OK Computer: the creation and integration of AI in Europe

Bernardo S. Buarque\textsuperscript{a}, Ronald B. Davies\textsuperscript{b}, Ryan M. Hynes\textsuperscript{b} and Dieter F. Kogler\textsuperscript{a}

\textsuperscript{a}Spatial Dynamics Lab, University College Dublin, Richview Campus, Belfield, D04 V1W8, Dublin 4, Ireland, bernardo.buarque@ucdconnect.ie, dieter.kogler@ucd.ie
\textsuperscript{b}School of Economics, University College Dublin, John Henry Newman Building, Stillorgan Rd, Belfield, D04 V1W8, Dublin 4, Ireland, ronbdavies@gmail.com, rhynes57@gmail.com

Received on May 9, 2019; editorial decision on October 29, 2019; accepted on October 29, 2019

This article investigates the creation and integration of artificial intelligence (AI) patents in Europe. We create a panel of AI patents over time, mapping them into regions at the NUTS2 level. We then proceed by examining how AI is integrated into the knowledge space of each region. In particular, we find that those regions where AI is most embedded into the innovation landscape are also those where the number of AI patents is largest. This suggests that, to increase AI innovation, it may be necessary to integrate it with industrial development, a feature central to many recent AI-promoting policies.

Keywords: artificial intelligence, geography of innovation, knowledge space, technological change, regional studies
JEL Classifications: O31, O33, R11

Introduction

Artificial intelligence (AI) has become one of the most prolific topics of discussion in the general media, policy circles and the academic community. It has captured a degree of attention that outstrips other less sensational but no less important fields of research, like green technologies. In academic literature, AI has already generated a body of research on its possible implications. Studies highlight that AI has the potential to automate cognitive non-routine tasks and replace humans in certain economic activities (Acemoglu and Restrepo, 2019; Frey and Osborne, 2017).\textsuperscript{1} AI also has the potential to spark positive economic change. Recent research suggests that AI has the capacity to enhance productivity (Brynjolfsson et al., 2018), innovation processes (Cockburn et al., 2018), international trade (Goldfarb and Trefler, 2018) and economic growth (Aghion et al., 2017). The rise of AI can lead to significant shifts in employment and income distributions across and within society, particularly when these gains are concentrated in AI-producing regions or sectors (Agrawal et al., 2019; Arntz et al., 2016; Atack et al., 2019; Korinek and Stiglitz, 2017).

Furman and Seamans (2019) sum up this sentiment nicely, commenting that the growth of...
AI “has led both to excitement about the capacity of technology to boost economic growth and to concern about the fate of human workers in a world in which computer algorithms can perform many of the functions that a human can” (Furman and Seamans, 2019, 161). This has led to the implementation of various policies encouraging AI development. Canada paved the way in 2017 by including a C$125 million budget for the development of its AI research capabilities. This was quickly followed by Germany, the USA, the UK, China and multilateral institutions, including the European Union (EU).

Despite all this interest, little is known about the development of AI technologies. While we observe a surge in interest regarding AI, we lack data to correctly appraise the creation and integration of AI technology and its many economic consequences. Empirical evidence on the creation and integration of AI technologies at the regional scale is very sparse if not missing altogether, and insights into this subject matter remain mostly theoretical and speculative. This constitutes a significant gap in the relevant literature, which the present contribution aims to tackle.

Our goal is 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. Based on information derived from full-text patent documents provided by the European Patent Office (EPO), these data provide a sophisticated visualisation and measure of the development of AI technology in Europe over time. This analysis is exploratory and largely descriptive. It is beyond the scope of the present study to provide all-inclusive evidence on the determinants of AI knowledge creation or its many implications (such as the impact of automation on labour markets). We instead provide a first look at where AI knowledge is created, which we hope will serve as a foundation for further and more detailed empirical investigations into this significant general-purpose technology. This is something that so far has not been achieved despite the excitement surrounding AI technology and its far-reaching consequences.

Our exploration has three steps: (i) identifying AI patents; (ii) discussing their distribution across space, time and technology and (iii) analysing their embeddedness in a region’s knowledge space. To the best of our knowledge, there is not yet a full picture of how AI creation is distributed across European regions and how it interacts with the overall technological knowledge production structure within a given region. This is the gap we aim to address.

The initial step in the analysis employs a keyword and technology matching algorithm on the abstracts of over 1.4 million European patent applications to identify those related to AI. We use a list of keywords and Cooperative Patent Classification (CPC) codes provided by the World Intellectual Property Organization (WIPO, 2019). We identify approximately 5300 AI patent families using this approach.

After identifying the AI patents, we map them into space (230 European NUTS2 regions, as defined by Eurostat), time (patent applications from 1980 to 2013) and technology sectors (four-digit CPC classes). AI patents tend to be clustered in 12 four-digit CPC codes out of the approximately 650 four-digit codes across the entire classification scheme. Although these 12 codes contain non-AI technologies as well, the bulk of AI patents fall into a small number of codes focussed on computing. Classes G06K—Recognition of Data; Presentation of Data; Record Carriers; Handling Record Carriers, G06T—Image Data Processing or Generation, In General and G06F—Electric Digital Data Processing are all particularly common. After mapping AI patents, we identify which regions are “AI superstars” and produce the most AI patents over our panel. These superstars are especially important as the top 10% of regions create 39.3% of AI patents in Europe.
Perhaps unsurprisingly, these superstar regions also dominate our AI-containing CPC codes. This suggests that the factors that support the development of computing-related technology may be a necessary condition for the development of AI knowledge as well. This concept of prerequisite knowledge is consistent with previous studies of regional technological evolution (Boschma et al., 2015; Kogler et al., 2013). Guided by the Evolutionary Economic Geography (EEG) framework (Kogler, 2015, 2016), this literature highlights how regional technological change is shaped by path-dependent processes (Martin and Sunley, 2006). These processes hinge on the composition of knowledge and technology. This sets the stage for regional diversification, with the direction dependent on cognitive relatedness between existing and new technologies (Boschma, 2017; Kogler, 2017). Thus, the feasibility of entering a new technology domain like AI is largely determined by the set of existing capabilities, such as computing, which are themselves largely historic and place specific. While these theories have been tested at the aggregate level (Kogler et al., 2017), our analysis extends the approach to the technology-specific sectoral level.

Finally, we develop a method to estimate how integral AI patents are to a regional technology portfolio. We follow the knowledge space methodology of Kogler et al. (2013, 2017) and use the co-occurrence of CPC codes on individual patent documents to construct a series of indicators describing the inventive network of a region. We then estimate how these indicators change when omitting AI patents, which shows how embedded those patents are in a region’s knowledge space. The main finding is that regions with the most AI patents also tend to be regions where AI is most connected to the overall knowledge space. If AI inventions were removed from these regions, there would be a significant shift in the inventive network structure. This finding suggests that AI is best developed when well connected to other research and development activities within the larger regional knowledge production ecosystem. Developing AI hand-in-hand with applications (such as image recognition for use in autonomous automobiles) may be more fruitful than developing AI as an isolated technology specialisation. This confirms insights from EEG-driven empirical research highlighted above. Cross-disciplinary development forms a central pillar of many of the government-led AI development policies discussed in Dutton (2018), as well as so-called Smart Specialization Strategies (S3) overall. Our results suggest that these policies may be on the right track. We also find that those regions with a specialisation in AI-containing CPC codes, and arguably the greatest AI potential, are also those where AI is most central. Targeting policies to these specific regions may increase their effectiveness.

Albeit exploratory, this study contributes to the literature by providing a path towards understanding the development of AI and its possible consequences. By identifying AI innovation, its distribution across key dimensions and its relationship to overall innovation patterns, our contribution provides a launching point for future work that seeks to describe the effects of AI on economic outcomes, the links between AI and industrial usage and the effectiveness of government AI initiatives.

The article proceeds as follows. The next section describes the identification of AI patents as well as the main CPC codes and regions where they are found. The third section analyzes how those AI patents fit into the knowledge spaces of the main AI regions, while the last concludes and offers future research directions.

**Identifying AI patents**

A significant challenge in identifying and measuring AI is defining the technology concept itself. “AI” can cover existing methodologies, such as machine learning; developing...
applications, such as autonomous automobiles; and the future possible creation of a sentient machine. Further complicating the understanding of how AI innovation fits into the overall research and development picture is the fact that AI is not the result of a single coherent technological base but rather a construct emanating from the recombination of many previously existing technological foundations. Taking advantage of the fact that patents are considered the best option for intellectual property protection in AI technology development, one way forward is to investigate the technological codes that underpin the development of an invention (EPO, 2018). Investigating these CPC codes has proven useful in previous studies where the identification and subsequent diffusion of a novel product or process of economic value was the object of inquiry (Feldman et al., 2015). Unfortunately, there is no single CPC technology class specific to AI, so this line of research must be expanded to address this issue.

Using patents has two additional limitations that are not unique to AI. First, not all innovations are patented. If AI is more common in unpatented innovations relative to other technologies, it will be under-represented in the sample of patents. That said, so long as there is not a systematic pattern of unpatented AI innovations across regions, time or technology, this will not affect the qualitative nature of our findings. Second, patents do not provide a systematic way of distinguishing between product and process innovation, a key factor influencing the economic impact of innovation. This latter distinction might be particularly hazy for AI because it is intended to be adaptable and applicable across different problems. For example, a machine-learning algorithm may be developed to speed up handwriting recognition in a smartphone but may also find its way into an application that suggests products to a consumer. In this way, the initial process innovation finds its way to a new application. Distinguishing between patented and non-patented AI and between product and process AI will be critical for future analysis of the impacts of AI development but is outside the current exploration.

With these limitations in mind, we proceed by applying a keyword identifier to European patents from 1980 to 2013 as reported by the EPO’s PATSTAT database. PATSTAT provides the patent text itself along with patent and inventor metadata, such as priority year, inventor address and CPC codes. To identify AI patents, we take the abstract text of each patent and perform a keyword search to identify AI-related terms. We use this search to classify each patent in our data set as “AI” or “non-AI.” Obviously, this approach relies on the choice of identifiers. For the results presented here, we use a combination of keywords and CPC codes (well beyond the four-digit aggregate level) that are indicative of AI proposed by WIPO (2019). We believe WIPO’s identification gives a more recent and specific definition of AI than alternatives. In robustness checks, we use the identifying keywords developed by Cockburn et al. (2018) and reach comparable conclusions.

This process identifies 5312 AI patents out of the 1.4 million patent documents in our version of the PATSTAT database. Despite the excitement surrounding AI, very few have successfully patented the technology so far. The first issue to explore is in which major CPC codes AI patents tend to fall. This is illustrated in Figure 1, where we aggregate AI patent CPC codes at the four-digit level. The most common CPC codes fall under the parent category G06—Computing, Calculating, and Counting. In particular, G06K—Recognition of Data; Presentation of Data; Record Carriers; and Handling Record Carriers is the most common four-digit CPC code, with G06T—Image Processing in a somewhat distant second place. Although appearing far less frequently than the G06 codes, technologies under H04—Electricity are also fairly common. Thus, AI
The creation and integration of AI in Europe

The creation and integration of AI in Europe

patents seem to be appropriately classified in terms of technological knowledge foundations. The second objective is to allocate AI knowledge production in space. We geo-code the addresses of AI inventors and assign them to their respective NUTS2 region. We then group the data into 5-year periods from 1980 to 2013, using the patents’ priority year following accepted conventions in the literature. For robustness, we conduct a yearly analysis of knowledge spaces but find little significant change in the underlying network structure. Larger conceptual trends are more readily apparent at the 5-year period level, something in line with earlier studies (Kogler et al., 2013, 2017). After performing this mapping, we have a count of AI patents per region-period.

As illustrated in Figure 2, the distribution of AI patents across regions is highly skewed. Panel A shows that most regions do not have any AI patents at all and those that do have very few. This is unsurprising, given that we only have 5300 patents distributed over 230 regions, and also follows previous insights that have shown that highly specialised knowledge production is concentrated in a small number of locations (Feldman et al., 2015). Even within the top quartile of AI patenting regions, Panel B indicates that the distribution remains skewed. AI patents are very concentrated in a handful of regions, with a mere 23 regions accounting for 39.9% of all patented AI knowledge production. The concentration of AI patents can also be seen in Figure 3, which compares the accumulated number of AI patents in each NUTS2 region in the 5-year periods ending in 1988 and 2013.

Figure 1. Count of AI patents by CPC code.
Year-to-year changes in the annual number of patents published are illustrated in Figure 4. The first AI patent applications we identify were published in 1987 when seven were published. Following a period of relatively little activity, there was a jump in the annual number of patents in the early 1990s. Around 2010, there was another shift where not only the number of AI patents increased but also the annual growth rate. This is potentially indicative of the achievement of a critical mass across both time and regions, as Figure 3 shows.

As the above figures indicate, certain regions have emerged as hotbeds for AI innovation. Table 1 lists the top 10% of AI-producing regions, along with the number of both AI and non-AI patents they produced during the sample. While large regions containing cities like Paris and Munich top this list, smaller regions, such as Mittelfranken (DE) and East Anglia (UK) also appear. Size aside, what sets these regions apart is their disproportionately large shares of expertise in AI-related CPC codes (% AI CPCs) in their local knowledge spaces. Dividing shares equally across the 230 NUTS2 regions gives an average of 0.43%. In comparison, the CPC share of the top five
The creation and integration of AI in Europe

The biggest AI-producing regions is at least five times larger. This suggests that regions at the forefront of innovation in computing overall are also those with an advantage in producing AI patents.

AI and the knowledge space

While the above describes the distribution of AI patents, of particular interest in the present study is how AI technology connects to and is embedded in a region’s overall knowledge-producing and innovative environment. This is because many AI-promoting policies specifically seek to develop AI with an eye towards its use in a multitude of commercial product and process applications. In order to develop a measure of AI knowledge production embeddedness in a region, we first follow the “knowledge space” methodology originally developed and applied to US metropolitan areas by Kogler et al. (2013). This methodology was also recently employed in a study that focussed on the evolution of technological specialisation among European NUTS2 regions (Kogler et al., 2017). The knowledge space is based on a co-occurrence matrix of patent CPC classes. Every patent is assigned to at least one CPC class, and most patents are assigned multiple CPC classes. These CPC classes indicate the fundamental knowledge building blocks that contributed to the development of the invention. When two CPCs occur frequently together, they are likely highly related in cognitive proximity (Nooteboom, 2000). Because it is possible to trace the changing patterns of knowledge domains’ relatedness scores to all other technology classes over time, this methodology is often employed in evolutionary studies that

Figure 4. AI patents over time

![Graph showing AI patents over time](https://academic.oup.com/cjres/article/13/175/1767/5719675)
Buarque et al.

The usual approach is to generate an overall knowledge space, or co-occurrence matrix, for the entire system under investigation. In Kogler et al. (2013, 2017), this was the North American or pan-European knowledge space, respectively. The derived relatedness measures between CPC classes are then averaged out in order to generate an overall “average relatedness score.” This score can be calculated in different ways based on the presence and quantity of CPC classes in a region (Leydesdorff et al., 2017). While this approach excels at illustrating the differences between regions, it does not consider that each region has its own unique knowledge-relatedness matrix that also changes over time. Changes in the local matrix can provide further insights into how technological knowledge gets produced and recombined in the local context (Kogler et al., 2018).

The original knowledge space methodology serves as a point of departure in the present investigation. Unlike other studies that use co-occurrence measures for invention, industries or occupations (Whittle and Kogler, 2019), we create a unique knowledge space for each region-period. We then calculate a series of descriptive network statistics for these regional knowledge spaces. We use these statistics to build an index that captures how embedded AI knowledge production is in the regional inventive environment.

For each knowledge space we construct the following descriptive network variables: (i) number of edges; (ii) network density; (iii) network diameter; (iv) average path length; (v) average clustering coefficient; (vi) average eigenvector centrality; (vii) average degree centrality and (viii) average betweenness centrality. Number of edges is the count of total

Table 1. Top AI-producing regions

| NUTS 2 | Region                              | AI Patents | Non-AI Patents | % AI CPCs | AICI Change |
|--------|-------------------------------------|------------|----------------|-----------|-------------|
| FR10  | Île de France                       | 293        | 173,250        | 6.93%     | 2.359       |
| DE21  | Oberbayern                          | 280        | 134,106        | 5.64%     | 4.852       |
| DE25  | Mittelfranken                       | 172        | 46,153         | 2.25%     | 0.908       |
| NL41  | Noord-Brabant                       | 137        | 84,923         | 5.45%     | 1.481       |
| DE11  | Stuttgart                           | 120        | 131,917        | 2.74%     | 1.303       |
| DE12  | Karlsruhe                           | 106        | 80,240         | 2.57%     | 0.954       |
| DE71  | Darmstadt                           | 77         | 108,034        | 1.81%     | 2.420       |
| UKH1  | East Anglia                         | 72         | 28,897         | 1.45%     | 2.405       |
| DEA2  | Köln                                | 72         | 87,022         | 1.72%     | 1.186       |
| ITC4  | Lombardia                           | 70         | 66,708         | 1.07%     | 1.328       |
| DE30  | Berlin                              | 68         | 40,251         | 1.48%     | 0.804       |
| FR71  | Rhône-Alpes                         | 66         | 78,988         | 1.93%     | 1.388       |
| DE14  | Tübingen                            | 66         | 44,424         | 1.02%     | 1.000       |
| DE13  | Freiburg                            | 65         | 51,076         | 1.72%     | 1.060       |
| SE11  | Stockholm                           | 64         | 40,235         | 2.95%     | 0.911       |
| UKJ2  | Surrey, East and West Sussex        | 63         | 21,303         | 1.32%     | 0.330       |
| DE92  | Hannover                            | 60         | 28,012         | 1.06%     | 1.274       |
| FR82  | Provence-Alpes-Côte d’Azur          | 54         | 25,405         | 1.82%     | 0.689       |
| SE22  | South Sweden                        | 52         | 22,225         | 1.41%     | 1.018       |
| UKJ1  | Berkshire, Buckinghamshire and Oxfordshire | 48    | 31,575         | 1.41%     | 0.279       |
| FI1B  | Helsinki-Uusimaa                    | 44         | 34,030         | 2.52%     | 0.406       |
| DE27  | Schwaben                            | 43         | 28,100         | 0.74%     | 1.378       |
edges in the network. If more than one patent contains the same link between nodes, edges are counted more than once. Network density measures the number of actual connections out of the total possible number of connections within a network. Network diameter measures the shortest path that spans the entire network. Average path length measures the average distance between nodes. Average clustering coefficient is the number of three-way node connections or clusters out of the total possible number of clusters. Average eigenvector centrality is a measure of the average centrality of nodes in a network. Average degree centrality is the average number of edges to which any given node is connected. Average betweenness centrality measures the number of times a node is used as an intermediate step in a path between two nodes. Thus, each measure gives a characteristic of the overall network structure of each region-period’s knowledge space.

Figure 5 illustrates two regional knowledge spaces. On top, we have the Dutch region of North Brabant (a significant AI producer) and, on the bottom, the Italian region of Veneto (a minor AI producer). The figures were created using a force-directed graph drawing (Kamada–Kawai layout), where the position of the nodes is proportional to graph distance and the size of the nodes reflects their degree centrality. While these regions’ network structures are similar, they differ in terms of the position and relative importance of AI-related nodes, which are the dark black circles. These nodes are more central, closer to one another and better connected in the North Brabant knowledge space.

For each NUTS2 region and 5-year time period, we construct two knowledge spaces. In the first, we use all of the region-period’s patents. In the second, we omit the AI patents and recalculate the knowledge space’s network statistics. Note that we omit only the AI patents, not entire CPC codes. Doing so affects the network by reducing the number of connections between CPC nodes and, in some cases, may delete CPC nodes entirely if they are not used in non-AI patents. These changes affect regional knowledge spaces differently depending on how embedded or central AI is to a given knowledge space. For example, a network in which AI is deeply embedded will suffer significant changes in terms of overall connectedness. The knowledge space will lose valuable links that might lead to recombinant knowledge production, that is, the use of two CPC classes or knowledge domains as input into the development of an invention that have not been previously used together. Conversely, in regions where AI knowledge production is absent or isolated from other inventive activities in the region, there will be little to no change in the knowledge space’s network statistics.

Because no single CPC contains all aspects of AI, we cannot measure the centrality of AI based on the centrality of one CPC code. Since
AI does not belong to one unique node but rather spans several CPCs, there is no simple method to estimate how integrated AI is to a region’s knowledge space. To obtain an approximate measure of AI embeddedness, we compute how much a region’s knowledge space changes when local AI patents, and not the entire CPC codes that contain them, are omitted from the knowledge space calculation.

European regions differ substantially regarding their patenting potential, which might affect the structure of their respective knowledge spaces. Namely, larger regions (for example, Paris) tend to patent in many technologies and have denser knowledge spaces. Thus, our network statistics will depend significantly on regional characteristics, which in turn would make it difficult to directly compare regions of different sizes. We net out this size effect by comparing each network with itself and taking the differences in network statistics. Thus, the estimated differences in the network statistics are not overly sensitive to regional patenting potential and our proposed index only captures how integral AI is to a region’s knowledge space. In summary, the method adopted in this analysis permits us to roughly control for regional differences in patenting potential and to focus on AI embeddedness.

We construct the AI Centrality Index (AICI) by taking the difference in network statistics with and without the AI patents. We then use principal component analysis (PCA) across those differences in the various network statistics. We identify the first component to collect weights and then analyse a linear combination of the eight network statistics mentioned above. This produces an equally weighted, normalised index of changes in the network statistics. The AICI is the weighted sum of these statistics. We use the first component because it accounts for the greatest possible variance in the data (Abdi and Williams, 2010). Table 2 provides the weights used to construct the AICI.

When AI patents are more embedded in the regional knowledge space, we generally expect their removal to cause a decline in the level of overall centralisation, a more dispersed network and greater path lengths. Because we calculate the AICI as the difference between network statistics with and without AI patents, we presume the variables “∆Avg_Path_Length” and “∆Avg_Clustering_Coef” to be positive. In contrast, since the network should have more links when we include the AI patents, “∆Net_Density” and “∆Avg_Degree_Centrality” should be negative. We should then observe larger AICI values for regions where AI is significantly embedded in the knowledge space.

Table 2. PCA of network characteristics

| Variables                        | Weight (PC1) |
|----------------------------------|--------------|
| ∆No_Edge                         | 0.09         |
| ∆Net_Density                     | 0.54         |
| ∆Net_Diameter                   | −0.13        |
| ∆Avg_Path_Length                 | −0.26        |
| ∆Avg_Clustering_Coef             | 0.29         |
| ∆Avg_Eigenvector_Centrality     | −0.41        |
| ∆Avg_Degree_Centrality          | 0.60         |
| ∆Avg_Betweeness_Centrality      | −0.05        |

Figure 6 shows a histogram of the average AICI for all regions and all periods. Similar to the distribution of AI patents in Figure 2, the distribution of the AICI values is highly skewed. Most region-periods show little to no change in the AICI. This is to be expected since most region-periods do not produce significant amounts of AI knowledge and thus we omit very few AI patents in those regions.

Table 3 gives details on changes in each of the components in the index as well as the total change in the AICI. The mean AICI across all region-periods is 0.564. This average is pulled down by the regions with little to no AI knowledge production, so we also report the changes for the 75th, 90th and 95th percentiles of AI-producing regions. The mean AICI grows considerably when we focus on those areas with the most AI patents, mostly due to changes in the number of edges.
A different approach is to plot the number of AI patents (vertical axis) against the AICI (horizontal axis) at the 75th percentile. This is shown in Figure 7a, where we also plot the estimated trend from a bivariate regression (the shaded area in the figure is the confidence

**Table 3. Knowledge space changes when omitting AI**

| Variables                        | All          | 75th percentile | 90th percentile | 95th percentile |
|----------------------------------|--------------|-----------------|-----------------|-----------------|
|                                  | Mean         | SD              | Mean            | SD              | Mean            | SD              | Mean            | SD              |
| ∆No_Edge                         | 2.487        | 2.788           | 4.540           | 3.616           | 6.068           | 4.073           | 7.913           | 4.617           |
| ∆Net_Density                     | -0.000       | 0.001           | 0.000           | 0.000           | 0.000           | 0.000           | 0.000           | 0.000           |
| ∆Net_Diameter                    | -0.008       | 0.190           | -0.002          | 0.108           | 0.012           | 0.137           | -0.019          | 0.068           |
| ∆Avg_Path_Length                 | 0.001        | 0.027           | 0.0004          | 0.020           | 0.002           | 0.020           | -0.002          | 0.002           |
| ∆Avg_Clustering_Coef            | -0.001       | 0.004           | -0.000          | 0.001           | -0.000          | 0.000           | 0.000           | 0.000           |
| ∆Avg_Eigenvector_Centrality     | 0.001        | 0.010           | -0.000          | 0.001           | -0.000          | 0.000           | -0.000          | 0.000           |
| ∆Avg_Degree_Centrality          | -0.001       | 0.002           | -0.000          | 0.001           | 0.000           | 0.000           | 0.000           | 0.000           |
| ∆Avg_Betweenness_Centrality     | -0.001       | 0.004           | -0.000          | 0.001           | 0.000           | 0.001           | -0.000          | 0.001           |
| AICI                             | 0.564        | 0.833           | 1.014           | 1.059           | 1.343           | 0.950           | 1.722           | 1.102           |
| Obs                              | 224          | 224             | 60              | 60              | 24              | 24              | 13              | 13              |
There is a positive correlation between those regions where AI patents are most prevalent and those for which AI is most embedded. This suggests that AI development is most successful when it is embedded into a region’s overall inventive landscape. As the policies discussed by Dutton (2018) often provide incentives for the hand-in-hand development of basic AI and its applications, this may well indicate that policymaking is on the right track.

One feature visible in Figure 7a is the presence of three outlier regions: ITC1 (Torino), FR10 (Paris) and DE21 (Munich). We remove them and re-estimate the correlation in Appendix Figure A1. There is still a clear positive relationship between the size of the AICI and the number of AI patents showing that, even for moderate AI-producing regions, this relationship holds.

Note that these figures compare the AICI with the absolute number of AI patents. This raises the issue of whether this relationship is simply one of scale. Since the regions with the most AI patents have the most removed when constructing the AICI, this could bias the value of the AICI. An alternative is to compare the AICI with some relative measure of AI patents. The challenge in doing this with our AI-identified patents is two-fold. First, if we compare the AICI to the share of AI patents we would be deleting a greater share of patents in regions specialising in AI and would exacerbate the possibility that the result is driven by scale. Second, recall that AI patents are few in number even in those regions that develop the bulk of them. Because of these small numbers, measures of specialisation can experience large changes from a small change in the number of AI patents, making them unsuitable to describing specialisation patterns. Together, this creates the possibility of misleading inference when using the share of AI patents in a region’s total.

As an alternative, in Figure 7b, we compare the AICI to the share of a region’s patents that fall in the main AI CPC codes (as in Table 1). This measure sidesteps these concerns both because it contains non-AI patents and because it does not suffer from the small number problem. Thus, here, we compare the change in the network to what can be thought of as the region’s potential for developing AI patents. The relationship is similar to that shown in Figure 7a. Those regions that specialise in AI-containing CPC codes are also those where the AICI is greatest. This provides some assurance that our finding is not merely the result of scale and further reinforces the notion that AI superstars are also those where AI is more embedded.

**Conclusion**

AI is clearly one of the buzzwords in contemporary innovation processes, and the rush to develop and market AI is palpable. As governments scramble to implement policies to
support “Industry 4.0”; it becomes increasingly necessary to develop an understanding of how, where and why AI development is taking place. We contribute to this in three key ways. First, to overcome the lack of CPC codes capturing AI, we use a text-matching algorithm to identify AI patents, finding just over 5300 patent families in Europe from 1987 to 2013. Second, we describe the regions that are most successful in generating AI patents and find that success in computing-related innovation activities likely feeds into success in AI. Finally, we position the AI patents into each region’s knowledge space and develop a methodology for describing how embedded AI is in the region’s inventive network. When doing so, we find that the AI superstar regions also tend to be those for which AI is most embedded in its knowledge space.

This last point also contributes methodologically to efforts to measure how embedded a particular technology is in a local knowledge space. When a technology is identifiable with a single CPC class, the approach is already laid out in work such as in Feldman et al. (2015), which investigates the spatial diffusion and adoption of rDNA methods.12 We extend this by providing a framework for doing so when the technology is not clearly identified with a single CPC class, something which can extend beyond AI. For instance, Haščič et al. (2015) offer a text-matching algorithm to identify “green patents”. By combining their algorithm along with the method adopted in this research, one could provide an indication of how integral “green technologies” are to a region’s knowledge space and how that integration relates to the development of those technologies.

From a policy perspective, our results suggest two things. First, if AI is best developed by linking it to other technological innovations, our findings support the application-driven AI promotion policies currently in use by many governments. Second, it suggests that those policies alone are unlikely to be successful since a large part of AI development capabilities build from the historic strengths of a region. Thus, even if “smart specialization” targets AI, only some regions may be able to achieve their policy goals. In addition, if AI has the economic returns predicted by its supporters, this suggests that, as the technology grows, it is likely to add to the unequal distribution of income growth across regions. Since our data indicate that the development of AI entered a new, rapid phase of development starting in 2010, such shifts may well already be underway. We therefore hope that our results combined with incoming data can provide a framework for identifying which factors feed into the development of AI and which policies can be used to best capture its opportunities. Further, by providing an initial foray into identifying AI innovation, we hope that our work acts as a stepping stone towards more accurate analysis of its economic and social impacts.

### Endnotes

1 Frey and Osborne (2017) measure the likelihood of occupations becoming automated with the growth of AI by breaking each occupation down into its component tasks, then assigning a probability of automation to each task. Acemoglu and Restrepo (2019) take a similar task-based approach. The easier the component tasks are to automate, the more likely an occupation will see humans replaced by AI. Some occupations, like analysing medical tests, are under significant risk.

2 For an overview of the various policies that have been implemented, see Dutton (2018). Details on the EU’s strategic AI policy can be found in European Commission (2018). The effects of China’s emphasis in robotics can be seen in Cheng et al. (2019).

3 Papers such as Acemoglu and Restrepo (2017) or Graetz and Michaels (2018) use data from the International Federation of Robotics to estimate sectoral and regional exposure to automation via the concentration of robots in those areas. However, the data set adopted by the authors simply uses the presence of robots in an industry as a proxy for AI and growth, a rather imperfect measure that offers only limited inference.
The CPC is a patent classification system jointly developed by the EPO and the United States Patent and Trademark Office (USPTO); see here for further details: https://www.cooperativepatentclassification.org/.

Appendix Tables A1 and A2 list the CPCs and keywords used to identify the AI patents.

This alternative found roughly one-third as many AI patents as the WIPO approach; however, this does not alter our fundamental conclusions: AI is concentrated in certain CPC codes and regions, with the major AI producers being those where AI technology is most central to their knowledge space. While we use a combination of keywords and CPC codes to identify AI patents, other identification strategies exist in the literature. Mann and Püttmann (2017) apply a supervised machine-learning algorithm (Naïve Bayes) to the text of USPTO patents to sort them into automation and non-automation patents. Haščič et al. (2015) develop an indicator to identify environmentally related patents, and Arts et al. (2018) use a text-matching algorithm to construct an alternative measure of patent similarity.

As the first AI patents we identify were in 1987, our first period runs from 1980 to 1988.

The spatial allocation of AI patents follows the fractional inventor share allocation applied in previous studies (Kogler et al., 2013, 2017).

Table 3 shows the average values for the eight network descriptive variables. It mainly confirms the general expectations.

Table 1 provides the AICI change for the top AI-producing regions.

In unreported results where we attempted this despite these concerns, we find no significant relationship between the share of AI patents and the AICI. This may be precisely due to these issues. When using percentage change in network characteristics rather than the AICI, we find a relationship similar to that in Figure 7(a and b), that is, the percentage change in the network is greatest where the share of AI patents is highest.

It is worth recognising that, comparable to what we find for AI, they document that regional cognitive proximity to certain technological classes is fundamental for the development of rDNA method-related innovations.

Acknowledgements

BB and DFK would like to acknowledge funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 715631, ERC TechEvo). DFK would also like to acknowledge further funding for this project from the Science Foundation Ireland (SFI) under the SFI Science Policy Research Programme (grant agreement No. 17/SPR/5324, SciTechSpace).

References

Abdi, H. and Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2: 433–459.

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://papers.ssrn.com/sol3/papers.cfm?abstract_id=2941263

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

Aghion, P., Jones, B. F. and Jones, C. I. (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: 31–50.

Arntz, M., Gregory, T. and Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. Technical report. OECD Social, Employment and Migration Working Papers No. 189. Paris: OECD Publishing. Available online at: https://doi.org/10.1787/5jlz9h56dvq7-en.

Arts, S., Cassiman, B. and Gomez, J. C. (2018). Text matching to measure patent similarity, Strategic Management Journal, 39: 62–84.

Atack, J., Margo, R. A. and Rhode, P. W. (2019). Automation of manufacturing in the late nineteenth century: the hand and machine labor study, Journal of Economic Perspectives, 33: 51–70.

Boschma, R. (2017). Relatedness as driver of regional diversification: a research agenda, Regional Studies, 51: 351–364.
Boschma, R., Balland, P.-A. and Kogler, D. F. (2015). Relatedness and technological change in cities: the rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010, *Industrial and Corporate Change, 24*: 223–250.

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

Cheng, H., Jia, R., Li, D. and Li, H. (2019). The rise of robots in China, *Journal of Economic Perspectives, 33*: 71–88.

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](https://www.nber.org/papers/w24449)

European Commission (2018). *Artificial Intelligence: An European Perspective*. Technical report. Luxembourg: Publications Office of the European Union. doi:10.2760/936974.

European Patent Office (EPO). (2018). *Conference Summary of the EPO Conference on Patenting Artificial Intelligence, European Patent Office, Munich, 30 may 2018*. Available online at: [http://documents.epo.org/projects/babylon/acad.nsf/0/D9F20464038C0753C125829E0031B814/$FILE/summary_conference_artificial_intelligence_en.pdf](http://documents.epo.org/projects/babylon/acad.nsf/0/D9F20464038C0753C125829E0031B814/$FILE/summary_conference_artificial_intelligence_en.pdf)

Feldman, M. P., Kogler, D. F. and Rigby, D. L. (2015). rKnowledge: the spatial diffusion and adoption of rDNA methods, *Regional Studies, 49*: 798–817.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerization? *Technological Forecasting and Social Change, 114*: 254–280.

Furman, J. and Seamans, R. (2019). AI and the economy, *Innovation Policy and the Economy, 19*: 161–191.

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

Graetz, G. and Michaels, G. (2018). Robots at work, *Review of Economics and Statistics, 100*: 753–768.

Haščič, I. and Migotto, M. (2015). *Measuring Environmental Innovation Using Patent Data*. Technical report. Paris: OECD Environment Working Papers, No. 89, OECD Publishing. doi:10.1787/5js009f48xw-en.

Kogler, D. F. (2015). Editorial: evolutionary economic geography— theoretical and empirical progress, *Regional Studies, 49*: 705–711.

Kogler, D. F. (ed.) (2016). *Evolutionary Economic Geography: Theoretical and Empirical Progress (Regions and Cities)*. Abingdon: Routledge.

Kogler, D. F. (2017). Relatedness as driver of regional diversification: a research agenda—a commentary, *Regional Studies, 51*: 365–369.

Kogler, D. F., Essletzbichler, J. and Rigby, D. L. (2017). The evolution of specialization in the EU15 knowledge space, *Journal of Economic Geography, 17*: 345–373.

Kogler, D. F., Heimeriks, G. and Leydesdorff, L. (2018). Patent portfolio analysis of cities: statistics and maps of technological inventiveness, *European Planning Studies, 26*: 2256–2278.

Kogler, D. F., Rigby, D. L. and Tucker, I. (2013). Mapping knowledge space and technological relatedness in US cities, *European Planning Studies, 21*: 1374–1391.

Korinek, A. and Stiglitz, J. E. (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](https://www.nber.org/papers/w24174)

Leydesdorff, L., Kogler, D. F., Yan, B. (2017) Mapping patent classifications: portfolio and statistical analysis, and the comparison of strengths and weaknesses, *Scientometrics, 112*: 1573–1591.

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](https://ssrn.com/abstract=2959584).

Martin, R. and Sunley, P. (2006). Path dependence and regional economic evolution, *Journal of Economic Geography, 6*: 395–437.

Nooteboom, B. (2000). Learning by interaction: absorptive capacity, cognitive distance and governance, *Journal of Management and Governance, 4*: 69–92.

Whittle, A. and Kogler, D. F. (2019). Related to What? Reviewing the Literature on Technological Relatedness: Where We Are Now and Where Can We Go? Papers in Regional Science, 1–17.

World Intellectual Property Organization (WIPO) (2019). *Background Paper for WIPO Technology Trends 2019: Artificial Intelligence*. Technical report. Geneva: World Intellectual Property Organization.
Figure A1. AI patents and AICI—outliers omitted.
### Table A1. List of AI CPCs

| AI CPCs | G06N 003 | G06N 005 - 027 | G06T 2207/20081 |
|---------|----------|----------------|-----------------|
| Y10S 706 | G06N 099/005 | G06T 009/002 | G06T 013/0275 |
| G06N 007/005 - 060 | G05B 013/027 | G05B 013/029 | G05D 001/0088 |
| G06T 2207/20084 | G05B 013/0285 | G10L 017 | G06F 017/3029 - 30035 |
| G06F 017/16 | G05B 2219/33002 | G06F 017/3043 | G06F 017/30663 |
| G05B 013/028 | G10L 015 | G06F 017/30672 | G06F 017/3069 |
| G05B 013/0295 | G06F 017/28 - 289 | G06F 017/30731 - 30737 | G06F 019/24 |
| G06K 009 | G06F 017/30401 | G01N 2201/1296 | G01N 2201/1296 |
| G06F 017/27 - 2795 | G06F 017/30654 | G01N 031/3651 | G06F 011/1476 |
| G06F 017/3024 - 30262 | G06F 017/30669 | G06F 015/18 | G06N 007/046 |
| G06F 017/30522 - 3053 | G06F 017/30687 | G10H 2250/311 | H01M 008/04992 |
| G06F 017/30666 | G06F 017/30705 - 30713 | H02P 023/0018 | H04L 2012/5686 |
| G06F 017/30684 | G06F 017/30784 - 30814 | H04L 025/0254 | H04L 045/08 |
| G06F 017/30702 | G01R 031/2846 - 2848 | H04Q 025/0384 | H04Q 2113/1334 |
| G06F 017/30743 - 30746 | G01N 033/0034 | G08B 029/186 | B60G 2600/1879 |
| G06F 019/707 | G06F 011/2263 | G08B 2600/1879 | B29C 007/986 |
| G01N 029/4481 | G06F 007/1482 | G10H 2250/311 | B29C 007/986 |
| G01S 007/417 | G10H 2250/151 | H01M 008/04992 | B29C 007/986 |
| G06F 011/2257 | H01J 2237/3047 | H02P 023/0018 | H04L 2012/5686 |
| G06F 2207/4824 | H02P 021/0014 | H04L 025/0254 | H04L 045/08 |
| G11B 020/10518 | H03H 2222/04 | H04Q 2113/1334 | B60G 2600/1879 |
| G10K 2210/3024 | H04L 2025/03554 | G08B 029/186 | B29C 007/986 |
| H02H 001/0092 | H04L 041/16 | G08B 2600/1879 | B29C 007/986 |
| H03H 2017/0208 | H04Q 2213/054 | B23K 031/006 | B29C 007/986 |
| H04L 2025/03464 | H04R 025/0507 | B25J 000/161 | B29C 007/986 |
| H04L 025/03165 | B60G 2600/1878 | B29C 007/986 | B29C 007/986 |
| H04N 021/4662 - 4666 | E21B 0241/0028 | E21B 2041/0028 | B29C 007/986 |
| H04Q 2213/3343 | F03D 007/046 | F03D 007/046 | B29C 007/986 |
| B60G 2600/1876 | F01H 2061/0081 | F01H 2061/0081 | F05B 2270/707 |
| B64G 2001/247 | B60W 030/10 - 12 | B60W 030/10 - 12 | F05B 2270/707 |
| B29C 2945/76979 | G06T 2207/30248 - 30268 | G06T 2207/30248 - 30268 | G06T 2207/30248 - 30268 |
| A61B 005/726 - 7267 | Y10S 128/924 | Y10S 128/924 | Y10S 128/924 |
| F02D 041/1405 | F03D 007/046 | F03D 007/046 | F03D 007/046 |
| F05B 2270/709 | F01H 2061/0081 | F01H 2061/0081 | F01H 2061/0081 |
| B60W 030/06 | B60W 030/10 - 12 | B60W 030/10 - 12 | B60W 030/10 - 12 |
| B62D 015/0285 | G06T 2207/30248 - 30268 | G06T 2207/30248 - 30268 | G06T 2207/30248 - 30268 |
| G05D 001 | A61B 005/7267 | F05D 2270/709 | F05D 2270/709 |
| G06T 2207/20084 | G10K 2210/3038 | G10L 025/30 | G10L 025/30 |
| H04N 021/4666 | A63F 013/67 | G06F 017/2282 | G06F 017/2282 |

Source: [WIPO, 2019](https://www.wipo.int/pctdb/en/).

Downloaded from https://academic.oup.com/cjres/article/13/1/175/5719587 by guest on 30 January 2022
Table A2. List of AI keywords

| Artificial Intelligence | Computation Intelligence | Neural Network Bayesian Network | Chatbot |
|-------------------------|--------------------------|-------------------------------|---------|
| Network                 | Bayesian Network         | Deep Learning Genetic Program|
| Data Mining             | Decision Model           | Inductive Program             |
| Algorithm               | Inductive Logic          | Natural Language              |
| Machine Learning        | Natural Language Generation|                               |
| Processing              | Supervised Training      | Swarm Intelligence            |
| Reinforcement Learning  | Unsupervised Training    | Expert System Fuzzy           |
| Unsupervised Learning   | Semi-Supervised Connections| Learning Algorithm            |
| Semi-Supervised Training| Transfer Learning        | Random Forest Decision        |
| Logic                   | Support Vector Machine   | XGBoost                       |
| Learning Model          | Gradient Tree Boosting    | Logistic Regression           |
| Tree                    | RankBoost                | Latent Semantic Analysis      |
| Adaboost                | Multilayer Perceptron    | Hidden Markov Model           |
| Stochastic Gradient Descent | Multi-Agent System       |                                |

Stemming and wildcards were applied to all terms. Source: [WIPO, 2019](https://academic.oup.com/cjres/article/13/1/175/5719587).