




Article

Data Science for Industry 4.0 and Sustainability: A Survey and Analysis Based on Open Data

Hélio Castro ^{1,2,*} , Filipe Costa ¹, Tânia Ferreira ³ , Paulo Ávila ^{1,2} , Manuela Cruz-Cunha ⁴, Luís Ferreira ⁴ , Goran D. Putnik ⁵  and João Bastos ^{1,2}

¹ ISEP—School of Engineering, Polytechnic of Porto, Rua Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal

² INESC TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

³ Data CoLAB, Av. de Cabo Verde, 1, 4900-568 Viana do Castelo, Portugal

⁴ 2Ai—Applied Artificial Intelligence Lab, School of Technology, IPCA, 4750-810 Barcelos, Portugal

⁵ School of Engineering, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal

* Correspondence: hcc@isep.ipp.pt

Abstract: In the last few years, the industrial, scientific, and technological fields have been subject to a revolutionary process of digitalization and automation called Industry 4.0. Its implementation has been successful mainly in the economic field of sustainability, while the environmental field has been gaining more attention from researchers recently. However, the social scope of Industry 4.0 is still somewhat neglected by researchers and organizations. This research aimed to study Industry 4.0 and sustainability themes using data science, by incorporating open data and open-source tools to achieve sustainable Industry 4.0. To that end, a quantitative analysis based on open data was developed using open-source software in order to study Industry 4.0 and sustainability trends. The main results show that manufacturing is a relevant value-added activity in the worldwide economy; that, foreseeing the importance of Industry 4.0, countries in America, Asia, Europe, and Oceania are incorporating technological principles of Industry 4.0 in their cities, creating so-called smart cities; and that the industries that invest most in technology are computers and electronics, pharmaceuticals, transport equipment, and IT (information technology) services. Furthermore, the G7 countries have a prevalent positive trend for the migration of technological and social skills toward sustainability, as it relates to the social pillar, and to Industry 4.0. Finally, on the global scale, a positive correlation between data openness and happiness was found.

Keywords: Industry 4.0; data science; sustainability; social sustainability open data; open-source tools



Citation: Castro, H.; Costa, F.; Ferreira, T.; Ávila, P.; Cruz-Cunha, M.; Ferreira, L.; Putnik, G.D.; Bastos, J. Data Science for Industry 4.0 and Sustainability: A Survey and Analysis Based on Open Data. *Machines* **2023**, *11*, 452. <https://doi.org/10.3390/machines11040452>

Academic Editor: Antonio J. Marques Cardoso

Received: 27 January 2023

Revised: 10 March 2023

Accepted: 30 March 2023

Published: 3 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sustainability and Industry 4.0 are probably the most important topics in the manufacturing field, both in academic and industrial environments. In the context of sustainability within Industry 4.0, as regards the triple bottom line, profit (economic) and planet (environmental) are being studied extensively by the research community, while people (social) are not receiving the same attention. The wider perspective of the social aspects of Industry 4.0 is neglected by the academic community [1]. Industry 5.0 emerged to overcome problems with the techno-centered approach of Industry 4.0 [2], namely the absence of the human being as the key player in an industrial digital environment and the influence on sustainability as it relates to the social pillar. Nowadays, the challenges are related to the way I4.0 is implemented and managed to achieve the desired outcomes—economic, environmental, and social [3].

Within the industrial digital environment, data act as an enabler or creator of value in Industry 4.0 [4], and most industrial organizations should have integrated low-cost analytical systems to collect and measure results [5] in order to stay competitive, especially small to medium-sized enterprises (SMEs). Industry 4.0 projects are perceived by SMEs

as cost-driven initiatives [6], and lack of capacity to invest in advanced technologies and the need for specialist expertise [7] to implement them are constraints SMEs face in the adoption of Industry 4.0 and, probably as a consequence, sustainability practices.

Besides these limitations, those SMEs that are capable of incorporating technologies are unable to access sufficiently complex data due to limitations of scale. Leveraging data access, treatment, and analysis and creating intelligence are elements of a company's preparedness to improve its decision-making processes. For example, the relevance of big data and analytics to competitive intelligence has been demonstrated [8], and for this reason, data science is gaining relevance in contemporary society.

However, in small-scale organizations, such as SMEs, the cornerstone of data science is data availability, and a possible solution to overcome this problem is data access via an open approach, so-called open data. The open approach is a movement that tends to be focused on involving communities and leveraging social elements.

Due to this framework, and based on an inductive research method, the purpose of this work is to use the open approach to data science, by using open data, and the open approach in terms of analytical tools, by using open-source software, in the study of Industry 4.0 and sustainability concepts, filling a research gap in this domain. Therefore, a meta-analysis is presented in this study, based on quantitative data that is openly available and free of macro-indicators. The study will use induction analysis to generate conclusions and contributions that will lead to a better understanding of these topics relevant to in manufacturing, with a special emphasis on the technological elements of Industry 4.0 and the social aspects of sustainability.

This paper is organized as follows: Section 2 presents the literature review of the core topics in this work (data science, open data, Industry 4.0, and sustainability); Section 3 describes the research methodology; Section 4 demonstrates the analysis and the main results and discusses these; and, finally, the main conclusions are presented in Section 5.

2. Literature Review

This section provides an overview of the four main themes (data science, open data, Industry 4.0, and sustainability) that are explored across the research, and a brief bibliographic review of each of those themes.

2.1. Data Science and Open Data

In a world that constantly produces and consumes data, it is essential to understand the value that can be extracted from it. Mikalef et al. [9] consider data science and big-data domains as the next frontier for both practitioners and researchers as they embody significant potential in exploiting data to sustain competitive advantage. Data science is an interdisciplinary field that supports and guides the extraction of useful patterns from raw data by exploring advanced technologies, algorithms, and processes [10]. The actual extraction of knowledge from data is defined as data mining, and it can be applied to a broad set of business areas, such as marketing, customer relationship management, supply chain management, or product optimization [11]. Data science should be seen as the domain that originates from the merging of big-data technologies with data-management skills and behavioral disciplines [12]. Data science and big data can be combined with co-creation and data-sharing technologies to enable organizations to leverage creativity outside their organizational boundaries [13]. The development and operation of software have become increasingly dependent on data [14], and this data can be made more accessible to organizations and individuals through data-sharing and open-source technologies. Runeson [15] highlights the need for the adoption of co-creation and collaboration principles to harness innovation potential and manage costs in the age of data.

Today, data volumes are exploding, and not only is the rate of data generated per individual increasing but so also is the rate at which we share information. Lawmakers and organizations worldwide are trying to envision the future of data ownership. Information

remains largely centralized, but the trend is shifting toward a distributed and open model of data sharing [16].

2.2. Open Data for Industry 4.0

As described by Tim Hall [17], one of the key drivers for the adoption of Industry 4.0 across the globe is the ability to use the power of data to revolutionize manufacturing. Open-data platforms provide innovative services with high impact on innovation [18], and data sources based on open data allow for the evaluation of Industry 4.0 readiness by regions [19].

However, the manufacturing sector has been slow to benefit from Industry 4.0 drivers evenly across different industries, enterprise sizes, and geographies. Since most Industry 4.0 technologies require substantial investment to be successfully implemented, the economic factor is undeniably crucial if they are to be adopted. Nevertheless, while differences in the economic situations of enterprises and countries have an obvious bearing on the speed and rate of success of Industry 4.0 adoption, they cannot be considered the only factor involved. Smart factories and smart cities are another relevant study theme, as technological advancements and digitalization are changing how companies operate their business and organizations reshape communities. All those changes and advancements require big R&D investments and qualified researchers and workers. Since there are many economic challenges as well as difficulties in recruiting the most qualified workers, the adoption of those technologies might be slow and unoptimized for SMEs, which need to adapt to technological changes in order to grow and compete.

Besides, the integration of open data is still oriented to applications, websites, and platforms [20], whereas it is necessary for it to be oriented toward product development. Only recently, a case-study project based on open data from academia and companies was applied in the development of Industry 4.0 technologies in additive manufacturing [21].

2.3. Sustainability in Industry 4.0

Wee et al. [22] reiterate that there is a need for deeper research on sustainability as it relates to Industry 4.0, since it has received little attention from academics and researchers. In Kamble et al.'s [23] framework for Industry 4.0 sustainability, the three sustainable outcomes that ideally should be accomplished from Industry 4.0 technologies and process integration are economic benefit, process automation, and safety and environmental protection. The sustainability pillars in manufacturing companies have a strong relationship with Industry 4.0 [24], and competitiveness and social and environmental advantages are potentialized in manufacturing companies that adopt Industry 4.0 [25].

Other models include open innovation and collaboration as guiding principles for sustainability in industry. In this research, analyzing the progress toward accomplishing sustainability goals using the open data available is considered an overall evaluation of sustainability across the three pillars. Since these are broad goals established not only in countries but also in organizations and companies, successful progress toward accomplishing them is also progress toward accomplishing sustainability in Industry 4.0. UN members need to collaborate across all established goals, even more so because Goal 17 itself—Partnerships for the Goals—focuses on evaluating members' progress toward economic, social, and environmental collaboration between them. For that reason, it is reasonable to assume that progressing in Goal 17 is essential to accomplish successful collaboration in the remaining goals. Social indicators are neglected within manufacturing companies [26], and Industry 4.0 still needs to improve intra-organization mechanisms for achieving social sustainability [1]. Within the social pillar of sustainability, one of the main social issues relating to the digitalization and automation of industry is how employment and skills requirements will be affected. The common understanding regarding this issue is that automation eliminates the need for human workers, which will bring unemployment and social dissatisfaction. However, researchers such as Shet and Pereira [27] believe that Industry 4.0 generates new job prospects in the emerging domains of science, technology,

and engineering. Those domains usually require a higher level of skills and specialization than traditional jobs, leaving unskilled workers more vulnerable to the gradual increase in demand for qualified workers.

Recently, to evolve the concept of Industry 4.0 by stressing the social and human aspects, Industry 5.0 emerged. This approach combines the technologies of Industry 4.0, sustainability, and the human factor [28].

3. Research Methodology

The research methodology section aims to present the researchers' design, with all key design choices being detailed and justified logically, according to the research theme.

The appropriate research type for this study is the inductive and quantitative approach since the study aims to explore quantitative data relevant to the research themes and afterward form appropriate conclusions and contributions, instead of pre-establishing hypotheses or theories about those themes.

The research strategy was, from the starting point of the established research themes, to collect and aggregate data from different relevant organizations that had information relevant to the subject, namely the World Bank, Data World, Open Data Watch and Our World in Data. However, the datasets of these organizations could only be useful if their content was fully open for being downloaded, modified, and published, which is one of the main characteristics of open data. During the process of cleaning and organizing the data available within these institutions, some challenges were encountered regarding the datasets. For instance, specific information was quite hard to find, so it was necessary to select topics that by induction could lead us directly to the research themes and provide insights and conclusions. Another challenge was the different time horizon between datasets: data on the selected topics were provided by different institutions and had different objectives, and the time period for each topic was different. As a result, this study could not normalize a constant time period across all analyses, but instead used the longest period of time possible for each analysis made.

The datasets gathered contained data grouped according to regions, countries, industries, and enterprise size. For the data analysis and visualization, the most prevalent techniques across the study were frequency graphs and visualizations for inferential statistics that analyze correlations between selected variables, as correlation and regression analysis are techniques frequently used in inferential statistics [29]. In keeping with the open-approach emphasis of this study, the open-source software tools used to analyze the data were R and Python. R is a free open-source programming language that provides a statistical analytics computing environment. R provides a variety of statistical and graphical techniques that can be used by importing useful packages. These techniques can be used to handle raw data and to retrieve information in order to have a sense of how the data is distributed or whether there are patterns that are masked [30]. The R packages used were *arules* and *arulesViz* for rule association, and *RQDA* for quantitative analysis. Python is currently the fastest-growing programming language in the world, thanks to its open accessibility, ease of use, fast learning curve, and numerous high-quality packages for data science and machine learning. Together with R, Python is extremely useful for identifying correlations between variables and creating powerful visualizations such as graphs, matrixes, plots, or maps [31]. The main Python libraries used were *Matplotlib*, *Numpy*, and *Seaborn*. During this stage, the graphs produced were critically analyzed for the observations and conclusions that comprise the findings of this work.

Table 1 summarizes the research methods adopted during the present study, namely research type, research strategy, sampling strategy, data collection method, and data analysis tools.

Table 1. Research Methodology.

Research Design	Method
Research Type	Inductive and Quantitative
Research Strategy	<ol style="list-style-type: none"> 1. Establishing the research themes 2. Collecting and Aggregating Open Data 3. Cleaning and Organizing Data 4. Data Analysis and Visualization 5. Results and Discussion
Sampling Strategy	Probability Sampling within groups such as regions, countries, industries, and enterprise size
Data Collection Method	Open Datasets
Data Analysis Tools	Open-Source software tools such as Python and R

4. Data Analysis, Results and Discussion

This section presents the data analysis (point 4 of the research strategy), followed by the results from the selected datasets and the discussion thereon (point 5 of the research strategy). Each research theme identified in the previous subsection is represented by several relevant visualizations and their respective critical analyses in the context of the research.

4.1. Open Data for Industry 4.0

The first part of the results and analysis obtained from the data treatment is the representation of the open data that was available for Industry 4.0 themes. This subsection deals with Industry 4.0 themes previously referred to, such as manufacturing value added to gross domestic product (GDP), smart cities, and R&D efforts for innovation.

4.1.1. Manufacturing Value Added to GDP

Manufacturing is one of the main sectors of industry around the world and one of the main adopters of Industry 4.0 [32]. By analyzing available open data and using it alongside other relevant variables that measure development, such as a country's GDP, this research intends to give a clearer perspective on the issues examined as part of the research model.

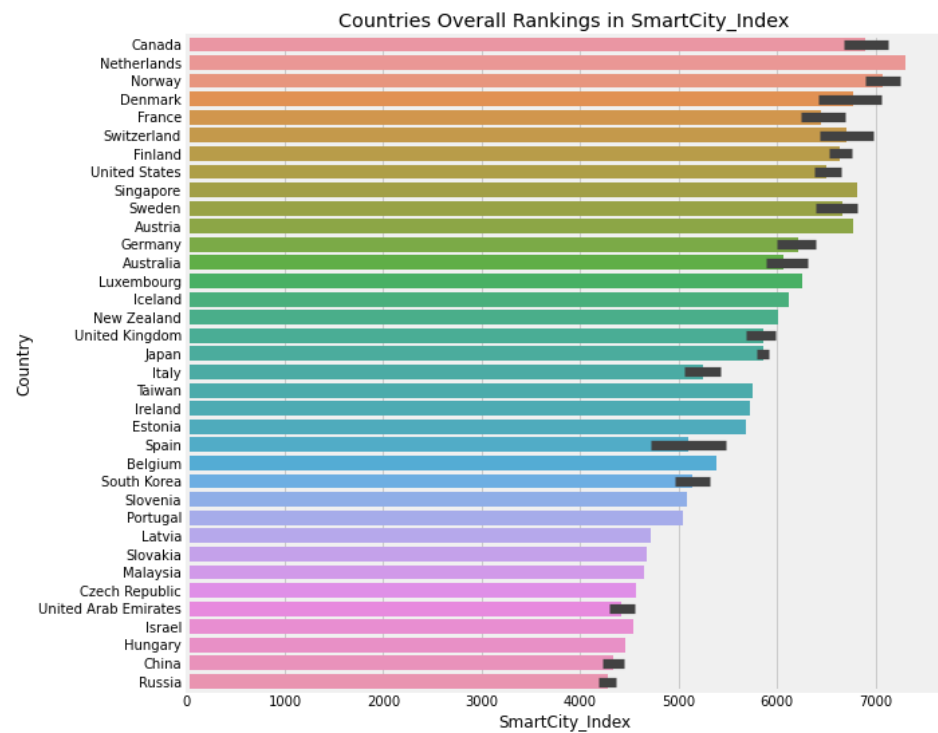
Since GDP characterizes the economic output of a country, according to the data on the global manufacturing value added to GDP in 2020 [33], Asia is the continent with the most countries in which manufacturing value added to GDP is greater than 30%, followed by Europe and America, with just one country each. In Oceania and Africa, no countries are in this situation. By country, China is one of the countries in the world with a big share of its GDP allocated to manufacturing—around 40%. Most countries appear to have manufacturing value added to GDP of between 10% and 20%. The continents with a larger share of countries that have less than 10% of their GDP value added from manufacturing are Africa and Oceania, while in Europe, America, and Asia there are few countries with less than 10% manufacturing value added to GDP.

With regards to the Industry 4.0 environment and the available open data, it can be observed that Asia is the leading continent in terms of countries with a high percentage of manufacturing value added to GDP, followed by Europe and America. China stands out as the country with the largest share of GDP derived from manufacturing.

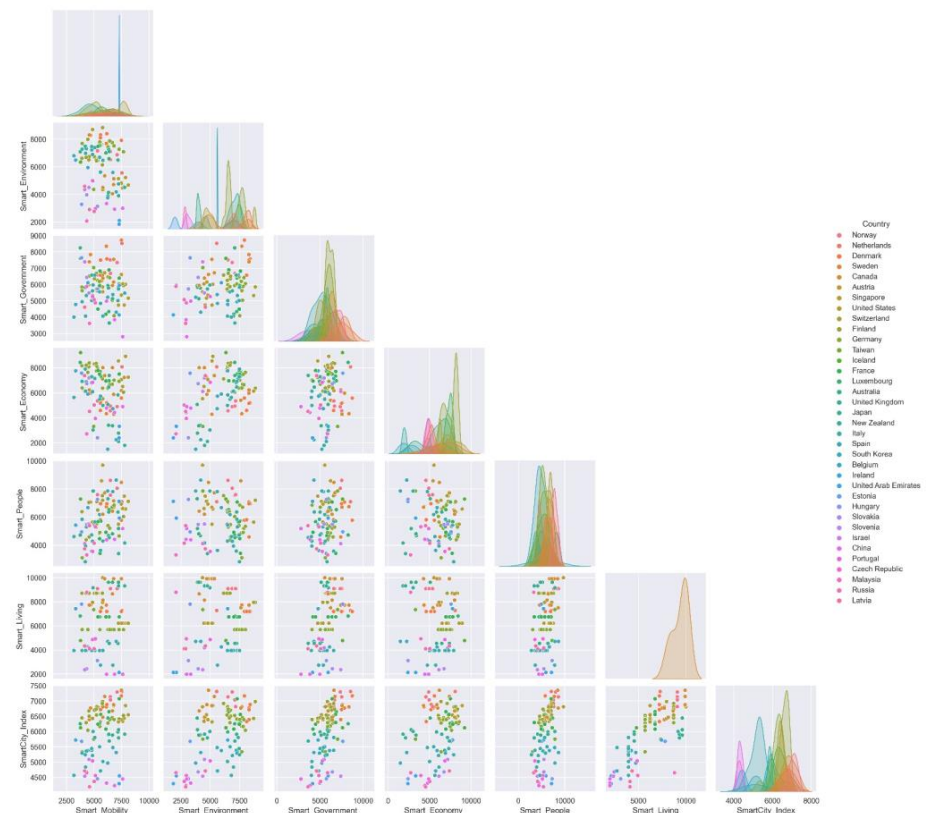
4.1.2. Smart Cities

A smart city uses information and technology to improve operational efficiency, share information, and provide a better quality of life for its citizens and workers [34]. Implementing smart technologies and processes within factories and services through Industry 4.0 adoption also tends to promote economic growth, social integrity, and environmental sustainability in industrial sectors, creating new jobs in the high-tech and creative industries [17]. The dataset evaluates cities across six smart categories: mobility, environment, government, economy, people, and living. The aggregation of those scores translates to a city's ranking along IMD's Smart City Index [35]. Below, Figure 1a ranks countries by

overall smart city scores and (b) presents a plot for correlations between all six variables and the overall score.



(a)



(b)

Figure 1. (a) Countries' Overall rankings in Smart City Index (b) Smart City Index categories and correlations with overall Smart City score.

Taking into account the first 20 countries of the overall ranking, the continent with most countries in this ranking is Europe, with 13 countries (the Netherlands, Norway, Denmark, France, Switzerland, Finland, Sweden, Austria, Germany, Luxembourg, Iceland, the United Kingdom, and Italy), followed by Asia, with three countries (Singapore, Japan, and Taiwan), and America and Oceania, with two each (Canada and the United States, and Australia and New Zealand, respectively).

The countries with the highest overall scores are Canada, the Netherlands, Norway, Denmark, and France. The lowest-scored countries are Russia, China, Hungary, Israel, and the United Arab Emirates. The fact that a country has a high number of smart cities does not necessarily mean that the country itself has a high smart score.

By analyzing each category within the overall Smart City Index, it looks as if the key factors that seem to correlate most strongly with the overall index are smart living and smart economy. Since Industry 4.0 is such a big driver for digitalization and automation in the global economy, it makes sense for developed and developing nations that seek to develop their cities technologically and sustainably to accelerate the transition to a smart economy and a smart way of living.

In terms of smart city rankings, therefore, the majority of the top 20 countries are located in Europe, followed by Asia, with Canada having the highest score.

4.1.3. R&D Efforts for Innovation

One of the main drivers of innovation, particularly in the technological and industrial fields, is the financing of research and development (R&D) by enterprises, academic researchers, and scientists [36]. However, because of the uncertainty around the level of return and the payback period, this kind of investment is not equally accessible to all countries, industries, and sizes of enterprises. Assessing which countries benefit the most from R&D investments by their enterprises and which industries allocate most expenditure to R&D from World Bank data [37] might provide a representation of the efforts to implement Industry 4.0. Figure 2 shows the expenditure in R&D and the patent share of various industries.

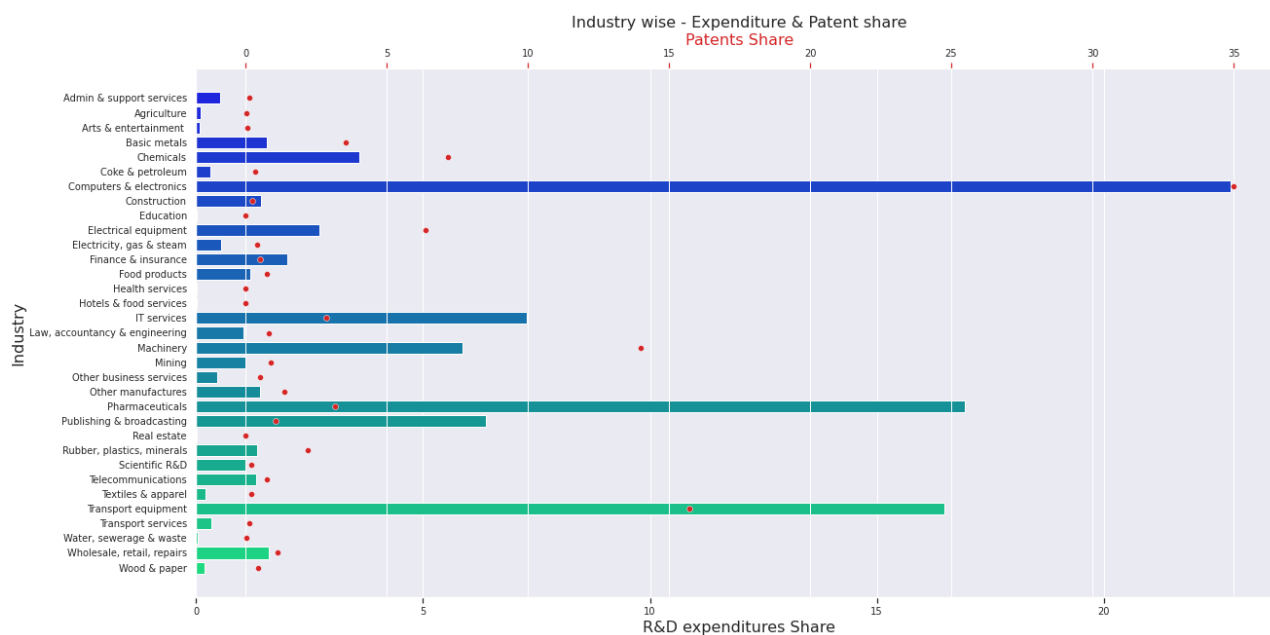


Figure 2. Industry expenditure in R&D and patent share.

It is possible to identify computers and electronics as the industry with the highest R&D expenditure share (close to 25%). As expected, this is also the industry with the highest patent share (35%). Pharmaceuticals appear as the second industry with the highest R&D expenditure, with around 17% of the share, followed by transport equipment, IT

services, and publishing and broadcasting, with 16.5%, 7.5%, and 6% share, respectively. Surprisingly, the patent share does not follow the R&D expenditure distribution so closely in these industries. The pharmaceuticals sector ranks only seventh in the patent share even though it ranks second in the R&D expenditure share. This might be caused by other factors, such as regulation and difficulties in innovating existing solutions. IT services and transport equipment also reveal a much higher R&D expenditure share compared with patent share. On the other hand, machinery is the third-ranked sector for patent share, at almost 15%, even though it occupies sixth position in R&D expenditure. Electrical equipment, chemicals, and basic metals are other sectors with a much larger patent share compared with their R&D expenditure share.

4.2. Open Data for Sustainability

The second part of the results and analysis obtained from the data treatment is the representation of the open data that was available for sustainability. This subsection deals with sustainability themes such as skills migration and the relationship between openness and happiness.

Although sustainability is based on three known pillars, as was evidenced in the literature the social pillar is arguably one of the most pressing sustainability issues, and yet it is the one that receives the least attention from researchers. One of the main issues regarding the relationship between the adoption of Industry 4.0 and the future of work is job shortages. The increasing digitalization and automation of business and service tasks often leads to worries about the permanent replacement of the human labor force by machines. However, the literature shows that that can be a misconception concerning the future of work. Shet and Pereira [27] argue that Industry 4.0 can generate job prospects by creating new employment opportunities in emerging domains, such as science, technology, engineering, and mathematics. While technological advancements and automation tend to minimize employment prospects in some sectors, they also bring about the simultaneous emergence of new businesses and services linked with economic growth and new markets, which leads to a rise in new job opportunities. Shet and Pereira [27] also warn that those jobs created by digitalization and automation require high levels of skill, knowledge, competence, and specialization that are not required by traditional jobs, leaving unskilled workers more vulnerable to the gradual increase in the demand for a qualified workforce.

4.2.1. Skills Migration

To study sustainability from the social perspective, this work considered data on skill migration across different countries and industries, to compare supply and demand trends for skilled workers. Skills migration can be defined as the trends in both supply and demand for professional skills over a number of years. As economies and labor markets change, largely because of the evolution of consumer behavior and the adoption of new technologies, so do the skills that are demanded from enterprises and public services. Better education also means better qualified workers who migrate from traditional industries to more technological and digitalized ones [38]. This dynamic is also accelerated by Industry 4.0 adoption. However, since not all countries are equal in terms of economic growth, technology adoption, and industry digitalization, naturally skills migration varies not only between industries but also in terms of geography.

This study, based on data from the World Bank [39] analyzed three categories of skills migration, namely specialized industry skills, soft skills, and disruptive-tech skills. The specialized industry skills category could indicate the overall migration of workers. The soft skills category could provide information on perceptions of important social skills, such as problem-solving, leadership, teamwork, communication, time management, persuasion, and negotiation, which are essential skills for workers, independent of location or industry sector. The disruptive-tech skills category could demonstrate trends in high-tech jobs.

For this analysis, three groups were considered in both categories: Group a), comprising China and the United States (US), as these are two of the most influential economies

in the world; Group b), the G7 (international Group of 7) countries, comprising seven advanced economies (Canada, France, Germany, Italy, Japan, the United Kingdom and the US); and Group c), BIC, comprising Brazil, China, and India due to their status as emerging economies in the BRIC bloc, with Russia's data not being available. The period of analysis was from 2015 to 2019. The purpose of the analysis of the graphs was to determine whether each skill saw an increase, a decrease, or a stabilization in demand, as plotted by their respective values.

For the specialized industry skills component, the analyses for the three groups (China and the US, the G7, and BIC) are represented in Figure 3a–c, respectively.

The plot in Figure 3a, which considers China and the US, presents China as having dynamic migration in relation to specialized industry skills, whereas the US presents a very stable graph, demonstrating that there was no migration of specialized industry skills. The overall dynamics of China showed that the dots are more located in negative than in positive values, suggesting an overall negative migration.

Excluding the US, which has been mentioned previously, for the G7 countries in Figure 3b, a stable-toward-positive migration trend for specialized industry skills was demonstrated, with very few exceptions. Among those exceptions, Germany and Japan are highlighted in the national security, army, and navy categories as having lost skilled workers in those categories over the years, revealing a possible antimilitaristic policy.

On the other hand, excluding China, which has been mentioned previously, the developing BIC economies of Brazil and India demonstrated an overall negative skills migration trend in specialized industries, which contrasts with the positive trend of the G7 countries.

For the soft skills component, the analysis for the three groups (China and the US, the G7, and BIC) are represented in Figure 4a–c, respectively.

Concerning the findings from the graph representing Group a), the first output is that the US was more stable as regards soft skills migration than China. For practically all skills during the studied period, values for the US are near zero and China showed different behavior. In China, some skills had become stable (active learning, communication, flexible approach, leadership, negotiation, problem solving, and writing), persuasion was floating from negative (2015, 2019) to positive (2016, 2018), social perception showed a negative trend throughout the period, and teamwork and time management displayed a positive trend.

Regarding the graph plotted for Group (b), besides the US, already analyzed in Group (a), there was a weak positive trend in the other G7 countries for active learning, communication, leadership, negotiation, social perception, and writing. In the flexible approach skill, most of the countries during the studied period showed fluctuation between a positive and a negative trend, but overall there was a positive trend (in 2019, most of the countries showed a positive value). The skill of persuasion also showed a positive trend in most of the countries, except Italy and the United Kingdom, where the trend was negative. Problem solving, teamwork, and time management skills showed a significant positive tendency, particularly in Germany and Japan.

For Group (c), besides China, already analyzed in Group (a), Brazil and India show a completely negative trend in all skills during the period 2015 to 2019. This tendency demonstrates a depreciation of the soft skills in Brazil and India, meaning a different approach compared with China and the G7 countries.

For the disruptive-tech skills component, the analyses for the three groups are characterized in Figure 5a–c.

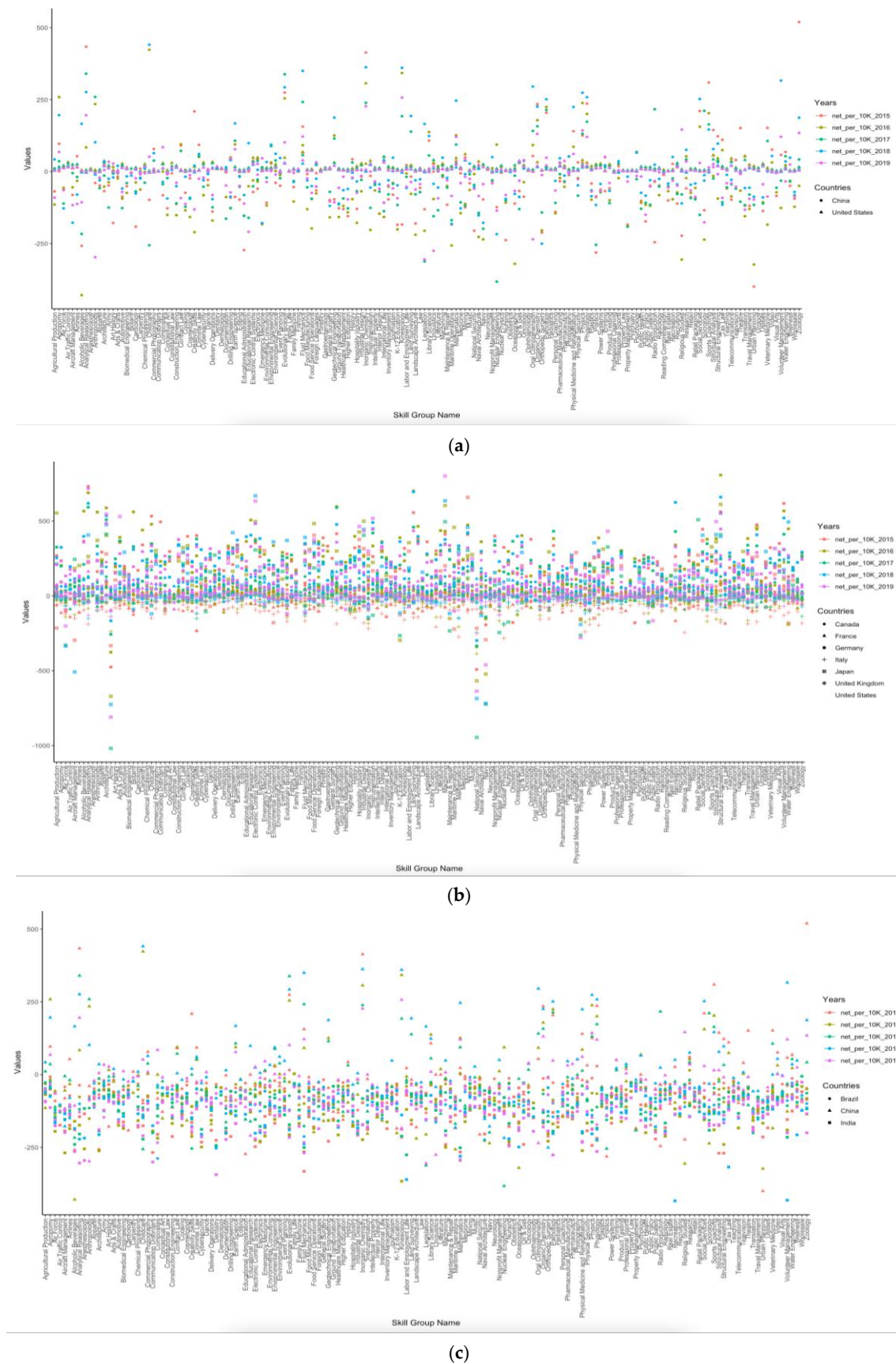


Figure 3. Specialized industry skills migration in (a) China and the US, (b) G7 countries, and (c) BIC countries.

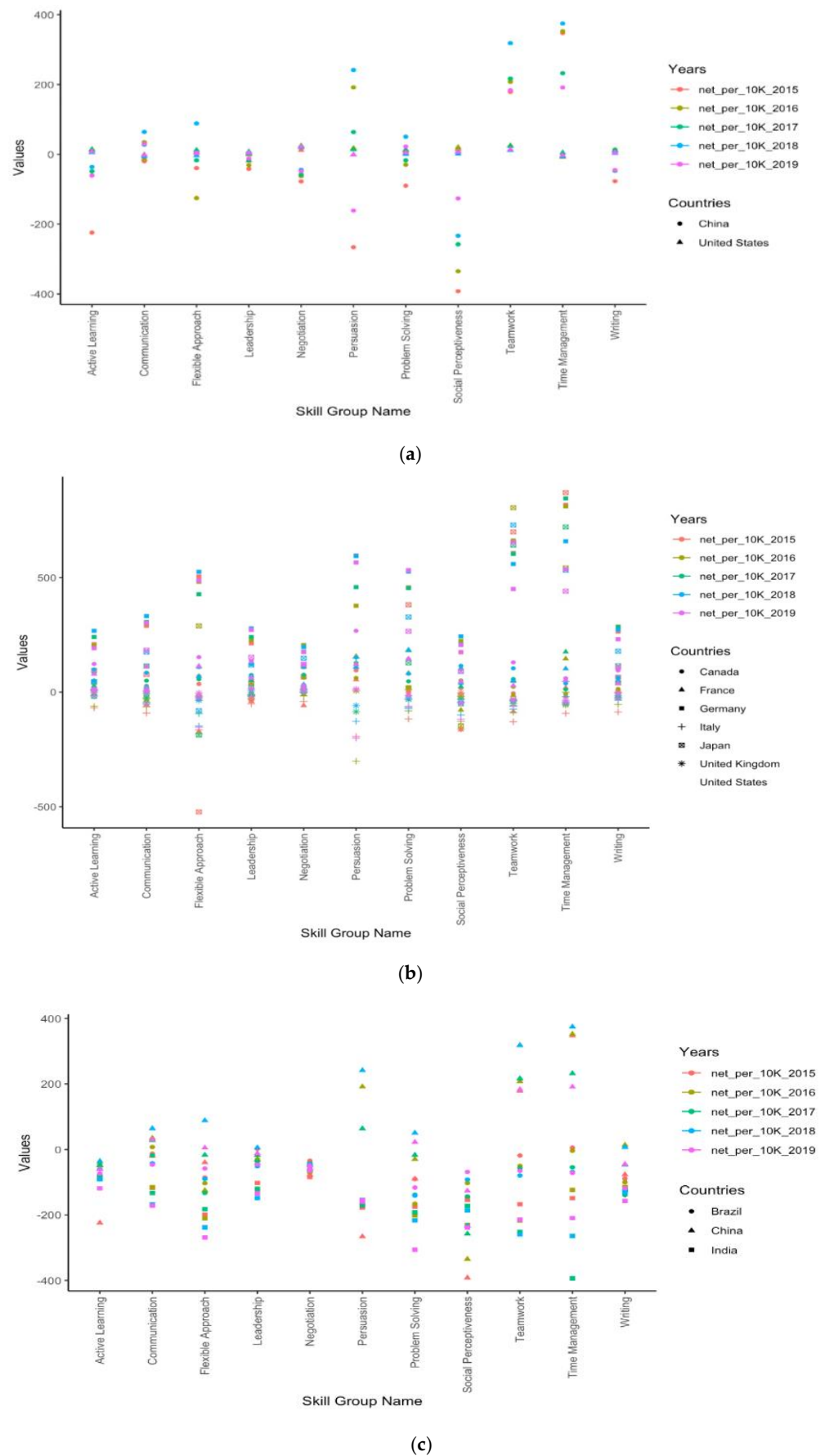


Figure 4. Soft skills migration in (a) China and the US, (b) G7 countries, and (c) BIC countries.

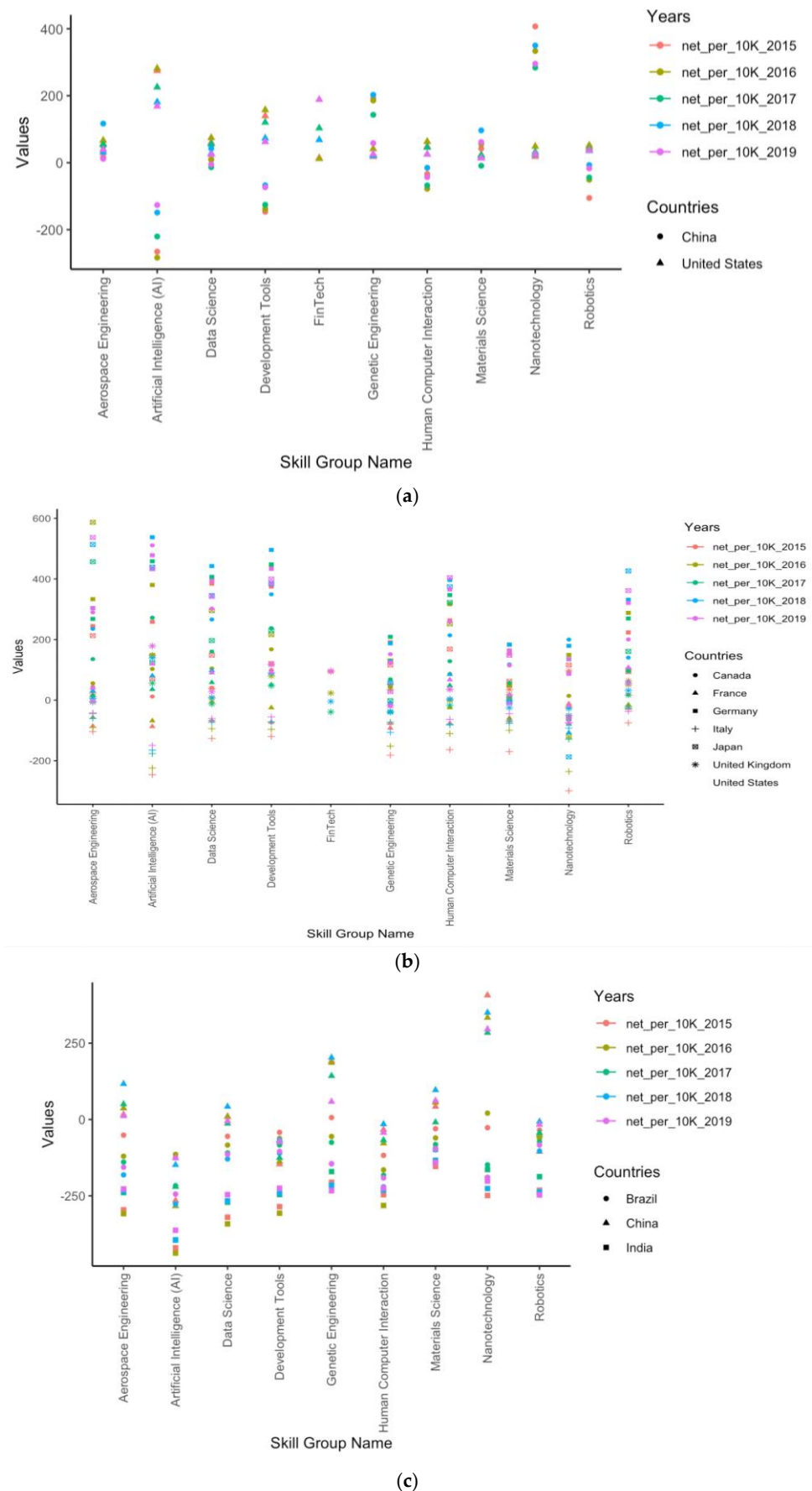


Figure 5. Disruptive-tech skills migration in (a) China and the US, (b) G7 countries, and (c) BIC countries.

The overall trend for disruptive-tech skills migration is positive for China and the US (Figure 5a). A weak positive tendency was found in aerospace engineering, data science, fintech (only in the US, Chinese data was not available), genetic engineering, and materials science. In the remaining skills (artificial intelligence (AI), development tools, human computer interaction, and robotics) a countercurrent was found between these countries, China having a negative migration trend and the US having a positive trend. For example, in China, AI showed a negative migration trend, while, in contrast, a positive migration trend was observed in the US.

Most of the G7 countries showed a very positive trend for migration of disruptive-tech skills, demonstrating the importance of this skill in high-tech jobs. The leading countries with higher values in most skills were Germany and Japan. Only Italy presented negative migration tendency in all skills. The fintech skill was presented only in the United Kingdom's data.

Similar to Italy, Brazil and India tended more to the negative position. When comparing the G7 countries with the BIC countries, the G7, except for Italy, had very high values for positive migration for these high-tech skills, while the BIC countries showed mixed results, tending more to the negative position, particularly Brazil and India, with China showing a mixed tendency, with both positive and negative skills migration.

Focusing on the disruptive-tech skills that are directly related to Industry 4.0, namely artificial intelligence, data science, development tools, human computer interaction, and robotics, the US and the G7 countries, except Italy, showed positive migration trends for these skills.

Table 2 presents a summary of the trends for the three categories of skills migration within the three defined groups.

Table 2. Summary of the skills migration trend analysis.

		Groups		
		China and the US	G7	BIC
Skills	Specialized Industry	China US	G7 except the US	All BIC
	Soft	China and US	G7 except the US	Brazil and India China and
	Disruptive Tech	China and the US	G7 except Italy	Brazil and India China
	Disruptive Tech in Industry 4.0	China US	G7 except Italy	BIC

4.2.2. Openness and Happiness

As open data platforms are providers of innovative services, this work correlated the extent of available openness in data with people's happiness. Open-data-friendly countries can establish happier sustainable societies and serve as an example of social success for the rest of the globe. To study this topic, the Open Data Scoring from the ODIN [40] dataset was used to evaluate openness across different countries, with scores from 0 to 100 and considering the values for the year 2020. From the sustainability perspective of the social pillar, the World Happiness Report from 2020 [41] and its respective dataset were used to evaluate social happiness across different countries, with a score from 0 to 10. This report is a survey published by the UN that ranks 156 countries, reviewing their state of social happiness. To better understand how different regions of the globe experience this correlation, Figure 6 outlines the clusters of the openness and social happiness scores from different regions for 2020 and the respective trendlines.

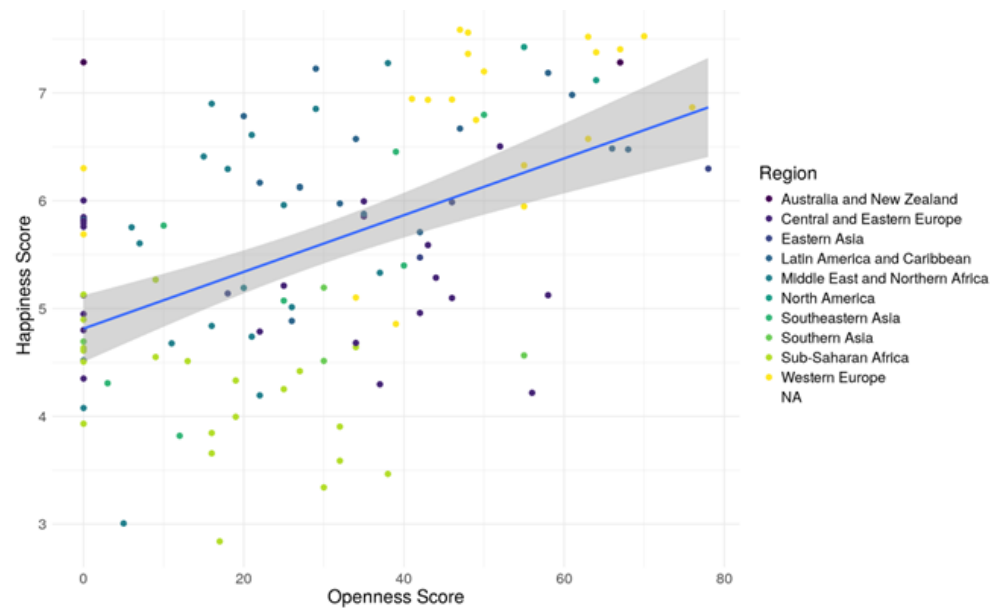


Figure 6. Openness and social happiness clustering and correlation in different regions.

Additionally, Figure 7 shows the correlation between openness and social happiness, for three established associations of countries: the G7 countries (Canada, France, Italy, Germany, Japan, the United Kingdom, and the US), the BRIC countries (Brazil, Russia, India, and China), and the Southeast Asian ASEAN countries (Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam).

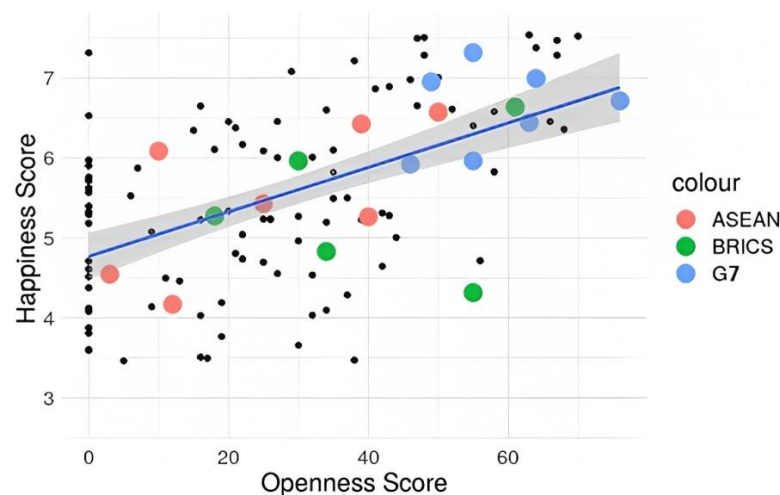


Figure 7. Clustering for the correlation between openness and happiness in the ASEAN, BRIC, and G7 countries.

The G7 countries have much higher openness in their data policies, as well as bigger indices of social happiness than the BRIC and ASEAN countries, which places the G7 clusters in the upper-right corner of the plot. The BRIC countries have a somewhat contradictory behavior, since the cluster with the second-highest openness score is also the one with the lowest happiness score, while the second lowest in openness is the second highest in happiness. Finally, ASEAN countries show clusters in the bottom-left corner, as well as clusters closer to the upper-right corner. Like the G7, this group closely matches the trend line, which means that countries in this group with high openness also tend to have high social happiness.

These visualizations allowed this study to conclude that a positive correlation exists between openness and happiness, both relevant aspect of Industry 4.0 and sustainability, especially the social pillar.

4.3. Relating the Open Data for Industry 4.0 and the Open Data for Sustainability

The third part of the analysis and results presentation involves establishing a connection between the findings from the results of the open data for Industry 4.0 and those from the open data for sustainability. To this end, Table 3 was developed to elucidate some links between the results obtained, namely the findings more relevant to the items considered in the scope of Industry 4.0 and those relevant to sustainability for the two economic blocs G7 and BIC. This table only considers the existence of simultaneous high significant values for both classes of parameters. This means that when G7 or BIC country(ies) appear in a specific square of the table, this indicates that for the country(ies) in these economic blocs significant values were verified for both entries of the table. When this apparent relationship was not verified, this is signified in the table by use of the term “No evidence”.

Table 3. Matrix relating the Open Data for Industry 4.0 and Open Data for Sustainability.

		Industry 4.0		
		Manufacturing Value Added to GDP	Smart City Index	R&D Efforts for Innovation
Sustainability	Specialized Skills	Germany, Italy, Japan, and the US.	Canada, France, Germany, Italy, Japan, and the United Kingdom.	No evidence
	Soft Skills	Germany, Italy, Japan, and the US.	Canada, France, Germany, Italy, Japan, and the United Kingdom.	No evidence
	Disruptive Tech	Germany, Italy, Japan, and the US.	Canada, France, Germany, Japan, the United Kingdom, and the US.	No evidence
	Openness and Happiness	Germany, Italy, Japan, and the US.	Canada, France, Germany, Italy, Japan, the United Kingdom, and the US.	No evidence

According to Table 3, it can be seen that there are a considerable number of countries (belonging to the G7 and BIC) that demonstrated a positive tendency for both entries (Industry 4.0 and sustainability) of the table. However, this observation should be seen as a possible positive correlation between the entries of the table for these countries. At this moment, this work does not have the scope to evaluate the correlation, but the aim is to open up ideas for future research on the scope of the relationship between Industry 4.0 and sustainability. From the tables presented, data can be observed that seem to suggest the existence of some positive relationships between specialized skills, soft skills, and disruptive-tech skills, as well as openness and happiness, on the one hand and manufacturing value added to GDP, Smart City Index, and R&D efforts for innovation on the other hand.

5. Conclusions

In this study, the theme of Industry 4.0 and social sustainability was approached through the lens of openness. Data Science using open data, together with analytical open-source tools, provided the method for generating an inductive meta-analysis, with the resultant conclusions and contributions providing a better understanding of the technological and economical elements of Industry 4.0 and the social aspect of sustainability.

A broader conclusion from this work is that society, with its computational means, open-source tools, and data science know-how, when combined with an environment that offers open data, could facilitate the treatment and analysis of data that would offer new solutions for a happier knowledge society.

Manufacturing still carries significant weight in worldwide value added to GDP; countries in Asia, particularly China with a share around 40%, as well as Europe and the

United States, depend on this sector, and for this reason the concept of Industry 4.0 must be considered quite relevant. Due to the technological premise upon which Industry 4.0 is based, a relevant showcase of countries' willingness to embrace smart environments is the smart city. When considering the 20 leading countries in the overall rankings for 2020, Europe led, with a number of its countries in the leading group, followed by Asia, America, and Oceania, while Canada was the highest scoring smart city. Investment plays an important role in the successful implementation of Industry 4.0, and according to the data on industry expenditure on R&D and patent share, the industries investing most in Industry 4.0 are computers and electronics, pharmaceuticals, transport equipment, and IT (information technology) services.

In relation to the need for more attention to be paid to the social pillar within the concept of sustainability, the migration of technical and social skills was examined in this study. Most specialized industry skills revealed a positive migration tendency in most of the G7 countries, there was a particularly stable situation in the US, and a negative migration trend in the BIC countries. In soft skills migration, China and the G7 countries presented a positive situation, but Brazil and India maintained a negative position. The soft skills with the highest positive migration were problem solving, teamwork and time management. For all disruptive-tech skills, the BIC countries showed negative migration movement, China assumed a mix tendency depending on the skill, and almost all G7 countries revealed a moderate-to-high positive migration trend in all skills. Italy was the only exception to this trend. Germany and Japan showed a high migration rate in these high-technology jobs. Linking with the Industry 4.0 concept, five of these skills were identified: artificial intelligence, data science, development tools, human computer interaction, and robotics. A positive migration tendency was found in the US and the G7 countries, except Italy. The open approach is known as a movement that tends to become oriented toward involvement of communities and the leveraging of social elements as part of the concept of social sustainability, and in this study, based on two approaches in different worldwide regions, a positive correlation between openness and happiness is concluded to exist.

Some positive relationships might exist between specialized skills, soft skills and disruptive-tech skills, as well as openness and happiness, on the one hand, and manufacturing value added to GDP, Smart City Index, and R&D efforts for innovation on the other hand.

Regarding the Industry 4.0 and sustainability concepts, the above are the main theoretical contributions considering the limitations of an inductive study for this extremely complex context and topics. Concerning the practical and managerial implications of this study, it stresses the role of open data as used in data science for Industry 4.0, sustainability and their relationship, identifying which industries, countries, economic blocs, and continents provide an attractive environment for Industry 4.0, sustainability and their relationship. Researchers, policy makers, technological investors, and manufacturing industry owners and managers could make better research and business decisions based on the findings presented throughout this study.

In future work, the investigation of smart factories by industry, country, economic bloc, and world region might prove important, should this data become available.

Author Contributions: Conceptualization, H.C. and F.C.; methodology, H.C. and F.C.; software, F.C., T.F. and L.F.; validation, H.C., P.Á., M.C.-C., L.F., G.D.P. and J.B.; formal analysis, H.C.; investigation, H.C. and F.C.; resources, F.C.; data curation, F.C., T.F. and L.F.; writing—original draft preparation, H.C. and P.Á.; writing—review and editing, P.Á., M.C.-C., L.F., G.D.P. and J.B.; visualization, H.C.; supervision, H.C.; project administration, H.C.; funding acquisition, P.Á. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by national funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020.

Data Availability Statement: The data presented in this study are openly available and presented along the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Grybauskas, A.; Stefanini, A.; Ghobakhloo, M. Social Sustainability in The Age of Digitalization: A Systematic Literature Review on The Social Implications of Industry 4.0. *Technol. Soc.* **2022**, *70*, 101997. [\[CrossRef\]](#)
- Sindhwan, R.; Afridi, S.; Kumar, A.; Banaitis, A.; Luthra, S.; Singh, P.L. Can Industry 5.0 Revolutionize the Wave of Resilience and Social Value Creation? A Multi-Criteria Framework to Analyze Enablers. *Technol. Soc.* **2022**, *68*, 101887. [\[CrossRef\]](#)
- Putnik, G.D.; Ávila, P.S. Manufacturing System and Enterprise Management for Industry 4.0: Guest Editorial. *FME Trans.* **2021**, *49*, 769–772. [\[CrossRef\]](#)
- Klingenberg, C.O.; Borges, M.A.V.; Antunes, J.A.V., Jr. Industry 4.0 as a Data-Driven Paradigm: A Systematic Literature Review on Technologies. *J. Manuf. Technol. Manag.* **2019**, *29*, 910–936. [\[CrossRef\]](#)
- Boehmke, B.; Hazen, B.; Boone, C.A.; Robinson, J.L. A Data Science and Open Source Software Approach to Analytics for Strategic Sourcing. *Int. J. Inf. Manag.* **2020**, *54*, 102167. [\[CrossRef\]](#)
- Moeuf, A.; Pellerin, R.; Lamouri, S.; Tamayo, S.; Barbaray, R. The Industrial Management of Smes in the Era of Industry 4.0. *Int. J. Prod. Res.* **2017**, *56*, 1118–1136. [\[CrossRef\]](#)
- Ingaldi, M.; Ulewicz, R. Problems with the Implementation of Industry 4.0 in Enterprises from the SME Sector. *Sustainability* **2020**, *12*, 217. [\[CrossRef\]](#)
- Ranjan, J.; Foropon, C. Big Data Analytics in Building the Competitive Intelligence of Organizations. *Int. J. Inf. Manag.* **2020**, *56*, 102231. [\[CrossRef\]](#)
- Mikalef, P.; Boura, M.; Lekakos, G.; Krogstie, J. Big Data Analytics Capabilities and Innovation: The Mediating Role of Dynamic Capabilities and Moderating Effect of the Environment. *Br. J. Manag.* **2018**, *30*, 272–298. [\[CrossRef\]](#)
- Provost, F.; Fawcett, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **2013**, *1*, 51–59. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bilal, M.; Oyedele, L.O.; Qadir, J.; Munir, K.; Ajayi, S.O.; Akinade, O.O.; Owolabi, H.A.; Alaka, H.A.; Pasha, M. Big Data in the Construction Industry: A Review of Present Status, Opportunities, and Future Trends. *Adv. Eng. Inform.* **2016**, *30*, 500–521. [\[CrossRef\]](#)
- Saritha, B.; Bonagiri, R.; Deepika, R. Open Source Technologies in Data Science and Big Data Analytics. *Mater. Today Proc.* **2021**, withdrawn. [\[CrossRef\]](#)
- Runeson, P.; Olsson, T.; Linåker, J. Open Data Ecosystems—An Empirical Investigation into an Emerging Industry Collaboration Concept. *J. Syst. Softw.* **2021**, *182*, 111088. [\[CrossRef\]](#)
- Gandomi, A.; Haider, M. Beyond the Hype: Big Data Concepts, Methods, and Analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [\[CrossRef\]](#)
- Runeson, P. Open Collaborative Data—Using OSS Principles to Share Data in SW Engineering. In Proceedings of the IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER), Montreal, QC, Canada, 25–31 May 2019; pp. 25–28. [\[CrossRef\]](#)
- Hickin, R.; Bechtel, M.; Golem, A.; Erb, L.; Buscalno, R. Technology Futures: Projecting the Possible, Navigating What’s Next. April 2021. Available online: https://www3.weforum.org/docs/WEF_Technology_Futures_GTGS_2021.pdf (accessed on 26 December 2022).
- Hall, T. The Role of Data in Industry 4.0. 20 May 2020. Available online: <https://industrytoday.com/the-role-of-data-in-industry-4-0/> (accessed on 16 December 2022).
- Cammarano, A.; Varriale, V.; Michelino, F.; Caputo, M. Open and Crowd-Based Platforms: Impact on Organizational and Market Performance. *Sustainability* **2022**, *14*, 2223. [\[CrossRef\]](#)
- Czvetkó, T.; Honti, G.; Abonyi, J. Regional Development Potentials of Industry 4.0: Open Data Indicators of the Industry 4.0+ Model. *PLoS ONE* **2021**, *16*, e0250247. [\[CrossRef\]](#)
- Sołtysik-Piorunkiewicz, A.; Zdonek, I. How Society 5.0 and Industry 4.0 Ideas Shape the Open Data Performance Expectancy. *Sustainability* **2021**, *13*, 917. [\[CrossRef\]](#)
- Gronle, M.; Grasso, M.; Granito, E.; Schaal, F.; Colosimo, B.M. Open Data for Open Science in Industry 4.0: In-Situ Monitoring of Quality in Additive Manufacturing. *J. Qual. Technol.* **2022**, *55*, 1–13. [\[CrossRef\]](#)
- Wee, D.; Kelly, R.; Cattel, J.; Breuning, M. *Industry 4.0-How to Navigate Digitization of the Manufacturing Sector*; McKinsey & Company: New York, NY, USA, 2015; Volume 58.
- Kamble, S.S.; Gunasekaran, A.; Gawankar, S.A. Sustainable Industry 4.0 framework: A Systematic Literature Review Identifying the Current Trends and Future Perspectives. *Process Saf. Environ. Prot.* **2018**, *117*, 408–425. [\[CrossRef\]](#)
- Varela, L.; Araújo, A.; Ávila, P.; Castro, H.; Putnik, G. Evaluation of the Relation between Lean Manufacturing, Industry 4.0, and Sustainability. *Sustainability* **2019**, *11*, 1439. [\[CrossRef\]](#)
- Kumar, V.; Vrat, P.; Shankar, R. Factors Influencing the Implementation of Industry 4.0 for Sustainability in Manufacturing. *Glob. J. Flex. Syst. Manag.* **2022**, *23*, 453–478. [\[CrossRef\]](#)

26. Contini, G.; Peruzzini, M. Sustainability and Industry 4.0: Definition of a Set of Key Performance Indicators for Manufacturing Companies. *Sustainability* **2022**, *14*, 11004. [[CrossRef](#)]
27. Shet, S.V.; Pereira, V. Proposed Managerial Competencies for Industry 4.0—Implications for Social Sustainability. *Technol. Forecast. Soc. Change* **2021**, *173*, 121080. [[CrossRef](#)]
28. Grabowska, S.; Saniuk, S.; Gajdzik, B. Industry 5.0: Improving Humanization and Sustainability of Industry 4.0. *Scientometrics* **2022**, *127*, 3117–3144. [[CrossRef](#)]
29. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *MIS Q. Manag. Inf. Syst.* **2004**, *28*, 75–105. [[CrossRef](#)]
30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019; Available online: <https://cran.r-project.org/doc/manuals/r-devel/fullrefman.pdf> (accessed on 30 May 2022).
31. Vallat, R. Pingouin: Statistics in Python. *J. Open Source Softw.* **2018**, *3*, 1026. [[CrossRef](#)]
32. Thames, L.; Schaefer, D. Industry 4.0: An Overview of Key Benefits, Technologies, and Challenges. In *Springer Series in Advanced Manufacturing*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 1–33.
33. UN. *Goal 9—Industry, Innovation and Infrastructure, Sustainable Development Goals*; UN: New York, NY, USA, 2022.
34. Angelidou, M. Smart City Policies: A spatial approach. *Cities* **2014**, *41*, S3–S11. [[CrossRef](#)]
35. IMD. Smart City Observatory. Available online: <https://www.imd.org/smart-city-observatory/home/> (accessed on 20 September 2022).
36. Mansfield, E.; Lee, J.-Y. The Modern University: Contributor to Industrial Innovation and Recipient of Industrial R&D Support. *Res. Policy* **1996**, *25*, 1047–1058. [[CrossRef](#)]
37. The World Bank. World Development Indicators. Available online: <http://data.worldbank.org/data-catalog/world-development-indicators> (accessed on 15 September 2022).
38. Kerr, S.P.; Kerr, W.; Özden, Ç.; Parsons, C. High-Skilled Migration and Agglomeration. *Annu. Rev. Econ.* **2016**, *9*, 201–234. [[CrossRef](#)]
39. The World Bank. Skills | LinkedIn Data. Available online: <https://datacatalog.worldbank.org/search/dataset/0038027/Skills> (accessed on 9 September 2022).
40. ODW. Open Data Inventory 2020/21 Annual Report. Available online: <https://opendatawatch.com/publications/open-data-inventory> (accessed on 11 May 2022).
41. Helliwell, J.F.; Layard, R.; Sachs, J.D.; Neve, J.-E.D. World happiness report 2020. Available online: <https://worldhappiness.report/ed/2020/#appendices-and-data> (accessed on 2 October 2022).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.