using a natural language processing (NLP) approach[3], we estimate a similarity measure associating different LS patents to one or more occupations, based on the occupation description contained in the 2008 International Standard Classification of Occupations (ISCO-08). This allows us to identify the occupations that are more likely to be affected by LS developments, and contrast this with data about employment levels by occupation over time, for a set of 31 OECD countries over the period 2011-19.[4]

The patent-based analysis draws heavily on the methodology developed by Montobbio et al. (2021, 2022) and proceeds in a general-to-specific fashion, according to the following steps. First, all patents relevant to robotics are singled out from the universe of patents filed at the EPO since its inception in 1978 and up to 31 December 2019. Within the scope of this paper, robotics is intended in a broad sense, in order to encompass not only core technological advances in the field, but also closely related enabling technologies and applications to production processes. Second, patents that explicitly mention a LS heuristic in their full-text are identified by means of a text-mining algorithm, ahead of further manual validation aimed to remove false positives. Third, LS patents are matched to their current corporate assignee(s) via the ORBIS® IP database. This allows uncovering the identity of underlying innovative actors, their geographic location, and their sector of economic activity. Finally, LS patents are mapped to job occupations through a text similarity algorithm between patents' full-text and ISCO-08 4-digit definitions. The similarity score is then used to rank occupations by relevance to the overall corpus of LS innovations.

We find that, despite the steady increase in the number of robotics patents observed since 1978, and the especially fast pace characterising the last decade, the share of LS patents has been quite stable over time.

The majority of LS patents published by the EPO are not filed by firms located in European countries or other members of the European Patent Convention (EPC), but rather by traditional robotics competence centres located in Japan and the United States (US). Among patents filed from companies located in EPC countries, Italy has the largest number of LS patents, with about twice as many patents as Germany, which ranks second.

Robot manufacturers, especially Japanese conglomerates, account for most LS product innovations, although LS heuristics are found to emerge along the entire supply chain.

To the extent that LS innovations are implemented into firms' production processes, among the occupations most likely to be impacted one would find both low-skilled and blue-collar jobs (e.g. vehicle drivers and cleaners), along with highly cognitive and specialised professions (e.g. system analysts and application programmers). While labour displacement of these occupations may occur in the future, a preliminary analysis does not find an appreciable negative effect on employment shares in selected OECD countries during the past decade. Further research to econometrically investigate the relationship between labour-saving technological developments and employment would be helpful.

The remainder of the paper is organised as follows. Section 2 provides some theoretical background and reviews the relevant literature. Section 3 describes our data sources and the workflow of our methodology. Section 4 provides a firm-level characterisation of innovative actors engaged in LS technologies. Section 5 identifies the occupations more exposed to LS innovations. Finally, Section 6 summarises the main results of the paper and concludes.

2. Framing the problem in the related literature

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 *focussing devices*, in the sense of Rosenberg (1976), namely the ensemble of technological bottlenecks, market incentives, and ultimately the cognitive *loci* and the behavioural patterns of those engaged in creating these technologies (Dosi and Nelson, 2010, 2013).

Although it is generally hard to identify invariant and ex-ante search heuristics or inducement effects, search efforts aimed at the reduction of human inputs in production appear to be invariant throughout the history of capitalist societies (Dosi, 1988; Rosenberg, 1976; von Tunzelmann, 1995). In what follows, we shed light on whether recent technological innovations pursue, or are even dominated, by such heuristics.

In our framework, robotics, and indirectly AI (which sees a number of applications in robotics, see e.g. Baruffaldi et al., 2020), are seen as pervasive *general purpose technologies* (GPTs)[5], 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). Moreover, patents, as loci of explicit codified knowledge, are generally considered as a good empirical instrument to proxy the rate and direction of innovative activity (Pavitt, 1985; Griliches, 1990). Although, in principle, there is no guarantee that patented innovations are actually implemented into production processes, it is worth noting that firms generally incur substantial R&D and patenting expenses, and it is unlikely that they would do so unless there is a reasonable expectation to make a profit out of them (Griliches, 1990). By looking at the textual contents of robotic patents, we thus aim at isolating the ones explicitly embedding a labour substituting trait.

As mentioned, our analysis is primarily based on the methodology developed by Montobbio et al. (2021, 2022) which, to the best of our knowledge, constitutes the sole contribution in the literature aimed at uncovering the presence of explicit LS heuristics within patents' full-text and computing a *direct* measure of occupational exposure. This paper extends the contributions of Montobbio et al. (2021, 2022) by mapping European LS patents to official occupational definitions and building a rank of occupations exposed to potential labour displacement.

Contributions close to the spirit of the present work include Mann and Püttmann (2018), which classifies US patents as automation or non-automation related, and links them to the industries of their use and, through local industry structure, to commuting zones. According to their estimates, advances in automation technologies have a positive influence on employment in local labour markets, as manufacturing employment declines but is more than compensated by job growth in the service sector.

Dechezleprêtre et al. (2019) analyse the relationship between higher wages and innovative activity in automation. They use the frequency of certain keywords in patent texts to identify automation innovations in machinery and build a firm-level panel combining macroeconomic data from 41 countries and information on geographical patent history to build firm-specific measures of low-skill and high-skill wages. They find that an increase in low-skill wages leads to more automation innovation with an elasticity between 2 and 4, while an increase in high-skill wages tends to reduce automation innovation.

Acemoglu and Restrepo (2020) study the impact of industrial robots on US labour markets, estimated as variations in exposure to robots, defined from industry-level advances in robotics and local industry employment. They obtain robust negative effects of robots on employment and wages, with one additional robot per thousand workers reducing the employment-to-population ratio by 0.2 percentage points and wages by 0.42%.

Frey and Osborne (2017) use a mix of the Delphi method and a Gaussian process classifier to assess the probability of computerisation for a set of 702 occupations. According to their estimates, 47% of total US employment is at high risk of potential automation over some unspecified number of years.

Arntz et al. (2016) estimate the automatability of jobs for 21 OECD countries based on a task-based approach and find that the threat from technological advances is much less pronounced compared to the occupation-based approach of Frey and Osborne (2017). Arntz et al. (2016) find that, overall, 9% of jobs are automatable across OECD countries, extreme examples being 6% in South Korea and 12% in Austria.

Nedelkoska and Quintini (2018) focus on the risk of automation and its interaction with training and the use of skills at work for individual jobs across 32 OECD countries. They find that close to one in two jobs are likely to be affected by automation, based on the tasks they involve, and 14% of jobs display a probability of automation above 70%. These are especially concentrated in Eastern and Southern European countries, Germany, Chile, and Japan (the upper extreme being Slovakia with 33% of jobs at high risk of automation), while jobs in Anglo-Saxon and Nordic countries, and the Netherlands, appear relatively less automatable (the lower extreme being Norway with 6%).

Georgieff and Milanez (2021) employ the risk measure of automation of Nedelkoska and Quintini (2018) to evaluate jobs deemed at high risk of automation and find that, over the period 2012-2019, employment growth had been in fact substantially lower for jobs at high risk of automation (6%), compared to jobs at low risk of automation (18%).

Finally, building on Montobbio et al. (2022), Staccioli and Virgillito (2021a) provide empirical evidence on the history of LS innovations back to the early 19th century, tracking their time evolution, clustering, eventual emergence of periodic behaviour, and their co-movements with long-term GDP growth (see also Staccioli and Virgillito, 2021b). Santarelli et al. (2021) extend the search to artificial intelligence patents and map the underlying knowledge base, both as a whole and distinguishing core and related innovations, along a 4-level core-periphery architecture.

3. Data and methodology

Our analysis covers the entire set of 6 109 462 patent applications and grants published by the European Patent Office (EPO) since its inception in 1978 up to 31 December 2019. Full-texts have been downloaded from the EPAB platform[6], an expert tool designed to monitor European patents, offering advanced query capabilities and full-text search.

Our methodology unfolds along the lines of Montobbio et al. (2021, 2022). By means of textual analysis and natural language processing, we aim at pinpointing a subset of patents that relate to robotics technology in a broad sense, and then look for LS heuristics therein. Next, we match our data with ORBIS® IP (Bureau van Dijk) firm-level database in order to characterise the relevant innovators in LS technologies in terms of industry and geographic location (see Section 4). Finally, we construct a text similarity measure that maps LS patents to ISCO-08 occupational definitions (see Section 5; mathematical definitions are reported in the Appendix).

In what follows, Section 3.1 shows how we reduce the universe of EPO publications to a set of robotics-related patents, whereas Section 3.2 identifies patents, within the robotic subset, which explicitly mention some sort of LS heuristics.

3.1. Robotic patents

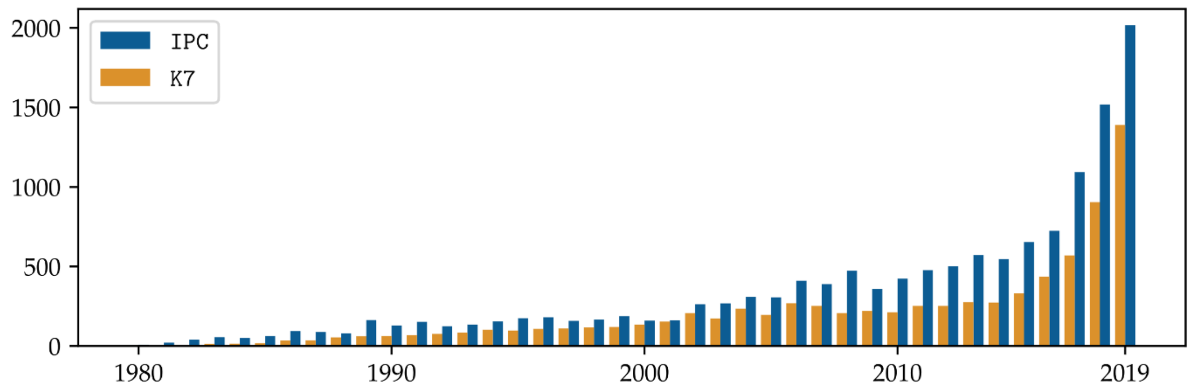
We begin with restricting the universe of EPO patents to those published in English, to avoid biases that could arise from the textual analysis of documents written in different languages.[7] In particular, we find 4 382 445 such documents within the EPAB database.

In order to identify the subset of robotics patents, we implement a twofold approach. First, we leverage patent classification codes known to be relevant to robotics. Montobbio et al. (2022) construct a list of 124 full-digit CPC (Cooperative Patent Classification) codes associated with robotic technologies. Here, we adopt an identical scheme, except that EPAB only allows classification queries for IPC (International Patent Classification) codes. We make use of the official CPC to IPC concordance table[8] to reconcile the approach of Montobbio et al. (2022) with our data. The concordance table maps the 124 original CPC codes into 109 distinct full-digit IPC codes. A patent is deemed "robotic" if it is assigned at least one of these 109 IPC codes.

Second, we perform a keyword search across all publications, excluding the ones already identified by the first criterion, in order to pinpoint other inventions which are closely related or tightly complementary to robotics but were not officially classified as such by patent examiners. A patent is deemed robotic if it contains the morphological root `'robot'` repeated at least 7 times across its abstract, description or claims sections. Given the low degree of ambiguity, patents exhibiting a copious repetition of this morphological root have been found to be tightly linked to robotics and its implementation within industrial manufacturing processes. The number 7 was decided based on a sensitivity analysis assessing the probability of false positive and false negatives obtained using a different number of repetitions of the same morphological root.[9]

In total, we find 21 977 robotics patents, 13 852 of which are selected by the first criterion and labelled `'IPC'`, and 8 125 selected by the second criterion and labelled `'K7'`.[10] Their evolution over time is depicted in Figure 1.

Figure 1. Number of robotics patents by year



Note: Blue bars denote the absolute number of 'IPC' robotics patents, identified through relevant IPC codes. Orange bars denote the absolute number of 'K7' robotics patents, identified through multiple occurrences (at least 7 times) of the keyword 'robot' in their full texts. Overall, there exist 21 977 robotics patents, of which 13 852 'IPC' patents and 8 125 'K7' patents.
Source: Authors' own compilation based on EPO data.

At a first glance, it is apparent that the overall number of robotics patents has steadily increased over time, reflecting the general surge in patenting activity. Moreover, it is possible to observe that the number of IPC robotics patents is consistently greater than the number of K7 robotic patents, in line with the relative size of the two subsets. This marks a first difference with respect to USPTO patents studied by Montobbio et al. (2022), for which patents selected by the keyword criterion are twice as many as those selected by classification codes.

A second difference with respect to American patents is given by the frequency of robotic patents overall. While Montobbio et al. (2022) find close to 30 000 total robotic patents in the period from 2009 to 2018, here we find roughly one third less within a much wider time frame, although, as we have seen, most robotic patents are concentrated in the past decade (cf. Figure 1).

3.2. Labour-saving patents

Our second methodological challenge lies in the discovery of the set of labour-saving patents. From the set of robotic patents identified in the previous section, we now want to single out those explicitly claiming a LS effect of the underlying innovation. We do this by performing a multiple *word* co-occurrence query at the *sentence* level, along the lines of Montobbio et al. (2022).

We start by pre-processing our textual corpus along the following steps. First, we subdivide, technically *tokenise*, the full text of each robotics patent (a single string concatenating the abstract, the description, and the claims sections) into a list of sentences.[11] 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.[12]

This allows us to look for the presence of specific words (actually, morphological roots) within the whole corpus of robotics patents. We aim at eliciting heuristics, 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 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.

Figure 2. Structure of the labour-saving textual query

$$\underbrace{\begin{bmatrix}\texttt{'reduc'}\\ \texttt{'replac'}\\ \texttt{'elimin'}\\ \texttt{'save'}\\ \texttt{'lower'}\\ \texttt{'substitut'}\\ \texttt{'autom'}\\ \texttt{'less'}\end{bmatrix}}_{\text{verbal predicate}} \times \underbrace{\begin{bmatrix}\texttt{'labo[u]r'}\\ \texttt{'worker'}\\ \texttt{'human'}\\ \texttt{'employe'}\\ \texttt{'manpow'}\\ \texttt{'job'}\\ \texttt{'manual'}\end{bmatrix}}_{\text{direct object}} \times \underbrace{\begin{bmatrix}\texttt{'cost'}\\ \texttt{'expenditure'}\\ \texttt{'expens'}\\ \texttt{'hour'}\\ \texttt{'intens'}\\ \texttt{'task'}\\ \texttt{'time'}\\ \texttt{'skill'}\\ \texttt{'effort'}\\ \texttt{'intervent'}\end{bmatrix}}_{\text{object attribute}}$$

Note: Words in the *verbal predicate*, *direct object*, and *object attribute* lists undergo the same stemming procedure as patent full texts. If a patent full text includes one word from each of the three lists within the same sentence, the patent is marked as *potentially* labour-saving and is then manually validated.
Source: Authors' own compilation.

Figure 2 shows the selected words that we use in our query, which are an augmented version of those appearing in Montobbio et al. (2022). First, we disambiguate `'labo[u]r'` spelled both with and without the `'u'` in order to better capture patent documents from British firms and inventors. Second, we add the stem `'less'` to the predicate group, `'manual'` to the direct object group, and `'effort'` and `'intervent'` to the attributes group. 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 (640, given the Cartesian product of the three sets) triplets.

This textual query singles out 1 662 inventions, 858 of which belong to the `IPC` class and 804 to the `K7` class. 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 1 545 *truly* LS patents (hereafter referred simply as LS patents), i.e. approximately 7% of all robotic patents, suggesting our methodology exhibits an accuracy of ≈93%. Of these, 814 (≈52.7%) come from the `IPC` group and 731 (≈47.3%) from the `K7` group.

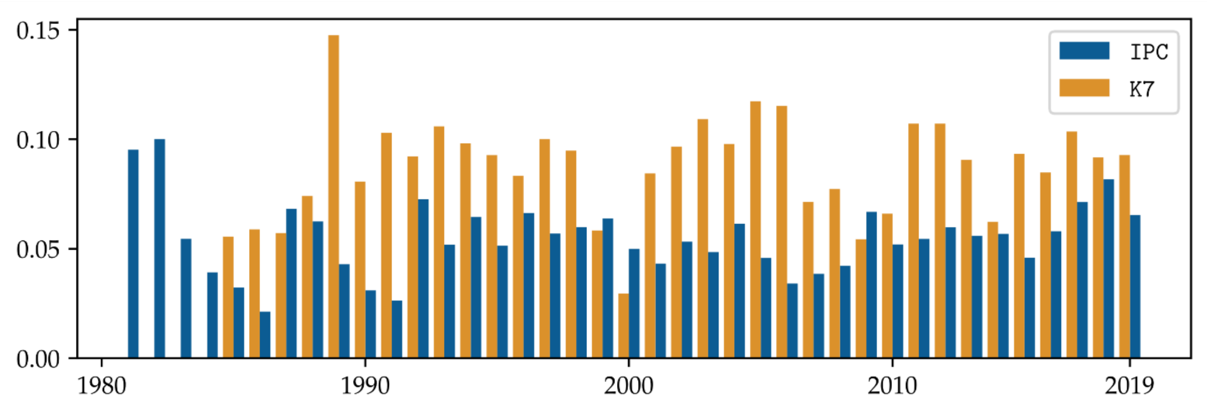
LS patents are found to make quite sharp statements of economic relevance, which somewhat go beyond the thorough technical description legally required for their enforcement. While it is possible that our selection of LS patents is a by-product of a specific writing style of a small group of individuals, Montobbio et al. (2022) show that, for USPTO patents, on average only a handful of patents are managed by the same patent attorney. Although no public data on entities that have been granted power of attorney for

EPO patents is readily available, their findings suggest that the writing style bias is likely to be negligible. A few selected excerpts are reported below in chronological order, along with the publication number and year of the patent considered.

> *"Robots […] [satisfy] the demand for saving labor and rationalization of work in view of the current rise in labor cost." [EP0068026A1, 1980]*
>
> *"The need for skilled labor, along with the attendant costs in training and replacement is reduced and, furthermore, if the skills involved constitute more an art than a skill, the call for such talent is avoided." [EP0778957B1, 1995]*
>
> *"Automated machining stations can be used to manufacture large quantities of pieces quickly and completely without human intervention." [EP2475501B1, 2009]*
>
> *"The main function of a robot arm is to act as a substitute for human arms and do repetitive or demanding works so as to increase productivity and reduce labor costs." [EP3379410A1, 2017]*

These examples suggest that the economic incentives behind LS innovations pervade even earlier waves of innovative efforts in the realm of mechanisation.

Figure 3. Number of labour-saving patents by year, as a fraction of robotics patents.



Note: Blue and orange bars denote the share of robotics patents of the 'IPC' and 'K7' types, respectively, which are found to be labour-saving according to a search procedure entailing a text-mining query (see Figure 2) and manual validation.
Source: Authors' own compilation based on EPO data.

Figure 3 shows the evolution in the number of LS patents over time, as a fraction of all robotic patents. A substantial share of LS patents come from technological fields which do not belong to the standard robotics-related IPC codes, although the first LS inventions in early 1980s are of the latter type. No clear trend is detectable in the picture, suggesting that the underlying LS heuristics have remained quite stable over time, hinting at a mature and established pattern. To sum up, we detect both an increasing innovative effort devoted to robotics technology, and a plateaued search heuristic guided towards labour-displacement, in relative terms.

4. Who develops labour saving technologies? A firm level analysis

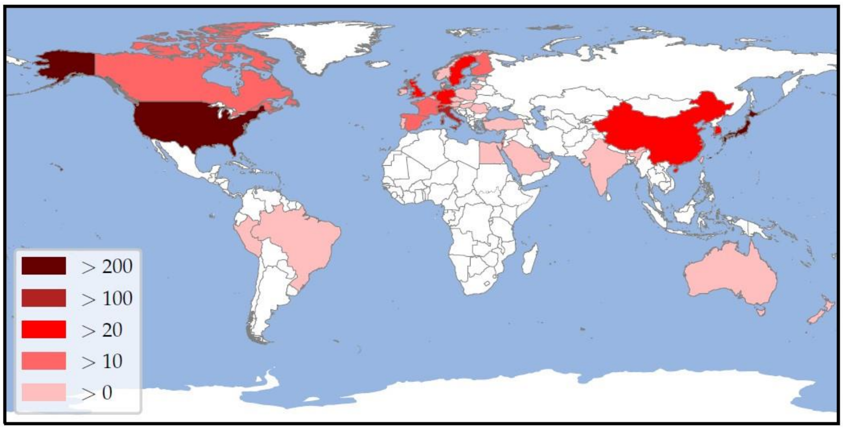
In the present section, we characterise LS patents in terms of identity, geographic location, and industrial sector of their current corporate assignee(s). To this purpose, we match our data to the ORBIS® IP (Bureau van Dijk) database through the relevant publication numbers. Of the 1 545 LS patents found in the previous step, 1 322 (≈85.6%) find a match with at least one firm. In total, there are 787 firms which hold at least one LS patent (hereafter, LS firms). Note that patents assigned to more than one firm are deliberately double counted, since we aim at grasping the degree of dispersion of the underlying LS heuristics. With respect to the American case, in which 1 276 LS patents are concentrated in only 408 firms (Montobbio et al., 2022), implying that each firm on average holds at least three patents, the European case suggests a wider dispersion of LS heuristics, since the average number of LS patents per firm, of about 1.7, is much lower.

We now proceed by characterising labour-saving firms by their geographic location (Section 4.1), identity and industrial classification (Section 4.2).

4.1. Labour-saving firms' geographic location

Montobbio et al. (2022) show that the primary source of LS patents at the USPTO is domestic. In other words, American LS patents are mainly produced by firms located in the US. This does not appear to be the case for European patents. The world map in Figure 4 offers an overview of the geographic distribution of LS patents at the EPO, given the location of their current assignee(s). Japan dominates the picture, with 400 LS patents, followed by the United States, with 394 LS patents. When combined together, member countries of the European Patent Convention only account for 403 LS patents at the EPO. Table 1 reports the number of LS patents for each member country, distinguishing between `IPC` and `K7`. Somewhat surprisingly, Italy, with 107 LS patents, not only ranks first among European countries, but tops the worldwide chart as the third sovereign country in a strict sense, surpassing other traditional robotics hubs and competence centres such as Germany (which ranks 4th) with 54 LS patents and China (5th) with 50.

Figure 4. Geographic location of labour-saving patents in absolute terms



Note: Colour coding represents the absolute number of labour-saving patents held by firms currently incorporated in the underlying country. There are 44 countries overall which hold at least one labour-saving patent.
Source: Authors' own compilation based on EPO and ORBIS IP data.

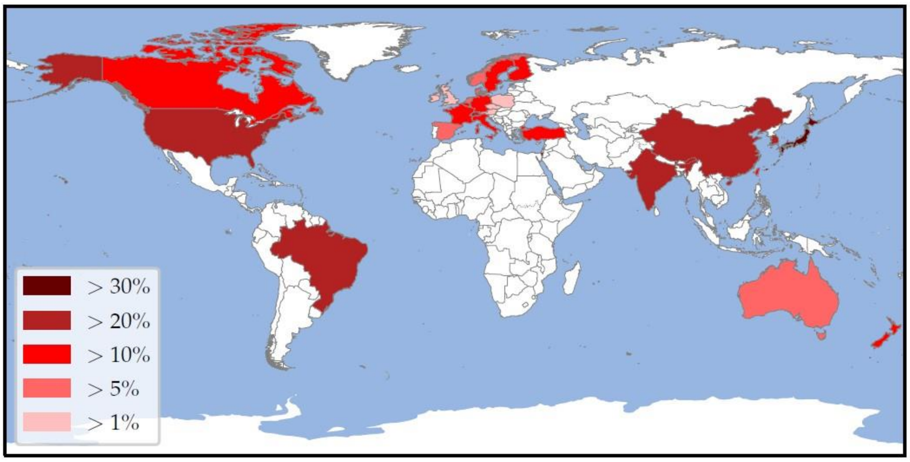
Table 1. Number of labour-saving patents assigned to firms located in member countries of the European Patent Convention.

	LS patents		
Country	**IPC**	**K7**	**Total**
Italy	73	34	107
Germany	36	18	54
Netherlands	14	30	44
Switzerland	27	13	40
Sweden	18	19	37
United Kingdom	18	10	28
France	6	12	18
Spain	9	9	18
Finland	15	1	16
Denmark	1	9	10
Belgium	5	4	9
Portugal	3	1	4
Norway	1	2	3
Estonia	0	2	2
Ireland	0	2	2
Turkey	2	0	2
Austria	1	0	1
Czech Republic	1	0	1
Hungary	0	1	1
Liechtenstein	0	1	1
Lithuania	1	0	1
Malta	0	1	1
Poland	1	0	1
Romania	1	0	1
Slovenia	1	0	1
Total	234	169	403

Note: Among the 38 current member countries of the European Patent Convention, the following have not filed any labour-saving robotics patent application: Albania, Bulgaria, Croatia, Cyprus, Greece, Iceland, Latvia, Luxembourg, Monaco, Republic of North Macedonia, San Marino, Serbia, and Slovakia.
Source Authors' own compilation based on EPO and ORBIS IP data.

Looking at absolute LS patents figures only provides a partial understanding of the associated international patenting activities. Focussing on a relative measure of propensity, i.e. rescaling the number of LS patents by the total number of robotics patents assigned to firms in a given country, allows inference on where LS search efforts are more intensive, compared to the ex-ante capability of producing a robotics patent. This new measure is represented in Figure 5, where countries holding fewer than 10 robotics patents are discarded.

Figure 5. Geographic location of labour-saving patents, as a fraction of robotics patents



Note: Colour coding represents the share of labour-saving patents, out of robotics patents, held by firms currently incorporated in the underlying country. There are 44 countries overall for which this share is strictly positive. Countries with a share below or equal to 1% are not shown.
Source: Authors' own compilation based on EPO and ORBIS IP data.

Among the countries with more than 10 robotic patents, Japan again leads the ranking, with more than 35% of its robotics patents also being LS. Second comes Israel (about 33%) and third comes India (about 30%). The US stands at about 22.7% while China shows levels of about 20.2%.

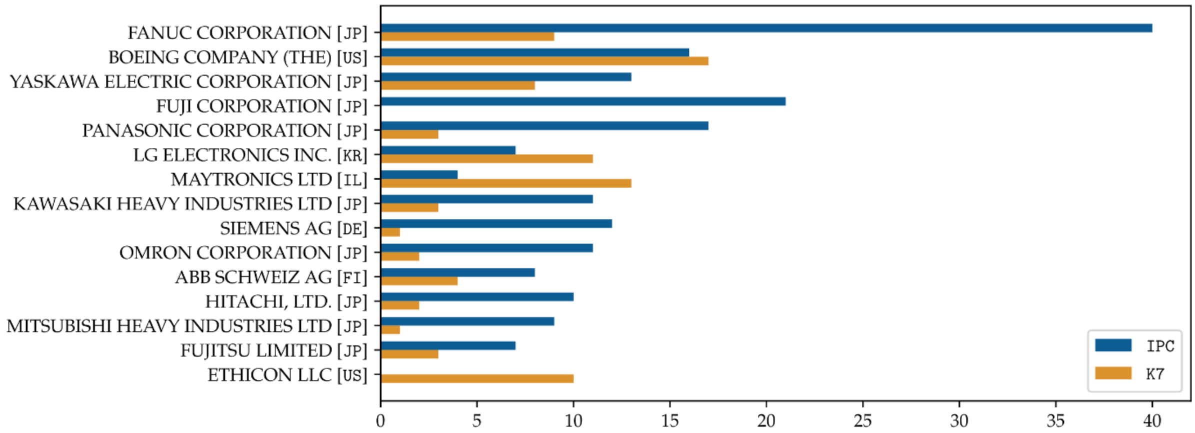
Domestically, robotic patents from member countries of the European Patent Convention are also LS in approximately 10% of cases. Italy, which ranks first among them when only the absolute number of patents is considered, is also first when it comes to this rescaled measure, with a ratio of about 16.5%. Other EPO countries featuring a position above average are Sweden (about 16.0%), Finland (about 15.8%), Turkey (12.5%), Germany (about 12.4%), and Belgium (about 10.6%).

It is worth noting that, while it is possible to trace a patent to a certain location, given the address of its corporate holder, it is not possible to distinguish whether the innovation is used or generates revenues domestically or at a different location. This could be especially an issue for multinational enterprises.

4.2. Labour-saving firms' identity and industrial classification

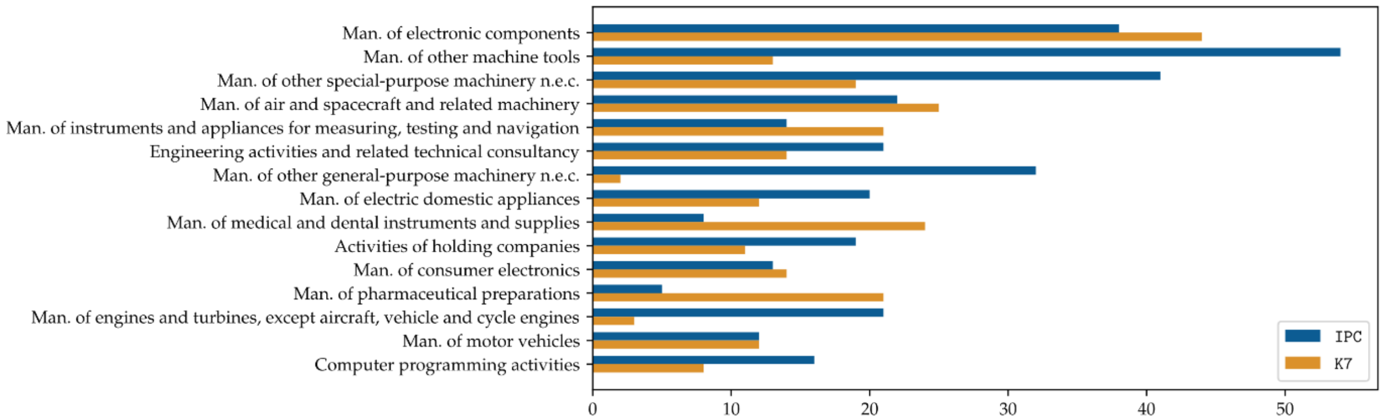
We now proceed by revealing the identity of the top LS patents holding firms and then characterise their sectoral dispersion. Figure 6 lists the top 15 holders by absolute number of LS patents, while Figure 7 lists the description of the top 15 primary sectors, identified as 4-digit NACE Rev. 2 assigned to the holders.

Figure 6. Top 15 firms holding labour-saving patents



Note: Blue (respectively orange) bars denote the absolute number of labour-saving patents of the 'IPC' (respectively 'K7') type currently held by the underlying firm. The chart is sorted by the sum of both types of patents, in decreasing order. There exist 787 firms overall which hold at least one labour-saving patent.
Source: Authors' own compilation based on EPO and ORBIS IP data.

Figure 7. Top 15 NACE industry descriptions of labour-saving patents' holders



Note: Blue (respectively orange) bars denote the absolute number of labour-saving patents of the 'IPC' (respectively 'K7') type currently held by firms whose economic sector is classified as the underlying NACE 4-digit code. The chart is sorted by the sum of both types of patents, in decreasing order. There exist 164 4-digit NACE codes whose associated firms hold at least one labour-saving patent (see Figure 8).
Source: Authors' own compilation based on EPO and ORBIS IP data.

Both pictures detail the underlying IPC and K7 composition. As expected, following the analysis in Section 4.1, Japanese high-tech companies are prevalent, ranging from industrial robots manufacturers (such as Fanuc, Yaskawa), to producers of heavy machinery and transport equipment (e.g. Kawasaki, Mitsubishi), other general purpose machinery equipment (e.g. Fuji), semiconductors and electr(on)ic components (e.g. Yaskawa, Omron, Hitachi), computers and peripheral equipment (e.g. Fujitsu), and electronic domestic appliances (e.g. Panasonic). The only US firms making it to the list are Boeing, the aircraft manufacturer, positioned second among the top 15, and Ethicon, a subsidiary of Johnson & Johnson, positioned last.

This is not entirely surprising, as Montobbio et al. (2022) show that Boeing happens to be the overall largest holder of LS patents at the USPTO. South Korea is represented by LG

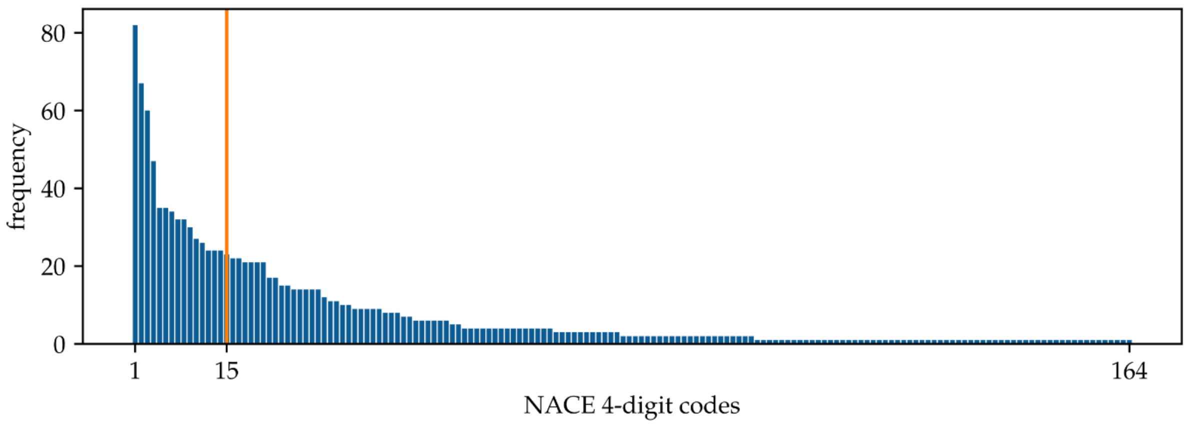
(6th position), the consumer electronics manufacturer, and Israel is represented by Maytronics (7th), which manufactures cleaner robots. Strikingly, European firms feature at the bottom of the chart, with Siemens (Germany), an industrial conglomerate, raking 9th, and ABB (Switzerland), a robots and electronics producer, ranking 11th.

In general, this picture suggests that major LS patents holders belong to producers of robots and related technologies, rather than adopters, who seek to implement robotics in their own production processes. This result is also in line with the picture emerging from Figure 7, which highlights a tight link between LS patents and manufacturing of inherently high-tech products.

There are exceptions to this dualistic view, however. Boeing has been found to develop robotic innovations for internal use, such as fuselage mounting platforms, and thus constitutes a producer and adopter at the same time (Montobbio et al., 2022). Possibly, Kawasaki follows a similar approach, since their innovations mainly deal with welding robots, which can be fruitfully employed within their motor vehicles production line. Given the conglomerate nature of many of the aforementioned Japanese groups, there may be even more examples of in-house horizontal synergies between the R&D department and manufacturing assembly lines.

All in all, the distinction between robots producers (i.e. the 'upstream' sector) and adopters (the 'downstream' sector) is sharper in the case of LS patents protected at the USPTO (Montobbio et al., 2022) and less so in the case of European patents.

Figure 8. Rank-frequency distribution of NACE codes across labour-saving firms



Note: The first 15 NACE codes, to the left of the orange vertical line, are described in better detail in Figure 7.
Source: Authors' own compilation based on EPO and ORBIS IP data.

The frequency distribution of NACE codes assigned to LS patent holders, depicted in Figure 8, is also worth noting. In particular, it reveals that, while most of the (787) LS firms are concentrated in a few industries (already shown in Figure 7), LS patents are overall present in as many as 164 distinct sectors, corresponding to 54 distinct NACE 2-digit specifications and covering all Level 1 codes except sections S ("Other Service Activities "), T ("Activities of Households as Employers [. . .]"), and U ("Activities of Extraterritorial Organisations and Bodies"). In other words, the distribution exhibits a 'long tail' across a wide support of NACE codes. This ultimately suggests that the LS heuristics embedded in robotic technology is quite widespread across the value chain.

5. Occupational exposure

In the present section, we map the knowledge base embedded in LS patents to actual employment categories. To this end, inspired by the methodology used in Montobbio et al. (2021), we construct a similarity measure between the full-text of LS patents and occupational definitions contained in the official ISCO-08 classification[13]. From the latter, we exclude group `10` "Armed forces occupations" and we focus on definitions at the 4-digit unit group level. Overall, we obtain a corpus of 430 definitions, which undergo the same text pre-processing workflow described in Section 3.2.

From a methodological point of view, we adopt the so-called *bag-of-words* model and measure textual proximity between patents and occupations by means of *cosine similarity* (see e.g. Aggarwal, 2018). The bag-of-words model entails the representation of text as a *multiset* of underlying words, which disregards any grammar structure and the order in which terms appear but keeps their multiplicity. The underlying assumption is that occupation-patent pairs whose descriptions consist of the very same words (actually, stems, see Section 3.2), and possibly repeatedly, are more associated to one another than pairs which share few common words, or their frequency is negligible.

Each piece of text, either a patent or an occupational definition, is transformed into a vector of frequencies of the underlying words. The number of vector components reflects the common dictionary of terms across the two whole corpuses. In other words, all vectors belong to the same vector space, whose dimension equals the number of distinct words in the common dictionary. The similarity of an occupation-patent pair is then quantified as the cosine of the angle between the two underlying vectors.

As opposed to simply counting the occurrences of each term in each piece of text, we adopt the customary *tf-idf* (i.e. term frequency–inverse document frequency) term-weighting scheme for computing relevant frequencies (see Definition 1 in the Appendix). The tf-idf statistics reflects how important a specific term is to a certain document, relative to other documents in the collection. The tf-idf value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus mentioning that word. This helps adjusting for the fact that some words appear more frequently in general.

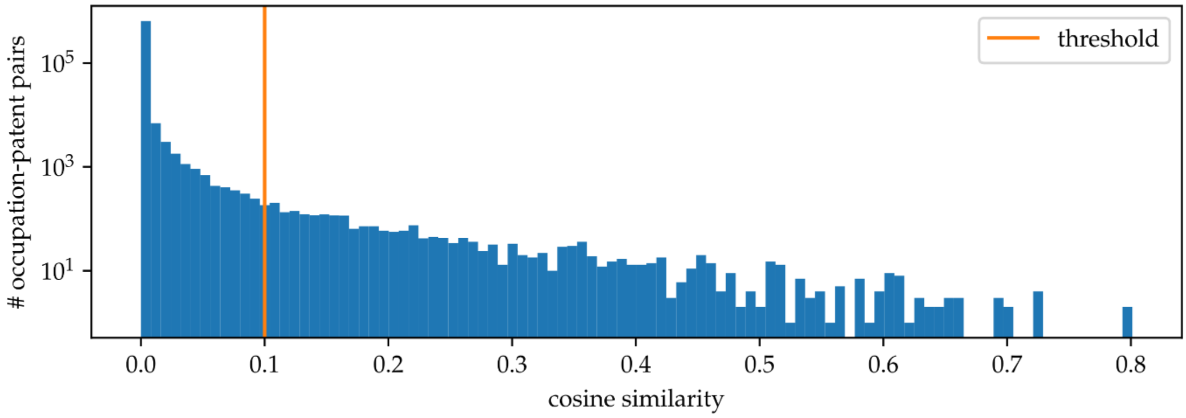
Extending the reasoning to the corpus level, we construct two document-term matrices, $\mathcal{D}^{ISCO}$ and $\mathcal{D}^{LS}$, whose rows contain the aforementioned tf-idf frequency vectors, for each ISCO occupation and each LS patent, respectively. Both matrices are based on the dictionary of terms from ISCO definitions, namely the smaller between the two collections, which consists of 5633 terms. Therefore $\mathcal{D}^{ISCO}$ has dimension 430 × 5633 and $\mathcal{D}^{LS}$ has dimension 1545 × 5633. To ensure that only words that are specific enough to the underlying pieces of text are retained, we introduce an additional restriction. We require words whose tf-idf is below a cut-off value of 0.4 not to be considered in the subsequent analysis.[14] Put simply, we substitute values below the cut-off in the document-term matrices with zeros.

Finally, we construct a cosine similarity matrix $\mathcal{C}$ - to measure how similar two documents are, irrespective of their size - containing the cosine similarity between all pairs of row vectors from the document-term matrices $\mathcal{D}^{ISCO}$ and $\mathcal{D}^{LS}$ (see Definition 2 in the Appendix):

$$\mathcal{C} = \cos(\mathcal{D}^{ISCO}, \mathcal{D}^{LS}).$$

Matrix $\mathcal{C}$ has dimension 430 × 1545, i.e. one row for each occupation and one column for each LS patent, and each cell contains the similarity score of the underlying occupation-patent pair. There exist 664 350 such pairs, whose values are summarised by the histogram in Figure 9.

Figure 9. Cosine similarity values across all occupation-patent pairs



Note: The computation of cosine similarity values includes 430 ISCO occupations and 1,545 labour-saving patents. The overall number of occupation-patent pairs is 664,350. The vertical axis of the histogram is in logarithmic scale. Occupation-patent pairs to the left of the orange vertical line (with a value of 0.1) are discarded from subsequent analysis.
Source: Authors' own compilation based on EPO and ISCO data.

For the sake of robustness, we discard occupation-patent pairs whose similarity score is below a threshold of 0.1 (marked by a vertical line in Figure 9). In what follows, the analysis is based on the remaining 2 413 pairs, corresponding to the top 0.36% of the distribution (note that the vertical axis of Figure 9 is in logarithmic scale). However, in separate experiments, we find that either retaining the whole set of pairs or setting a higher threshold of 0.5 (102 pairs corresponding to the top 0.02% of the distribution), yields remarkably similar results.

In order to rank the occupations by similarity score with the entire ensemble of LS patents, we compute the row sums of matrix $\mathcal{C}$ across all patents (columns) and sort the resulting values in decreasing order.

Table 2. Top 40 ISCO 4-digit occupations by exposure to labour-saving patents

ISCO	sim.	mean	stdev	# pat.	Title
2511	1.00	0.05	0.06	256	Systems Analysts
4132	0.83	0.03	0.07	161	Data Entry Clerks
9331	0.79	0.02	0.09	100	Hand and Pedal Vehicle Drivers
9122	0.79	0.02	0.09	100	Vehicle Cleaners
7231	0.53	0.02	0.06	104	Motor Vehicle Mechanics and Repairers
2514	0.33	0.02	0.05	63	Applications Programmers
8159	0.32	0.02	0.04	73	Textile, Fur and Leather Products Ma...
8154	0.32	0.02	0.04	73	Bleaching, Dyeing and Fabric Cleanin...
8156	0.32	0.02	0.04	73	Shoemaking and Related Machine Opera...
7223	0.32	0.02	0.04	73	Metal Working Machine Tool Setters a...
8183	0.32	0.02	0.04	73	Packing, Bottling and Labelling Mach...
3254	0.22	0.01	0.04	34	Dispensing Opticians
3131	0.21	0.01	0.04	37	Power Production Plant Operators
7549	0.21	0.01	0.05	32	Craft and Related Workers Not Elsewh...
7213	0.19	0.01	0.04	30	Sheet Metal Workers
5165	0.18	0.01	0.04	28	Driving Instructors
1321	0.18	0.01	0.04	34	Manufacturing Managers
7311	0.17	0.01	0.04	21	Precision-instrument Makers and Repa...
3155	0.15	0.01	0.03	34	Air Traffic Safety Electronics Techn...
6320	0.14	0.00	0.04	19	Subsistence Livestock Farmers
6121	0.14	0.00	0.04	19	Livestock and Dairy Producers
2250	0.14	0.00	0.04	19	Veterinarians
5164	0.14	0.00	0.04	19	Pet Groomers and Animal Care Workers
9332	0.14	0.00	0.04	19	Drivers of Animal-drawn Vehicles and...
6129	0.14	0.00	0.04	19	Animal Producers Not Elsewhere Class...
6130	0.14	0.00	0.04	19	Mixed Crop and Animal Producers
7312	0.13	0.01	0.03	22	Musical Instrument Makers and Tuners
4322	0.13	0.01	0.03	26	Production Clerks
2523	0.13	0.01	0.03	24	Computer Network Professionals
3513	0.13	0.01	0.03	24	Computer Network and Systems Technic...
2522	0.13	0.01	0.03	24	Systems Administrators
2351	0.11	0.00	0.03	22	Education Methods Specialists
9121	0.11	0.01	0.03	26	Hand Launderers and Pressers
9321	0.11	0.01	0.03	26	Hand Packers
3240	0.10	0.00	0.03	16	Veterinary Technicians and Assistants
8311	0.09	0.01	0.03	16	Locomotive Engine Drivers
3122	0.09	0.01	0.03	19	Manufacturing Supervisors
4323	0.08	0.00	0.02	19	Transport Clerks
8157	0.08	0.00	0.02	17	Laundry Machine Operators
5243	0.08	0.00	0.03	15	Door-to-door salespersons

Note: The first column reports relevant ISCO 4-digit occupational codes. The second column reports the aggregate similarity measure to the underlying occupational code, computed as the sum of associated cosine similarity values across all labour-saving patents and normalised between 0 and 1. The third and fourth columns report the mean and standard deviation of the distribution of cosine similarity values for each relevant occupation-patent pair. The fifth column reports the absolute number of labour-saving patents taken into account, i.e. for which the cosine similarity value is above the 0.1 threshold (see Figure 9). The last column reports the official definition of the underlying ISCO occupational title. The table is sorted by the normalised similarity value, in decreasing order.
Source: Authors' own compilation based on EPO and ISCO data.

Table 2 reports the top 40 exposed occupations, where the overall similarity measure has been normalised between 0 and 1. The mean and average of the similarity distribution across patents, and the number of underlying patents whose cosine similarity is greater than the 0.1 threshold, are also shown.

As can be noticed, the chart includes both blue- and white-collar occupations. It should nevertheless be considered that, since exposure is quantified by text similarity of underlying occupation-patent pairs, our measure is able to capture which occupations are inherently relevant to the corpus of LS patents, while it does not distinguish between those which will likely witness an increase or a decrease in demand. Accordingly, system analysts (ISCO code 2511), data entry clerks (4132), and applications programmers (2514), which all constitute "enabling" professions of Industry 4.0 technological artefacts and rank remarkably high for exposure, are not likely to belong to the set of occupations threatened by LS innovations, at least in the near future.

Once these are discarded, highly exposed occupations predominantly include a range of low- to high-skill blue-collar jobs. It is worth keeping in mind that, according to some routine intensity estimates by Marcolin et al. (2016) based on 20 OECD countries, in 2012 on average 46% of employed persons worked in non-routine (18%) or low (28%) routine intensive occupations. Interestingly, highest levels of exposure belong to occupations in the service sector. Hand and pedal vehicle drivers (9331), such as delivery riders, seem to be threatened by LS innovations in the logistics sector (which also ranks first in the human-machine taxonomy of Montobbio et al., (2022) for American patents). Vehicle cleaners (9122) seem to be threatened by cleaning robots' producers, among which Israeli Maytronics Ltd. ranks 7th in the chart of top LS patents holders (cf. Figure 6).

Next come a group of occupational titles mainly engaged in shop floor and warehouse jobs in a diversified range of sectors. Focussing on the top half of Table 2, these include: the engineering and automotive industry, with motor vehicle mechanics and repairers (7231), metal working machine tool setters and operators (7223), and sheet metal workers (7213); the clothing/garment industry, with shoemaking and related machine operators (8156), textile, fur and leather products machine operators (8159), and bleaching, dyeing and fabric cleaning machine operators (8154); logistics and general warehouse workers, with packing, bottling and labelling machine operators (8183).

Another cluster of occupations at risk of future displacement because of LS technological developments is associated to livestock breeding and related activities, with animal producers (6129), mixed crop and animal producers (6130), subsistence livestock farmers (6320), livestock and dairy producers (6121), pet groomers and animal care workers (5164), and veterinarians (2250).

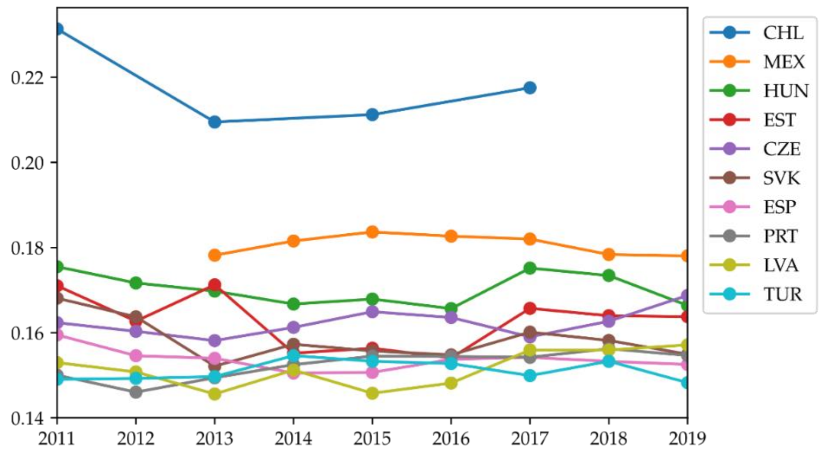
In order to better grasp the employment magnitude underlying jobs exposed to LS innovations, we aggregate the similarity data to ISCO occupations at the 2-digit level and match them with the ILO (International Labour Organization) employment database for 31 OECD countries (since data regarding Australia, Canada, Colombia, Japan, South Korea, and New Zealand are not available). For each country, we compute the overall employment shares associated to the top ten 2-digit occupations by aggregate similarity. Table 3 reports the normalised similarity scores for the first ten ISCO 2-digit occupations, while Figure 10 shows the time evolution of the aggregate employment shares for the top ten countries where this measure is highest, on average.

Table 3. Top 10 ISCO 2-digit occupations by exposure to labour-saving patents

Sim.	ISCO	Title
1.0	25	Information and Communications Technology Professionals
0.99	81	Stationary Plant and Machine Operators
0.72	72	Metal, Machinery and Related Trades Workers
0.63	93	Labourers in Mining, Construction, Manufacturing and Transport
0.58	91	Cleaners and Helpers
0.53	31	Science and Engineering Associate Professionals
0.48	41	General and Keyboard Clerks
0.26	51	Personal Services Workers
0.25	61	Market-oriented Skilled Agricultural Workers
0.25	73	Handicraft and Printing workers

Note: The first column reports the aggregate similarity measure to the underlying occupational code, computed as the sum of associated cosine similarity values across all labour-saving patents and normalised between 0 and 1. The second column reports relevant ISCO 2-digit occupational codes. The last column reports the official definition of the underlying ISCO occupational title. The table is sorted by the normalised similarity value, in decreasing order.
Source: Authors' own compilation based on EPO and ISCO data.

Figure 10. Time evolution of employment shares for top 10 OECD countries by aggregate similarity over top 10 ISCO 2-digit occupations



Note: For each country, a weighted average is computed over ISCO 2-digit occupations reported in Table 3.
Source: Authors' own compilation based on EPO, ISCO, and ILO data.

The country with the largest fraction of labour force exposed to LS technologies is Chile, followed by Mexico, Hungary, and Estonia. Overall, between 2011 and 2019 (the widest timespan available using ILO data) all employment shares have experienced negligible changes. An analogous dynamics also applies to other OECD countries included in the analysis (not shown here). This may suggest that either LS technologies have yet to be implemented into production processes in these countries, or that their effect has been offset, on average, by complementary labour-*augmenting* technologies or by increases in related demand (Bessen, 2019). Further research to econometrically investigate the relationship between labour-saving technological developments and employment would be helpful.

6. Concluding remarks

This work informs the discussion about developments in robotics substituting workers or otherwise complementing human activities. To this end, we implement a natural language approach to detect the presence of explicit labour-saving heuristics in inventive efforts, as proxied by patented technologies, which exploits the full text of patents published by the European Patent Office (EPO) between 1978 and 2019. We further link patent data to firm data from ORBIS® IP, to characterise the leading actors in the robotics space and their LS innovations. Again, using a natural language processing (NLP) approach, we estimate a similarity measure associating different LS patents to one or more occupations, based on the description contained in the 2008 International Standard Classification of Occupations (ISCO-08).

This allows identifying the occupations that are more likely to be affected by LS developments, and contrasting this with data about employment levels by occupation over time, for a set of 31 OECD countries over the period 2011-19.

A first result of our study is that labour-saving patents are found to make quite sharp statements of economic relevance, which somewhat go beyond the thorough technical description legally required for their patent protection.

A second finding is that, despite the number of robotics patents having been steadily increasing since 1978, and at especially fast pace in the last decade, the share of labour-saving patents has been quite stable over time. This confirms that labour-saving heuristics behind technological innovations are not a new phenomenon, but rather a quite established one.

A third finding is that only about a quarter of all labour-saving patents published by the EPO are filed by firms located in European countries or other members of the European Patent Convention (EPC). The remaining three quarters are roughly equally distributed between Japan, the US, and other non-EPC countries. Overall, traditional robotics competence centres, such as Japan, US, and Italy, dominate the picture, although a few developing countries are also present, such as China, India, Turkey, and Brazil.

A fourth finding is that robots' manufacturers, especially from Japan, account for most labour-saving product innovations. However, labour-saving heuristics are found to emerge along the entire supply chain. Industries such as aerospace, mining, retail, and food processing patent labour-saving robotics technology as process innovation, which could lead to employment disruption once implemented.

A fifth finding is that, to the extent that labour-saving innovations are implemented in firms' production processes, occupations most exposed to them would include not only low-skilled and blue-collar jobs (e.g. vehicle drivers and cleaners), but also highly cognitive and specialised professions (e.g. system analysts and application programmers). While labour displacement of these occupations may be expected to occur in the future, a preliminary analysis does not find an appreciable negative effect on employment shares in selected OECD countries during the past decade.

Overall, our findings point to the need to have a balanced policy discussion about the possible threat posed by automation to employment. The descriptive evidence provided, which for the first time relies on real technological developments as compared to estimates of hypothetical automatability of tasks and occupations (as is the case of e.g. Frey and Osborne, 2017), suggests that labour-saving developments happen in a context of stable employment levels of the same occupations that these labour-saving technologies have the potential to offset. Demand-related dynamics as well as an expansion of production may well explain these findings and call for future research to econometrically investigate the relationship between labour-saving technological developments and employment.

Endnotes

1 The term "Industrie 4.0" (shortened to "I4.0" or "I4"), was first proposed in 2011 in a high-tech strategy project of the German government, promoting the computerisation of manufacturing, and has been widely adopted since to refer to a new production paradigm or revolution.

2 Heuristics is a term that refers to the mental strategies that an agent employs to reduce the cognitive demand associated with certain decision-making tasks. Heuristics generally means "rules" that people use to make judgments or estimates of probabilities and frequencies, also in situations of uncertainty.

3 NLP is a subfield of artificial intelligence (AI) that deals with the interaction between computers and humans using the natural language. The ultimate objective of NLP is to read, decipher, understand, and extract information from natural language in a manner that is valuable.

4 It is important to acknowledge that during the global COVID-19 pandemic, still ongoing at the time of writing, robots and other autonomous systems, thanks to their intrinsic immunity to the virus, have played a crucial role, allowing the continuation of certain pivotal economic activities, especially in the healthcare sector, that otherwise would have been partially disrupted (Javaid et al., 2020; Shen et al., 2020).

5 According to Bresnahan and Trajtenberg (1995, p. 84) a GPT is "characterized by the potential for pervasive use in a wide range of sectors". In other terms, it is a single technology that underpins other technologies and multiply their value, with the ability of generating generalised productivity gains.

6 Available here: https://www.epo.org/searching-for-patents/technical/ep-full-text.html

7 The EPO publishes patent documents in either English, French or German.

8 Available here: https://www.cooperativepatentclassification.org/cpcConcordances

9 This entailed a non-trivial effort, as we had to read a representative sample of patents identified using different numbers of repetitions of the morphological root `'robot'`, namely between 5 and 10.

10 Duplicate publications, which can arise when an application is republished or gets granted, are removed and only the latest version is retained.

11 We do this by means of a punctuation regular expression, which is a is a sequence of characters that specifies a textual search pattern.

12 When running a search, it is key to find relevant results not only for the exact expression typed but also for other possible forms of the words used. This can be achieved via two methods, i.e. *stemming* and *lemmatization*, both aimed at reducing the words into a common base or root. Stemming algorithms work by means of removing the end or the beginning of the word, following common prefixes and suffixes that can be found in a word. Lemmatisation conversely takes into account the morphological analysis of the words. In the present work, we use the so-called `Porter2` stemmer, an improved version of the pioneering Porter (1980) algorithm.

13 Available here: http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm

14 This value has been fine-tuned to maximise the average cosine similarity across occupation-patent pairs.

References

Acemoglu, Daron and Pascual Restrepo (2019). '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 (2020). 'Robots and Jobs: Evidence from US Labor Markets'. *Journal of Political Economy* 128(6), pp. 2188–2244. DOI: 10.1086/705716

Aggarwal, Charu C. (2018). *Machine Learning for Text*. Springer.

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

Baruffaldi, Stefano, Brigitte van Beuzekom, Hélène Dernis, Dietmar Harhoff, Nandan Rao, David Rosenfeld, and Mariagrazia Squicciarini (2020). *Identifying and measuring developments in artificial intelligence: Making the impossible possible*. OECD Science, Technology and Industry Working Papers, No. 2020/05, OECD Publishing, Paris. DOI: 10.1787/5f65ff7e-en

Bessen, James (2019). 'Automation and jobs: when technology boosts employment'. *Economic Policy 34(100), 589–626. DOI:* 10.1093/epolic/eiaa001

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

Cockburn, Iain M., Rebecca Henderson, and Scott Stern (2018). *The Impact of Artificial Intelligence on Innovation*. NBER Working Paper No. 24449. URL: https://www.nber.org/papers/w24449

Dechezleprêtre, Antoine, David Hémous, Morten Olsen, and Carlo Zanella (2019). *Automating Labor: Evidence from Firm-level Patent Data*. Available at SSRN. URL: https://ssrn.com/abstract=3508783

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 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: 10.1111/j.1468-0297.1997.tb00064.x

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.

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

Georgieff, Alexandre and Anna Milanez (2021). *What happened to jobs at high risk of automation?* OECD Social, Employment and Migration Working Papers, No. 255, OECD Publishing, Paris. DOI: 10.1787/10bc97f4-en

Griliches, Zvi (1990). 'Patent Statistics as Economic Indicators: A Survey'. *Journal of Economic Literature* 28(4), pp. 1661–1707.

Javaid, Mohd, Abid Haleem, Abhishek Vaish, Raju Vaishya, and Karthikeyan P. Iyengar (2020). 'Robotics Applications in COVID-19: A Review'. Journal of Industrial Integration and Management 5(4), pp. 441–451. DOI: 10.1142/S2424862220300033

Mann, Katja and Lukas Püttmann (2018). *Benign Effects of Automation: New Evidence from Patent Texts*. Available at SSRN. URL: https://ssrn.com/abstract=2959584

Manyika, James, Michael Chui, and Mehdi Miremadi (2017). *A future that works: AI, automation, employment, and productivity*. McKinsey Global Institute Research, Tech. Report 60.

Marcolin, Luca, Sébastien Miroudot, and Mariagrazia Squicciarini (2016). *The Routine Content Of Occupations: New Cross-Country Measures Based On PIAAC*. OECD Trade Policy Papers, No. 188, OECD Publishing, Paris. DOI: 10.1787/5jm0mq86fljg-en

Montobbio, Fabio, Jacopo Staccioli, Maria Enrica Virgillito, and Marco Vivarelli (2022). 'Robots and the origin of their labour-saving impact'. *Technological Forecasting and Social Change* 174, 121122. DOI: 10.1016/j.techfore.2021.121122

Montobbio, Fabio, Jacopo Staccioli, Maria Enrica Virgillito, and Marco Vivarelli (2021). *Labour-saving technologies and occupational exposure*. LEM Working Paper Series 43/2021. URL: http://www.lem.sssup.it/WPLem/2021-43.html

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

OECD (2017). *The next production revolution and institutions for technology diffusion*. OECD Publishing, Paris. DOI: 10.1787/9789264271036-en

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

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

PwC (2014). *The new hire: how a new generation of robots is transforming manufacturing*. PricewaterhouseCoopers report. URL: https://www.nist.gov/document/thenewhirepdf

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

Santarelli, Enrico, Jacopo Staccioli, and Marco Vivarelli (2021). *Robots, AI, and Related Technologies: A Mapping of the New Knowledge Base*. LEM Working Papers series n. 01/2021. URL: http://www.lem.sssup.it/WPLem/2021-01.html

Shen, Yang, Dejun Guo, Fei Long, Luis A. Mateos, Houzhu Ding, Zhen Xiu, Randall B. Hellman, Adam King, Shixun Chen, Chengkun Zhang, and Huan Tan (2020). 'Robots Under COVID-19 Pandemic: A Comprehensive Survey'. *IEEE Access* 9, pp. 1590–1615. DOI: 10.1109/ACCESS.2020.3045792

Staccioli, Jacopo and Maria Enrica Virgillito (2021a). 'Back to the past: the historical roots of labour-saving automation'. *Eurasian Business Review* 11, pp. 27–57. In press. DOI: 10.1007/s40821-020-00179-1

Staccioli, Jacopo and Maria Enrica Virgillito (2021b). 'The present, past, and future of labor-saving technologies'. In: *Handbook of Labor, Human Resources and Population Economics*. Ed. by Klaus F. Zimmermann. Springer. DOI: 10.1007/978-3-319-57365-6_229-1

Trajtenberg, Manuel (2018). *AI as the next GPT: A Political-Economy Perspective*. NBER Working Paper No. 24245. URL: https://www.nber.org/papers/w24245

von Tunzelmann, G. Nick (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

Annex A. Appendix

Term frequency – inverse document frequency

Definition 1 Let D be a collection of documents d, each composed of an ensemble of terms t from a dictionary T. The tf-idf measure of term t appearing in document d is defined as follows:

$$\text{tf-idf}(t,d,D) \coloneqq \text{tf}(t,d) \cdot \text{idf}(t,D) \quad \forall d \in D, \quad \forall t \in T$$

$$\text{tf}(t,d) \coloneqq \mathbf{1}_d(t) = \begin{cases} 1 & \text{if } t \in d \\ 0 & \text{otherwise} \end{cases}, \quad \forall d \in D, \quad \forall t \in T$$

$$\text{idf}(t,D) \coloneqq \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right), \quad \forall t \in T$$

The associated $|D| \times |T|$ document-term matrix $\mathcal{D}^D$ is the collection of *tf-idf* measures for all documents in the collection and for all terms in the relevant dictionary. In other words,

$$\mathcal{D}^D_{d,t} = \text{tf-idf}(t,d,D), \quad \forall d \in D, \quad \forall t \in T.$$

Cosine similarity

Definition 2 Given two vectors $X, Y \in \mathbb{R}^{|T|}$, their cosine similarity is defined as the cosine of the angle between them, which is also equal to the inner product of the same vectors normalised to unit length, as follows:

$$\cos(X,\ Y) \coloneqq \frac{X \cdot Y}{\|X\|\|Y\|} = \frac{\sum_{t=1}^{|T|} x_t y_t}{\sqrt{\sum_{t=1}^{|T|} {x_t}^2}\sqrt{\sum_{t=1}^{|T|} {y_t}^2}}$$

Where x_t and y_t denote the components of vectors X and Y, respectively, and $\|\cdot\|$ denotes the Euclidean norm.

Since row vectors of document-term matrices are non-negative valued, their cosine similarity is bounded by the unit interval, i.e. $\cos(X,Y) \in [0,1]$.

Moreover, when term frequency is measured by tf-idf, the normalisation denominator in the aforementioned equation is redundant and $\cos(X,Y) \equiv X \cdot Y$. By the same token, given document-term matrices $\mathcal{D}^{ISCO}$ and $\mathcal{D}^{LS}$, and extending the cosine similarity computation to the matrix level, it holds

$$\cos(\mathcal{D}^{ISCO}, \mathcal{D}^{LS}) \equiv \mathcal{D}^{ISCO} \cdot (\mathcal{D}^{LS})'\ .$$