Title	The Science Space of Artificial Intelligence Knowledge Production
Authors(s)	Kogler, Dieter Franz, Whittle, Adam, Buarque, Bernardo S.
Publication date	2022-02-16
Publication information	Kogler, Dieter Franz, Adam Whittle, and Bernardo S. Buarque. "The Science Space of Artificial Intelligence Knowledge Production." Routledge, 2022.
Publisher	Routledge
Item record/more information	http://hdl.handle.net/10197/24459
Publisher's version (DOI)	10.4324/9780429351921-17

Regional Disparities in the Race for the Machine
Handbook of Smart Technologies

November 2020

The Science Space of Artificial Intelligence Knowledge Production: Global and Regional Patterns, 1990-2016

Dieter F. Kogler [a]

Adam Whittle [b]

Bernardo Buarque [b]

[a] Spatial Dynamics Lab & Insight Centre for Data Analytics,
University College Dublin, Dublin, Ireland.
dieter.kogler@ucd.ie
https://orcid.org/0000-0002-6744-5632

[b] Spatial Dynamics Lab, University College Dublin

The present contribution seeks to map the development of Artificial Intelligence (AI) knowledge as indicated through scientific publications. A set of principal keywords is employed in order to identify those publications that are related to AI in the Web of Science and the metadata provided by the database is then utilized to map the evolution of the field across different scales, i.e. the national, country and regional level. This analysis allows for the identification of potential AI "hotspots", while also establishing places that have been leading in the development of AI knowledge from the onset vis-à-vis those that have managed to catch up over time.

In addition, a network that illustrates international collaborative efforts in AI knowledge creation via co-authorships across nations and via the evolution of keywords' co-occurrence across three decades is illustrated. It is evident that these networks have become denser with time, and that they have changed across regions. The objective of the present analysis is to enhance our understanding about where and how scientific AI knowledge is created, which in turn should encourage and assist future research efforts looking to study AI knowledge and its consequences.

Keywords: Artificial Intelligence, Knowledge/Science Space, Web of Science, Multi-Scale Collaboration, Evolutionary Economic Geography.

Dieter F. Kogler

Dr. Dieter F. Kogler is the Academic Director of the Spatial Dynamics Lab at University College Dublin (www.ucd.ie/sdl). His research focus is on the geography of innovation and evolutionary economic geography, with particular emphasis on knowledge production and diffusion, and processes related to technological change, innovation, and economic growth. He is currently an ERC Starter Grant holder with the following project title: Technology Evolution in Regional Economies (TechEvo)

Dr. Adam Whittle

Dr. Adam Whittle is a postdoctoral research fellow at the Spatial Dynamics Lab, University College Dublin (UCD). He received his Ph.D. in economic geography from the School of Geography (UCD) and since then has been a visiting researcher at both the School of Human Geography and Spatial Planning at Utrecht University (January – May 2017) and the Agglomeration and Social Networks Research Lab in Budapest (January 2019 – present). His research interests are primarily as an evolutionary economic geographer. In particular, his research focuses on evolutionary connotations of technological change, knowledge complexity, network analysis, and regional diversification.

Bernardo Buarque

Bernardo Buarque is a PhD candidate at the Spatial Dynamics Lab, University College Dublin (UCD). His doctoral research project investigates the determinants of local knowledge creation. Using data from scientific publications and patents he seeks to estimate the competence of regions to produce certain technologies. Bernardo utilizes a selection of dynamics models to study the patterns of regional specialization, and to appraise which places have the core competence to produce new emergent technologies.

1. Introduction

How will automation affect human wellbeing? This question arguably dates back as far as the first industrial revolution, i.e. the invention of the steam engine, and beyond. It is also a question that has more recently recaptured the attention of social scientists following the advent of Industry 4.0, and in particular the emergence of Artificial Intelligence (AI). Most noticeably from an economics perspective, several recent publications have sought to evaluate how AI will transform labour productivity (Brynjolfsson *et al.,* 2018), economic growth (Aghion *et al.,* 2017), international trade (Goldfarb and Trefler, 2018), and employment (Arntz *et al.,* 2016; Agrawal *et al.,* 2019).

To put it succinctly, while AI has the potential to result in economic growth, prosperity, and positive change, it could just as easily produce job displacement and income inequality (Korinek and Stiglitz, 2017). That is, despite the excitement over its many promises, the recent debate regarding AI seems to focus on the future of work "in a world in which computer algorithms can perform many of the functions that a human can" (Furman and Seamans, 2019, p.161). Indeed, comparable to the emergence of any other disruptive general-purpose technology (GPT), the rise of AI will have a profound impact on our daily lives and well-being. As Buarque *et al.,* (2020, p.175) state, it will eventually "lead to significant shifts in employment and income distributions across and within society, particularly when the gains are concentrated in AI-producing regions or sectors".

Irrespective of whether your point of view of AI is optimistic or pessimistic, it is unquestionable that policies enacted today will shape how AI impacts society tomorrow (Agrawal et al., 2019). As interests around AI have begun to increase, countries around the world have started developing their own AI programmes with an eye on becoming a market leader in this disruptive technology (Dutton, 2018). Equally, scholars in world class institutions are also actively working to examine, propose and implement policies that can enhance the benefits offered by AI, while mitigating against any of its negative consequences. Instrumental in this arena was a conference organised by the OECD in 2017 "AI: Intelligence Machines, Smart Policies" with the sole aim of mobilising social, economic and political responses to the transformation of society brought on by the advent of AI technologies. Therefore, we are not only observing a mere surge in AI advancement, but also rising political concern about how to respond to it, and moreover how it should be managed. However, before it is possible to design fit-for-purpose policies, we must first understand, in an in-depth manner, the evolution and diffusion of AI systems as well as their many socio-economic consequences that are just now unfolding.

It is strikingly obvious that AI has become a hot topic and a frequently used buzzword. Nevertheless, despite its growing popularity, one dimension which remains unclear is how we can accurately measure the creation and diffusion of AI. In fact, whilst there seems to be a

consensus that AI will transform our daily lives, it remains to be seen how these transformations will manifest in space *i.e.* through economic growth or productivity. As a consequence, AI risks becoming a policy ahead of the theory initiative, based primarily on speculative analysis and anecdotal evidence. Moreover, to the best of our knowledge, the relevant literature continues to even lack a precise definition of AI. Therefore, it is critical for AI's successful implementation that there is an accurate depiction of its creation and development. Further still, to produce reliable inferences about how AI impacts our economy and society, we also need to develop robust data on its spatial and temporal diffusion; otherwise, our understanding of AI will remain speculative at best.

Against this backdrop, the present handbook chapter is a structured attempt to inform AI discourse and provide both a review of the relevant literature as well as a novel methodological axiom to analyse the creation and diffusion of AI technology. Essentially, our objective is to construct a relational database composed of academia publications on AI derived from Web of Science (WoS) data. Thereafter, this information is used to graph the distribution of AI knowledge in both the global and EU scientific communities. In doing so, it is possible to address the issue of when and where AI is created, as well as to identify potential trends in the evolution of this new disruptive technological domain.

As previously mentioned, very few empirical studies have managed to accurately disentangle the relationship between AI and socio-economic outcomes, primarily due to the lack of necessary data required. Furthermore, those notable exceptions that have addressed this issue, have primarily looked at the impact of automation on labour outputs by using proxy data for the local exposure of AI methods. For instance, Acemoglu and Restrepo (2017) and Graetz and Michaels (2018) use data from the International Federation of Robotics (IFR) to estimate the regional and industrial exposure to robots and thus determine their impact on the local economy. Yet, their dataset by no means captures all the dimensions of AI, it merely proxies a fast-growing technology for the presence of robots in the industry. Hence, it does not actually allow the authors to correctly infer the multiple aspects or consequences of Artificial Intelligence in an economy.

Another common approach is to measure the likelihood that certain occupations will become automated by advances in the field of AI. Frey and Osborne (2017) famously pioneered this method when they gathered data on the probability that different cognitive tasks would become "computerized" in the future. Thereafter, they combined this information with the O*NET dataset, which describes the dependency of 702 distinct occupations on each of these tasks. Using both sets of data, they were able to estimate the prospect of automation for all 702 occupations. The task-based approach to measuring the *risk of automation* has since become a popular strategy for scholars looking to evaluate the relationship between AI and the labour market (OECD 2017; Acemoglu and Restrepo, 2019).

For example, both Arnitz *et al.* (2016) and Nedelkoska and Quitini (2018) applied this approach to study the likelihood of automation for the OECD member nations.

However, despite its merits, these task-based approaches all suffer from the same empirical shortcoming in that they do not enable a thorough and detailed investigation into the presence and diffusion of AI across local economies. As a consequence, they are not suitable methods for evaluating the determinants of AI knowledge production and its many implications on society. In fact, these task-based approaches only permit a high-level view of how AI driven automation 'might' affect one specific aspect of the economy, i.e. the local labour market.

Given the above, it therefore appears that existing data are not sufficient to fully appraise the spread of AI, and as a consequence we are severely limited in our capacity to understand the drivers of this change and its impact on our society. Notwithstanding these numerous shortcomings, authors have begun exploiting recent advances in the field of text-analytics to circumvent these issues. Namely, previous research used text-analysis to classify patents and other documents into unique "technological" groups and used this information to infer the extent of innovation in a given domain *i.e.* measuring environment-related technologies (Haščič and Migotti, 2015). Following this logic, Mann and Puttmann (2018) applied a machine learning algorithm to a dataset consisting of texts from American patents to sort them into automation and non-automation innovations. Once they identified the automation patents, the authors could (geo)locate them in time and space. Further still, they could begin to expose the relationship between the volume of automation patents and local employment outputs.

Similarly, Cockburn *et al.* (2018) conducted a keyword search on a corpus of publications and American patents to distinguish which "inventions" should be classified as symbolic systems, robotics, and deep learning. Thereafter, the authors compared the diffusion of these methodologies across fields, regions, and time. Moreover, at least for the scope of this analysis, Buarque *et al.* (2020, p.176) employed a list of technological classes and keywords to identify European patents that are associated with AI methods. In doing so, their goal was to "build a comprehensive data set of AI patents, which will enable us to study AI knowledge production and how it is distributed across the different regions and technological sectors of the European economy".

Following Buarque *et al.* (2020), the present investigation employs a list of identifying keywords from WIPO (2019) to map the creation of scientific knowledge. Therefore, unlike most prior studies the focus here is not on patents, but rather on the scientific literature that concerns Artificial Intelligence. As the initial step, we created a subsample of academic documents from the Web of Science (WoS), *i.e.* our primary source of bibliographic data. WoS indexes approximately 280,000 scientific journals, as well as several conference proceedings

and books.[1] As such, it provides valuable information on academic publications, authors, institutions, and citations. Most importantly, however, the WoS also collects data on the keywords for each document - as provided by the authors. Exploring the information available in these keywords, we performed a search algorithm to identify and classify all the AI relevant documents. More precisely, we looked within the publications for keywords that describe an AI method, like "neural networks" or "genetic algorithms." We then classified and sampled a document as AI whenever it includes at least one keyword associated with the technology. Next, we used the metadata of these AI publications to graph the development of Artificial Intelligence in space and time. Thus, providing a valuable map of the creation and diffusion of Artificial Intelligence among the global scientific community.

While the present investigation should prove very useful, it does not provide empirical evidence on the determinants of AI creation, nor does it provide estimates on its potential implication for local economies. Instead, the objective is to offer a first glance at the creation and diffusion of AI methods in the scientific world and on a variety of spatial scales, i.e. global, national, and regional. In turn, we hope to inspire and support more detailed empirical investigations into this emergent and meaningful technology.

Taking advantage of the proposed data and methods, future investigations in the field might shine a light on how to foster the development of AI as well as produce essential estimates on the impact of AI on social inequality and human wellbeing. Following this approach, further analysis will potentially contribute not only to our understanding of this general-purpose technology, but also inform policymakers seeking to design "smart policies" in the age of Artificial Intelligence. Most policy briefings currently emphasize the economic opportunities brought about by AI systems, and the need to educate displaced workers for the jobs of the future (OECD, 2017). The methodological framework developed by Buarque *et al.* (2020), which we also employ here, will allow one to touch - at least marginally - on both of those issues. First, mapping the evolution of AI would allow for the recognition of the sectors/regions with related core competencies most suited for building a development pathway into this rapidly growing technology field (Hidalgo *et al.,* 2007; Kogler *et al.,* 2017; Whittle, 2020). That is, mapping the diffusion of AI across regions and sectors would enable one to recognize opportunities to "invest in and develop AI for its many benefits" (The White House, Executive Office of the President, 2016). Second, the AI "knowledge-space" (Kogler *et al.*, 2013) could identify which sectors/regions are more likely to be affected by this expanding technology. In other words, studying the diffusion of AI could help to identify those more vulnerable to job-displacement and other negative consequences.

---

[1] The present analysis is based on data retrieved from the following Web of Science bibliographic databases: "1980-2017 – annual Science Citation Index Expanded and Proceedings-Science combined".

2. Identification Strategy

The first phase of our analysis involves identifying AI documents. The database which supports our analysis is a raw Web of Science (WoS) corpus containing articles from over 46m journals, books, and conference proceedings. Following, we will use all documents in that database that have been published over the period 1990-2016, which is about 38m records. While WoS shares commonalities with other bibliometric databases, including Google Scholar, Scopus, or Microsoft Academic Knowledge Graph, two noticeable differences are pertinent for our investigation. Firstly, WoS has a proclivity to favour journal articles over other outlets. Further, WoS has a bias towards the "Hard Sciences", *i.e.* the Natural Sciences, Engineering, and Biomedical Research, at the expense of Social Science and Arts and Humanities (Mongeon and Paul-Hus, 2016). Nevertheless, for the purpose of the present investigation these issues are not that detrimental. In fact, given the substantive nature of AI and the fact that journal articles are the preferred outlet for dissemination across relevant disciplines, this bias may even serve to our advantage and increase our overall coverage.

Turning to the information contained within each document, WoS lists the titles, journal names, year of publication, authors, and their affiliations, among other data. Most importantly, WoS provides a list of keywords for each document, as determined by the authors. For these reasons, WoS can rightfully be considered a strong medium for analysing the creation, integration, and evolution of Artificial Intelligence throughout space and time.

Given its inherently fuzzy nature, there is no easy way to identify the AI documents. To tackle this problem, we adopt a commonly used technique in bibliographic studies and apply a keyword identifier to the WoS database. Namely, we search across the keywords section of each document for AI-specific terms, such as "machine learning" or "supervised learning". In turn, this approach enables us to classify all documents in our database as either AI or Non-AI.

Naturally, the choice of AI identifiers will heavily influence our results. For this analysis, we follow the example set by Buarque *et al.,* (2020) and borrow a list of AI-related keywords produced by the World Intellectual Property Organization (WIPO, 2019). We believe WIPO's identification gives a more recent and specific definition of AI than other alternatives. Seeking to identify AI-related patents in Europe, Buarque *et al.,* (2020) employed a list of 43 n-grams[2] that are indicative of modern AI technology. Thus, we adopt the same list of words to recognize AI-specific knowledge in the WoS database, which includes the terms: *artificial intelligence, data mining, and learning algorithm* (see Appendix A for a full list of terms).

---

[2] A n-gram consists of a list of "n" items from a sample text. For example, in this analysis, we say that "machine learning" is a 2-gram.

Using the methodology described above, we identified 260,351 documents as AI - out of the 38 million possible documents in our WoS database. Thus, although everyone seems to be discussing AI, very few have been able to seriously engage with the technology so far. To explore this issue further, **Figure 1** plots the twenty most frequent keywords associated with AI. The histogram is positively skewed with "*Neural Networks*" being the most frequently occurring keyword. This is immediately followed by "*Genetic Algorithm*", "Data Mining", and "*Support Vector Machine*."
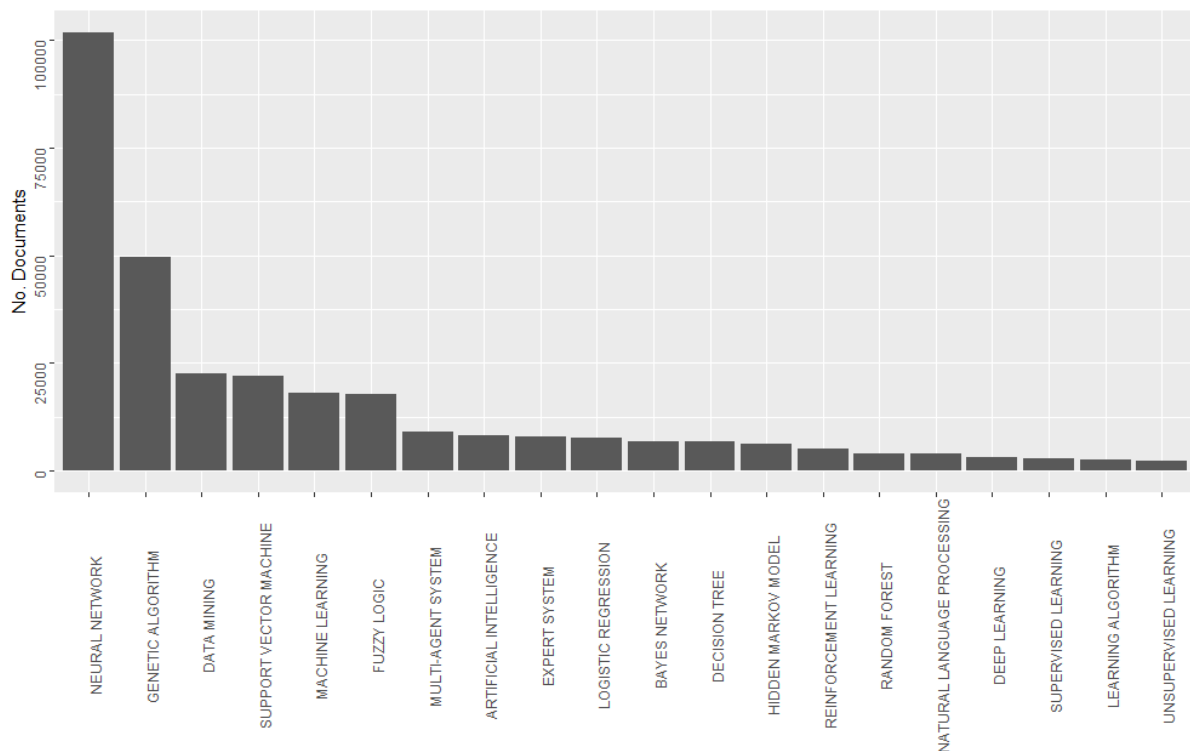
Figure 1. Frequency of AI keywords in searched publications.
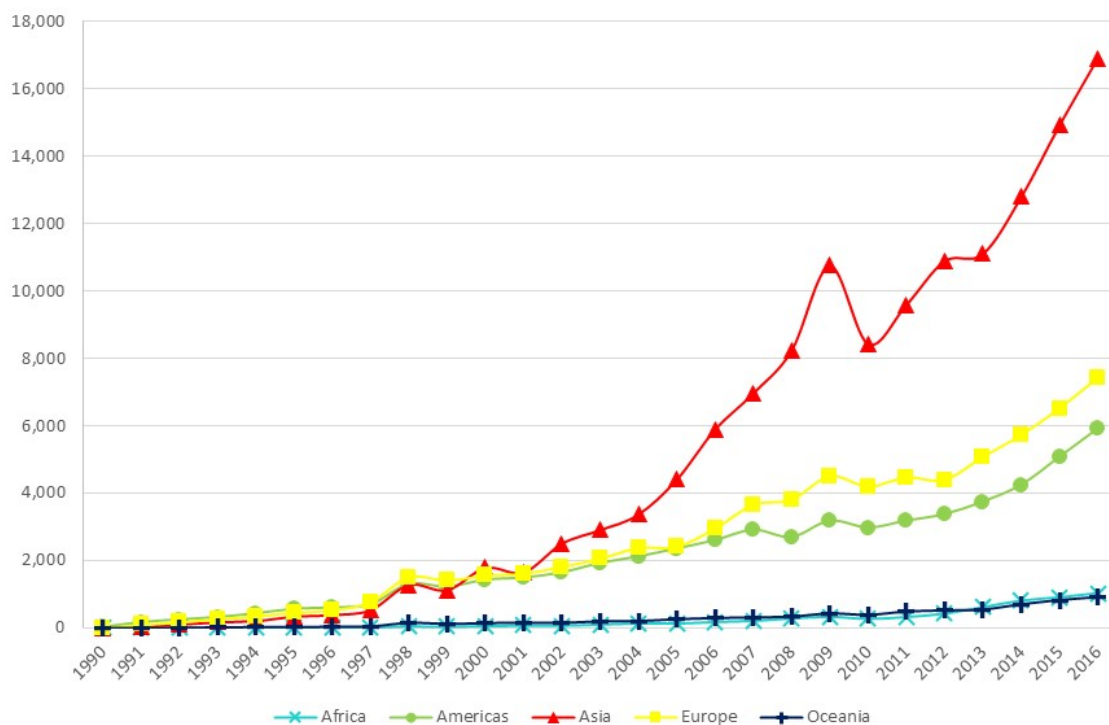Source: Authors' calculation.

It should be noted that we have carried out a thorough and careful stemming and cleaning process on our entire list of AI documents. These processes are necessary given the fact that inconsistencies are a common feature with any large-scale dataset. Moreover, one of the most common difficulties is that different elements actually represent the same thing. For example, the same keyword may be reported in a variety of ways (e.g. neural network or neural networks or "NN"). For the analysis reported here, we have applied a stemming technique to all of the keywords with the aim to minimize duplications.

Additionally, we consider a document to be AI if at least one of its keywords matches any of our 43 unique n-gram identifiers.[3] Therefore, documents which we classify as AI can also include non-AI keywords. As such, our final database, which contains only the "AI" documents also includes non-AI keywords. Indeed, we expect every document in our database to have a mixture of AI-specific keywords, as well as terms that are not related to AI. That being said, this is valuable information since it allows us to study the integration of AI knowledge across space, time, and subject matter.

3. Data and Methods

3.1. Global Focus

Once we identified the AI documents within our WoS sample, we plotted the total volume of AI publications in space and time to obtain a better picture of the evolution of Artificial Intelligence. Along these lines, **Figure 2** provides a first glance at the growth in the production of AI documents by continent over the past three decades.

Figure 2. Number of AI documents by continent.
Source: Authors' calculation.

---

[3] Over the period of analysis, the mean number of keywords listed on journal articles remained constant at 4.8. Therefore, for an article to be identified as an AI document, one of its keywords would need to match our identifier, but the remaining three or four keywords does not. This approach enables us to create a robust picture of how AI is beginning to integrate with other research fields.

Initially, the number of AI publications grew very slowly, which is not surprising due to the novelty of that field paired with uncertainties of how such a radically new technology could be applied in the market place. This is evident once we considered that between 1990-1999 the global number of AI publications was equal to 21,531 - which represents only about 8% of all documents in our sample and 22% of documents produced in the following decade. Nevertheless, even at this primitive stage, it is still possible to identify the Western economies of Europe and the Americas as key regional players, whereby they account for 5,632 (26%) and 5,555 (25%) respectfully. Finally, although not as prominent in the early years of the first period, it is also possible to identify a nascent cluster in Asia (19%).

In the second period (2000–2009) the first 'real' surge in AI publications is evident. During this time, a total of 95,813 papers were published, signifying an almost five-fold increase over the first period. From a path-dependence perspective, Europe and America continue to dominate the initial years of the decade, accounting for 28% and 23% respectfully However, this position is transformed following the emergence, and thereafter the dominance, of Asia. Throughout the decade, the Asian economy accounts for almost half (50%) of the global output, and from 2002 onwards reports a doubling of its publication output relative to its western counterparts.

Equally important, during this period, AI begins to formalize as a discipline. Frequency analysis based on the journals' keywords (see Section 3.2) indicates that topics such as "Neural Networks", "Machine Learning", "Genetic Algorithms", "Pattern Recognition", "Fuzzy Logic" and "Data Mining" began to emerge on documents during this time. More formally, these advances served to establish the foundation on which modern AI is based and thus can be regarded as setting up the pre-conditions for the final stage.

The third and final stage refers to the years 2010–2016, and it is during this time period that the bulk of the data lies. During this stage, we observe a further increase in the divergence between Eastern and Western economies. Focusing on the global crisis of 2008, we see that Asia was initially impacted by this crisis, but quickly recovered. In contrast, whereas Europe and the Americas appeared to be less affected in the years preceding and during the crisis, they have struggled to increase their total number of publications since then.

3.2. Country-Level Analysis

Expanding the initial analysis beyond the continental level, it is possible to use the authors' affiliations on each WoS document to map the spatial distribution of Artificial Intelligence across countries. Doing so enables us to view continents in terms of their countries rather than as a collective. To illustrate, Figure 3 (Panels a, b, and c) displays the spatial distribution of AI-specific documents across countries for the three time periods
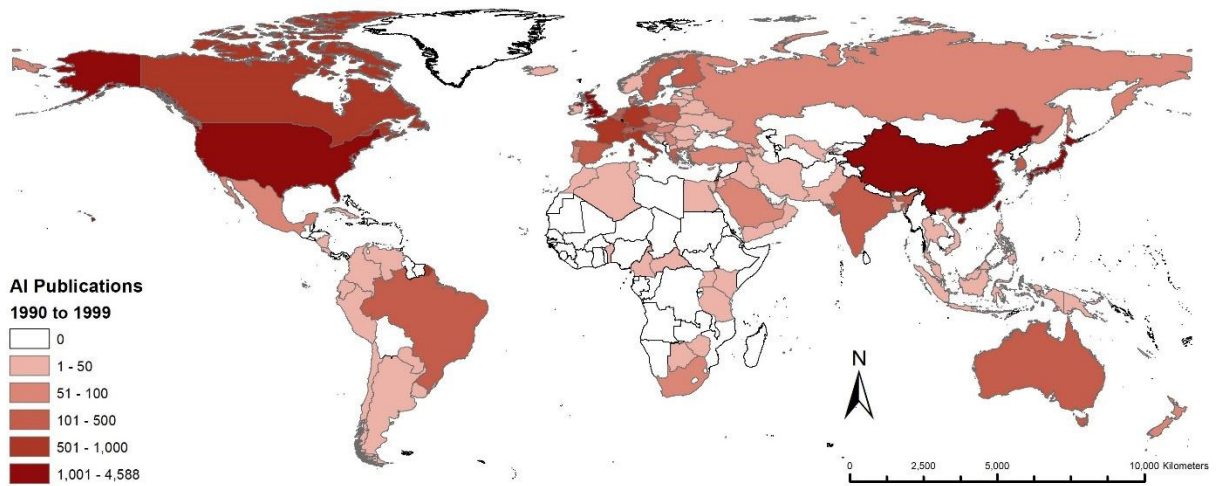
discussed previously. Using this approach, we observe that Europe's dominance in AI is partially the result of five countries (United Kingdom, France, Spain, Germany, and Italy), which make up nearly 62% of the continent's total output. A similar pattern emerges in Asia, where China (47%) clearly dominates while other countries like Taiwan (7%), India (12%), Iran (7%), and Japan (7%) also make a noticeable contribution. The same pattern does not emerge for the Americas where the United States is continually the primary producer of AI with 70% of the continent's contribution.[4]

Comparing between periods, we can further observe the path dependency process mentioned in the previous section. Namely, those countries with a historical advantage in AI production, i.e. those leading the development of AI in our first period, continue to dominate in terms of the overall share of documents. Our results, thus, seem to corroborate past work in Evolutionary Economic Geography (EEG), which highlights that regional innovation is driven by a path-dependent process (Martin and Sunley, 2006, Kogler, 2016). Furthermore, it is in line with the findings of Buarque et al., (2020) who showed the same tendency amid the AI-specific patents of Europe, i.e. those regions that excel in computing technology in the early stages of the analysis are also those with the largest share of AI patents in the end period.
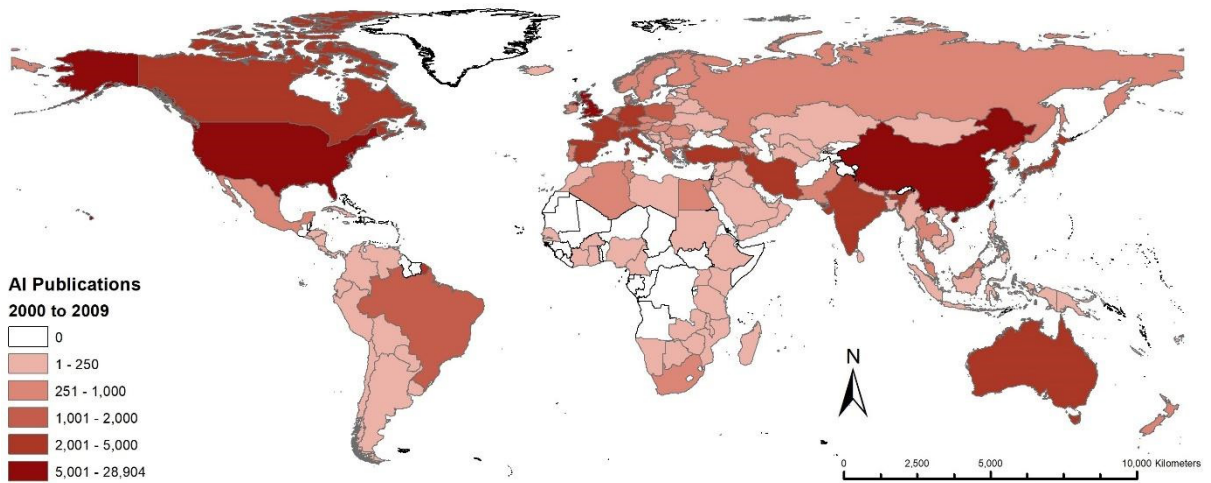
On the other hand, our maps show the emergence of Asian economies, China in particular, as key players in the scientific development of AI technologies. One might explain the rapid rise of China as a by-product of global geopolitics. After all, as China marches to become the largest economy in the world, it is only natural that, it will control the technologies of the future, e.g. AI. Nevertheless, it is more likely that the growth in China is a result of the country's ambitious policies regarding the development of AI (Dutton, 2018). Since 2017, when the "New Generation of Artificial Intelligence Development Plan" was implemented, China has even invested more in the production of AI for the future than it has in the past. Thus, China's continuous dominance in this technology domain will most likely prevail for quite some time.
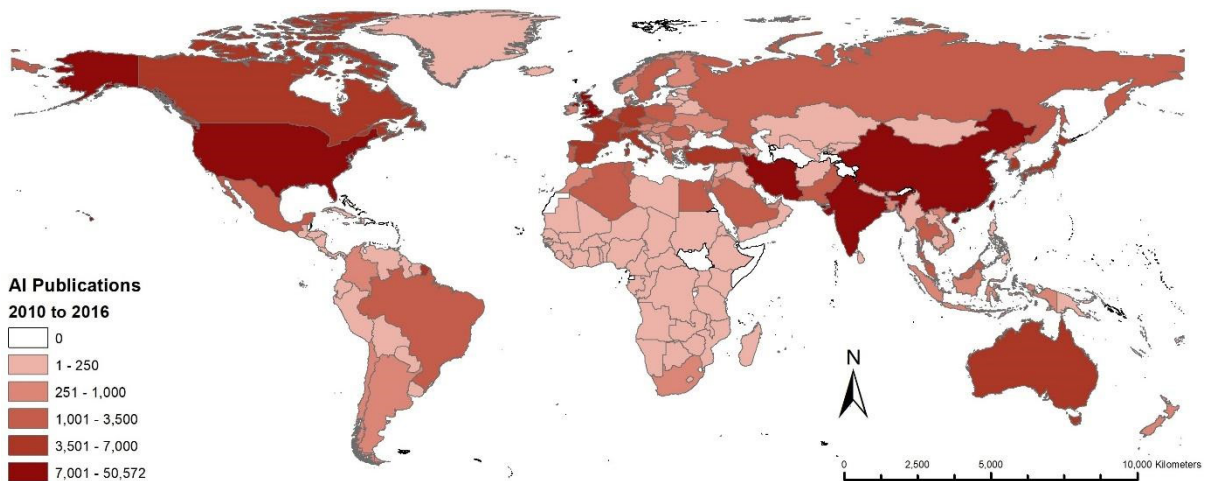
---

[4] As with any country level analysis the problem with the United States is that its sheer size distorts the innovative potential of the individual US States. Given the space constraints here, we do not go into detail as to which States in the United States are responsible for its dominance in AI, but future work should engage with these questions.

AI Publications 1990 to 1999
- 0
- 1 - 50
- 51 - 100
- 101 - 500
- 501 - 1,000
- 1,001 - 4,588

(a) 1990-1999

AI Publications 2000 to 2009
- 0
- 1 - 250
- 251 - 1,000
- 1,001 - 2,000
- 2,001 - 5,000
- 5,001 - 28,904

(b) 2000-2009

AI Publications 2010 to 2016
- 0
- 1 - 250
- 251 - 1,000
- 1,001 - 3,500
- 3,501 - 7,000
- 7,001 - 50,572

(c) 2010-2016

Figure 3. Distribution of AI documents by country and time period.

Source: The authors; based on Web of Science records derived from the "Science Citation Index Expanded" and the "Conference Proceedings Citation Index". GIS shapefile source is EUROSTAT.

3.3. Country Collaboration Network

Measuring the distribution in time of AI documents across countries is surely informative. Nonetheless, to obtain a better picture of the creation and diffusion of AI, we also need to account for international collaborative efforts. The seminal contribution of Wuchty et al., (2007) highlights how journal articles are increasingly found to be the result of collaborative efforts involving teams of researchers. Taking an Evolutionary Economic Geography stand, one possible reason for their findings is that there is an upper echelon limiting the extent an individual scientist, firm, region, or country can create all the knowledge they require internally; this seems to be particularly valid for the creation of general purpose or complex technologies whose production is the result of the recombination of multiple parts (Whittle, 2019). Besides, Buarque et al. (2020) have further commented that "AI is best developed when well connected to other research and development activities within the larger regional knowledge production ecosystem" (p.177). From this perspective, one might expect that collaboration between institutions, regions, and countries is at the heart of developing Artificial Intelligence knowledge. Further, to understand the creation of AI, it is paramount that we study cross-country collaboration networks in our WoS subsample.

Along these lines, **Figure 4** uses the information on the co-location of authors listed within the same WoS document to generate a global collaboration network for AI. Countries are coloured according to the continent they belong to, whereas their size indicates the number of AI publications in that country.[5] To draw the networks, we used a force-directed algorithm to ensure the position of the nodes is proportional to their graph distances. Hence, from this network, we can deduce that countries that are closer together collaborate more frequently than those further apart. Moreover, as one might expect, the major AI producers we identified in the last sections (e.g., China and the United States), also occupy the centre of the collaborative network.[6] That is, these big AI producers also are among the most influential nation-states in the international collaboration network.

Beyond this, we observe that a large proportion of international collaboration occurs between countries within the same continent. Cultural and social arguments from economic geography may further help to explain these trends. Particularly, we expect that collaborations and interactions are more likely to occur between agents (e.g., individuals, firms, countries) that share the same language, customs, and routines (Boschma, 2005). Finally, we must acknowledge those countries scattered on the periphery of the collaboration network who have yet to establish a serious footing in AI and may have only published a few articles. Many of these are either African (turquoise) or those smaller Asian (red) economies.

---

[5] Americas = Green, Asia = Red, Europe = Yellow, Oceania = Navy and Africa = Turquoise.

[6] One potential reason why Europe might appear as less innovative to America (which is in contrast to Figures 1 and 2) is because Europe is a collection of many individual countries whereas the Americas are largely dominated by the United States.

From an evolutionary perspective, it will be interesting to see how these economies develop over the coming years and whether they will enter the global AI collaboration network.
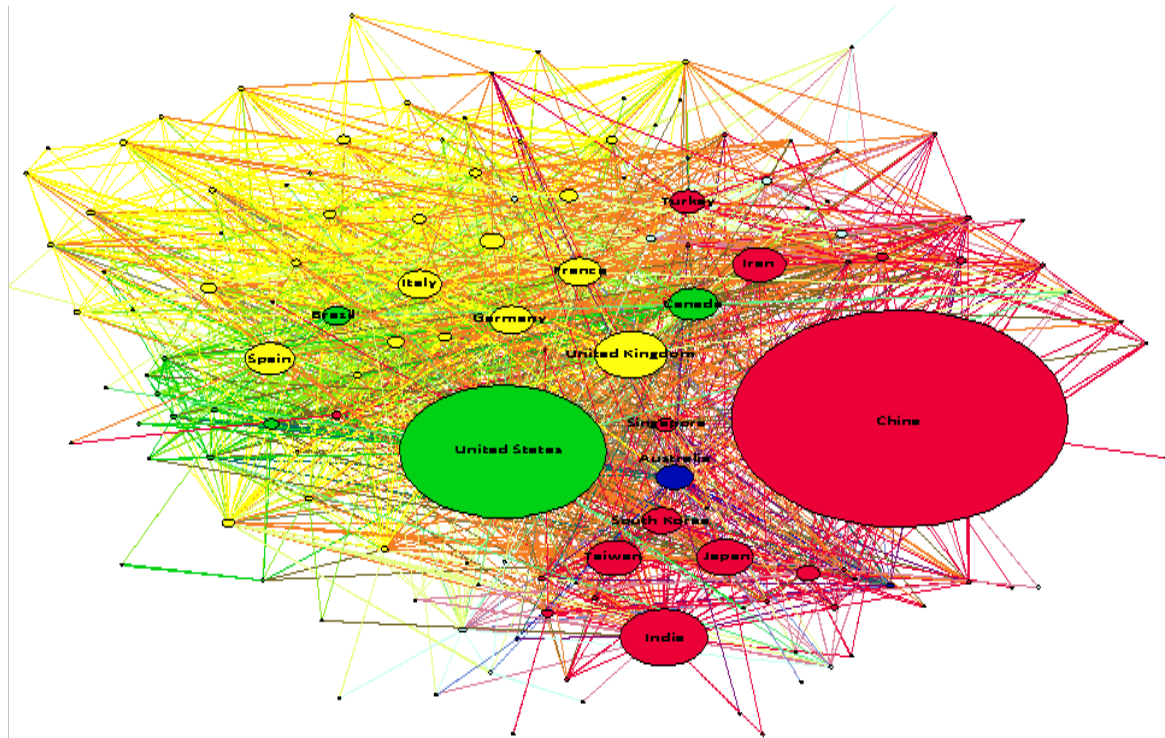

Figure 4. Global AI science/knowledge collaborator network.
Source: Authors' calculation/illustration.

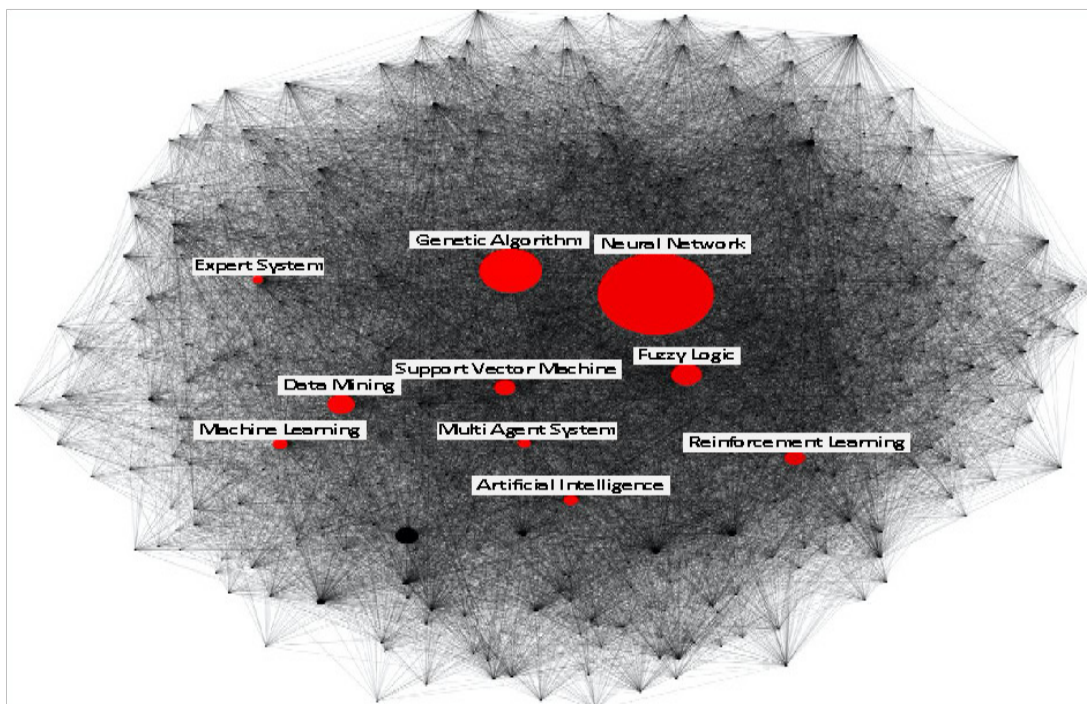## 3.4. Evolution of Keywords

As a methodological axiom, co-occurrence analysis is a valuable research strategy. It has found residence in a wide variety of fields, including economic geography (Kogler *et al.,* 2013), regional development (Hidalgo *et al.,* 2007), and scientometrics (Leydesdorff, 2007). Likewise, with the advent of big-data, co-occurrence analysis based on the frequency of (key)words that occur in the same publication has also been identified as a burgeoning research field. Indeed, while earlier research sought to measure similarity across authors and fields using the co-citations networks (McCain, 1990), more recent analysis has drawn on advances in text-mining and text-analytics to map knowledge structures using the co-occurrence of words. These developments are particularly helpful when tracing the evolution and intellectual structures of emerging new fields, such as the Internet of Things (Yan *et al.,* 2015) and Infometrics (Sedighi, 2016). Further, the methodology can also be used to produce bibliometric data on particular journals (Ravikumar *et al*., 2015), and to produce systematic literature reviews (Zhu *et al.,* 2018).

Following this line of inquiry, **Figure 5** illustrates the keyword co-occurrence network for the AI documents in the sample. It splits the analysis into the previously discussed periods

in order to examine the changing research frontier of Artificial Intelligence. Here, each node represents a keyword with its size being proportional to the frequency at which it occurs in journal articles. Like before, when drawing these networks, we ensured that keywords that frequently co-occur across our AI data sample are closer together than those that do not. Doing so reveals a core-periphery structure with the most focal concepts at the centre. At last, for a better visualization, the ten most common nodes are highlighted in red and have been labelled.
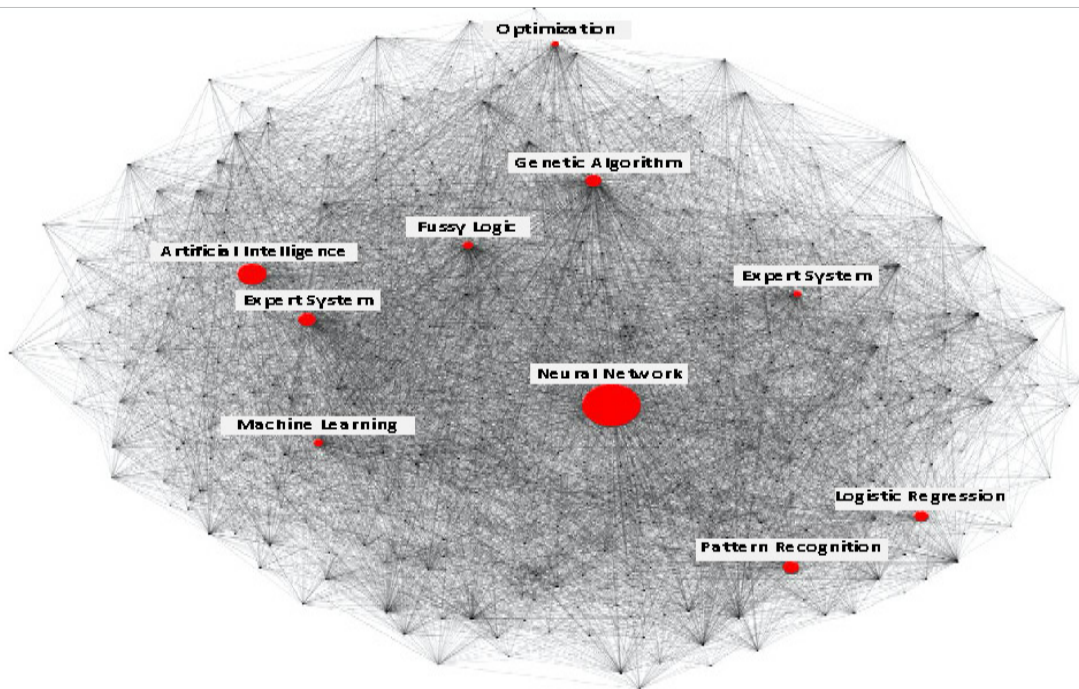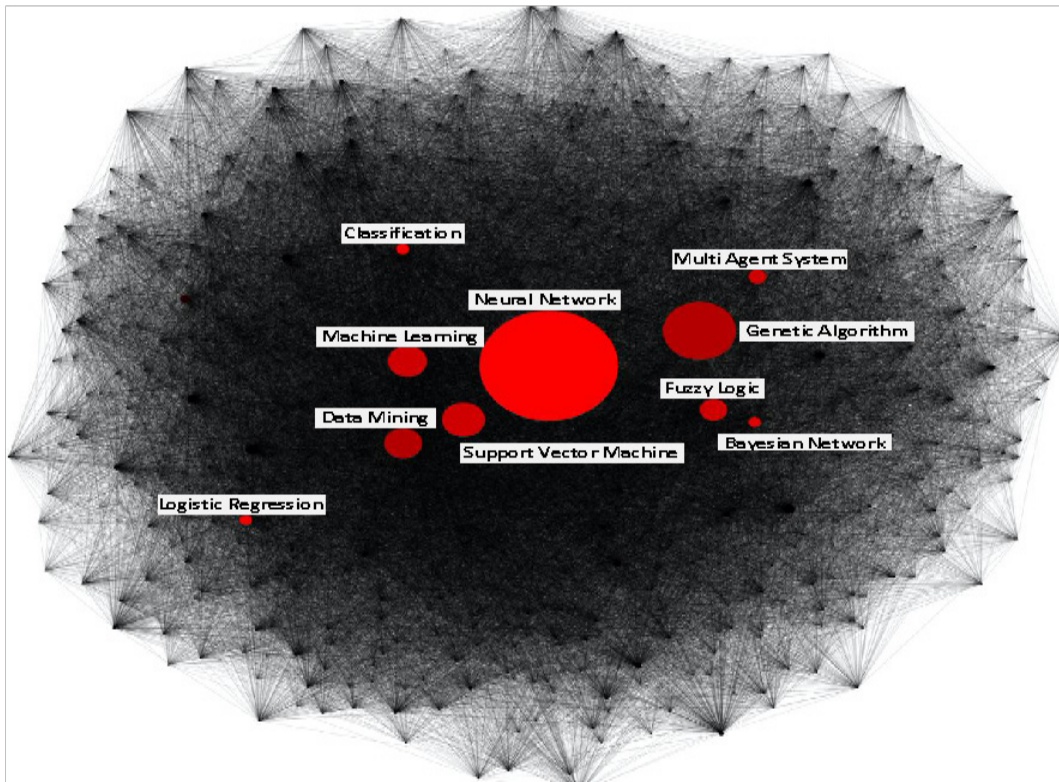

(a) 1990 – 1999

(b) 2000 – 2009

(c) 2010 – 2016

Figure 5. The changing research frontier of AI and the evolution of AI keywords.
Source: Authors' calculation/illustration.

Although Artificial Intelligence initially developed slowly (see Section 3.1), it still produced a very dense network of approximately 1,600 nodes and 32,000 edges. A plausible explanation is that, due to its infancy, many researchers were actively experimenting and trying to find applications for such a radical technology. In **Figure 5(a)**, the three largest nodes are *Neural Network, Genetic Algorithm,* and *Pattern Recognition*, which, to this day, are considered frontier concepts in the field. Furthermore, other concepts such as *Optimization, Fuzzy Logic,* and *Classification* have also begun to emerge - although in a far smaller capacity. Like before, these concepts are also inherent to AI and, in particular, its trial-and-error processes. Lastly, whilst it is often very informative to know which nodes are present, it is equally valuable to know which nodes have yet to appear. To these ends, concepts relating to *Decision Trees, Data Mining,* or *Unsupervised Learning* are still missing at this first stage.

Moving to **Figure 5(b)**, what immediately becomes clear is that there are significantly more nodes. Indeed, even after filtering we observe 400 more nodes than previously, which indicates that the network has grown. Likewise, we observe a significant shift regarding how said networks are wired. That is, comparing the two periods, not only the keyword co-occurrence network has increased in size, but it also has become denser and more

concentrated. As with the period before, *Neural Network* remains the largest node, which is unsurprising given its focal positioning in the study of artificial intelligence (see **Figure 1**). Other noticeable changes include the introduction of keywords like *Data Mining, Support Vector Machines,* and *Reinforcement Learning*. Colloquially, these tokens are commonly used to explain Artificial Intelligence, so it is not that surprising they appear here. Beyond this, there was also an obvious concentration around the concepts of F*uzzy Logic, Classification,* and *Optimization,* which again have a strong resonance with AI.

However, perhaps the most important shift between the two time periods was the concentration around the keyword *Genetic Algorithm*. Although inherently different from *Neural Network*, the two terms co-occur rather frequently on the same documents.[7] Indeed, throughout the period in consideration, a little over 2,000 documents listed both terms as keywords - which represents nearly 10% of all "Genetic Algorithm" articles in the period. In turn, this event highlights the potential for combining different AI methods to achieve faster and better algorithms, particularly for these two wide-spread methods.

Moving to the last stage, one can see that the network has become increasingly dense. This is largely because while the shift from **Figure 5(b)** to **5(c)** reports only a small increase in the number of nodes, it sees a massive increase in the overall number of edges, *i.e.* connections between nodes, which now stands at over 157,000. The surge in edges implies that AI has become increasingly intertwined with other disciplines. Otherwise stated, the theory and methods which traditionally have underscored AI are now finding residence in new areas, including mechanical engineering, medicine, finance, and automation. These changes are the driving factors behind the self-driving car, smart home technologies, and mechanical medicine. Finally, keywords such as *Neural Network* (50,237) *Genetic Algorithm* (26,242), and *Support Vector Machine* (15,250) remain vitally important. We also see they are now produced closer together in the network; that is, they appear together on the same publications, which is very different than observed in the initial time period where they were distinct and further apart.

3.5. Regional Focus

In this final section, we shift the focus of our analysis to the subnational level. That is, we will focus on the temporal evolution and spatial distribution of AI across 318 European NUTS2 regions. In this context, **Figure 6** compares the total number of AI articles in our dataset over the NUTS2 areas in the initial period 1 (1990-1999) compared to the final period

---

[7] Neural networks are a sub-form of deep-learning where the algorithms are inspired by the structure of the human brain. In short, neural networks are trained to identify patterns in data (text, audio, visuals etc.) and then predict outputs for a new set of similar data. Genetic Algorithms on the other hand reflect the processes of natural selection where the fittest mutations are selected for producing the next generation.

3 (2010-2016). Supplementary information about the top AI producing regions of Europe is also provided in **Table 1**.
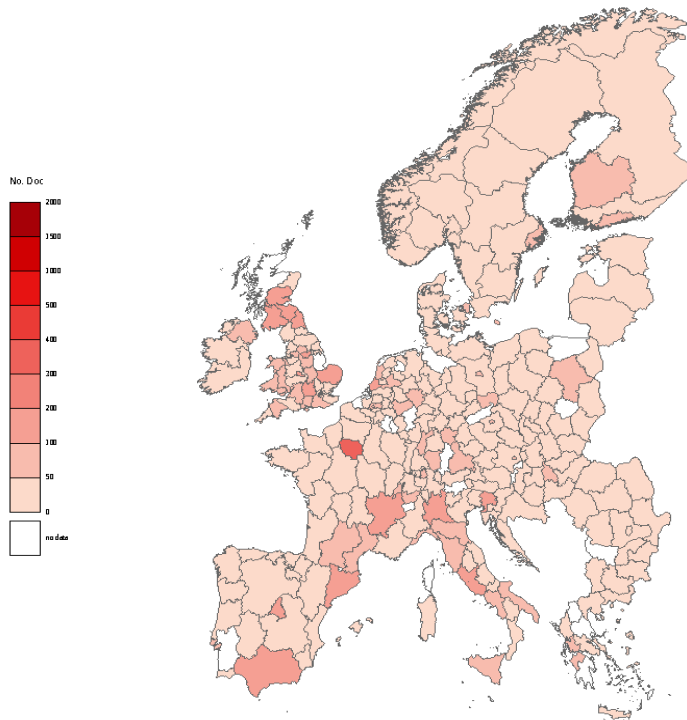
(a) 1990 – 1999
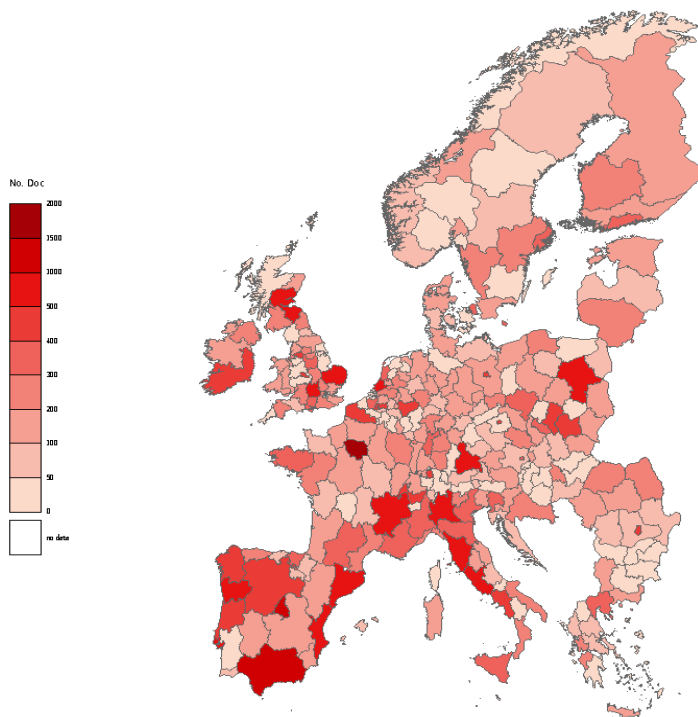
(b) 2010 – 2016

Figure 6. The spatial and temporal evolution of Artificial Intelligence science across European regional economies.
Source: Authors' calculation/illustration

Surprisingly**, Figure 6(a)** shows that most regions appear to have had an early start publishing in AI. During the first period, the median number of AI documents was 68, which in part illustrates the technology's novelty. At the same time, it is possible to identify some early AI "hotspots" - with the Île de France (FR10) appearing as a driving force. By the same token, there are clusters of activity in the South-East of England (UK13), Northern Italy (ITC4), and in Central Spain (ES30).

Moving between **Figures 6(a)** and **6(b)**, several differences are immediately apparent. Early hotspots like Île de France, Madrid, Lombardy, and London retained their status as leaders. However, new regions have emerged indicating a restructuring among the most influential AI producing regions.[8]

Table 1. Top AI Producing Regions

NUTS2 Region	Period 1	NUTS2 Region	Period 3
Île de France (FR10)	250	**Île de France** (FR10)	1,516
Lombardia (ITC4)	127	**Madrid** (ES30)	1,264
Inner London – West (UKI3)	118	**Andalucía** (ES61)	1,123
Madrid (ES30)	101	**Inner London – West** (UKI3)	1,021
South Yorkshire (UKE3)	96	**Cataluña** (ES51)	863
South Holland (NL33)	92	**Norte** (PT11)	765
West Midlands (UKG3)	91	**Lombardia** (ITC4)	714
Rhône-Alpes (FR71)	87	**Fermo** (ITI4)	684
Southern West. Scotland (UKM3)	84	**Rhône-Alpes** (FR71)	662
Attica (EL30)	83	**Mazowieckie** (PL12)	645

Source: Authors' calculation

More generally, the median number of AI documents per region rose from 68 to 344, which represents nearly a fivefold growth. Further still, this rise is potentially indicative of achieving a critical mass across both time and regions. Indeed, beyond the hotspots listed in **Table 1**, many regions are actively attempting to establish themselves as centres of excellence and as key players in the production of AI in Europe. Visually, Dublin (IE02), South Holland (NL33), Eastern Scotland (UKM2), and Oberbayern (DE21) have all moved into the foreground of scientific AI research, despite their weak starting point.

---

[8] The results concerning the spatial distribution of AI scientific knowledge production across European regions, in particular for period 3 (2010-2016), are not necessarily what we would have expected at the onset based on prior research on the distribution of AI technical knowledge production (Buarque et al., 2020). We've conducted some further investigation into the validity of these results, e.g. all records for region PT11 (Portugal, North) were reviewed manually to ensure consistency in our geocoding approach, which confirmed the findings.

Nonetheless, one must be careful when interpreting the results above. Foremost, because we are using the total number of AI documents, surely there are other decisive factors behind the patterns observed. Namely, Ile de France - our chief AI producer - is also the largest metropolitan area in the continent. It has the largest population density, the most researchers and academic publications, as well as more patents and firms. Thus, declaring the region as the AI hotspot in Europe seems premature. Indeed, if we consider how innovation scales with the urban population, it might just be that Ile de France is an average AI producer (Bettencourt et al., 2010).

Still, we believe the table and graphs above do reveal worthy patterns. Continuing from the above paragraph, the largest NUTS2 economies in Europe seem to concentrate most of the AI production on the continent. These results are perhaps unsurprising given those are the places with the most resources to invest in the development of this nascent technology; but, it also mirrors recent evidence that academic research "concentrates disproportionately in large cities" (Balland et al., 2020, p.248).

However, in contrast with the data provided by Buarque et al. (2020) which uses patents, we find that scientific publications in Artificial Intelligence are far more diffused throughout the continent. Smaller regions, and those often considered marginal in the European market, are producing a lot more scientific knowledge on AI than one might expect from the patent data. In turn, this could reveal inherent differences on how far and fast scientific or practical knowledge travels.

While the above paragraphs describe the distribution of AI documents in Europe, the present analysis seeks to go beyond this and examine how AI-specific knowledge connects to other sectors of the regional economy. That is, we wish to estimate how AI is embedded in the local knowledge-producing and innovative environment. Since many AI-promoting policies seek to specifically develop with an eye towards multiple commercial products and processes (Dutton, 2018), we need to understand how AI interacts with other domains of knowledge.

To construct a measure of AI embeddedness, we follow the approach introduced by Kogler *et al.,* (2013; 2017) and produce a regional knowledge space for each European region. Using data contained in patent documents, these authors discern a measure of technological relatedness based on the co-occurrence of "Cooperative Patent Classification" (CPC) classes. Focusing on the U.S. metropolitan areas, they discovered a link between higher levels of technological relatedness and faster rates of patenting per worker. In a subsequent analysis, Rigby (2015) found that technologies related to a region's pre-existing knowledge were more likely to enter the said region than those that were further apart from the region's expertise. Since then, geographers and regional scientists have applied these general principles to

examine the innovative ability of cities, regions, and countries using a variety of indicators (see Whittle and Kogler, 2020 for an overview of these methods).

Nonetheless, for the present examination, we are particularly interested in the recent works by Feldman *et al.,* (2015) and Buarque *et al.,* (2020). The first examined the diffusion of rDNA technology and illustrated how cognitive and geographical proximity affects the spread of this revolutionary method. Whereas Buarque *et al.,* (2020) illustrated how bibliographical analysis alongside embeddedness studies can be used to measure the creation and integration of Artificial Intelligence in Europe.

However, despite their value these above contributions have exclusively used patent data to compute the innovative performance of regions. Whilst earlier research has recognized patents as an excellent proxy for innovation, especially on a regional scale (Acs *et al.,* 2002), they also have significant limitations. Particularly, it has been argued that patents are the result of R&D and therefore reflect the innovative output potential of a region. In turn, we wish to study more succinctly the inputs of knowledge creation. To this end, by looking at academic publications, which form the bedrock on which many patents are created, we have a more accurate picture of the creation and diffusion of AI in the regional economies.

Using detailed information contained within each journal article, we can produce a *scientific space,* a graph that maps the co-occurrence of keywords across our full stock of documents. Every article has at least one keyword, but most have between four and five. These keywords are signifiers and provide a snapshot of the document's underlying knowledge. Thus, by examining the frequency by which individual keywords occur together in our WoS data sample, we can generate individual matrices of how related these AI keywords are to one another. Namely, as we did for the complete sample of AI documents (Section 3.2), we can assume that keywords that often co-occur together are more related than those that do not.

Hence, we prepared a visual graph of the AI scientific space for all 318 European regions in our sample. To be specific, we used all the documents flagged as Artificial Intelligence for each region-period pair. Next, we plotted the data from these documents in a network, where each keyword is a node, and each time that two keywords are listed on the same article an edge is created between the two nodes. Our network is then constructed by minimizing the edges' length, with more weight given to the most traversed links. As such, the region-by-region optimization places the most frequently used and tightly integrated keywords at the centre of its scientific space.

To understand how the production and integration of AI knowledge varies across Europe, we collected several network characteristic datasets for the entire region-period scientific space. That is, we measured how dense, how clustered, how centralized, and how

long it takes to transverse each AI regional network. In turn, these statistics allow us to infer how embedded Artificial Intelligence knowledge is across the European regions. Particularly, we assume that where the AI scientific space is denser and longer, the technology is more rooted in the local innovation environment, as it seems to be more connected to other valuable sectors.

Following, **Figure 7** shows the distribution of four network statistics obtained from the AI scientific spaces. As one may expect, the European regions vary largely in respect of their "AI embeddedness." Indeed, it seems like AI-specific knowledge is far more centralized in some regions than others. Overall, the region's AI scientific space is more concentrated in a few keywords, and it lacks some potential applications and alternative methods. However, on the other hand, some regions have a denser network, thus demonstrating a more connected AI knowledge space. The result is not surprising in light of research by Buarque *et al.,* (2020), who measured a very skewed distribution for the AI centrality index in patents.
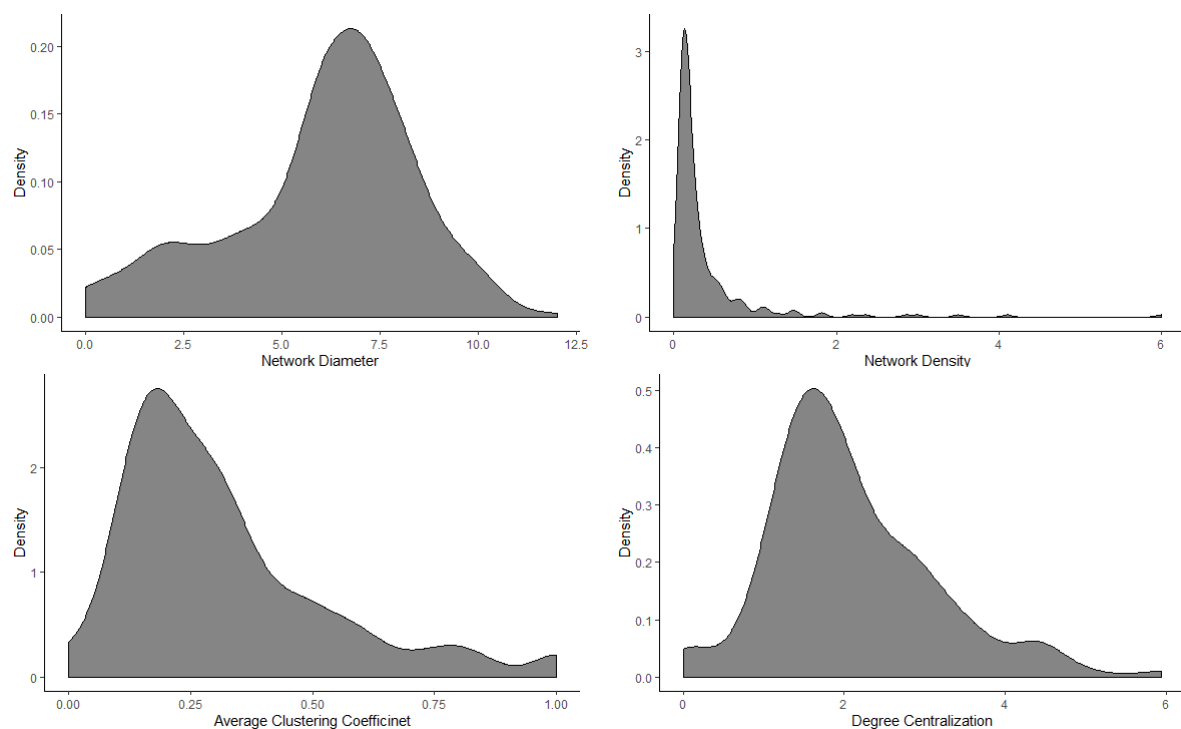

Figure 7. AI Science Space Network Statistics.
Source: Authors' calculation/illustration.

Given the significant divergence across the European regions, one might propose to examine how the level of AI embeddedness relates to the production of new AI products. To this point, Buarque *et al.,* (2020) sought to expose the correlation between the number of AI patents and their relative importance for the region's knowledge space. Accordingly, they demonstrate "there is a positive correlation between those regions where AI patents are most prevalent and those for which AI is most embedded" (p.186).

We must note, however, that our methodology diverges from Buarque *et al.,* (2020) in one vital detail. They estimated the value of AI patents in the regional knowledge-space by artificially removing said patents when building their networks and observing the impact of this exclusion on the network characteristics. That is, they measure how the absence of AI patents affects the region's knowledge space centrality and efficiency. In contrast, in the present contribution we are exclusively illustrating the scientific space derived from AI documents. We only consider the co-occurrence of keywords within articles that we identified as AI, and so we don't estimate the region's overall scientific space; *i.e.* focus is only directed at the AI-specific knowledge. As such, our methodology provides a different and original measure of how embedded or diffused AI knowledge is in the regional knowledge-producing capacity.

Further, we also highlight that examining the correlation between AI production and its embeddedness is beyond the scope of this analysis. Albeit, to obtain a better illustration of the uses of our network statistics, **Figure 8** illustrates two regional AI science spaces for the period 2010-2016. On the top, we have Dublin (IE02) and on the bottom Vienna (AT13). For comparative purposes both regions belong to high-income countries, have a similar number of universities, and enjoy a very high standard of living. As depicted above, both regions are significant producers of artificial intelligence and have roughly the same amount of journal articles, IE02 (422) and AT13 (381). However, despite their commonalities, these regions produce very different network structures with AI occupying a more central position in Dublin.
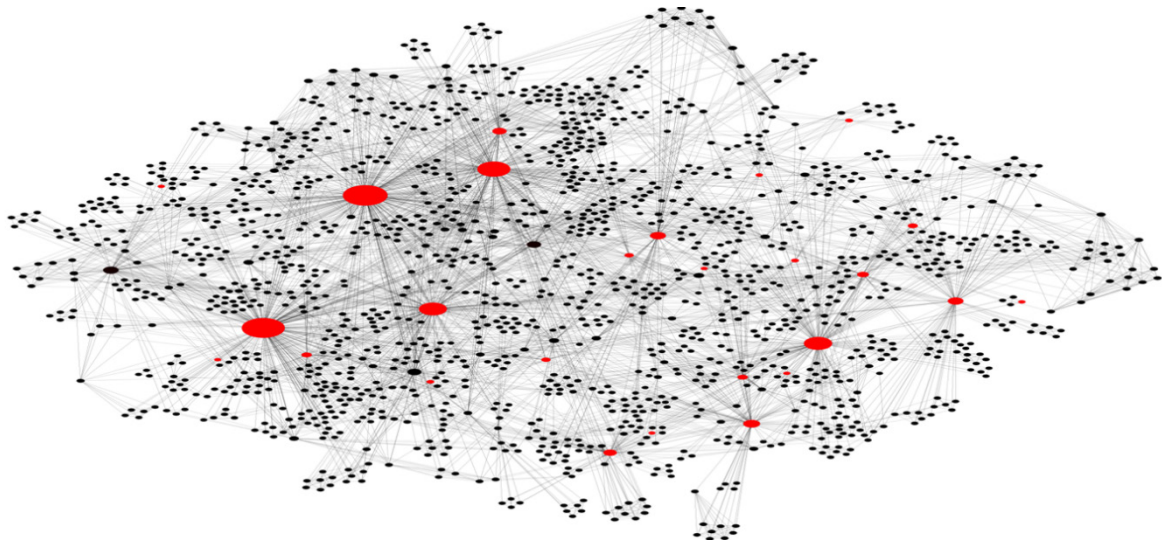
(a) Dublin (IE02)
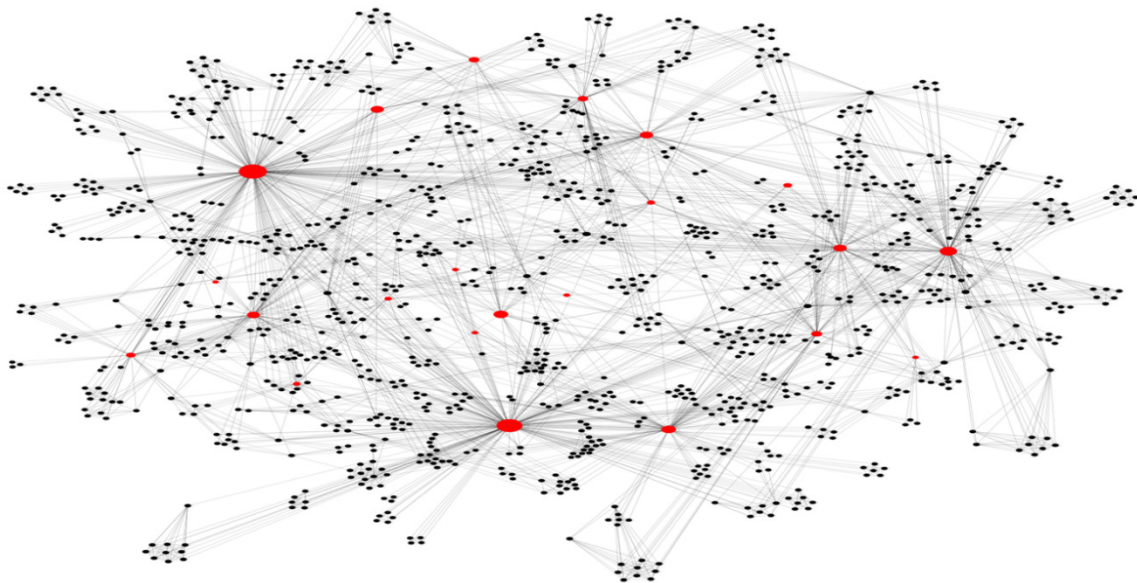
(b) Vienna (AT13)

Figure 8. Science Space of AI. Red nodes indicate AI keywords.
Source: Authors' calculation/illustration.

In terms of the sheer number of AI-specific keywords, Dublin and Vienna are once again very similar. Of the 36 keywords[9] listed by the WIPO (2019), Vienna has published in 23 of them, whereas Dublin published in 26. Although the volume of documents and keywords

---

[9] The World Intellectual Property Organization (WIPO, 2019) provides 43 n-grams, which we used to identify the AI documents. We grouped the different n-grams into 36 keywords. Namely, we grouped together terms that refer to the same or very similar method - such as, "supervised training" and "supervised learning."

in a region is indicative of its capacity to produce AI knowledge, in order to understand the region's full potential we must also account for the links between keywords and the interconnections among the different kinds of knowledge produced. Along these lines, comparing the two networks shown in **Figure 8**, we first observe that Dublin's scientific AI space has more nodes overall and is denser than Vienna's. That is, visually, we can conclude that Dublin combines more industries and sectors into its AI network; thus leading to a more diverse and applied technology when compared to Vienna.

Furthermore, in both networks under consideration, we highlight "Neural Networks" and "Genetic Algorithms" as the most traversed nodes, i.e. the most relevant keywords. For Dublin, you can see that these keywords are closely connected, indicating they frequently occur in the same publications. Moreover, these keywords are also tightly surrounded by other nodes (both AI and Non-AI), further demonstrating their re-combinatorial potential. Recall from the previous sections how AI has become increasingly intertwining with other sectors of the economy (Frey and Osborne, 2017) and technological frontiers (Buarque *et al.,* 2020); it is precisely this recombination that is driving AI policy and regional development (Clifton *et al.,* 2020; Acemoglu and Restrepo, 2020).

Conversely, whilst "Neural Network" and "Genetic Algorithm" are also the most connected nodes in Vienna's network, they are not as embedded in the region's scientific space. Thus, it seems that the region has been unable to connect distinct research frontiers in AI, which significantly hampers its ability to harness the capabilities of Artificial Intelligence. Insights from Evolutionary Economic Geography (Kogler, 2016) further substantiate this point, illustrating that though Vienna might have the necessary building blocks, it fails to connect them in a meaningful way and as a result their network remains sparsely connected.

4. Conclusion

Artificial Intelligence is currently one of, if not the most, widely debated science-technology field in business and policy circles, and the rush to develop and market AI related technology is palpable. Since its emergence in the early 1990s, governments around the world have been keen to develop strategies in order to harness and capitalise on its obvious societal and economic potential. In this context, the purpose of the present contribution is to provide insights into the spatial and temporal evolution of AI scientific knowledge production over the past three decades. Following this vision, the objective we set out with was to make a series of connected contributions to the relevant literature, both theoretical and empirical, all of which should inspire and support further work on this relevant topic. In particular, the study provides insights into three aspects that should advance this line of inquiry:

Firstly, to overcome a lack of precise definition, we augmented the methodology developed of Buarque *et al.,* (2020) to identify AI-based journal articles and indexed approximately 260,000 such documents. To establish a solid foundation, we began by conducting an exploratory analysis in order to examine the path-dependence of countries (continents), as well as to identify the jurisdictions that were the leading AI producers throughout each period in time.

Secondly, utilizing the keywords listed on journal articles we examined the changing research frontier of AI. Through a detailed analysis of their co-occurrence, it was possible to explore how AI is simultaneously becoming more concentrated and diverse. Concentrated by virtue of the fact that the core concepts, *i.e.* neural networks, genetic algorithm and machine learning, that define AI are appearing more frequently on journal articles over time. Diverse in terms of the number of non-AI keywords that are also appearing alongside them. This indicates AI's recombinatory potential whereby theories and methods that traditionally have underscored and defined core AI research are now also finding residence in new areas, including mechanical engineering, medicine, finance, and automation. As mentioned previously, it is precisely these principles that have led to the creation of the self-driving car, smart home technologies, and mechanical medicine.

Finally, we positioned these AI documents into the scientific knowledge space (Kogler et al., 2013; 2017) of to two capital EU regions and developed a methodology for describing how embedded AI is in these regions. The results reveal that although Dublin (IE02) and Vienna (AT13) are very similar in terms of their overall number of publications and AI keywords, by the end these two places produce very different scientific knowledge production networks. A preliminary finding here is that AI knowledge production is more central in Dublin's network, and as such, that Dublin might be better equipped to further harness its capabilities. On the other hand, whilst Vienna has the necessary building blocks to potentially exploit AI scientific knowledge, it has yet to connect these in a meaningful way to its broader network structure, i.e. other non-AI subjects.

In terms of next steps, an obvious direction would be to extend the methodology that was utilized in this study and to include information embedded in the relevant publications regarding authors and their institutions. The addition of this micro-dimension would permit a more thorough and detailed analysis of both the creation and diffusion of AI focusing specifically on those actors involved. In doing so, it would be possible to analyse not just which countries are collaborating, but also the institutions and individuals embedded in these countries. Such an analysis would be of critical importance in identifying those institutions that are at the forefront of AI scientific knowledge production and could be used by policy-makers and funding agencies when targeting specific investment opportunities. Similarly, by focusing on authors, it would possible to discern (at the institutional/departmental level) who is collaborating with whom? For example, if a researcher in computer science, i.e. AI research,

is engaged in a collaborate process with a colleague in medicine, the expectation then would be that scientific AI knowledge gets applied to a specific problem, and in turn provides inputs to generate a solution.   It is these inter-disciplinary collaborations that provide the opportunity to produce recombinant knowledge with the potential to push forward technological change and the research frontier in the science/technology knowledge space (Kogler et al., 2013; 2017; Kedron et al., 2020).

This point also speaks more broadly to a crucial methodological contribution in the present investigation. In particular, we further substantiate the viability of text-matching and text-analysis methodologies for identifying and analysing the creation and diffusion of science and technologies when they are not easily identifiable by traditional means. Haščič and Migotto (2015) provides an additional example that follows this approach where a text-matching algorithm is employed in order to identify "green" technologies. We hope that the present study will inspire other scholars to further explore AI scientific knowledge production as well as other non-standardised technology fields at the intersection of traditional domains.

5. Bibliography

Acemoglu, D. and Restrepo, P. (2017) "Robots and Jobs: Evidence from US Labor Markets" Technical report. NBER Working Paper No. w23287. Cambridge MA: *National Bureau of Economic Research*. Available online at: https://www.nber.org/papers/w23285.

Acemoglu, D. and Restrepo, P. (2020) "The wrong kind of AI? Artificial intelligence and the future of labour demand", *Cambridge Journal of Regions, Economy and Society*, 13 (1), pp. 25-35. https://doi.org/10.1093/cjres/rsz022.

Acemoglu, D. and Restrepo, P. (2019) 'Automation and new tasks: how technology displaces and reinstates labor', *Journal of Economic Perspectives*, 33 (2), pp. 3-30.

Acs, Z., Anselin, L. and Varga, A. (2002) "Patents and innovation counts as measures of regional production of new knowledge", *Research Policy*, 31 (7), pp. 1069-1085. https://doi.org/10.1016/S0048-7333(01)00184-6.

Aghion, P., Jones, B. and Jones, C. (2017) "Artificial Intelligence and Economic Growth" Technical report. NBER Working Paper No. w23928. Cambridge, MA: *National Bureau of Economic Research*. Available online at: https://www.nber.org/papers/w23928.

Agrawal, A., Gans, J. S. and Goldfarb, A. (2019) "Artificial intelligence: the ambiguous labor market impact of automating prediction", *Journal of Economic Perspectives*, 33 (2), pp. 31-50.

Arntz, M., Gregory, U. and Zierahn, U. (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. Available online at: https://doi.org/10.1787/5jlz9h56dvq7-en.

Boschma, R. (2005) 'Proximity and innovation: a critical assessment', *Regional studies*, 39(1), pp.61-74. https://doi.org/10.1080/0034340052000320887.

Brynjolfsson, E., Rock, D., and Syverson, C. (2018) "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics". In Agrawal, A. Gans, J. and Goldfarb, A. (eds.) *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, pp. 23-66.

Buarque, B., Davies, D., Hynes, R. and Kogler, D. F. (2020) "OK Computer: the creation and integration of AI in Europe", *Cambridge Journal of Regions, Economy and Society, 13* (1), pp. 175-192. https://doi.org/10.1093/cjres/rsz023.

Clifton, J., Glasmeier, A. and Gray, M. (2020) "When machines think for us: the consequences of work and place", *Cambridge Journal of Regions, Economy and Society, 13* (1), pp 3–23. https://doi.org/10.1093/cjres/rsaa004.

Cockburn, I. M., Henderson, R. and Stern, S. (2018) "The Impact of Artificial Intelligence on Innovation". Technical report. NBER Working Paper No. w24449. Cambridge, MA: National Bureau of Economic Research. Available online at: https://www.nber.org/papers/w24449.

Dutton, T. (2018) *An Overview of National AI Strategies*, *Medium*. Available online at: https://medium.com/politics-ai/an-overview-of-nationalai-strategies-2a70ec6edfd.

Feldman, M., Kogler, D.F. and Rigby, D. (2015) "rKnowledge: The spatial diffusion and adoption of rDNA methods", *Regional studies*, 49 (5), pp. 798-817. https://doi.org/10.1080/00343404.2014.980799.

Frey, C. and Osborne, M. (2017) "The future of employment: How susceptible are jobs to computerisation?", *Technological Forcasting and Social Change,* 114 pp. 254-280. https://doi.org/10.1016/j.techfore.2016.08.019.

Furman, J. and Seamans, R. (2019) "AI and the economy", *Innovation Policy and the Economy*, 19 (1), pp. 161–191. https://doi.org/10.1086/699936.

Goldfarb, A. and Trefler, D. (2018) "AI and international trade". In Agrawal, A. Gans, J. and Goldfarb, A. (eds.) *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, pp. 463-493.

Graetz, G. and Michaels, G. (2018) "Robots at work", *Review of Economics and Statistics*, 100 (5), pp. 753-768. https://doi.org/10.1162/rest_a_00754.

Haščič, I. and Migotto, M. (2015) "Measuring Environmental Innovation Using Patent Data". Technical report. Paris: OECD Environment Working Papers, No. 89, OECD Publishing. Available online at: https://doi.org/10.1787/19970900.

Hidalgo, C., Klinger, B., Barabási, A. and Hausmann, R. (2007) "The product space conditions the development of nations", *Science*, 317 (5837), pp. 482-487.

Leydesdorff, L. (2007) 'Betweenness centrality as an indicator of the interdisciplinarity of scientific journals' *Journal of the American Society for Information Science and Technology*, 58 (9), pp. 1303-1319. https://doi.org/10.1002/asi.20614.

Kogler, D. F. (2017) "Relatedness as driver of regional diversification: a research agenda–a commentary", *Regional Studies*, 51 (3) pp. 365–369. https://doi.org/10.1080/00343404.2016.1276282.

Kogler, D. F. (2016) (Ed.) *Evolutionary Economic Geography: Theoretical and Empirical Progress*. New York: Routledge.

Kogler, D. F. Essletzbichler, J. and Rigby, D. (2017) "The evolution of specialization in the eu15 knowledge space", *Journal of Economic Geography*, 17 (2) pp. 345–373. https://doi.org/10.1093/jeg/lbw024

Kogler, D. F., Rigby, D. and Tucker, I. (2013) "Mapping knowledge space and technological relatedness in us cities", *European Planning Studies*, 21 (9) pp. 1374–1391. https://doi.org/10.1080/09654313.2012.755832.

Korinek, A. and Stiglitz, J. (2017) "Artificial Intelligence and its Implications for Income Distribution and Unemployment". Technical report. NBER Working Paper No. w24174. Cambridge, MA: *National Bureau of Economic Research*. Available online at: https://www.nber.org/papers/w24174.

Mann, K. and Püttmann, L. (2017) "Benign Effects of Automation: New Evidence from Patent Texts". Technical report. SSRN 2959584. *Social Science Research Network*. Available online at: https://ssrn.com/abstract=2959584.

Martin, R. and Sunley, P. (2006) "Path dependence and regional economic evolution", *Journal of economic geography*, 6 (4), pp. 395-437. https://doi.org/10.1093/jeg/lbl012.

McCain, K. (1990) "Mapping authors in intellectual space: A technical overview", *Journal of the American society for information science*, 41 (6), pp. 433-443.

Mongeon, P. and Paul-Hus, A. (2016) "The journal coverage of Web of Science and Scopus: a comparative analysis", *Scientometrics*, 106 (1), pp.213-228. https://doi.org/10.1007/s11192-015-1765-5.

Nedelkoska, L. and G Quintini (2018) "Automation, Skill Use and Training", *OECD Social, Employment, and Migration Working Papers* No 202, OECD Publishing, Paris. Available online at: https://doi.org/10.1787/1815199X.

OECD (2017) "Future of work and skills". OECD Employment Outlook Technical report, OECD. Available online at: https://doi.org/10.1787/19991266.

Sedighi, M. (2016) "Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of Informetrics)", *Library Review,* 65 (1), pp. 52-64. https://doi.org/10.1108/LR-07-2015-0075.

Ravikumar, S., Agrahari, A. and Singh, S. (2015) "Mapping the intellectual structure of scientometrics: A co-word analysis of the journal Scientometrics (2005–2010)", *Scientometrics*, 102(1), pp. 929-955. https://doi.org/10.1007/s11192-014-1402-8.

Rigby, D. (2015) "Technological relatedness and knowledge space: Entry and exit of US cities from patent classes", *Regional Studies*, 49 (11), pp. 1922-1937. https://doi.org/10.1080/00343404.2013.854878.

Whittle, A. (2019) 'Local and nonlocal knowledge typologies: technological complexity in the Irish knowledge space', *European Planning Studies*, 27 (4), pp. 661-677. https://doi.org/10.1080/09654313.2019.1567695.

Whittle, A. (2020) 'Operationalizing the knowledge space: theory, methods and insights for Smart Specialisation', *Regional Studies, Regional Science*, 7 (1), pp. 27-34. https://doi.org/10.1080/21681376.2019.1703795.

Whittle, A. and Kogler, D. F. (2020) 'Related to what? Reviewing the literature on technological relatedness: Where we are now and where can we go?', *Papers in Regional Science*, pp. 97-114. https://doi.org/10.1111/pirs.12481.

World Intellectual Property Organization (WIPO) (2019) "Background Paper for WIPO Technology Trends 2019: Artificial Intelligence". Technical report. Geneva: World Intellectual Property Organization. Available online at: https://www.wipo.int/publications/en/details.jsp?id=4386.

Wuchty, S., Jones, B. and Uzzi, B. (2007) "The increasing dominance of teams in production of knowledge", *Science*, 316 (5827), pp. 1036-1039.

Yan, B., Lee, T. and Lee, T. (2015) "Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): A co-word analysis", *Scientometrics*, 105 (2), pp. 1285-1300. https://doi.org/10.1007/s11192-015-1740-1.

Zhu, S., Jin, W. and He, C. (2019) "On evolutionary economic geography: A literature review using bibliometric analysis", *European Planning Studies*, 27(4), pp. 639-660. https://doi.org/10.1080/09654313.2019.1568395.

6. Appendix A

Table 2(a) List of AI Keywords

artificial intelligence	computation intelligence	neural network bayes	network
bayesian network	Rankboost	semi-supervised connections	decision model
deep learning genetic	data mining	semi-supervised training	inductive program
machine learning	natural language generation	natural language generation	reinforcement learning
unsupervised learning	unsupervised training	semi-supervised learning	algorithm
inductive logic	expert system	random forest decision tree	transfer learning
learning algorithm	learning model	support vector machine	adaboost
gradient tree boosting	Chatbot	natural language processing	xgboost
logistic regression	stochastic gradient descent	multilayer perceptron	latent semantic analysis
latent dirichlet allocation	multi-agent system	hidden markov model	fussy logic

Stemming and wildcards were applied to all terms.
Source: (WIPO, 2019).