JRC TECHNICAL REPORTS

AI Watch
TES analysis of AI Worldwide Ecosystem in 2009-2018

Joint Research Centre

EUR 30109 EN

Contact information
Address: Edificio Expo, C/ Inca Garcilaso 3, Sevilla 41092, Spain
Email: EC-AI-WATCH@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC120106

EUR 30109 EN

PDF ISBN 978-92-76-16661-0 ISSN 1831-9424 doi:10.2760/85212

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Samoili S., Righi R., Cardona M., López Cobo M., Vázquez-Prada Baillet M., and De Prato G., TES analysis of AI Worldwide Ecosystem in 2009-2018, EUR 30109 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-16661-0, doi:10.2760/85212, JRC120106.

Contents

Foreword

This report is published in the context of AI Watch, the European Commission knowledge service to monitor the development, uptake and impact of Artificial Intelligence (AI) for Europe, launched in December 2018.

AI has become an area of strategic importance with potential to be a key driver of economic development. AI also has a wide range of potential social implications. As part of its Digital Single Market Strategy, the European Commission put forward in April 2018 a European strategy on AI in its Communication "Artificial Intelligence for Europe" COM(2018)237. The aims of the European AI strategy announced in the communication are:

- To boost the EU's technological and industrial capacity and AI uptake across the economy, both by the private and public sectors
- To prepare for socio-economic changes brought about by AI
- To ensure an appropriate ethical and legal framework.

Subsequently, in December 2018, the European Commission and the Member States published a "Coordinated Plan on Artificial Intelligence", COM(2018)795, on the development of AI in the EU. The Coordinated Plan mentions the role of AI Watch to monitor its implementation.

AI Watch monitors European Union's industrial, technological and research capacity in AI; AI-related policy initiatives in the Member States; uptake and technical developments of AI; and AI impact. AI Watch has a European focus within the global landscape. In the context of AI Watch, the Commission works in coordination with Member States. AI Watch results and analyses are published on the AI Watch Portal (https://ec.europa.eu/knowledge4policy/ai-watch_en).

From AI Watch in-depth analyses we will be able to understand better European Union's areas of strength and areas where investment is needed. AI Watch will provide an independent assessment of the impacts and benefits of AI on growth, jobs, education, and society.

AI Watch is developed by the Joint Research Centre (JRC) of the European Commission in collaboration with the Directorate-General for Communications Networks, Content and Technology (DG CONNECT). This report addresses the following objectives of AI Watch: Developing an overview and analysis of the European AI ecosystem.

Acknowledgements

The authors would like to thank Alessandro Annoni, and Michael Lutz (JRC) for their support. The authors are grateful to Marc Bogdanowicz for the conceptualisation of the TES approach at early stages. The authors would also like to acknowledge Paul Desruelle, Emilia Gómez and Jesús Vega Villa for their useful comments and suggestions (JRC). The authors would like to express their recognition to the experts participating in the validation workshop of the JRC Flagship Report on Artificial Intelligence in Brussels on October 11th, 2018, in particular Prof. Barry O'Sullivan (University of Cork), Dr. Cosmina Dorobantu (Alan Turing Institute) and Prof. Yves Poullet (University of Namur), with whom some preliminary findings were discussed. Finally, thorough checking and editing of the text by Paula Galnares Rodríguez is gratefully acknowledged.

Authors

Samoili, Sofia

Righi, Riccardo

Cardona Melisande.

López-Cobo, Montserrat

Vázquez-Prada Baillet, Miguel

De Prato Giuditta

Abstract

This report analyses and compares countries and regions in the evolving international industrial and research landscape of Artificial Intelligence (AI). The evidence presented is based on a unique database covering the years 2009-2018. The database has been specifically built from a multitude of sources to provide scientific evidence and monitor the AI landscape worldwide. Companies, universities, research institutes and governmental authorities with an active role in AI are identified and analysed in an aggregated fashion.

The report presents a wide variety of indicators, allowing us to expand our knowledge on issues such as: the size of the AI ecosystem globally and at country level; which are the main global competitors of the EU; what is the level of industrial involvement per country; what are the firms' demographics, profiling of economic agents according to their strengths in innovation and take-up of AI, including their patenting performance; and the degree of internal and external collaborations between EU and non-EU firms and research institutions. The analysis of the AI activities developed by agents in the studied territories provides interesting insights on their areas of specialisation, highlighting the strengths of the EU and its Member States in the global landscape. Each section offers a focus on EU Member States.

Executive summary

This report analyses and compares countries and regions in the evolving international industrial and research landscape of Artificial Intelligence (AI). The evidence presented is based on a **unique database** covering the years **2009-2018**. The database has been specifically built from a multitude of sources to provide scientific evidence and monitor the AI landscape worldwide. The sources include AI specific business platforms and scientific conferences, and firm, patent and research data sources, from which AI-relevant activities are extracted. Companies, universities, research institutes and governmental authorities with an active role in AI are identified and analysed.

From the collected information, making use of topic modelling and semantic analysis **six different AI thematic areas are identified** within the AI domain. These are Computer vision, Connected and Automated vehicles (CAVs), Natural language processing (NLP), Robotics and Automation, AI Services, representing the commercial and application side of AI, and Machine learning representing the underlying theoretical work in AI.

In line with other studies, the **main geographic areas** in the AI technological landscape identified in our analysis during the last decade are the **US, China** and the **EU28**. The analysis undertaken in this report uncovers very different specialisations for these three geographic areas.

The **US** is the **leader** in the absolute **number of AI players** worldwide. This is strongly driven by the **very high number of firms** involved in AI processes. AI firms in the United States also show a wide spectrum of different characteristics. Compared to other countries, **US AI firms are more evenly distributed** among **younger** and **more established** companies, and **among** different **industrial sectors**. Additionally, apart from being home to some AI developing giants, the US also has large number of firms providing AI services without patenting. The US has a very high output per patenting firm, but a relatively small amount of firms patenting compared to China. Of all AI activities in the US, the country is **specialised** in the areas of **AI Services** and in **Robotics and Automation**, and **has a leading or strong presence in all the identified AI thematic areas** (first or second rank in every thematic area) Robotics and Automation.

China follows the US with the second largest number of total AI players. However, China has by far the **largest presence of governmental institutions** and the **highest number of research institutions in AI worldwide**. It also ranks first regarding **number of players** filing **patent applications**, though these produce less patents on average. China's specialisation reflects its **leadership** in **Computer vision**, as well as in **Machine learning** and **Connected and Automated vehicles**, hosting by far the highest number of activities worldwide in these thematic areas. Similar to the US, China is an important competitor for all of the remaining thematic areas as well, but comparatively less in AI Services. The leadership in the majority of thematic areas, in contrast to the US, is mainly carried by the research and governmental institutions. China dominates the activities worldwide performed by governmental institutions. Nearly 90% of all government institutions involved in AI worldwide correspond to players located in China. **Chinese AI firms** tend to be very large and over half of them belong to the **manufacturing sector**, in line with the traditional role of this country in Information and Communication Technologies manufacturing. China has a large quantity of firms that **contribute to the technological development of AI without having a core business focused on AI** (e.g. a petroleum corporation patenting in AI).

The **EU** also takes a prominent role in the global AI landscape, with a strong research area, having the **second highest number of research institutions** active in AI worldwide and the **highest number of research institutions with AI publications**. However, the individual propensity of each institution to develop patents or publications is modest. In addition, the EU has a comparatively **small percentage of AI firms filing patent applications**, especially compared to China, and those patenting display a low average number of patents, especially compared to the US. Hence, the large majority of EU AI firms have a core business that is AI-related, but are **not involved in patenting technological developments of AI**. European **AI firms are comparatively young** compared to those of the US and China and are often active in the "Information and Communication" or the "Professional, Scientific and Technical Activities" sectors. The areas of specialisation of the EU are **Robotics and Automation** and **AI Services**. In both of these thematic areas, the EU has a **strong revealed comparative advantage** (RCA), and takes the second position worldwide in number of AI activities. The EU's strong position in AI is represented by both the industrial and the research worlds, but it has a stronger foot in the research output. Another two thematic areas where the EU holds the third position worldwide in number of activities are Machine learning and Connected and Automated vehicles, the latter mainly driven by Germany, UK and Sweden.

Within Europe, the three countries with the largest number of AI-related players are the **United Kingdom[1], Germany and France**. After these, there is a second group of countries consisting of the Netherlands, Spain, Italy and Sweden. Estonia takes on a specific role with a very high ratio of AI players to GDP and very young and ICT-related firm demographics. Poland distinguishes itself with a high percentage of manufacturing and large firms involved in AI compared to other European countries. The UK accounts for about one third of the overall number of AI firms detected in the Union and for half of the EU's "Big Tech" companies. In addition, United Kingdom holds the first place regarding activities in every thematic area, apart from the Connected and automated vehicles, which is led by Germany.

Another strong country in the AI landscape is **India**, which ranks **fourth in number of players** and shows a **revealed comparative advantage in AI Services** and in **Robotics and Automation**, and with a very large percentage of young firms active in AI. **South Korea** and **Japan** (fifth and seventh rank in number of players) have very similar specialisation profiles, focused on **Natural language processing**, **Computer vision**, and **Connected and Automated vehicles**. Moreover, they have a similar distribution of firms across industrial sectors, with a comparatively high percentage in the **manufacturing sector**. This together with China corroborates Asia's strong position in AI manufacturing. South Korea has the highest value in the RCA indicator in Natural language processing among the top 10 countries worldwide, with a value that triples the world average level of specialisation in this thematic area. Indeed, South Korean research institutions present a large share in the technological subdomain of Natural language processing. Japan has a comparatively low presence of AI players in the portrayed landscape. Nevertheless, the patenting performance of the Japanese players is the highest: 3.7 patents per patenting firm. Furthermore noticeable is the largest percentage of AI firms in the oldest age group compared to other countries (50% of Japanese AI firms were founded at least 20 years ago).

Other important countries in AI are **Canada** (sixth by number of players), **Israel** (eighth rank) and **Russia** (ninth rank). Israel stands out with a very high AI intensity compared to the size of its economy. Canada has a good presence in AI compared to its size, while Russia lags behind in AI players in relation to the size of its economy. Canada appears strong in the areas of AI Services and in Robotics and shows a similar firm structure and specialisation profile to the US, while Israel's specialisation profile is comparable to the EU's (Robotics and Automation, and AI Services being the most important). Russia has a strong specialisation in Connected and Automated vehicles.

At **regional level**, the leadership of the two giants persists: out of the top-30 regions with highest activity level in AI, 15 are Chinese and 7 are from the US. California and Beijing are the top worldwide regions in AI in almost all thematic areas, along with Jiangsu and Guangdong.

Among EU regions, the big urban areas dominate. The most important regions by number of AI players are **Inner London-West** (UK), **Île de France** (FR), and **Berlin** (DE). Together they accumulate more than 25% of all EU AI players. Germany is the only country to have two regions, Berlin and **Oberbayern**, among the top five EU regions. The most influential region in the entire EU network of R&D collaborations is Île de France. This region leads based on the amount of activities in the Natural language processing, while Inner London-West leads in Machine learning, Robotics and Automation, and AI Services. Second in terms of capacity to influence other regions is Oberbayern and it has a relative specialisation in Connected and Automated vehicles. Other relevant AI regions in the EU are **Cataluña** (ES), **Noord-Holland** (NL), **Comunidad de Madrid** (ES) and **Berkshire, Buckinghamshire and Oxfordshire** (UK).

1 This study covers the period 2009-2018, when the United Kingdom was still a Member State of the European Union.

1 Introduction

The field of Artificial intelligence (AI) is experiencing a period of intense progress, due to the consolidation of several key technological enablers such as increased processing power, the growth of data availability and improved algorithms. Past industrial revolutions have had a profound impact on the economy with wide disruptions in society. Internationally, the industrial revolutions have transferred economic and technological power to the more advanced nations and regions. AI could be a similar game changer. Consequently AI has become an area of strategic importance and been identified as a potential key driver of economic development as underlined in the European strategy on AI (COM (2018) 237 on Artificial Intelligence for Europe) and in the related Coordinated Plan (COM(2018)795). Similarly, AI has become a clear target for national governments resulting in the formulation of national AI strategies.[2]

In view of the above, policy makers need quantitative evidence to monitor progress and impact of such policies on a regular basis. While data on different industrial sectors (e.g. ICT) and its R&D are available from various international and national sources, they are often incomplete and not comparable across countries. Moreover they do not reflect the wide spectrum of AI activities in diverse technological and scientific domains and different economic sectors.

This report offers insights into AI by monitoring the evolving international industrial and research landscape. The application of the Techno-economic segments (TES) analytical approach developed by the EC JRC maps the segment, monitoring and benchmarking its main players, their locations and R&D and industrial activities. By applying a wide range of methodological tools it sheds light on the dynamics of the worldwide networks of stakeholders and technological subdomains, their gatekeepers and expected future champions.

For this purpose a unique database has been built from a wide variety of sources, exploiting different types of factual data including e.g. patents, conference proceedings, research projects, segment specific firm repositories, business and funding platforms. A broad and externally validated keyword list extracts the relevant AI data to portray the wide array of the AI world. Machine learning based textual analysis identifies different thematic areas and network analysis allows to understand the relative importance of players in the ecosystem and the network's resilience over time. In summary, the TES analytical approach is conceived as an analytical framework and replicable methodology to analyse and describe the dynamics of a specific TES ecosystem, mapping, monitoring and benchmarking such segment with suitable unbiased scientific evidence.

The EU is, with China and the United States an important actor in AI research. The United States is clearly at the forefront of AI related activity in most industrial areas, hosting the biggest amount of firms getting venture capital or most start-ups active in the field and China is catching up rapidly (Perrault et al., 2019). The analysis presented in this report draws a similar picture, going into more detail at the regional level. The report provides results on multiple dimensions, differentiating between different research outputs, and studying collaborations in networks. It shows diverse business structures of AI companies and different thematic profiles in the analysed countries. Furthermore, every section includes a focus on EU Member States. This report highlights some indicators based on the extensive data collected, for further visual presentation of the data a dashboard is accessible in the AI WATCH portal.

The structure of this report is as follows: section 2 explains the applied methodology and data collection. Section 3 gives an overview of the AI landscape as captured by the TES approach by presenting indicators of AI presence in countries and regions. Section 4 looks into the different AI thematic specialisations, analysing country profiles and identifying thematic hotspots. Section 5 gives a more detailed picture of the AI industrial landscape worldwide and in the EU28, at country level, and section 6 focuses on the research world and presents the network analysis of collaborations. Section 7 includes concluding remarks.

2 E.g. Executive Order on Maintaining American Leadership in Artificial Intelligence (White House 2019), Villani Report for France (Villani mission, 2018), Strategie Künstliche Intelligenz der Bundesregierung (BMBF, 2018).

2 Methodology and objectives

2.1 Objectives and units of analysis

The general aim of this study is to map a techno-economic segment from a multidimensional perspective, providing an overview of the worldwide AI landscape in the last decade. The unit of the analysis is the economic agent, or hereinafter player, that can be a company, university, research institution or governmental authority. The player is expected to have an active role in the segment, with the capability to influence its economic development and future evolution. In this sense, the focus is set on the organisations, and not on individuals, namely the applicant organisation owning the invention in the case of patents, authors' affiliation in conference proceedings, companies, governmental entities, etc. To establish a comprehensive landscape, we target both industrial and R&D activities. This helps to capture economic agents that participate in the landscape with a variety of foci, interests and impact capacity. Therefore, players' economic activities of interest for the analysis of the TES ecosystem include R&D processes (research and innovative developments), general economic processes (industrial production, trade, marketing and other services), firms funding (venture capital funds or other types of investment).

The specific objectives of the study are the identification and analysis of: i) the players involved in the AI technological field globally, ii) the thematic type of R&D and industrial activities in the field, and iii) the players' interactions within the techno-economic space. The first objective is fulfilled by mapping the location of the AI players involved in R&D and industrial activities, in several geographic levels (geographic macro areas, country, region, sub-region). The second objective is addressed by the exploration and identification of technological subdomains that are related to the AI field, and the subsequent assessment of countries' and regions' specialisations and performance. This is achieved through natural language methods, semantic analysis and connection to the players' R&D and industrial activities information. The third objective is accomplished through the detection and evaluation of the connections between R&D players, and the analysis of potential cooperation in the AI landscape. In this scope we develop geo-based networks and employ network analytics to determine international and regional synergies. All the aforementioned objectives are in line with points and objectives set in the European strategy on AI[3] and its Coordinated Plan[4], and the Declaration of Cooperation on AI (2018)[5].

2.2 Methodological outline and data overview

Emerging technologies, as **AI, evolve across several technological knowledge subdomains**. With a semantic analysis performed on the textual part of the agents' activities, and through the agents' related geographical allocation, we: (i) detect the knowledge subdomains of an emerging technology (subsection 4.1), and (ii) identify the countries' specialisations and quantify the involvement of countries' activities to the detected subdomains (subsection 4.2). The procedure is explained subsequently and illustrated in Figure 1.

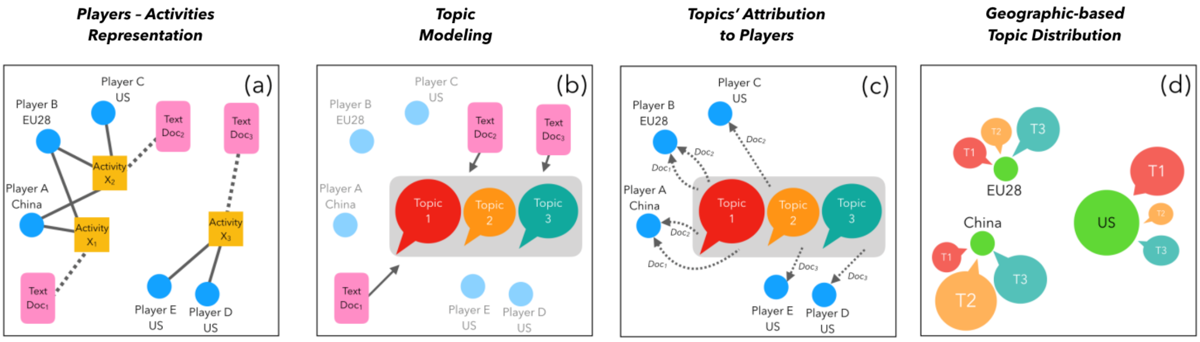
The conjecture is that players in the same or related knowledge subdomains mutually use a set of technical terms, originated by their activities (Figure 1 (a)). This use of a **common language** is used **to map the thematic subdomains** of an emerging, hence uncharted, technological territory. The analysis of these subdomains, or else thematic key areas, can indicate **knowledge trajectories**. The terms of the player's activities are clustered in thematic groups, referred to as topics (Figure 1 (b)), which are **proxies for thematic key areas** of the technological domain under study (see Annex 4 for a more formal explanation of the topic modelling). Topics are essentially overviews of the most pertinent themes of a large collection of documents (patents, publications, firms' descriptions). They are used to rapidly communicate the subjects on which the players are involved (Figure 1 (c)), and thus the thematic key areas of strength or specialisation of the associated geographical areas (Figure 1 (d)). Players grouped into the same thematic key area do not necessarily collaborate in common activities.

3 Explicitly for EU in COM (2018) 237 it is mentioned that "the public and private sectors must seize the opportunities that come both from developing innovative AI solution and applying them to a range of fields. Without efforts, the EU risks losing out on the opportunities offered by AI, facing a brain drain and being a consumer of solutions developed elsewhere".

4 In COM (2018) 795 one of the objectives mentioned is to "foster cooperation among the best research teams in Europe".

5 In the Declaration of Cooperation on AI (2018) the participating MS agreed among other points to cooperate on "boosting Europe's technology and industrial capacity in AI", and to "contribute to the establishment of a dense network of Digital Innovation Hubs at European level", so as to ensure the competitiveness of EU in AI.

Figure 1. Schematic explanation of the process of identification and allocation of key areas to players and geographic entities, based on the common technological language found in activities.

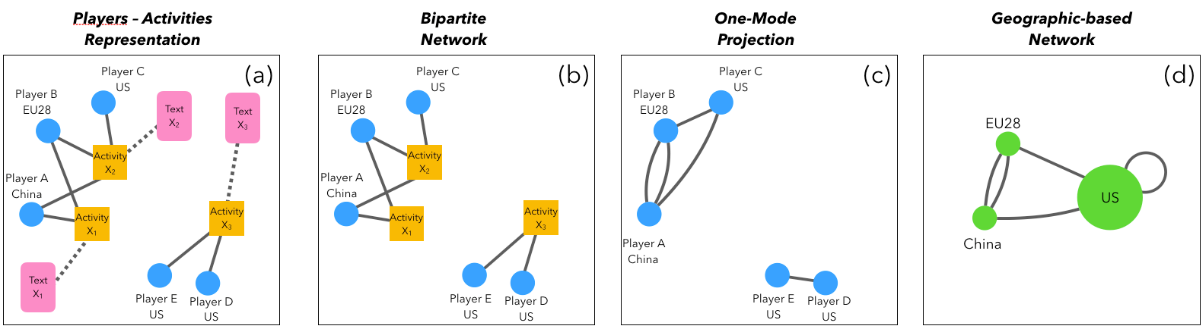


Note: AI-related activities are depicted as X1,X2,X3, their textual information as documents Doc1,Doc2,Doc3, and players as Player A, B, C, D, E, etc. In (a), the AI ecosystem is illustrated with the connections of the activities to the players and their geographic location. In (b), the thematic topics are extracted from the textual information of the activities through topic modelling. In (c), the activities are connected again to the players carrying the thematic information. In (d), the thematic profile of each country is presented.

Source: Authors' elaboration.

In order to investigate potential innovation dynamics we analyse the connective pattern that players structure while developing technological collaborations in the form of scientific publications or patent applications. According to existing literature on socio-economic adaptive complex systems, innovations are unleashed when agents interact and exchange information (Lane, Pumain, der Leeuw & West 2009). In the present report, instead of focusing the attention on the role of the individual players, we investigate the potential of geographic areas (according to the selected granularity level) to be influential information hubs. The bipartite network represented in Figure 2(b) and made of activities (in yellow) and their involved players (in blue) is therefore transformed in the corresponding geo-based network (Figure 2(d)). This allows the quantification of the collaborations internally and externally developed by the different areas, as well as the computation of centrality measures.

Figure 2. Methodological steps for the generation of a geo-based network aimed at investigating collaborations in relevant AI R&D activities.



Note: The initial landscape, made of players (blue nodes), activities (yellow nodes) and textual information (pink nodes) is represented in (a). In (b), the bipartite network consists of two types of nodes: players and activities. Players are connected to the activities in which they participate. In (c), the one-mode projection based on players is generated by connecting players who participate in the same activity. In case of multiple common activities, multiple edges are generated. Finally, in (d) the geographic-based network is built by merging together nodes belonging to the same geographical area. In case of collaborations involving players of the same area, self-loops are generated.

Source: Authors' elaboration.

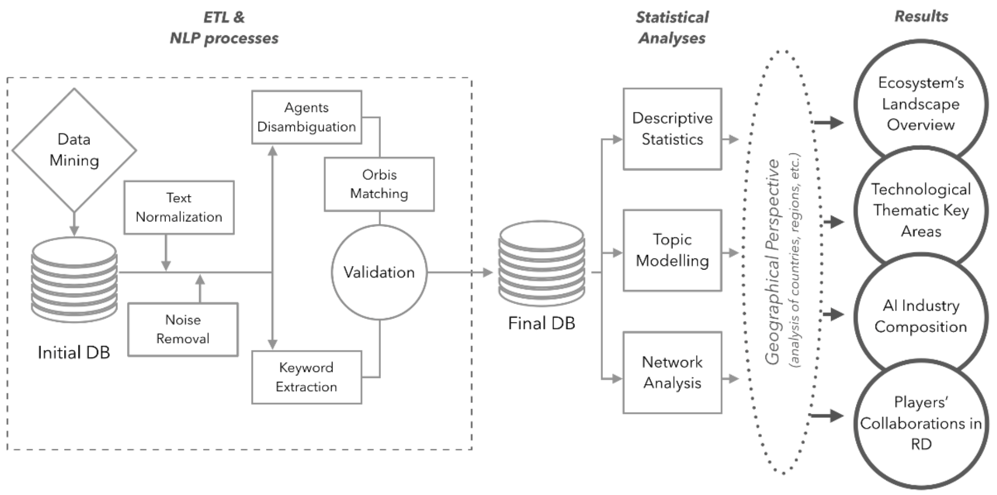
To target the AI TES ecosystem and collect representative data about AI-related R&D and industrial activities, we target a list of vertical sources (AI-specific sources), and a list of horizontal databases (general, not-AI specific sources) containing information regarding any technology. The latter are queried in order to identify those activities relevant to the selected technological domain, AI and their corresponding players. The sources that are investigated and collected are:

1. Vertical sources (AI specific sources):
    1.1. For production and trade (companies' repositories including startups, firms financed by venture capital, etc.):
        1.1.1. AI startups from AngelList, a platform joining about 27`500 companies and startups aiming at facilitating contacts with angel investors and job-seekers (angel.co/artificial-intelligence)
        1.1.2. The Global Artificial Intelligence Landscape by Asgard and Roland Berger 2018, a set of data and information on the AI segment collected for investment purpose and targeting mainly software startups (asgard.vc/global-ai//)
        1.1.3. CB Insights, a platform collecting data on companies, venture capital, startups, etc., in new or dynamic markets (cbinsights.com)
        1.1.4. Other sources: AAI Job Fair, AI & Big Data Expo Global 2018, Allianz Global Artificial Intelligence, AI Breakthrough Awards 2018
    1.2. For research activity:
        1.2.1. Frontier research: the articles and authorship information are collected from the top 10 AI international conferences. For the list and topics of the conferences see Annex 5. We consider fractional count to avoid double counting of publications submitted by several authors from different affiliations[6].
2. Horizontal sources (general sources):
    2.1. For innovation activity:
        2.1.1. EPO PATSTAT, a widely recognised database containing bibliographical data relating to more than 100 million patent documents from leading industrialised and developing countries (epo.org). The documents used in the study refer to priority patent applications. We consider fractional count to avoid double counting of patent applications submitted by several applicants.
    2.2. For research activity:
        2.2.1. CORDIS data repository collecting information on EU funded projects, from which projects from FP7 and H2020 have been selected (cordis.europa.eu/projects). This source is only considered for a deeper analysis in the European comparison, but is discarded in all worldwide comparisons to avoid a Eurocentric bias in this analysis. As with patent applications, fractional counting is applied to the EU-funded projects.
    2.3. For production and trade:
        2.3.1. BvD Orbis, the Bureau van Dijk (now a Moody's Analytics company) database collecting financial strength indicators, company reports and ownership information about around 300 million companies worldwide (bvdinfo.com)
        2.3.2. Crunchbase, a platform listing business and investment information about private and public companies (crunchbase.com)
        2.3.3. Venturesource by Dow Jones, a comprehensive global database on companies backed by venture capital and private equity, including information on venture capital transactions (dowjones.com/products/ venturesource-2)
        2.3.4. VentureRadar, a data collection including information and interpreting innovation and growth signals of companies discovered by means of their digital presence (ventureradar.com)

6 Because the unit of the analysis is the institutional economic agent, the focus is on the affiliation, not on the author.

The selected sources cover a wide range of types of economic activities and geographic areas. However, the fact that EU-funded projects were collected, while other national funded sources were not available, introduced an EU-centric bias. In order to avoid it in the worldwide analysis presented here, we have discarded EU-funded projects. Some minor sources have also been excluded from further consideration for the present report, as the textual part of these activities' description contained generic information and no technological details. Documents have been collected from all sources. Each document, collected from the described data sources, contains information about the activity (e.g. the description of the activity, date of the activity) to which they are associated and the player/s involved in it. In the initial part of the ETL (Extract, transform, load) process, database queries with terms relevant to the technology are performed to acquire activities and metadata of identified sources. The text-mining part is conducted with a scalable search for the technology in question, through a search engine based on the Lucene library (Elasticsearch engine version 2.4.6), and then results are evaluated based on a series of statistical indices. The objective is to detect and select the most representative R&D and industrial activities of an emerging technology. Based on the collected documents, players are identified in the first stage ("Data Mining" and "Initial DB" in Figure 3).

Figure 3. Workflow of TES



Note: The depicted workflow presents the main points of the analytical approach, from left to right. The initial phase of data mining is integrated with a series of procedures defining the data pre-process. Once the final database is structured, we implement statistical methods to orient the analyses towards specific types of outcome. We take a geographical perspective to analyse the ecosystem at different granularity levels, from macro areas to sub-regions.

Source: Authors' elaboration.

For the **textual information** of the constructed collection, natural language processing (NLP) processes are then employed to obtain information about the activities (R&D and industrial) and players. A text normalisation is firstly applied to reduce textual morphological variation. This step includes removal of diacritical marks, single case conversion, non-alphanumeric characters conversion, canonization (American English spelling). An unsupervised keyword extraction algorithm is integrated, in order to improve the subsequent identification of thematic key areas in the entire technological knowledge domain. To that end, a pre-trained language agnostic model (Straka, M., Straková, J., 2017) is used to annotate the text corpus, which includes text tokenization, parts of speech tagging (POS) and dependency parsing. Hence, the resulted semantic patterns can be used for the keyword extraction. The output is a concise corpus of n-grams with improved semantic relatedness in comparison to the standard TF-IDF index (Wartena et al, 2010), to which a limited noise removal is implemented. The removed number of sparse terms remains constrained, so as to avoid elimination of less frequent terms describing newly emerged or very specialized processes, and moreover creation of empty documents that would reduce the value of the formed technological knowledge representation corpus. Similarly, the removal of stop words is limited, in order not to remove potentially meaningful terms.

We also perform a phase of **disambiguation of agents** and their geographic information. The metadata of the aforementioned data are used, so as to identify the economic players involved in each activity. Initially, the identity of the players is defined based on information about their name and location. Following the name normalization, geographical coordinates are retrieved based on the data sources' available information about players' location. In the cases that location information is missing, it is inferred by a multi-step algorithm involving scraping of player's information from the web. Players with same names, but different geographic information, as e.g. multinational enterprises with several locations, are stored as distinct players. When all players' identities are defined, another algorithm is run to eliminate duplicated players. This algorithm detects duplicated players and merges them into a unique one, based on (i) the similarity of the normalized names, and (ii) their geographic proximity. This process is performed within each data source, and then for all data sources, so as to identify the presence of duplicated agents throughout the collected data.

Finally, we implement a manual validation of the data mining and pre-processing phase. Based on the described steps, the final database is created ("Final DB" in Figure 3).

To summarise, the main characteristics of the dataset are:

- information is collected at a **micro-level for players and activities**, meaning that the final database is made up of detailed **information about the participation of each detected player in the activities in which it is involved;**
- **each player's location is collected in form of geographic coordinates** (latitude and longitude), and this implies that all levels of geographical agglomeration may be considered for the analysis (e.g. country level, regional level, hotspots);
- **textual information** about activities allows the exploration of the **technological and thematic content of activities and players**, meaning that thanks to implemented techniques a **qualitative dimension of the collected documents is analysed in quantitative terms**;
- information about **co-participation of players in same activities** allows the definition of a **network of collaborations**, and based on this information the investigation of the **structure of connections in the considered system** is developed.

Thanks to the richness of the database, we can also analyse another relevant dimension of the ecosystem: the players' organisational type[7]. This allows the distinction between firms, research institutes, and governmental institutions. Regarding the category "governmental institutions", all the institutions owned by the state or with public administrative functions are included. In the category "research institutes", all players mainly devoted to research activity are encompassed, i.e. private research centres, public research centres, universities[8], university/academic spin-offs, and industrial research centres exclusively dedicated to research activities. This enables further analysis of the relationships between research and industry, research and government and industry and government in different geographic areas, and allows the assessment of different properties of the whole ecosystem and its local parts.

The information finally considered for the analyses of the AI worldwide ecosystem from 2009 to 2018 presented in this report contains 57,722 documents by means of which 34,009 players are identified.

As presented in the right part of in Figure 3 ("Statistical Analyses" and "Results"), for the work presented in this report the final database is used to develop four analyses: (i) descriptive statistics allowing an overview of the ecosystem's landscape (see section 3); (ii) topic modelling techniques allowing the study of thematic and technological key areas (see section 4); (iii) in-depth analysis of industrial players (see section 5); and (iv) network analysis allowing the investigation of players' R&D collaborations (see section 6).

7 This information is collected from the description of the companies' activities in business registers, or derived from information present in some of the activities to which they are associated, e.g. from patents, conference proceedings and research projects.

8 Departments of a same university are not considered as separate players.

2.3 Weights applied for the industry analysis

The analysis addressing the industrial activities requires another step in the data preparation before obtaining the final database: firm players detected in AI TES are matched with the Bureau van Dijk Orbis database in order to access company level information, e.g. on the age, size and industrial activity sector of the companies. Bureau van Dijk compiles this Orbis firm level data from administrative data. An advantage of this database compared to its competitors is the inclusion of data of companies that are non-listed in the main stock market, and therefore this data provide a better representation of smaller companies. For Europe, the database covers around 75-80% of the economy compared to Eurostat figures and matches the official firm size distribution (Kalemli-Ozcanet al., 2015). A drawback is that for countries with fewer regulatory requirements the data is less representative. Nevertheless, the database is widely used in academic literature and regarded as the most reliable source to study questions from productivity and innovation to knowledge spillovers worldwide.

After the process of data collection (in which Orbis is also one of the sources) and players' disambiguation, the players resulting to be active in the AI TES 2009-2018 landscape are then searched in Orbis. We have found an Orbis entry for 44% of AI firms from all data sources. In order to carry out an industrial analysis that is representative of the entire landscape, firms with an Orbis profile are weighted to correct for the over or under representation that certain firm groups may have in this data source. These weights, which are elaborated by accounting for specific known totals regarding the whole set of AI firms detected by TES (which are supposed to extensively cover the firms active in AI) allow us to draw conclusions from the economic variables that are present in the Orbis dataset for the whole set of AI TES firms. This procedure, typical of survey analysis, is here implemented in order to fill the informational gap determined by the fact that not all the firms have been detected in the Orbis database.

In this case, the calibration algorithm reweights the observations based on the combination of two variables. The first one, number of firms per geographic area, permits to account for possible biases in Orbis due to the geographic location of the firms. The second one, patenting class or number of patent applications filed, is here intended as a proxy of firms' performance and therefore permits to account for possible biases in Orbis regarding the relevance of the firms. The distribution of firms by patenting class and macro area is presented in Table 1. The left side of the table shows the distribution of all firms found in the AI TES dataset, while the right side shows the distribution of firms as found in the Orbis database. China, and to a lesser extent India and the US, are underrepresented in Orbis compared to the AI TES database, while the EU, South Korea and Japan are over represented. All the results presented in section 5 are calibrated based on the weights calculated from these two variables.

Table 1. Distribution of all AI TES firms and subset matched with ORBIS by patenting class and macro area, 2009-2018

	All AI TES firms					Firms matched with Orbis				
	Number of Patent					Number of Patent				
Area	0	1	2-5	6+	**Sum**	0	1	2-5	6+	**Sum**
Africa	0.75%	0.20%	0.03%	0.00%	**0.98%**	0.79%	0.14%	0.02%	0.00%	**0.95%**
Canada	2.75%	0.22%	0.04%	0.03%	**3.05%**	2.79%	0.13%	0.06%	0.06%	**3.04%**
China	4.68%	14.27%	3.46%	0.64%	**23.05%**	4.44%	5.28%	1.89%	0.60%	**12.21%**
EU28	16.33%	1.69%	0.38%	0.15%	**18.55%**	24.73%	1.72%	0.61%	0.29%	**27.36%**
India	4.12%	0.29%	0.03%	0.01%	**4.44%**	2.85%	0.10%	0.01%	0.02%	**2.97%**
Japan	1.21%	0.67%	0.31%	0.23%	**2.41%**	1.81%	0.72%	0.54%	0.38%	**3.45%**
Middle East	2.69%	0.18%	0.05%	0.01%	**2.93%**	3.84%	0.09%	0.04%	0.02%	**4.00%**
Oceania	1.28%	0.20%	0.03%	0.01%	**1.52%**	1.60%	0.24%	0.05%	0.02%	**1.91%**
Other American countries	1.65%	0.25%	0.06%	0.01%	**1.97%**	1.94%	0.15%	0.12%	0.03%	**2.24%**
Other Asian countries	2.17%	0.68%	0.31%	0.12%	**3.29%**	3.68%	0.89%	0.61%	0.27%	**5.44%**
Other European countries	1.90%	0.71%	0.12%	0.03%	**2.77%**	2.78%	0.28%	0.16%	0.04%	**3.26%**
South Korea	0.59%	1.71%	0.41%	0.16%	**2.87%**	0.87%	2.66%	0.77%	0.29%	**4.59%**
US	27.47%	3.42%	0.86%	0.42%	**32.17%**	24.96%	2.02%	0.94%	0.66%	**28.58%**
Sum	**67.59%**	**24.49%**	**6.11%**	**1.82%**	**100.00%**	**77.07%**	**14.43%**	**5.82%**	**2.68%**	**100%**

Note: Green cells identify overrepresented firm groups in the matched database, red cells identify underrepresented groups.

Source: JRC PREDICT- AI TES Dataset, 2019.

3 Competitive international landscape overview

In the present section, we offer some indicators addressing the international landscape of the AI ecosystem in the perspective of the objectives set by the EC communications in 2018 (European Strategy on AI, Coordinated Plan on AI), and the Declaration of Cooperation on AI signed on 10 April 2018 by 24 Member States (MS) and Norway for the EU competitiveness and (industrial) strength, and for R&D development in the domain of AI. In particular, the European Strategy on AI highlights "The EU can lead the way in developing and using AI for good and for all, building on its value and strengths [...], building on Europe's scientific and industrial strengths, as well as on its innovative start-ups to be in a leading position". The countries that signed the Declaration of Cooperation agreed to "work towards a comprehensive and integrated European approach on AI to increase the EU's competitiveness, attractiveness and excellence in R&D in AI, and where needed review and modernise national policies to ensure that the opportunities arising from AI are seized and the emerging challenges are addressed". Emphasising the importance of AI for the future competitiveness of the EU, the Coordinated Plan on AI foresaw that "building on Europe's strengths, it will aim at the development of a dynamic EU-wide AI innovation ecosystem, fostering close cooperation between industry and academia, and reinforcing competitiveness across the whole AI value chain".

3.1 Worldwide overview

The first subsection explores the AI worldwide landscape in the period 2009-2018. The size of the countries' economies is also considered for comparative reasons. We also analyse the types of institutions (firms, research institutes, universities, governmental institutions, laboratories, etc., involved in AI-related processes) and their patenting performance.

Note on location of AI players

In this subsection, we focus on country level to have a comparative worldwide landscape overview. The number of players detected in each country is used as an **indicator of involvement of a country in AI-related economic processes**. The following explanations on how the indicator is computed will assist in its interpretation. In this report, the players are considered as equally relevant from an economic point of view, irrespectively from their organisation type, and without considering a quantitative measurement of their economic activity. Moreover, players belonging to the same organisation but based in different locations are currently considered as different players (e.g. *JRC-Petten* and *JRC-Bruxelles* are currently two different players; similarly, branches of the IBM located in different countries or cities are also considered as different players). An exception to this rule is made for university departments, not considered as individual players, being considered the university a single player. The agglomeration of players following their ownership structure is an extension expected to be addressed among the future developments of the project.

3.1.1 *Countries' performance: absolute and relative indicators of AI presence*

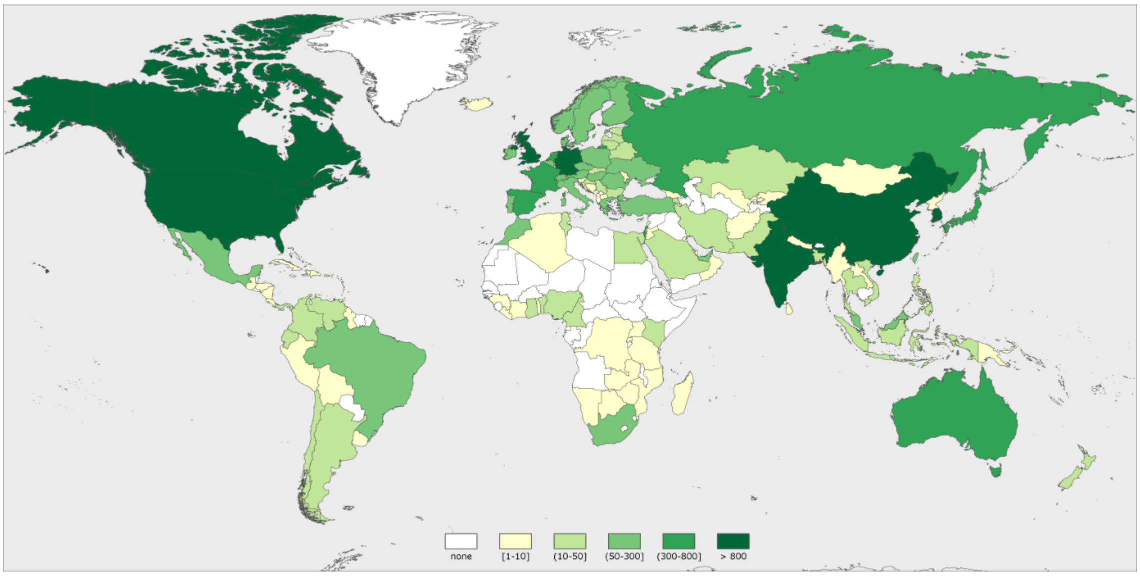
The mapping of the worldwide players involved in AI-related economic processes in the last decade (from 2009 to 2018), aggregated at country level, reveals mainly the **leading presence of US and China**, corroborated by several other studies and national policies' overviews (WIPO, 2019; Parliamentary Mission (Villani Mission), 2018; Steering Group of the Artificial Intelligence Programme, 2017; McKinsey, 2017a,b). To identify the degree of involvement of EU Member States in the AI worldwide landscape, they are separately represented in Figure 4. This allows the identification of prominent European countries, such as the United Kingdom[9], Germany and France, which are competitive at a global and European level. At a worldwide level, more countries are highly active apart from the US, China and the EU28. These are India, South Korea, Canada and Japan, with the latter mainly due to activities in the beginning of the study period[10]. Moreover, Israel, the Russian Federation, Singapore and Australia hold key positions in the AI ecosystem, hosting a large number of players involved in AI techno-economic processes (WIPO, 2019, Parliamentary Mission (Villani Mission), 2018).

When the EU28 is considered as a single entity, it ranks among the leading relevant areas. In Figure 5 the number of AI players is illustrated by blue bars. To assess the impact that EU-funded projects (FP7 and H2020) have on the presence of EU28 players in the landscape, the number of AI players participating in AI activities exclusively because of their involvement in FP7 and H2020 projects is shown in grey. The US, China

9 During the period covered by this analyses, 2009-2018, the United Kingdom is part of the European Union.

10 WIPO mentions that Japan had a strong AI sector based on the number of patents filings, which stagnated after 2000 (WIPO, 2019).

Figure 4. Worldwide distribution of players active in Artificial Intelligence, country level, 2009-2018



Source: JRC PREDICT- AI TES Dataset, 2019.

and the EU28 differ significantly compared to other worldwide areas (at least four times larger). The fact that the grey bar accounts for one third of EU28 AI players demonstrates the importance of the **contribution of EU research programmes in promoting the AI segment**. However, as AI activities supported by national policies are not available for the full worldwide landscape, the EU players considered in the rest of the report are those with at least one AI activity different to EU funded projects[11].

To evaluate the countries' performance, the number of players (bars) is compared with the number of players over GDP (circles). **The ratio** $r_i = \frac{N.\ of\ Players_i}{GDP_i}$, where *i* is the considered country, is used as an indicator to compare countries worldwide in terms of **relative presence of players**. Assuming that the number of players is correlated with the size of a country´s economy, this indicator measures if the number of AI players located in a country is higher or lower than expected considering the size of the corresponding economy (as measured by the GDP).

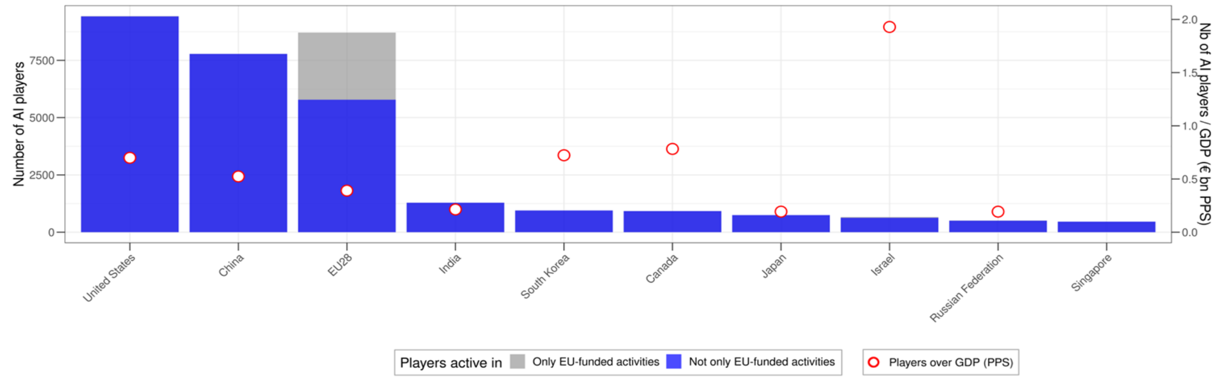
The worldwide AI landscape is mainly led by the **United States, China and EU28**. India and South Korea follow with significantly lower number of players. Considering the ratio **number of player divided by GDP-PPS**[12], Israel emerges with a considerable role in AI, a prominent position also reported by other studies (Statista, 2017; China Institute for Science and Technology Policy at Tsinghua University, 2018; WIPO, 2019). Moreover, **two groups of countries** are identified among the top-10 areas per number of AI players: one with a **medium ratio** (Canada, South Korea, US, China and EU28), and another with a **relatively low ratio** (India, Japan and Russian Federation). The low ratio can be explained by the delay of AI uptake in their national economies, with the exception of Japan, whose modest position is caused by the stagnation of the sector during the last two decades (WIPO, 2019). The analysis of the high-medium group suggests an **intense development of the AI national segment**. Hence, US, China and EU28, given also their number of AI players, appear as leaders in the AI segment. The position of EU28, namely the lowest of the considered group, could be partially explained by the complexity and heterogeneity of its economic and political structure. Canada and South Korea's presence in the top-10 is expected. Canada is the sixth area in number of AI enterprises after US, ChinaEU28, India and South Korea, and in the top 20 countries in number of patent

11 This means that only players that exclusively participated in EU funded projects are not considered for the rest of the work.

12 To allow cross-country comparability, national currencies are converted into euro Purchasing Power Standard (PPS), a unit based on current euros, to account for the effect of differences in price levels across countries and of movements in exchange rates.

applications (China Institute for Science and Technology Policy at Tsinghua University, 2018; WIPO, 2019). South Korea's position is mainly explained by its strong involvement in technologies related to Natural language processing, Computer vision and Connected and Automated Vehicles (CAVs) (Viereckl et al., 2016))

Figure 5. Top 10 world geographic areas by number of AI players: Absolute number and Relative to GDP (€ bn PPS), 2009-2018



Note: To allow cross-country comparability, national currencies are converted into euro Purchasing Power Standard (PPS), a unit based on current euros, to account for the effect of differences in price levels across countries and of movements in exchange rates.

Source: JRC PREDICT- AI TES Dataset, 2019. (AI Players), OECD, IMF and EUROSTAT (GDP).

Table 2. Main indicators of top AI countries by number of players: players' organisation types, performance in AI patenting, and relative economic ratios, 2009-2018

Code	Country	Number of AI players by organisation type				Performance in AI patenting					Number of players relative to economic indicators							
		N. of Firms	N. of Governamental Institutions	N. of Research Intitutes	TOTAL N. of Players	N. of Patenting Players (all types)	N. of Patenting Firms	N. of Patent Applications	Average N. of Patent Applications per Patenting Player (all types)	Average N. of Patent Applications per Patenting Firm	AI players / GDP (GDP in € bn PPS)		AI firms / BERD (BERD in € m PPS)		AI firms / GERD (GERD in € m PPS)		AI players / Population (Pop. in m Persons)	
US	United States	9,040	4	332	9,376	1,341	1,294	4,380	3.3	3.2	0.70		0.03		0.02		29.41	*
CN	China	6,477	187	1,085	7,749	6,382	5,160	17,942	2.8	1.6	0.52		0.03		0.02		5.59	**
EU28	EU28	5,213	3	560	5,776	717	616	1,115	1.6	1.6	0.39		0.03		0.02		11.36	
IN	India	1,248	2	35	1,285	101	91	76	0.8	0.7	0.21		0.08		0.03		1.03	**
KR	South Korea	807	2	139	948	766	640	2,225	2.9	2.3	0.72		0.02		0.01		18.80	*
CA	Canada	856	1	62	919	84	79	174	2.1	2.1	0.78	*	0.09		0.05		25.86	*
JP	Japan	678	-	68	746	385	336	1,443	3.7	4.0	0.19		0.01		0.01		5.86	**
IL	Israel	607	-	21	628	37	37	29	0.8	0.8	1.93		0.04	*	0.04	*	77.93	**
RU	Russian Federation	366	12	122	500	313	190	356	1.1	1.0	0.19		0.02		0.01		3.48	*
SG	Singapore	422	1	36	459	83	65	237	2.9	2.5	0.00		0.00		0.00		0.00	
AU	Australia	382	-	34	416	58	55	40	0.7	0.7	0.49		0.04		0.02		17.71	*
TW	Taiwan	228	1	54	283	217	165	503	2.3	2.0	0.34		0.01		0.01		0.00	
CH	Switzerland	200	-	21	221	31	27	37	1.2	1.2	0.56		0.03	***	0.02	***	26.83	
MX	Mexico	184	-	13	197	33	22	29	0.9	0.6	0.07		0.00		0.00		1.65	*
BR	Brazil	155	-	13	168	18	14	43	2.4	2.7	0.07		0.00		0.01	*	0.83	*

Note: * Base year 2014, ** Base year 2013, *** Base year 2012. GDP: Gross domestic product; BERD: Business expenditure on R&D; GERD: Gross domestic expenditure on R&D.

Source: AI related indicators: JRC PREDICT- AI TES Dataset, 2019. Economic Indicators: PREDICT - AI TES OECD, IMF and EUROSTAT.

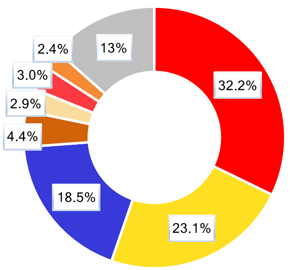
Table 2 shows a considerably **greater number of US firms than in any other country**: nearly 2,500 firms more than in China. This justifies the US leadership in the AI segment. In fact, the large number of firms involved in AI-related processes indicates that the **US economy is strongly involved in AI**, as private commercial initiatives are already mature and active in the AI market (China Institute for Science and Technology Policy at Tsinghua University, 2018; Ding, 2018). Regarding **China**, first, a **large presence of governmental institutions** suggests that the Chinese government is intensively promoting the AI economic segment through related institutions. Second, the **presence of research institutions is noticeable** (1,085): almost twice than in the EU28 (Steering Group of the Artificial Intelligence Programme, 2017; National Strategy for Artificial Intelligence #AIFORALL, 2018; China's State Council, 2017; China Institute for Science and Technology Policy at Tsinghua University, 2018). In the **EU28** we detect 560 research institutions involved in AI economic activities, exceeding by more than 200 the number of US research institutions. Even though players participating exclusively in EU funded projects are not considered, the **research area is prominent**. Finally, the presence of South Korean and the Russian research institutions is also noticeable**.**

3.1.2 ***Performance in AI patenting activity***

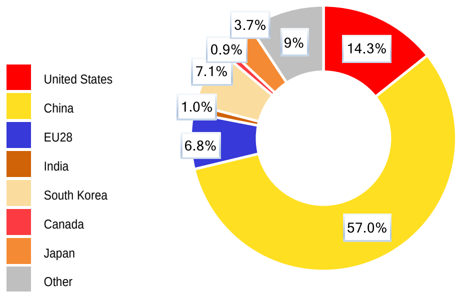
As shown in this subsection, the analysis of the organisation types of AI players reinforces the US as the most advanced country in AI. Only a relatively small part of US firms is filing patent applications (1,294 out of 9,040). However, the average number of patents they develop is very large. This performance indicator suggests that **US firms** are divided **between those commercialising AI-related goods and services**, and **those involved in the development of new technological solutions**. On the contrary, in **China**, we detect more patenting players, but with less patents on average. While in the US nearly 14% of AI **firms is developing patents**, in China this percentage reaches almost **80%**. As shown in Figure 6 (c), the **average number of patent applications filed by Chinese firms** (with at least one patent) **is 1.6 vs 3.2 for US ones**. This is a consequence of different strategies regarding the patenting activity (WIPO, 2019). Chinese tend to patent much more (17,941 patents in the last decade), but only 4% of these patents are then filed in a different jurisdiction (WIPO, 2019).

Figure 6. Overview of firms involved in AI related processes, 2009-2018

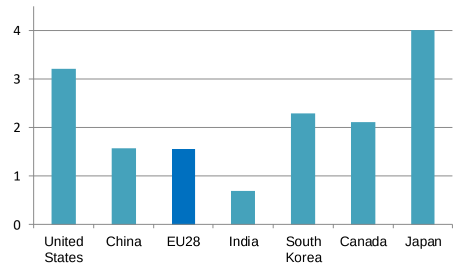
(a) AI firms by country (%)



(b) AI firms with at least one patent application filed by country (%)



(c) Average number of AI-related patent applications filed per patenting firm, by country



Source: JRC PREDICT- AI TES Dataset, 2019.

Another interesting aspect to highlight is the **productivity of Japanese players** Figure 6(c). As this report analyses the AI landscape from 2009 to 2018, Japan does not appear as a leading country. In fact, the modest presence of Japanese players in the portrayed landscape is due to a **stagnation of Japanese AI activities in the last two decades** (China Institute for Science and Technology Policy at Tsinghua University, 2018; WIPO, 2019). Nevertheless, the **performance** of the detected Japanese players is **the highest: 3.7 patents per patenting player**.

Finally, EU28 players' patenting activity is less developed. However, there is a noticeable number of players that, without being firms, have contributed to at least one patent. More specifically, **in EU28 14% of 717 patenting players are not firms**. In the US, this percentage is 3.5%, and 19% in China. Therefore, we can conclude that **EU28 firms** are less **patent-oriented (11.8% has at least one patent)** and, those patenting display a modest average number of patents (1.6). Nevertheless, the patenting activity developed by the EU28 research centres should be noted.

3.2 The AI landscape of EU28 Member States

In this subsection, we analyse EU28 Member States. As there are different patterns of AI activities and players between and within the MS, we consider different geographical levels. The first part of the analysis is at country level and the subsequent at regional level.

3.2.1 *EU Member States' performance: absolute and relative indicators of AI presence*

The three countries with the largest number of AI-related players are the United Kingdom, Germany and France (Figure 7). After these, there is a second group of countries consisting of the Netherlands, Spain, Italy and Sweden. A large group of EU28 countries presents a relatively small number of players: Ireland, Finland, Poland, Belgium, Denmark, Austria, Portugal, Romania, Czechia, Greece, Hungary, Estonia and Bulgaria. The AI players exclusively involved in EU funded projects are represented by the grey part of the bars in Figure 8. **When considering participants in EU funded projects**, the Netherlands is replaced by Italy in the in the Top 5 countries and **Spain and Italy almost double their number of AI players**. There are also other countries with similar relative increases (Belgium, Austria, Portugal and Greece) but the number of players they have is smaller than those of the aforementioned countries. The **UK** is the country with the largest amount of AI players and it shows a very **small increase due to EU-funded AI players**. We can conclude that **UK AI-related processes are developed, or still developing, creating an emerging position for the UK in the EU and in the world**.

In order to better compare the presence of players in each country, we take into account the values of the GDP as a measure of Member States' economy. The computed ratios[13], plotted as circles in Figure 8, enable the identification of four groups of countries[14]. The first group is formed by the countries presenting the **highest values**, namely **Estonia, the United Kingdom**. Estonian high ratio can be explained by the high level of digitalisation reached by this country (EDPR Estonia, 2017). The United Kingdom presents a high ratio, and, as mentioned, is also one of the countries where the relative increase of the number of players due to the consideration of EU projects is modest (grey bar).

The second group of countries, in terms of number of players over GDP, consists of the **Netherlands, Sweden, Ireland, Finland and Denmark**. In these countries the considered ratio is noticeable (close to the one of UK), showing a **gap between northern and southern Europe in the process of digitalization of the economy**.

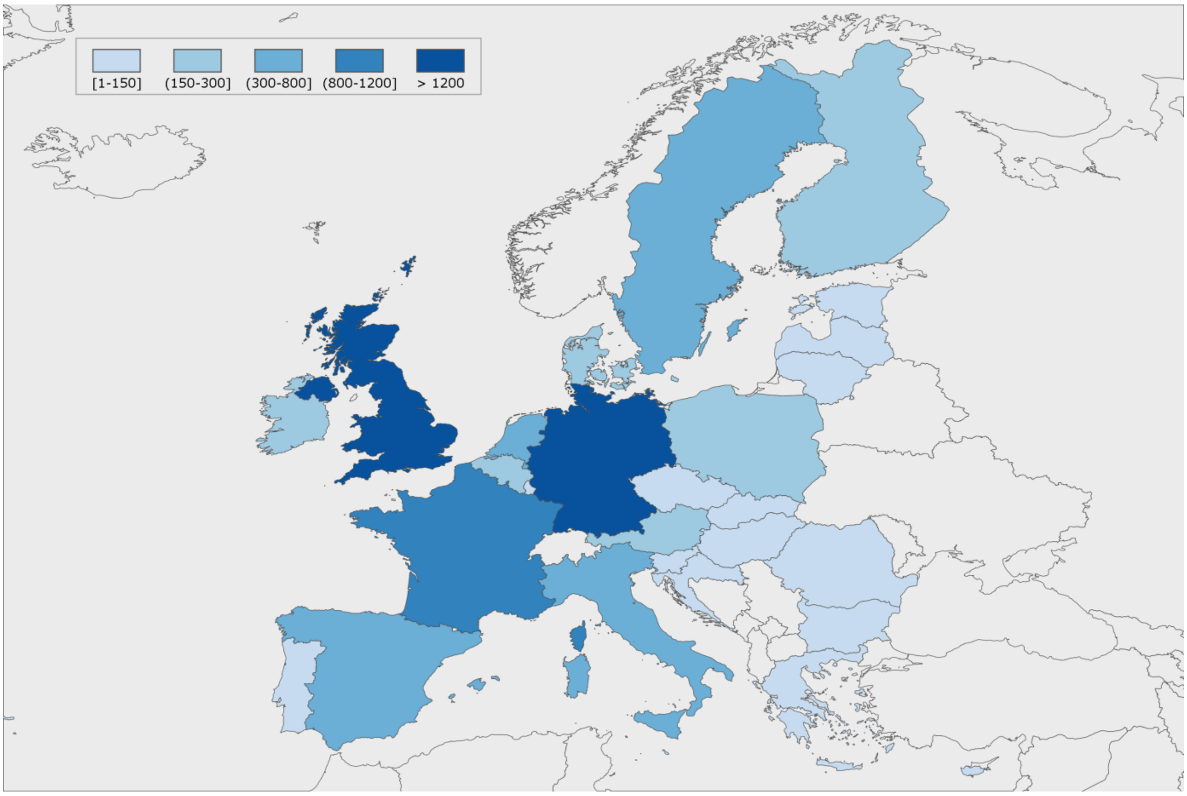
In addition, there is a group containing most of the large EU28 countries (in terms of population and GDP): **Germany, France, Spain** and, even if with a much smaller ratio, **Italy**. This suggests that the complete uptake **of AI in these countries requires time because of the size** of their economies. Nevertheless, these countries have recently launched structured national policies or initiatives towards national strategies, promoting AI development and this is likely to positively affect their position in the continental and worldwide landscape (Parliamentary Mission (Villani Mission), 2018; Plattform Industrie 4.0, 2015; BDI, 2015; AGID, 2018; Ministerio de Ciencia, Innovación y Universidades, 2019).

13 Number of players over GPD (€ bn PPS) (see subsection 3.1).

14 The three groups do not consider Malta, which is considered as an outlier because of its very small economy.

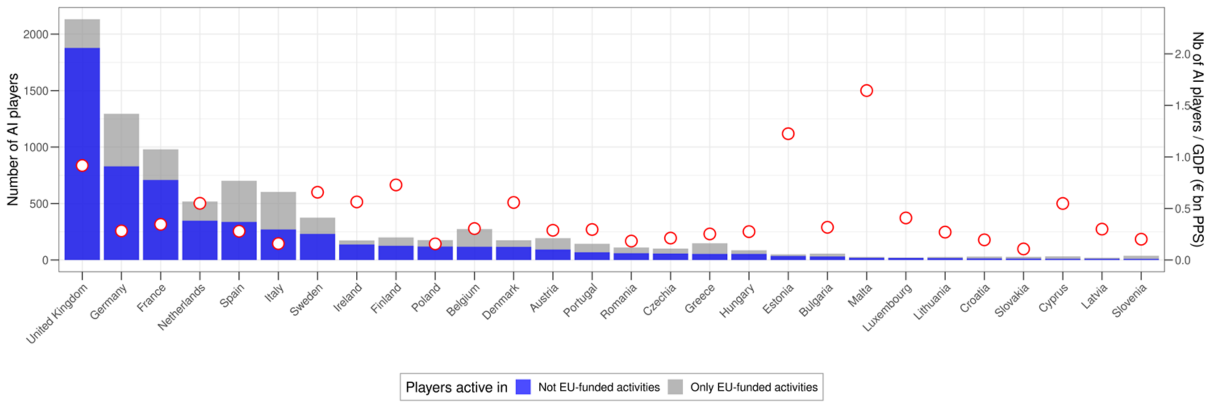
Finally, all the remaining MS form the fourth group. They all display a low number of players and have low ratios. Nevertheless, in many cases their ratio is higher than that of countries in the third group. This indicates that AI techno-economic activities are spreading and developing everywhere in the European Union.

Figure 7. Number of EU28 AI players by Member State, 2009-2018



Source: JRC PREDICT- AI TES Dataset, 2019.

Figure 8. AI players: Absolute number and Relative to GDP (€ bn PPS), European Union, 2009-2018



Note: To allow cross-country comparability, national currencies are converted into euro Purchasing Power Standard (PPS), a unit based on current euros, to account for the effect of differences in price levels across countries and of movements in exchange rates.

Source: JRC PREDICT- AI TES Dataset, 2019. (AI players), and EUROSTAT (GDP).

3.2.2 *EU28 Member States' organisation type and performance in AI patenting activity*

In Table 3 the players are distinguished by their organisation type. The presence of firms in the two countries with the largest amount of players, UK and Germany, is uneven. **UK hosts double number of AI firms than Germany**. This difference is substantial and confirms that AI is more integrated in the UK economy than in any other EU28 MS's economy. Moreover, a relatively **small percentage of UK firms is patenting** (149 out of 1,766). This suggests a **higher maturity level of AI in the UK, where AI-related processes are already oriented towards the use and commercialization of the technology**. Other MS, with less than one half of UK firms, seem to be still in the development and integration phases of AI technology in national economies, as some present a percentage of patenting firms (over the total number of firms in AI) higher than in the UK (8.4%)[15].

Table 3. EU28 Countries by number of players and organisation type, with economic indicators and ratios, 2009-2018.

Code	Country	Number of AI players by organisation type			Performance in AI patenting					Number of players relative to economic indicators			
		N. of Firms	N. of Research Intitutes	TOTAL N. of Players	N. of Patenting Players (all types)	N. of Patenting Firms	N. of Patent Applications	Average N. of Patent Applications per Patenting Player (all types)	Average N. of Patent Applications per Patenting Firm	AI players / GDP (GDP in € bn PPS)	AI firms / BERD (BERD in € m PPS)	AI firms / GERD (GERD in € m PPS)	AI players / Population (Pop. in m Persons)
UK	United Kingdom	1,766	107	1,875	163	149	201	1.2	1.2	0.92	0.08	0.05	28.90
DE	Germany	711	118	829	147	121	350	2.4	2.3	0.28	0.01	0.01	10.21
FR	France	633	74	708	76	67	88	1.2	1.2	0.35	0.02	0.01	10.65
NL	Netherlands	323	25	348	51	49	61	1.2	1.2	0.55	0.05	0.03	20.59
ES	Spain	299	38	337	41	31	41	1.0	0.9	0.28	0.04	0.02	7.26
IT	Italy	227	43	270	42	38	38	0.9	0.9	0.16	0.02	0.01	4.44
SE	Sweden	212	18	230	41	38	103	2.5	2.6	0.65	0.03	0.02	23.60
IE	Ireland	126	10	136	19	17	59	3.1	3.4	0.56	0.06	0.04	29.07
FI	Finland	119	7	126	10	8	43	4.3	5.2	0.73	0.04	0.02	23.03
PL	Poland	106	13	119	25	20	26	1.0	0.9	0.16	0.03	0.01	3.13
BE	Belgium	96	21	117	11	8	8.2	0.7	0.8	0.30	0.01	0.01	10.41
DK	Denmark	104	12	116	12	11	12	1.0	1.0	0.56	0.03	0.02	20.50
AT	Austria	72	21	93	6	4	4.4	0.7	0.6	0.29	0.01	0.01	10.83
PT	Portugal	61	7	68	3	3	6.5	2.2	2.2	0.30	0.05	0.02	6.55
RO	Romania	50	10	60	17	8	22	1.3	1.0	0.18	0.07	0.03	3.02
CZ	Czechia	50	7	57	11	7	15	1.4	1.7	0.21	0.02	0.01	5.41
EL	Greece	42	12	54	9	8	5.7	0.6	0.6	0.28	0.02	0.02	5.48
HU	Hungary	50	4	54	11	11	4.8	0.4	0.4	0.25	0.07	0.02	4.97
EE	Estonia	35	-	35	1	1	1	1.0	1.0	1.23	0.18	0.08	26.62
BG	Bulgaria	29	2	31	3	2	1.8	0.6	0.4	0.32	0.04	0.03	4.30
MT	Malta	18	1	19	-	-	-		-	1.64	0.39	0.20	43.21
LU	Luxembourg	16	2	18	3	3	3.3	1.1	1.1	0.41	0.06	0.03	31.97
LT	Lithuania	16	1	17	5	4	6.5	1.3	1.5	0.27	0.09	0.02	5.82
HR	Croatia	13	1	14	3	3	8	2.7	2.7	0.20	0.04	0.02	3.31
SK	Slovakia	12	1	13	2	2	2	1.0	1.0	0.11	0.03	0.01	2.40
CY	Cyprus	10	1	11	2	2	0.8	0.4	0.4	0.55	0.45	0.10	12.99
LV	Latvia	11	-	11	-	-	-		-	0.30	0.19	0.05	5.54
SI	Slovenia	6	4	10	3	1	2.3	0.8	1.0	0.20	0.01	0.01	4.85

Note: GDP: Gross domestic product; BERD: Business expenditure on R&D; GERD: Global expenditure on R&D.

Sources: AI related indicators: JRC PREDICT- AI TES Dataset, 2019. Economic Indicators: EUROSTAT.

15 Only exceptions are Finland (6.7%), Belgium (8.3%), Austria (5.6%), Portugal (4.9%), Estonia (2.9%) and Bulgaria (6.9%).

Among the countries with a GDP (in 2015) higher than € 1,000 bn PPS, **Italy has the highest proportion of research institutes** over the total amount of players (43 out of 270 players). This suggests a noticeable involvement of the Italian research in AI (China Institute for Science and Technology Policy at Tsinghua University, 2018). However, this large proportion is likely to be related to a relatively modest presence of Italian firms in AI.

Among the countries with the largest number of players, **UK and Ireland show the largest number of players over BERD and GERD.** These statistics allow the evaluation of the number of players with respect to the amount of overall expenditure in R&D (GERD) and only by business enterprises (BERD): large ratio indicate that the corresponding area presents a noteworthy number of players, given the resources that were invested. While for the UK the importance of the AI-related activities has already been stressed, in the case of Ireland this is potentially a positive consequence of its national policy towards digitisation, initiated in July 2013[16]. Netherlands and Spain are also ranking high according to the aforementioned ratios.

France is ranking third in number of AI players (708). Its percentage of firms over total number of players (89.4%) is behind UK (94.2%) and ahead of Germany (85.7%). The patenting intensity is not as high as in other leaders (1.2 patents per patenting firm), and the percentage of patenting firms is close to the one of UK (8.4%). The number of players over GDP (0.35) is higher than the one of Germany (0.28) but smaller than the one of UK (0.92).

Regarding patenting activity, the average number of patent applications filed by each firm is remarkable in **Finland (5.2), Ireland (3.4), Sweden (2.6), Germany (2.3), Portugal (2.2) and Croatia (2.7)**. This heterogeneity indicates different approaches to pursue AI industrial development within the EU28.

3.3 Regional diversity in the EU28

In this subsection, we analyse the presence of the AI-sector in European regions (NUTS2). As in the previous subsections, we present indicators regarding the ranking of AI players: number of players (absolute and relative to regional GDP), their institutional typology and their patenting performance.

3.3.1 ***AI Players in top 30 EU28 regions***

Some regions are densely populated by AI players performing R&D and industrial activities. As expected, the areas of **Inner London-West** (UK), **Île de France** (FR), and **Berlin** (DE) are leading in terms of number of AI players, according to our data collection (Figure 9). This is in line with the findings of the previous subsection, where UK, Germany and France are the EU leaders. Significantly, while at national level Germany is the second country after the UK, in the regional comparison the **French capital ranks second**. **However, Germany is the only country to have two regions**, Berlin and Oberbayern, **in the top 5** (respectively the 3rd and the 5th regions in EU28 by number of players). The next **set of regions** in terms of number of AI players is composed by Noord-Holland (NL), Cataluña (ES), Southern and Eastern Ireland (IE), Comunidad de Madrid (ES), Stockholm (SE) and Inner London-East (UK). Within this group, Cataluña presents the minimum number of firms (2.0% over total number of EU28 firms).

Another group of 20 **regions host an amount of players ranging between 1.6% and 0.7% of all EU28 AI players** (see Table 4). It is noteworthy that **8 out of these 20 are in the UK**. UK reveals a **balanced and distributed regional AI structure**, the area of London being just its most apparent region. Finally, the majority of **top 30 regions** are in Northern/Central Europe. Out of the regions not belonging to the UK, **55%** are **in Central or Northern Europe**, and only 25% in Southern Europe (Cataluña and Comunidad de Madrid in Spain, Lombardia and Lazio in Italy, and Área Metropolitana de Lisboa in Portugal).

3.3.2 ***Research institute's locations and regions' responsiveness***

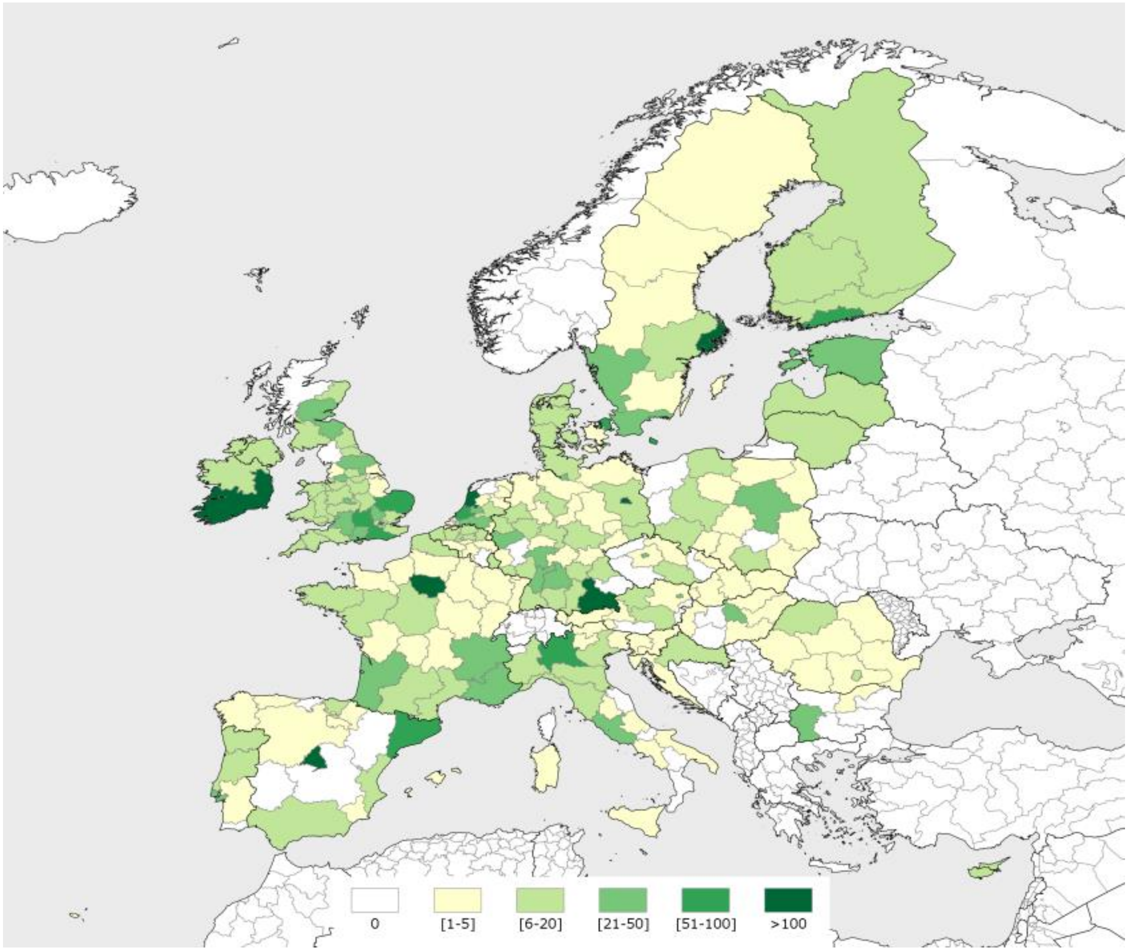
The organisational type of the players belonging to each EU28 region is shown in columns "N. of Firms of the Region (% over the total n. of EU28 Firms)" and "N. of Research Institutes of the Region (% over the total n. of EU28 Research Institutes)" in Table 4. In particular, few regions include more than 10 research institutions. Apart from the top 6 regions (Inner London-West (UK), Île de France (FR), Berlin (DE), Noord-Holland (NL), Oberbayern (DE), and Cataluña (ES)), only Berkshire, Buckinghamshire and Oxfordshire (UK), Rhône-Alpes (FR) and North Yorkshire (UK) have more than 9 research institutes. In terms of percentage of research institutes

16 National Digital Strategy (NDS) document "Phase 1 - Digital Engagement" was launched in July 2013.

over regional number of players (column "% of Research Institutes over Regional Players"), the most relevant regions are Cataluña (ES), Rhône-Alpes and Provence-Alpes-Côte d'Azur (FR), North Yorkshire and Eastern Scotland and West Midlands (UK), and Wien (AT), with nearly **18% of their players as research institutions**. These percentages indicate that these regions are relevant R&D poles for the development of AI. These findings are confirmed by the number of universities and research institutes present in the aforementioned regions by the profiles described by the Regional Innovation Monitor Plus of DG GROW (DG-GROW RIM).

Considering the **number of players over GERD** (column "N. of Players over GERD in € m PPS" in Table 4), **five regions from UK and one from Ireland** are **in the top 10**. This suggests that in these regions the gross domestic expenditure on R&D (GERD) was able to sustain and promote the activity of a larger amount of AI players. Finally, an additional confirmation of the stronger development of the AI sector in the UK in comparison with the rest of the EU28, is given by the **business expenditure in R&D (BERD)**. When this indicator is used to evaluate the number of players across the different EU28 regions (column "N. of Players over BERD in EUR m PPS" in Table 4), UK is still leading, as **five of its regions** are in the top 10 for number of players over BERD. The others are Noord-Holland (NL), Berlin (DE), Área Metropolitana de Lisboa (PL), Southern and Eastern Ireland (IE), and Cataluña (ES), regions with a large number of players given the level of business expenditure in R&D. This is an indicator that in these **regions** the AI **business activities are responding well to investments in R&D performed by enterprises**.

Figure 9. Number of EU28 AI Players at regional level (NUTS2), 2009-2018



JRC PREDICT- AI TES Dataset, 2019.

Table 4. EU28 Top 30 Regions by number of players and organisation type, with economic indicators and ratios, 2009-2018.

NUTS2 code	Region	Number of Players				Number of Players relative to economic indicstors		
		N. of Players in the Region (% over the EU28 Total N. of Players)	N. of Firms in the Region (% over the EU28 Total N. of Firms)	N. of Research Institutes in the Region (% over the EU28 Total N. of Research Institutes)	N. of Research Institutes (% over Players in the Region)	N. of Players / GDP in € bn PPS	N. of Players / BERD in € m PPS	N. of Players / GERD in € m PPS
UKI3	Inner London - West	15.4%	16.7%	4.1%	2.6%	4.487	0.915	0.273
FR10	Île de France	7.3%	7.4%	6.1%	8.1%	0.681	0.037	0.025
DE30	Berlin	4.2%	4.3%	2.9%	6.6%	2.002	0.137	0.057
NL32	Noord-Holland	2.8%	2.8%	2.1%	7.5%	1.194	0.159	0.070
DE21	Oberbayern	2.2%	2.1%	3.0%	13.2%	0.549	0.016	0.013
ES51	Cataluña	2.2%	2.0%	3.9%	17.6%	0.538	0.062	0.036
IE02	Southern and Eastern	2.1%	2.2%	1.6%	7.3%	0.575	0.072	0.052
ES30	Comunidad de Madrid	2.1%	2.1%	1.6%	7.4%	0.526	0.054	0.031
SE11	Stockholm	2.0%	2.1%	1.4%	6.9%	1.026	0.036	0.027
UKI4	Inner London - East	1.8%	2.0%	0.4%	1.9%	0.915	0.295	0.171
UKJ1	Berkshire, Buckinghamshire and Oxfordshire	1.6%	1.6%	1.8%	10.5%	0.920	0.041	0.026
FI1B	Helsinki-Uusimaa	1.6%	1.6%	0.9%	5.6%	1.325	0.055	0.037
UKH1	East Anglia	1.6%	1.6%	0.7%	4.4%	1.311	0.041	0.027
NL33	Zuid-Holland	1.5%	1.5%	0.7%	4.8%	0.621	0.061	0.029
DK01	Hovedstaden	1.4%	1.4%	1.6%	10.8%	0.979	0.031	0.021
ITC4	Lombardia	1.4%	1.4%	1.4%	10.0%	0.219	0.024	0.017
UKJ2	Surrey, East and West Sussex	1.1%	1.2%	0.7%	6.2%	0.693	0.054	0.044
FR71	Rhône-Alpes	1.0%	0.9%	2.0%	18.3%	0.303	0.018	0.012
SE23	Västsverige	1.0%	1.0%	0.7%	7.0%	0.831	0.028	0.022
UKE2	North Yorkshire	0.9%	0.8%	1.8%	18.5%	2.509	0.231	0.155
AT13	Wien	0.9%	0.8%	1.6%	17.6%	0.617	0.033	0.017
UKK1	Gloucestershire, Wiltshire and Bristol/Bath area	0.9%	0.8%	1.1%	12.0%	0.640	0.042	0.031
PL12	Mazowieckie	0.9%	0.9%	0.7%	8.0%	0.296	0.040	0.017
UKD3	Greater Manchester	0.8%	0.9%	0.5%	6.1%	0.661	0.157	0.070
HU10	Közép-Magyarország	0.8%	0.8%	0.7%	8.3%	0.529	0.038	0.028
UKM2	Eastern Scotland	0.8%	0.7%	1.4%	18.2%	0.718	0.085	0.034
FR82	Provence-Alpes-Côte d'Azur	0.7%	0.7%	1.4%	18.6%	0.303	0.021	0.013
PT17	Área Metropolitana de Lisboa	0.7%	0.8%	0.4%	4.8%	0.504	0.074	0.033
UKG3	West Midlands	0.7%	0.7%	1.4%	19.0%	0.587	0.048	0.031
ITI4	Lazio	0.7%	0.7%	0.7%	9.8%	0.220	0.037	0.014

Note: * Base year 2014, ** Base year 2013. GDP: Gross domestic product; BERD: Business expenditure on R&D; GERD: Global expenditure on R&D.

Sources: AI related indicators: JRC PREDICT- AI TES Dataset, 2019. Economic Indicators: OECD and EUROSTAT.

4 AI thematic key areas

To identify technological subdomains from the entire AI technological landscape and key areas of specialisation of countries and regions, we analyse the **textual content of worldwide R&D and industrial activities**. To that end, we implement a machine learning approach on the collected corpus of documents representing AI R&D and industrial activities. The **corpus encompasses documents describing worldwide production and trade activities** (description of companies' activities, including startups, firms financed by venture capital, etc.), as well as **research and innovation activities** (patent applications, conference publications and EU-funded research projects).

The section addresses the points arisen in: (i) the European Strategy on AI[17], (ii) its Coordinated Plan[18], (iii) the Declaration of Cooperation on AI, signed in April 2018 by EU countries' representatives[19], and (iv) *AI WATCH* project[20], regarding the opportunities and risks in AI for EU. Moreover, this section presents the thematic key areas of AI specialisation per country[21] and region with respect to other worldwide regions.

4.1 Occurrence and relevance of AI key areas

Following the semantic analysis approach described in section 2.2, we present in this subsection the resulted **six non-exclusive thematic key areas**, or else knowledge subdomains, **that are identified in the AI technological landscape** for 2009-2018:

- Natural language processing (NLP)
- Computer vision
- Machine learning (ML)
- Robotics and Automation
- Connected and Automated vehicles (CAVs)
- AI Services

Further description of the thematic areas can be found in Annex 1. The identified thematic areas represent effectively the AI domain and are in accordance to recent AI analyses (Parliamentary Mission (Villani Mission), 2018; Strategic Council for AI Technology, 2017; China's State Council; 2017 McKinsey Global Institute; 2017b; US Congressional Research Service, 2019; US Department of Defense: Govini, 2018; National Strategy for Artificial Intelligence #AIFORALL, 2018; OECD, 2018; China Institute for Science and Technology Policy at Tsinghua University, 2018; Stone et al.: AI100, 2016).

Table 5 shows the most relevant terms of each topic. The thematic areas are illustrated in Figure 10 by means of the first two principal components of the principal components analysis (PCA). The PCA is implemented to the topics that result from the topic model applied to the AI R&D and industrial activities over the entire study period. The correspondence between the topics are the thematic subdomains of Table 5 is the result of the following process (i) initial manual configuration of topics' titles inspired by the automatic summary of the top terms, (ii) correspondence evaluation to real-world topics through literature comparison (used also for the identification of the number of topics that is representative of the technology), (iii) modification of topics' titles in view of (ii), (iv) evaluation of the titles by experts, who are also provided with a list of the most relevant terms per topic, (v) final correspondence between topics and technological subdomains.

17 Explicitly for EU in COM (2018) 237 it is mentioned that "the public and private sectors must seize the opportunities that come both from developing innovative AI solution and applying them to a range of fields. Without efforts, the EU risks losing out on the opportunities offered by AI, facing a brain drain and being a consumer of solutions developed elsewhere".

18 In COM (2018) 795 it is highlighted that investments should follow regulations to ensure the best and well distributed AI use in a European socio-economic context

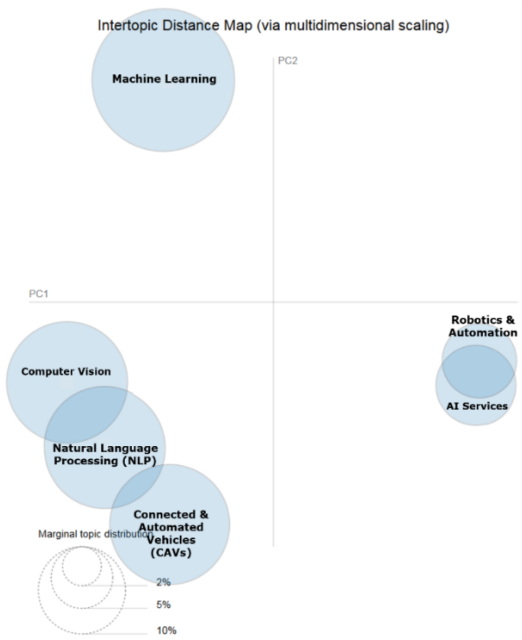
19 The Declaration on Cooperation on Artificial Intelligence signed by EU countries' representatives, highlights the priority of allocating R&D&I funds to support the AI development and distribution. Romania, Greece, Cyprus and Croatia have joined the initiative since it was launched (https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence)

20 https://ec.europa.eu/knowledge4policy/ai-watch_en

21 For the worldwide comparisons EU28 is considered as a single geo-political entity.

Table 5. AI thematic key areas and top keywords per area, 2009-2018.	
Key thematic area	**Top keywords**
Natural language processing	natural language; user information; speech device; recognition result; recognition device; recognition system; user data; voice recognition; user input
Computer vision	face recognition; face image; computer vision; convolutional neural network; image recognition; image information; image acquisition; image device; face information; face feature
Machine learning	neural network model; convolutional neural network; supervised learning model; support vector machine; deep neural network; prediction model; linear step; network training; data driven model
Robotics and Automation	robotic system; big data; computer vision; automated vehicle; control system; energy efficiency; management system; robotic technology; intelligent hardware software development
Connected and Automated vehicles	automated vehicle; control device/system; vehicle control system; control module; control unit; intelligent control system; unmanned vehicle; recognition module; communication module
AI Services	neural network model; convolutional neural network; service support business; software development company; deep neural network; big data; computer vision; software provider; business support; software application
Source: JRC PREDICT- AI TES Dataset, 2019.	

Figure 10. Identified AI key areas' map by thematic relatedness (Inter-topic distance map), 2009-2018.



Note: Topics inter-distance is analogous to their thematic relation, namely topics that are more related are closer mapped. Nodes' size is proportional to the number of terms relevant to the thematic area (conditional topic distribution based term). For the visualisation the LDAvis R package was used (Sievert and Shirley, 2014).

Source: JRC PREDICT- AI TES Dataset, 2019.

The thematic areas include applications and methods from interdisciplinary fields, which explain the closeness between certain identified key areas. Figure 10 charts the distance between topics, in order to show how the thematic key areas are related to each other, where the common inclusion of terms in topics illustrates semantic, or technological in this case, proximity. Adjacent key areas have more operations and terms used interchangeably than key areas of greater distance. Namely, the NLP key area is close to the Computer vision and the CAVs key areas because of the speech and face recognition applications ensuing from the two first mentioned subdomains, used in CAVs from the automobile industry. More specifically, a CAV is the synergistic outcome of the classic automobile parts and the voice and image processing applications and hardware (sensors, cameras) of the corresponding key thematic areas. On the contrary, the operations ensuing from the subdomain AI Services are not closely related to the CAVs subdomain, as e.g. decision management applications for profit increase are not related with CAVs decision management. Finally, Machine learning is the largest thematic area as all other key areas adopt the methodological terms of this area to describe the development of their corresponding activities. The distance of this topic from the other five on the second component (PC2), suggests a distinction between theoretical advancements and applications.

4.2 Worldwide AI specialisation and Thematic hotspots

The thematic **Revealed Comparative Advantage (RCA)** indicator is employed to explore and analyse the specialisation of geographic areas in the AI field. It measures a country's specialisation within the AI domain in comparison with the world average specialisation in that area.. Let A_{C_i,k_z} be the number of activities of country C_i in topic k_z, defined as the sum of activities of all the country's players in said topic: $A_{C_i,k_z} = \sum_j A_j^{C_i,k_z}$, then the RCA for country C_i and topic k_z is defined as:

$$RCA_{C_i,k_z} = \frac{\frac{A_{C_i,k_z}}{\sum_z A_{C_i,k_z}}}{\frac{\sum_C A_{C_i,k_z}}{\sum_{C,Z} A_{C_i,k_z}}} = \frac{\frac{sum\ of\ activities\ of\ country\ C_i\ in\ a\ topic\ k_z}{sum\ of\ activities\ of\ country\ C_i\ in\ all\ topics}}{\frac{sum\ of\ worlwide\ activities\ in\ a\ topic\ k_z}{sum\ of\ worlwide\ activities\ in\ all\ topics}}$$

The value RCA = 1[22] represents the world average or average specialisation in the thematic area when all countries are considered. It is the benchmark towards which all countries are compared, and it is illustrated by a black line in the plots. When a country C_i presents RCA>1 in a topic k_z, then this country is relatively specialised in this topic and has a **revealed comparative advantage**. Figure 11 illustrates the thematic key areas of attention or specialisation of top geographic areas during the period 2009-2018.

To further explore countries' performance globally on each AI thematic area, in terms of number of R&D and industrial AI activities, we develop the **Thematic Hotspot Indicator (THI)**. Let A_{C_i,k_z} be the number of activities of country C_i in topic k_z, defined as the sum of activities of all the country's players in said topic: $A_{C_i,k_z} = \sum_j A_j^{C_i,k_z}$, then the Thematic Hotspot Indicator (THI) for country C_i and topic k_z is defined as:

$$THI_{C_i,k_z} = \frac{A_{C_i,k_z}}{\sum_i A_{C_i,k_z}} = \frac{sum\ of\ activities\ of\ country\ C_i\ in\ topic\ k_z}{sum\ of\ activities\ of\ countries\ worldwide\ in\ topic\ k_z} x100$$

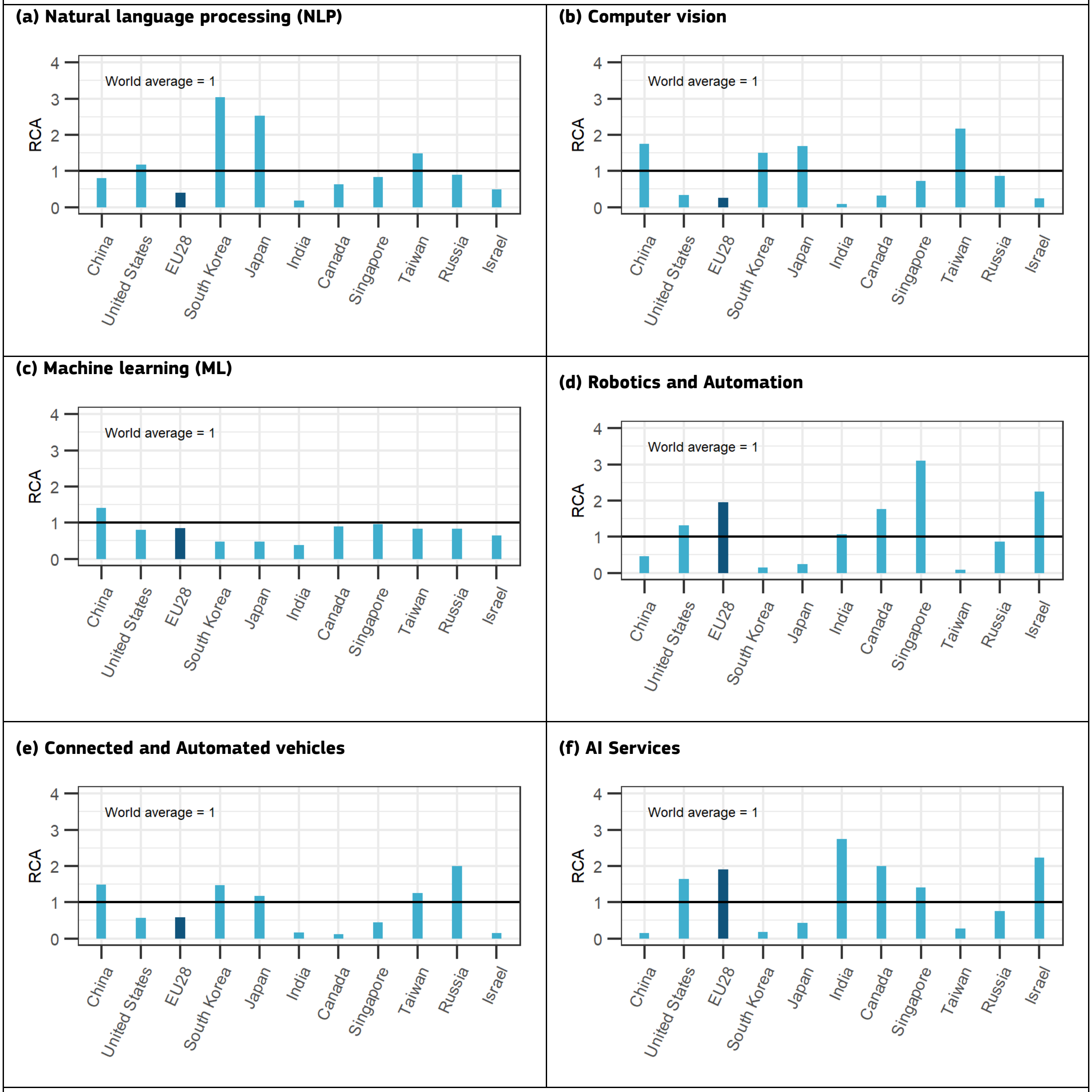
THI shows the intensity of a country's participation in a thematic area compared to all countries globally, computed as the proportion of country´s activities in this thematic area to the activities of all countries in the same area. Figure 12 shows the top-10 geographic entities by number of AI activities in each thematic area.

4.2.1.1 EU28

The RCA index of the **EU28** demonstrates a **comparative advantage in the thematic areas of Robotics and Automation, and of AI Services** with values of the indicator above world's benchmark (Figure 11). EU28's leading position in the robotics thematic area is justified by its strong presence in both the industrial and the research parts. In number of activities, the EU ranks second after the US (Figure 12) and is expected to lead in this subdomain in the future (Parliamentary Mission (Villani Mission), 2018). Europe has a strong robotics industry, hosting three of the world's largest producers of industrial robots (KUKA, ABB and Comau). More precisely, Europe produces "more than a quarter of the world's industrial and professional service robots (e.g. for precision farming, security, health, logistics)" (International Federation of Robotics). The European robotics research excellence is the pay-off of the allocation of € 700 million for research until 2020, through the Robotics Public Private Partnership (SPARC, the partnership for robotics in Europe), and further private investments that reach 2.8€ billion in total (Accenture, 2018; National Strategy for Artificial Intelligence #AIFORALL, 2018). Moreover, SPARC ensures a strong global position in the research part and interdisciplinary research (International Federation of Robotics). When considering all worldwide Robotics activities, also the US, Canada, Singapore and Israel have prominent roles, with revealed comparative advantage on this area.

22 The RCA indicator takes values ≥ 0, without an upper limit. In practical terms, most values lie below 5.

Figure 11. Revealed Comparative Advantage (RCA) of top AI geographical areas, 2009-2018



Notes:

The Revealed Comparative Advantage (RCA) for country C_i and topic k_z is defined as:

$$RCA_{C_i,k_z} = \frac{\frac{A_{C_i,k_z}}{\sum_z A_{C_i,k_z}}}{\frac{\sum_C A_{C_i,k_z}}{\sum_{C,z} A_{C_i,k_z}}} = \frac{\frac{sum\ of\ activities\ of\ country\ C_i\ in\ a\ topic\ k_z}{sum\ of\ activities\ of\ country\ C_i\ in\ all\ topics}}{\frac{sum\ of\ worlwide\ activities\ in\ a\ topic\ k_z}{sum\ of\ worlwide\ activities\ in\ all\ topics}}$$

where $A_{C_i,k_z} = \sum_j A_j^{C_i,k_z}$ is the sum of activities of all players in country C_i and topic k_z.

The RCA indicator takes values ≥ 0, without an upper limit. In practical terms, most values lie below 5. The value RCA = 1 reflects the benchmark represented by the world average. Values > 1 reveal a comparative advantage.

Source: JRC PREDICT- AI TES Dataset, 2019.

Regarding the **AI Services thematic area, the EU's advantage** may be explained by the growth of the software industry, which is five times the growth of the rest of the European economy (Atomico, 2018). The substantial role of Europe in developing e.g. platforms providing services that contribute to the "intelligent enterprise" and e-government schemes, is highlighted also in the European Strategy's Coordinated Plan.

Another strong area for EU28 is Machine learning, the "know-why"-oriented thematic area. Although the EU with an RCA of 0.9 does not reach the world average, it is among the three top world leaders just below the benchmark and China, which has a RCA above 1. This result is supported by the third position in the world ranking in terms of number of AI activities in this area (Figure 12). EU's strength in the AI research field is supported by other studies, showing that Europe hosts one third of the top 100 AI research institutions worldwide[23]. An example of these is the German Research Center for Artificial Intelligence (DFKI), specialised in innovative commercial AI-based software since 1988 (European Commission (2018b)).

Connected and Automated vehicles is not among the EU28's specialisations in relative terms (Figure 11), although in number of patent applications, research publications and firms activities, the EU ranks third after China and the US (Figure 12). However, another study reveals that until mid-2016, European and US activities were leading the development of CAVs, with predictions assessing that China would not follow closely due to traffic conditions, road conditions and driving behaviour (Viereckl et al., 2016). As it is discussed in the next sub-subsection for China, this landscape changed following the incentives that the Chinese government announced in mid-2015 with the strategic plan "Made in China 2025", and implemented in the following 3 years (China's State Council). China's strategic plan is compared to Germany's plan ("Plattform Industrie 4.0"), in terms of promoting SMEs network and production, although in a far broader scale, and not only using IoT (Li, 2018; Plattform Industrie 4.0, 2015; BDI, 2015). It should be also noted that based on patents' applicants location for autonomous driving, German (Audi, BMW, Bosch, Continental, Daimler, Volkswagen) and Swedish (Volvo) automobile companies dominate the worldwide autonomous automobile industry, (Bardt, 2017; Statista, 2017).

Based on the aforementioned aspects, Europe is one of the important actors in AI worldwide. As it is discussed further in section 4.3, the most AI-relevant European countries are **United Kingdom** as one of the AI emerging economies, **France** as one of the cyber colonies (Parliamentary Mission (Villani Mission), 2018), and **Germany**.

4.2.1.2 US

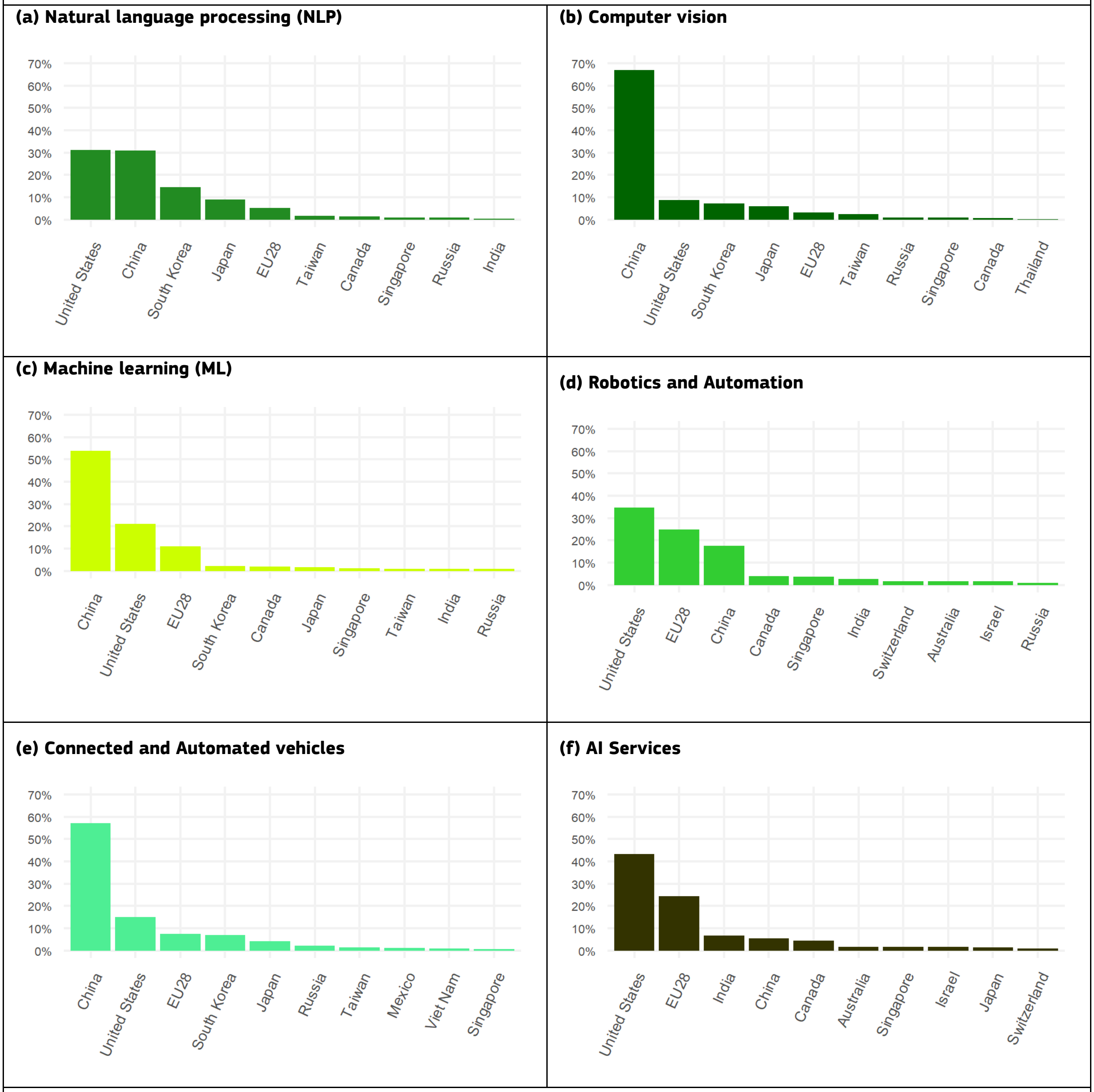
The **US** has a **comparative advantage in three AI areas: AI Services**, **Robotics and Automation,** and **NLP** (Figure 11). In these three areas, the US hosts between 30% and 45% of worldwide activities (Figure 12). The country has also a strong position in the theoretical part of AI, the area of ML, hence establishing its interest and specialisation in the "know-why" of the AI technology. The country is relatively better positioned in the CAVs area than in Computer vision, but both with an important part in absolute terms, since this country ranks second in number of activities in this area (Figure 12). **The US has strong presence, although not always leading, in almost all the identified AI key areas**. As one of the two colossi in AI, such an extensive presence in all areas is expected (Parliamentary Mission (Villani Mission), 2018).

4.2.1.3 China

In Figure 11, China's areas of specialisation are **Computer vision**, **CAVs**, and the theoretical aspect of AI development, **Machine learning**. In particular in ML, China appears as the only country among the top having a RCA indicator above the world benchmark. China is not only specialised in these areas, but is also the **world leader in number of AI activities** in the three of them (Figure 12), hosting more than half of the worldwide activities in these thematic areas. Computer vision and CAVs are among the targeted areas promoted according to the national Next Generation AI Development Plan (AIDP) (China's State Council, 2017). The country's engagement with the AI know-how (the ML thematic area), can be explained by the university program launched to educate approximately 500 teachers and 5,000 students on AI (China's State Council, 2017; National Strategy for Artificial Intelligence #AIFORALL, 2018). The country is also relatively well positioned in the thematic area of NLP. In Robotics, China is also present, though moderately active, with one fifth of the worldwide activities (Figure 12).

23 In the global top-100 for AI-related research paper citations 32 are European research institutions, 30 are from the USA and 15 from China [COM(2018) 237 final; Atomico, 2017].

Figure 12. National hotspots of thematic key areas: country activities as a percentage of worldwide activities in each key area, 2009-2018.



Notes:

The Thematic Hotspot Indicator (THI) for country C_i and topic k_z is defined as:

$$THI_{C_i,k_z} = \frac{A_{C_i,k_z}}{\sum_i A_{C_i,k_z}} = \frac{\textit{sum of activities of country } C_i \textit{ in topic } k_z}{\textit{sum of activities of countries worldwide in topic } k_z} x100$$, values 0 to 100%.

where A_{C_i,k_z} is the number of activities of country C_i in topic k_z, defined as the sum of activities of all the country's players in said topic: $A_{C_i,k_z} = \sum_j A_j^{C_i,k_z}$

Top 10 countries presented in decreasing order per key area. EU28 is aggregated as one geographic area.

Source: JRC PREDICT- AI TES Dataset, 2019.

China is among the top three global venture capital investors in unmanned vehicles (CAVs and drones), robotics and virtual reality among others (Pinho, 2019; Ke and de Diego, 2019). China's AIDP shows that the country targets leadership in AI until 2030 (China's State Council, 2017). This aim is supported by **governmental investments in Chinese companies, especially in CAVs** (Viereckl et al., 2016). The consequences of this support in the area is confirmed in Figure 12, where the overwhelming majority of activities on CAVs are led by Chinese governmental institutions. **Chinese companies become competitive**

in the automated driving area. Alibaba, Baidu and Tencent partnered with vehicle automobile companies to develop platforms for CAVs. In particular, Baidu agreed to provide a connectivity platform to BMW, Mercedes-Benz, Ford, Hyundai, General Motors, Audi, and the Chinese BYD (Statista, 2017). The other large Chinese digital company, Alibaba, collaborated with the Chinese automobile firm SAIC for the production of a type of CAV and became a stakeholder in the electric vehicle and technology start-up Xiaopeng Motors (Viereckl et al., 2016; Xiaopeng Motors). Tencent also became a minor stakeholder in Tesla in 2017 and invested in the Chinese BYD. The aforementioned examples explain to an extent the recent significant growth in the CAVs thematic area, which establishes its global leading position in the area, ahead from US, EU, South Korean and Japanese counterparts (Viereckl et al., 2016).

4.2.1.4 Other Asian countries

India shows a comparative advantage in **AI Services** and in the area of **Robotics & automation**.

South Korea, **Japan** and **Taiwan** are mainly specialised in **NLP**, **Computer vision**, and **CAVs** (Toyota, Nissan, and Hyundai respectively (Viereckl et al., 2016)). In particular in Computer vision, Taiwan has the highest comparative advantage among the top countries. Taiwan's activity in the CAVs area is expected to be more on the manufacturing side than on research, due the country's strength on manufacturing.

Singapore is the tenth country by number of active players in AI (Figure 5). Its areas of specialisation are **Robotics and Automation**, as well as **AI Services** (Figure 11).

4.2.1.5 Canada

Canada shows a comparative advantage mainly in the **AI Services** and **Robotics and Automation** areas. Canada's presence in the AI industrial landscape as fourth in the average number of patents applications (Figure 6 (c)), in the top 10 countries globally in number of firms active in AI by geographic area (Figure 15(a)), and in the top 10 countries in number of patenting firms (Figure 15 (b)), explains its leaning on the application side of AI. This finding is also corroborated by other analyses (China Institute for Science and Technology Policy at Tsinghua University, 2018; WIPO, 2019). **ML** and, to a lesser extent, **NLP** activities occupy a part of the country's focus (Figure 12). As the first country with an AI national strategy aiming to invest on AI research and talents, the country's thematic profile might soon change and increase its modest presence in the AI experts' landscape (Pan-Canadian Artificial Intelligence Strategy, 2017; China Institute for Science and Technology Policy at Tsinghua University, 2018; CIFAR, 2019).

4.2.1.6 Israel

Israel has a specialisation profile similar to Canada's, with higher comparative advantage in the application side of AI: **AI Services** and **Robotics and Automation**. Moreover, Israel's emerging role in the aforementioned key areas is corroborated by the findings presented in (Figure 12), where the country is considered a hotspot ranking among the top-10 countries with highest number of activities in these areas. This is in accordance with other studies stating that Israel is one of the **emerging economies in the AI field, holding a key position along with Canada and the United Kingdom** (Parliamentary Mission (Villani Mission), 2018; WIPO, 2019). It is included in the top 20 countries in number of patent applications by patent office and in the top 10 countries in number of acquisitions of AI companies (Statista, 2017; China Institute for Science and Technology Policy at Tsinghua University, 2018; WIPO, 2019).

4.2.1.7 Russian Federation

Russian Federation's RCA suggests that the its main specialisation is on the **CAVs** area, and it shows the highest value for the RCA indicator among the countries considered in Figure 11. Recent projects from Russian manufacturing firms towards the development of electric vehicles (Kalashnikov), unmanned vehicles as taxis (Yandex in collaboration with Hyundai), trams (PK Transportnye Systemy in collaboration with Cognitive Technologies) and trucks24 support this finding. The government's regulation allowing the testing of unmanned vehicles is expected to positively affect the developing activities in this thematic area (KPMG, 2019; Regulation 1415, 2018). Russia has also relatively high levels of specialisation in all other thematic areas: NLP, Computer vision, Robotics and Automation, ML and AI Services.

24 KAMAZ Russian truck and engine manufacturer, in collaboration with the Russian VIST Group mining industry management system developer, Zyfra industrial digital technology corporation, and the Kazakhstani Nazarbayev University.

To summarise, the combination of the Revealed Comparative Advantage and the Thematic Hotspot indicators per thematic area enables the **identification of the best performers**. These are countries that are specialised in the area and holding a high share of world activities, hence with a high influencing capacity on the development of AI in these areas. The results show that countries with RCA > 1 and a share of worldwide activities above 5% in the same area are: the **US** and the **EU28** in **Robotics and Automation** and **AI Services**; **China** in **ML**, **Computer vision** and **CAVs**; **South Korea** in **NLP**, **Computer vision** and **CAVs**; **Japan** in **NLP** and **Computer vision**; and **India** in **AI Services**.

4.3 EU28 national thematic hotspots

Figure 13 shows an overview of the EU MS's performance on each AI thematic area, based on the number of R&D and industrial activities. It represents the relative position of each country with respect to the whole EU28. Within EU, the United Kingdom and Germany are the leading countries in all thematic key areas, and only in CAVs Germany comes first. The countries thematic hotspot indicator (THI, see subsection 4.2 for definition) is used to assess the Member States' presence in AI key thematic areas (Figure 13).

In the thematic area of **NLP, the United Kingdom leads with more than 30% of all the EU28's activities**. Germany's presence is also significant, although with slightly more than half of UK's share in activities. France ranks third, holding a share of one tenth of EU28's activities. Ireland, Netherlands and Sweden follow with marginally similar performance.

In **Computer vision**, the **United Kingdom** hosts the majority of related activities, ranking first again, although much closer to **Germany, Ireland, and France**. The Netherlands ranks fifth, although with less than 10% share. Spain and Italy are the only Mediterranean MS among the top 10 in this thematic area.

The **United Kingdom** leads also in **Machine learning**, concentrating more than 35% of the activities among all EU28 MS. Germany and France hold again a substantial share of activities, with the Netherlands and Italy completing the top-5 ranking.

In **Robotics and Automation**, the behaviour is similar to other thematic areas for the **United Kingdom, Germany** and **France**, in this case followed by **Spain** and **Italy**.

Connected and Automated Vehicles has captured the focus of R&D and industrial **German** activities, holding more than 40% of EU28's total activities in this key area, and being ahead of the United Kingdom by more than 25 percent points. They are followed by Sweden and France.

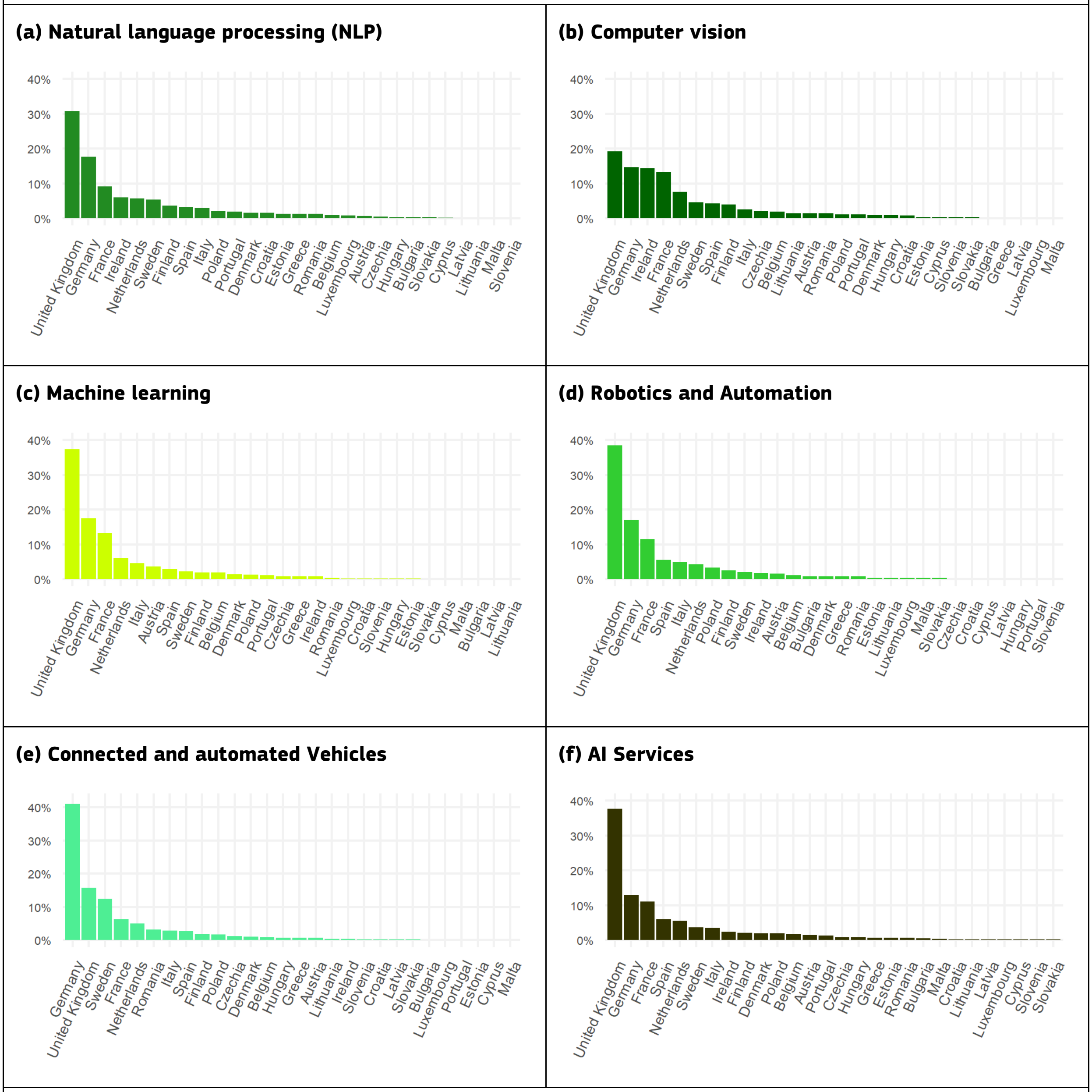
In **AI Services**, the **United Kingdom** is again in the spotlight with almost 38% of activities out of the total EU28 R&D and industrial activities in this thematic area. Germany, France, Spain and Netherlands follow.

Succinctly:

- **Germany and United Kingdom are the most active** and leading in **every thematic area**, based on their relative number of activities. Beyond an obvious size effect, this could be explained because they hold the majority of tech talent that migrates throughout Europe (4.6% and 3.9% respectively of the global AI experts) and they have an advanced manufacturing sector and defined AI national strategies. Additionally, they both appear in the top 10 countries worldwide in terms of number of AI firms, according to both our study and other repots (Atomico, 2018; Plattform Industrie 4.0, 2015; BDI, 2015; Federal Government, 2018; HM Government, 2017; China Institute for Science and Technology Policy at Tsinghua University, 2018).
- **The United Kingdom holds the first place in every area, apart from the *Connected and Automated vehicles***. UK's excellence is expected as 10% of its GDP is expected in 2030 to be sourced from AI activities (Indian National Strategy for Artificial Intelligence #AIFORALL, 2018). By 2025, the government plans to fund over 1,000 PhD researchers, and a Turing fellowship to establish a strong basis for AI studies and development (National Strategy for Artificial Intelligence #AIFORALL, 2018). Since 2010, it holds with Australia an approximately 8% of the worldwide AAAI conference publications, while China has approximately 20% and the US approximately 50% (Ding, 2018; Koshiba et al., 2016).
- **France is ranking among the top-5 in all thematic areas**, which is expected as it holds the 16.5% of the AI experts worldwide, one public research organisation in the top-20 research institution with AI activities (CNRS), and is present in the top-10 of the worldwide AI enterprises according to both our study and other sources (China Institute for Science and Technology Policy at

Tsinghua University, 2018; WIPO, 2019). Moreover, based on their AI national strategy, the AI activities are promoted by the allocation of 1.5 billion euros on AI during the next 5 years (Parliamentary Mission, 2018). France's strong global AI position can be also found subsequently in our study in the regional thematic hotspot analysis (subsection 4.4).

Figure 13. EU28 national hotspots of thematic key areas: country activities as a percentage of EU activities in each key area, 2009-2018.



Notes:

The Thematic Hotspot Indicator (THI) for country C_i and topic k_z is defined as:

$$THI_{C_i,k_z} = \frac{A_{C_i,k_z}}{\sum_i A_{C_i,k_z}} = \frac{\textit{sum of activities of country } C_i \textit{ in topic } k_z}{\textit{sum of activities of countries worldwide in topic } k_z} x100$$, values 0 to 100%.

where A_{C_i,k_z} is the number of activities of country C_i in topic k_z, defined as the sum of activities of all the country's players in said topic: $A_{C_i,k_z} = \sum_j A_j^{C_i,k_z}$

Source: JRC PREDICT- AI TES Dataset, 2019.

- **Netherlands** is in the top-5 leading EU countries in five of the six thematic areas, except for Robotics and Automation, where it ranks sixth.

- Another three EU MS appear among the top-5 positions in two areas: **Spain** in Robotics and Automation and AI Services; **Italy** in Robotics and Automation and Machine learning; and **Ireland** in NLP and Computer vision. The presence of Italy and Spain in the AI field could be explained by their industry's strong presence, as they appear in the top 20 countries with AI enterprises worldwide, according to both our study and other reports (China Institute for Science and Technology Policy at Tsinghua University, 2018).

4.4 Worldwide regional thematic hotspots overview

The regional **Normalised Thematic Hotspots Indicator** (nTHI) analyses the participation of a region in each thematic key area worldwide. In particular, it explores **which regions are leading in each key area.** The nTHI is based on the THI as defined in subsection 4.2, applied at regional level (instead of country level) and normalised in the interval [0,1][25].

Let A_{R_i,k_z} be the number of activities of region R_i in topic k_z, defined as the sum of activities of all the region's players in said topic: $A_{R_i,k_z} = \sum_j A_j^{R_i,k_z}$, then the Thematic Hotspot Indicator (THI) for region R_i and topic k_z is defined as:

$$THI_{R_i,k_z} = \frac{A_{R_i,k_z}}{\sum_i A_{R_i,k_z}} = \frac{N. of\ activities\ of\ region\ R_i\ in\ topic\ k_z}{sum\ of\ activities\ of\ regions\ worldwide\ in\ topic\ k_z} x100$$

and the regional normalised indicator, nTHI, is defined as:

$$nTHI_{R_i,k_z} = \frac{THI_{R_i,k_z} - min_i(THI_{R_i,k_z})}{max_i(THI_{R_i,k_z}) - min_i(THI_{R_i,k_z})}$$

Based on the nTHI (a detailed table can be found in Annex 2, coloured by intensity of the indicator for each region), the following points are made (Figure 14):

Chinese regions are the absolute leaders worldwide in most thematic areas, although the first position corresponds to California in two areas (Robotics and Automation and AI Services). Multiple Chinese regions are involved in AI, with the strongest being **Beijing**, **Guangdong**, **Jiangsu** and **Shanghai**.

The **most prominent states** from the **US** are **California** and **New York**, appearing among the top regions in every thematic area. Washington, Massachusetts, Michigan, Texas and Pennsylvania follow, with involvement in multiple key areas. These hotspots were expected, given the location of US' innovation hubs, universities and research institutes in these states. Moreover, in these states are located six universities and public research organisations included in WIPO's top 20 list (sorted by number of AI articles), namely University of California, Carnegie Mellon University, IEEE, Massachusetts Institute of Technology (MIT), Stanford University, and Georgia Institute of Technology (WIPO, 2019).

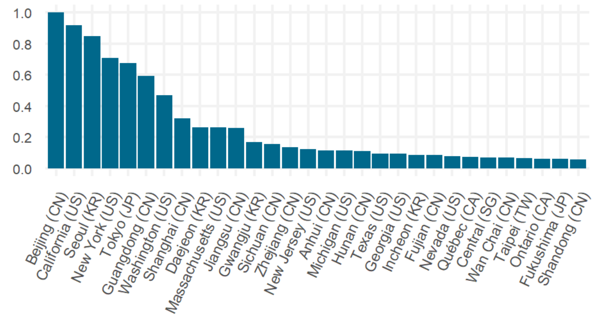
The EU regions among the top 30 per thematic area are from the **United Kingdom, France, Germany, Netherlands** and **Sweden. Switzerland** is also represented among the top worldwide regions by Zürich. The five EU regions present among the top 30 in more than one thematic area are: Inner London-West (UK), Inner London-East (UK), Berkshire, Buckinghamshire and Oxfordshire (UK); Île de France (FR); and Berlin (DE). This European presence shows a concentration in the areas of Robotics and Automation and Machine learning.

Seoul shows a strong leadership in Robotics and Automation in NLP, CAVs and Computer vision, and ranks among the top 30 regions in ML. Tokyo is very actively present in NLP in Computer vision, and less in CAVs, ML, and AI Services.

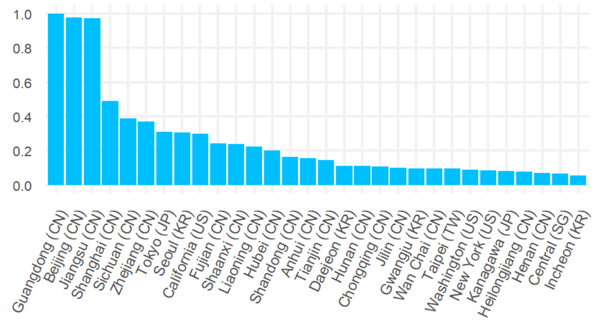
25 The values are normalised [0,1] within each thematic area with the min-max normalisation. The rationale behind the normalisation is the high variability on the number of worldwide regional activities that would impede the visualisation of regions with lower participation in each key area. With the normalisation, these regions are shown, while keeping the differences in the ranges of the original indicator.

Figure 14. Regional hotspots of thematic key areas: region activities as a percentage of worldwide activities in each key area (top-30 regions per thematic key area) (normalised [0,1]), 2009-2018

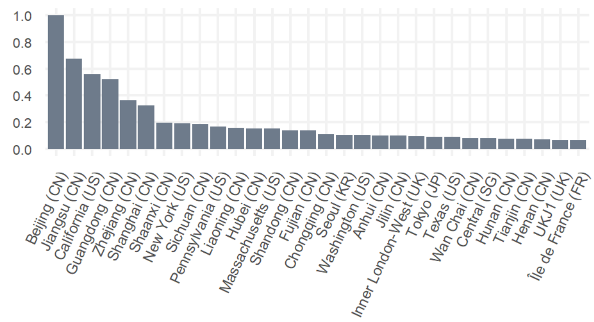
(a) Natural language processing (NLP)



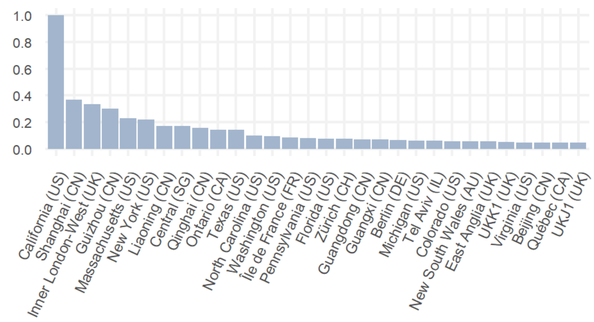
(b) Computer vision



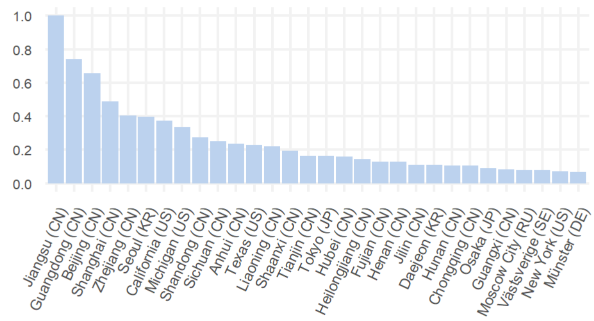
(c) Machine learning (ML)



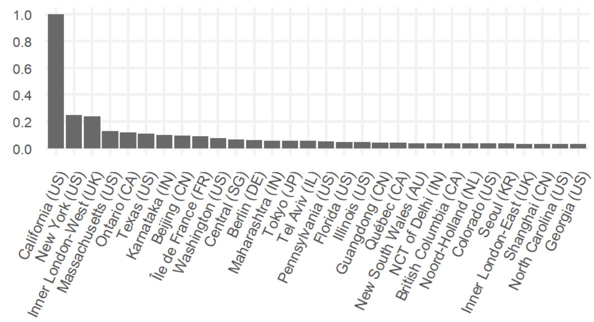
(d) Robotics and Automation



(e) Connected and Automated vehicles



(f) AI Services



Notes:

Top-30 region with highest number of activities per key area.

The region with code UKJ1 is Berkshire, Buckinghamshire and Oxfordshire, the region with code UKK1 is Gloucestershire, Wiltshire and Bristol/Bath area.

The regional Normalised Thematic Hotspot Indicator, nTHI, for region R_i and topic k_z is defined as:

$$nTHI_{R_i,k_z} = \frac{THI_{R_i,k_z} - min_i(THI_{R_i,k_z})}{max_i(THI_{R_i,k_z}) - min_i(THI_{R_i,k_z})}$$

Where THI_{R_i,k_z} is the regional Thematic Hotspot Indicator for region R_i and topic k_z, defined as:

$$THI_{R_i,k_z} = \frac{A_{R_i,k_z}}{\sum_i A_{R_i,k_z}} = \frac{sum\ of\ activities\ of\ region\ R_i\ in\ topic\ k_z}{sum\ of\ activities\ of\ regions\ worldwide\ in\ topic\ k_z} x100$$

where A_{R_i,k_z} is the number of activities of region R_i in topic k_z, defined as the sum of activities of all the region's players in said topic: $A_{R_i,k_z} = \sum_j A_j^{R_i,k_z}$

EU-funded activities are illustrated with grey when applicable.

Source: JRC PREDICT- AI TES Dataset, 2019.

5 AI industrial capacity

In this section, the attention is focused on the firms detected to be involved in AI. These, firstly, can be firms with a core business related to AI, so mainly supplying AI-related goods or AI-services. Secondly, the analysis also includes enterprises which use AI -as evidenced by AI research or patenting activities of the firm. This can be either to offer goods or services with embedded AI (e.g. film streaming service with an AI powered recommendation systems) or developing AI to improve their management or production processes. In order to infer the AI industrial capacity worldwide, we present results regarding the location of firms, their age and main sector of economic activity. In addition, we analyse the firms' involvement in different thematic areas and the use they make of AI with respect of their economic activity.

The first subsection is dedicated to a worldwide overview, in which the attention is mainly focused on the most relevant economic areas: the US, China, the EU28, India, Canada, Japan and South Korea. Then, the second subsection focuses on EU Member States.

5.1 Worldwide overview

5.1.1 *Firms' location and performance*

The AI industrial landscape is mainly dominated by the US, China and the EU28, hosting almost 75% of worldwide AI firms, as presented in Figure 15(a). Following the top 3 areas, India and Canada reveal to be relevant countries in the AI landscape, along with South Korea. The position of Japan, which in terms of number of firms involved in the supply of AI-related goods or services is modest, is very relevant regarding its patenting performance. The distribution of innovating firms, i.e. with at least one AI patent application filed during the observed period (Figure 15(b)), shows China in a very strong leading position. Almost 60% of all AI patenting firms are from China. Only 6.5% are from the EU28, considerably less than the US and just behind South Korea.

Figure 15. Worldwide AI firms by geographic area (%): total firms and patenting firms, 2009-2018.

(a) All AI firms

(b) AI firms with at least one patent application in AI

Source: JRC PREDICT- AI TES Dataset, 2019.

5.1.2 *AI industry composition: age, size and sector*

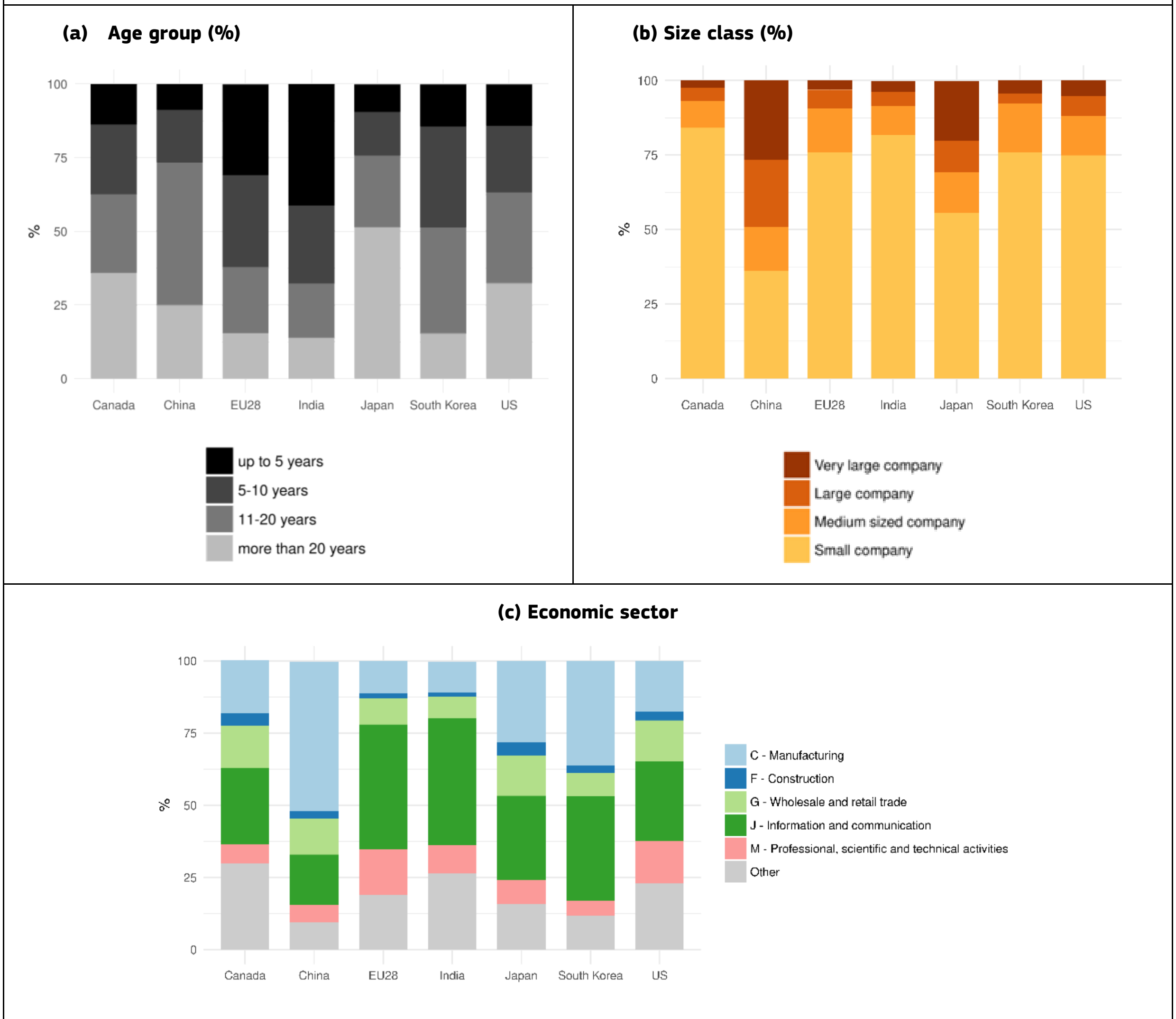
The set of firms involved in the supply of AI-related goods and services is here discussed with reference to their age, size and sector of economic activity. Figure 16(a) presents the corresponding values for the most important macro areas. India and the EU28 hold the largest share of youngest firms involved in AI (up to 5 years), which is around 40% and 30% respectively (Figure 16(a), colour black). The same age group, for all other areas, accounts for a much smaller percentage, and it is never higher than 15% (South Korea). Also the percentage of EU28 firms between 5 and 10 years is remarkable (30%) and presents values similar to those presented by South Korea (34%) and India (26%). In conclusion, the involvement of EU28 firms in AI seems to be a relatively recent phenomenon that strongly took off only in the last ten years.

When moving the attention to the firms founded between 11 and 20 years ago, the areas presenting the largest percentages are (in decreasing order): China, South Korea and the US, which therefore appear to have started their involvement in AI in the first decade of the new millennium. While South Korea and the US present more balanced proportions among the percentages of firms born in different periods, China presents a massive number of AI firms (almost 50%) born approximatively in the first decade of the century.

Finally, the largest percentage of old firms (born at least 20 years ago) is held by Japan (50%), which seems to have had a pioneering role in the development of AI industry. Another interpretation is that already established firms may have successfully taken up AI in later stages. This suggests that while in Japan AI is significantly spread in the economy, this process didn't lead to a wave of new firms foundation, as testified by the small number of active firms detected in the period 2009-2018 (Figure 15) and by the low share of Japanese firms detected to be born in the last 5 or even 10 years.

Regarding the size of firms (for the details on size categorisation see Annex 6), presented in Figure 16(b), the analysis reveals the presence of larger firms in China and Japan. In particular, half of Chinese companies are large or very large (dark orange and brown, respectively, in Figure 16(b)). The EU28, South Korea, Canada and the US, show a different size structure, with a higher involvement of small firms (light orange) that accounts for about 75%, in each of these areas. This gives evidence of a different pattern regarding the manner in which AI techno-economic processes and industry are structured and developed in these economies.

Figure 16. AI firms by geographical area and age group, size class and economic sector (%), 2009-2018



Source: JRC PREDICT- AI TES Dataset, 2019.

Finally, the consideration of the main economic sector in which AI TES firms are operating, also reveals different patterns among major areas, as presented in Figure 16(c). First of all, the large percentage (50%) of AI Chinese firms belonging to the Manufacturing sector (C) is substantially higher than in any other area. This result confirms the traditional strength of China in the ICT manufacturing sector in all macro-economic variables (Mas et al. 2019) and what observed on Chinese trade activity: the main involvement of China in AI is related to the production and development of hardware, not only in the form of complete computers but also chips, electronic components and sensors (Righi et al. 2017). This is also in line with what observed in section 4.2, where China reveals to have involved strong specialisation in Computer vision and CAVs, which are technological subdomains strongly related to manufacturing processes.

On the other hand, the EU28, the US and India present a sensibly large percentage of firms belonging to the Information and Communication sector (J) and to the Professional, Scientific and Technical Activities (M). This is in line with what is observed in the analysis of the key areas of specialisation, where the EU28, the US and India differ with respect to China especially because of their leading involvement in AI Services.

Regarding South Korea and Japan, they present a more balanced distribution between firms belonging to sector (C) and firms belonging to sectors (J) and (M). In particular, South Korea presents more than 30% of firms in the Manufacturing sector, reflecting the strength of the ICT manufacturing sector in this country (Mas et al. 2019). South Korea, as well as the EU28, appears to have only a modest percentage of firms from the Wholesale and Retail Trade (G), a sector in which the US, Canada, Japan and China have a more prominent presence.

5.1.3 *AI in the industrial processes: core businesses and technological development*

In this subsection we differentiate between firms according to their degree of developing capacity and use of AI. For this purpose, we have grouped AI firms into three categories based on their business description and patenting activity. This categorisation allows the distinction of different uses of the AI technology in the economic activities firms perform.

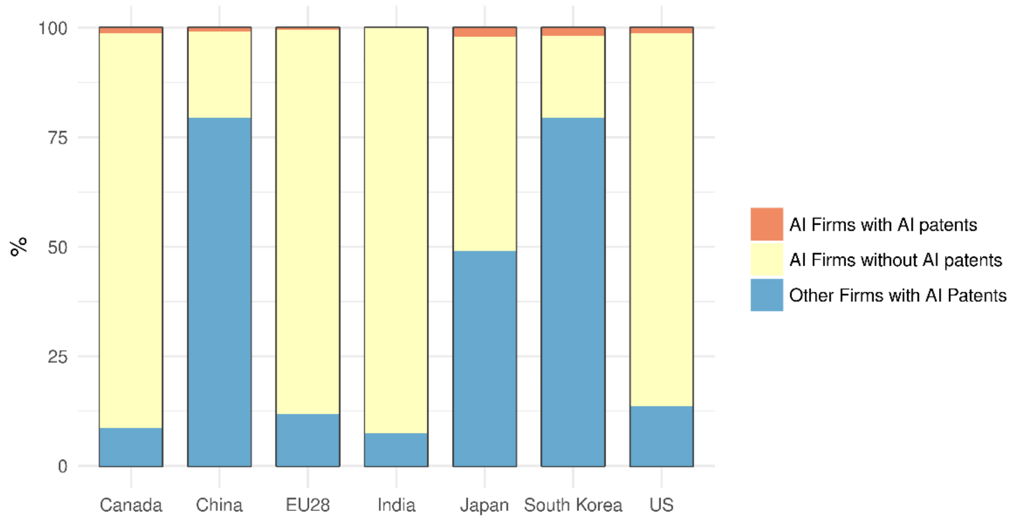
i. Firms having a **core business related to AI and developing AI patents** ("Big Tech", "AI firms with AI patents", orange in Figure 17). These are companies whose core business is strictly linked to AI, and which are producing technological developments in the field of AI. These are mainly large hi-tech companies that have placed AI at the core of their business. They make up 1% of all AI firms and more than 50% of them is active in the ICT sector. Examples of firms that qualify for this group are Alcatel and Deepmind in the EU; Tencent and Baidu in China; Google, Facebook, Microsoft or Salseforce in the US. While the US has the highest number of firms in this category, Japan and South Korea have the highest percentages of big AI tech firms.

ii. Firms having a **core business related to AI and not developing patents** ("AI Firms without AI patents", yellow in Figure 17). Firms which are mainly based on AI technologies and whose main economic activity deals with the supply of AI services, solutions and products, but that do not contribute with the innovative technological developments. They represent the majority of AI firms (67 %), they tend to be smaller and younger firms and are disproportionately active in the sector Professional, Scientific and Technical activities, though the majority (34%) belong to the Information and Communication sector.

iii. firms whose **core business is not related to AI but are developing AI-related patents** ("Other Firms with AI patents", blue in Figure 17). These are firms that are integrating AI technologies in their products or production processes. Even though AI is not their main economic activity, the firms in this group have a crucial role in contributing to the implementation and development of AI throughout the whole economy. 31% of AI firms belong to this category, out of which more than 40% are active in the manufacturing sector. Examples are Dyson, Shazam, Hisense, China Petroleum, Netflix and Bank of America.

Figure 17 shows a very varied distribution regarding business activity in AI across geographical areas. On the one hand, in the US and the EU28 the large majority of firms have a core business that is AI-related and they are not involved in the core development of the technology, as they do not patent. On the other hand, China and South Korea have a large quantity of firms that contribute to technological development without having a core business focused on AI, and much less firms that can be classified as core AI-firms. Considering the textual analysis presented in section 4 and the analysis of firms' economic sector presented in Figure 16 (c), this evidence confirms that the EU28 and the US are characterized by having much more economic activities providing AI services and solutions (e.g. data processing, ML platforms), both to private customers (B2C) and

Figure 17. AI firms by geographical area and relation between core business and use of AI (%), 2009-2018



Source: JRC PREDICT- AI TES Dataset, 2019.

to other businesses (B2B). On the other hand, Chinese firms reveal to be much more focused on the technological development of the fundaments of AI, e.g. concerning the integration of processors and sensors in more complex machines (for instance in NLP, Computer vision, CAVs) and the development of new algorithms (i.e. machine learning).

5.2 AI Industry in the EU28 Member States

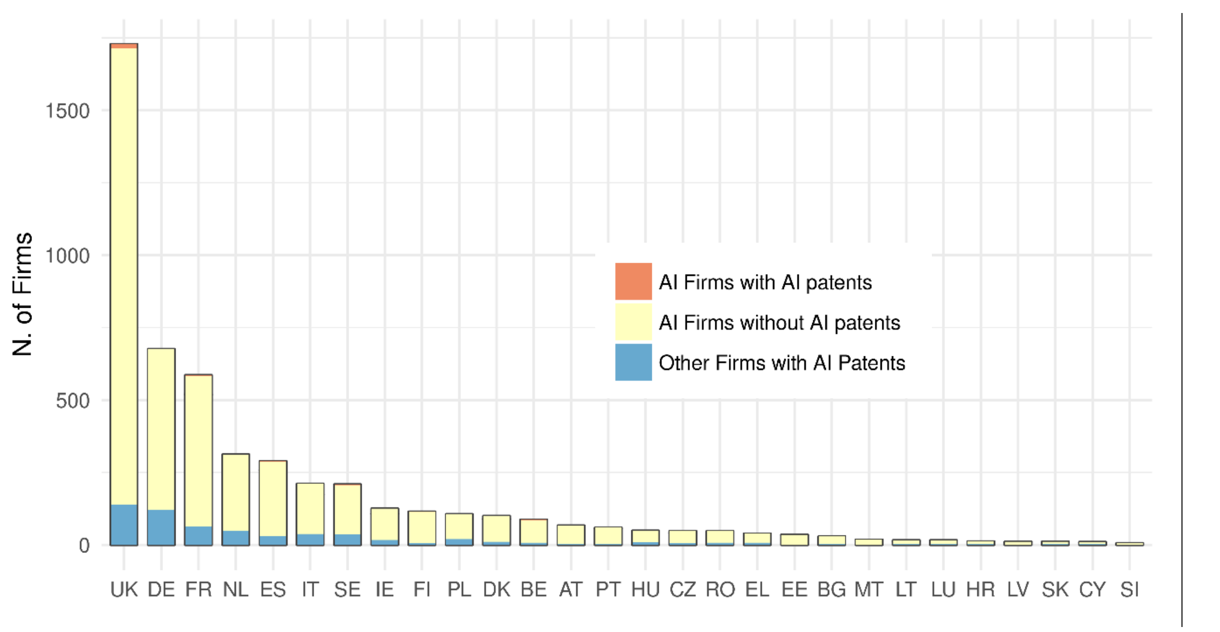
This subsection is devoted to the analysis of the characteristics of AI industry in the EU28 Member States. First we identify the countries with highest number of AI firms, then we discuss their use of AI and the main AI industry composition features.

5.2.1 *AI in the EU28 MS: location of firms and relation between AI and the core business*

Regarding the geographical distribution of AI firms in the EU28, it is observed to be very concentrated in few countries. Figure 18 shows that the country with largest amount of firms is UK, and accounts for about one third of the overall number of AI firms detected in the Union and for half of the EU's ''Big Tech" ("AI firms with AI patents") firms. UK AI firms almost equal the number of firms detected in Germany, France, and Netherlands summed together. It is therefore clear the relevance of UK in the landscape of the EU AI industry. Only France has otherwise a significant contribution to the European big tech firms (14%). Other countries with relatively high number of AI firms include Spain, Italy and Sweden.

The analysis of how the firms in the different Member States are dealing with AI technologies, presented in Figure 18, confirms for all countries what observed at aggregated level in the EU28: most firms have a core business in AI but not active in filing patents (colour yellow). Nevertheless, some differences can be observed among MS. In particular, it is relevant to observe that UK, which is the leader in terms of absolute number of firms, presents a modest share of firms developing AI patents. Among the most relevant countries in terms of number of AI firms, Germany (DE), Netherlands (NL), Italy (IT), Sweden (SE) and Poland (PL) present a share of firms developing AI patents (colour blue) that is almost twice the one of the UK. This shows that the general orientation of EU28 MS towards a large use of AI for the provision of services and solutions to customers (B2C) and other businesses (B2B) is even more pronounced than for the UK.

Figure 18. AI firms by relation between core business and use of AI (%). EU28 Member States, 2009-2018



Source: JRC PREDICT- AI TES Dataset, 2019.

5.2.2 ***Age, size and sector of EU28 Member States' AI firms***

The consideration of the age of the firms involved in AI, as presented in Figure 19, presents several interesting elements. First of all, Estonia appears as the most emerging country, in the sense that more than 50% of its AI firms have been founded during the last 5 years. This observation is not surprising considering the investments of Estonia (EE) in the development of public digital services and infrastructures with the aim of integrating a fully integrated information society (Margetts, H. & Naumann, A., 2016). Then, regarding the UK, it is remarkable that almost 75% of its AI firms were born in the last 10 years (black and dark grey). A similar pattern is also found for the Netherlands (NL), Spain (ES), Ireland (IE), and Finland (FI). Therefore, not only these countries reveal to be relevant in terms of the number of AI firms they host, but their involvement in AI is the result of a process that began recently and leads to new entrepreneurial activity. Other relevant countries, like Germany (DE), France (FR) and Italy (IT), present a larger percentage of old AI firms (grey and light grey). In particular, the presence of young AI firms in France is the smallest observed in the EU28 MS with at least 30 AI firms.

Regarding the size of AI firms, few differences can be noted among MS. All the countries have at least 60% of firms that are small (light orange). The presence of very large and large firms (brown and dark orange respectively) is modest, while the presence of medium size firms (orange) is slightly higher. Among the countries with the highest percentage in the largest firm group are Poland (PL), Sweden (SE) and Denmark (DK). The group of countries presenting the largest involvement of non-small companies are Sweden (SE), France (FR) and Italy (IT), though the small companies still account for over 60%. On the other end of the scale, UK presents a percentage of small firms that is only smaller than what is observed for Estonia (EE), Hungary (HU), Belgium (BE), Netherlands (NL) and Czechia (CZ). This reinforces the preceding findings about the UK industrial system related to AI, as being an emerging and widespread phenomenon mainly driven by young and small services-providing firms.

About the economic sectors in which AI EU28 MS firms are involved, the situation is more diverse. The main element constantly observed in all the economies of the MS is the presence of at least 25% of firms belonging to the Information and Communication sector (J), with the only exception of Hungary (HU). The highest percentage of 50% or more of AI firms in the ICT sector are found in Estonia (EE), Bulgaria (BG) and Finland (FI). The number of firms belonging to the Manufacturing sector (C) is modest, and related to those countries in which the Information and Communication sector (J) is not outstanding. This is the case of Austria (AT), France (FR), Greece (EL), Belgium (BE), Italy (IT), and Germany (DE).

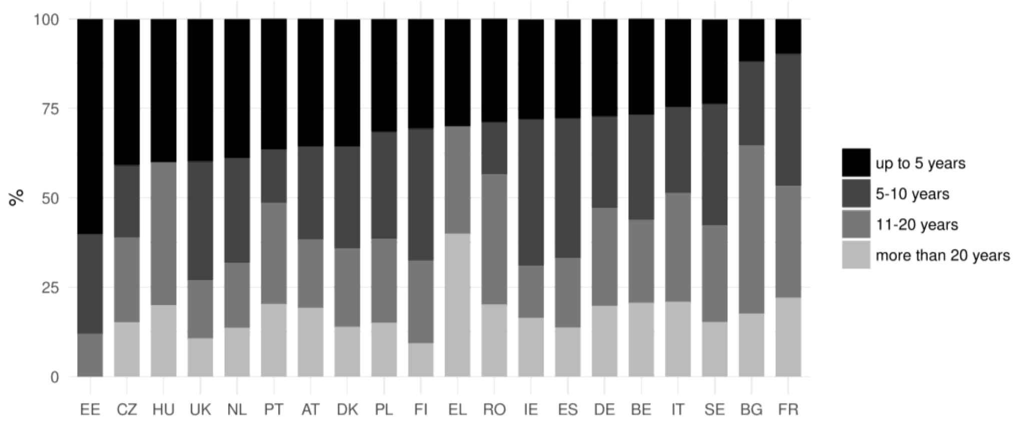
The presence of firms from the Wholesale and Retail trade sector (G) seems to couple with the presence of firms from the Manufacturing sector (C). In fact, in many cases, the two sectors appear simultaneously and with similar size. This is what is detected in Austria (AT), France (FR), Belgium (BE), Italy (IT), Czechia (CZ), Portugal (PT), and Denmark (DK).

A relevant sector for almost all EU28 MS countries is the Professional, Scientific and Technical Activities sector (M). Among the countries with at least 30 AI firms, only Bulgaria (BG), Ireland (IE), Poland (PL), Italy (IT), and Greece (EL) do not present firms in the Professional, Scientific and Technical Activities sector (M).

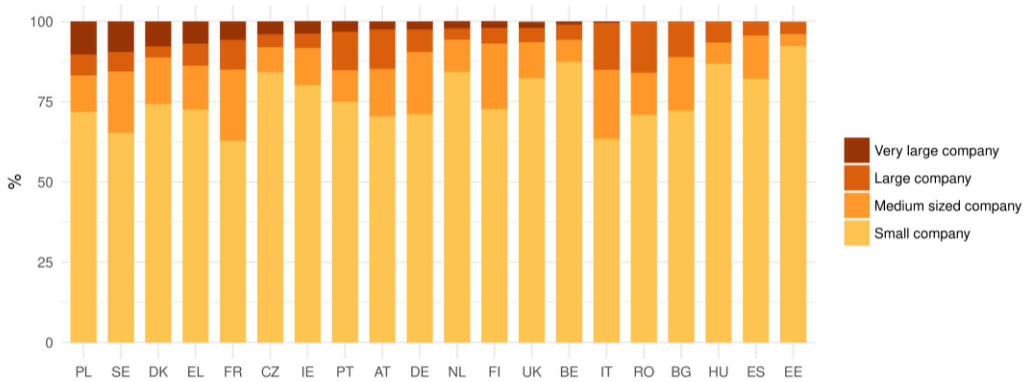
Figure 20 shows the structure of the European firms by age group. Not surprisingly, we can observe a strong positive correlation of firms size with age. It is older firms that generate large turnover and high numbers of employment (Figure 20 a). Patenting activity decreases with firm age (Figure 20 b), which is in line with the literature that states that older firms patent more, but on the other hand these patents tend to be more incremental while younger firms invest in riskier innovations with a higher potential of innovative quality (Balasubramanian, N. & J. Lee (2008), Coad, A. et al. (2016)). The other view of the same picture is that the younger the AI firm the more likely they will see AI as their core business ("AI firms with no patents"). The smallest group of firms, which patent and have AI at the core of their business activity ("Big Tech"), increase by percentage the younger the age group is. However, this group disappears completely within the youngest group (up to 5 years), which may be partially explained by the patenting lag. Another difference between younger and older EU AI firms is the industry sector. While AI firms older than 20 years are mostly manufacturing firms, the younger the firm the more likely they will be an ICT company, while manufacturing, construction and trade decrease with the firm age group. The new entrepreneurial activity in AI in the EU is, hence, largely driven by the ICT sector.

Figure 19. AI firms by age group, size class and economic sector (%). EU28 Member States, 2009-2018

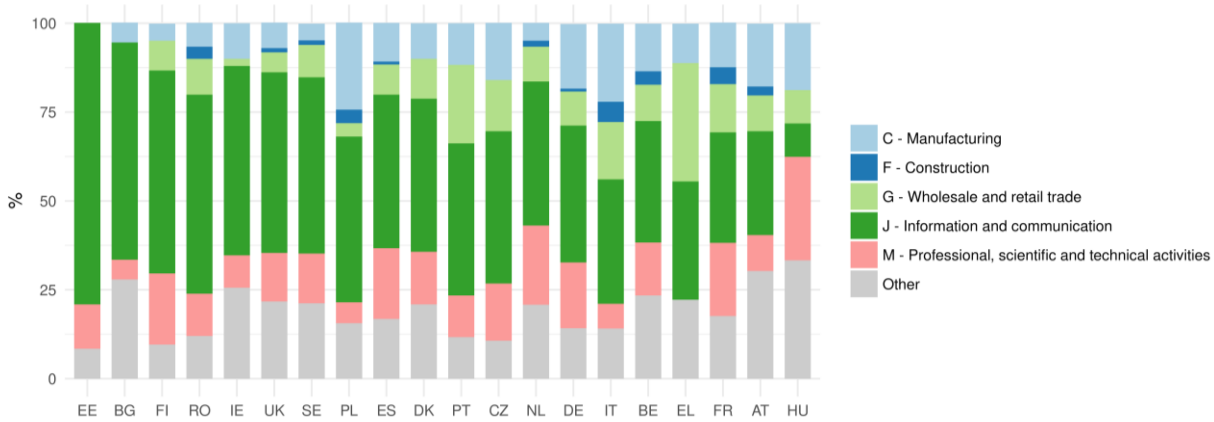
(a) Age group



(b) Size class



(c) Economic sector

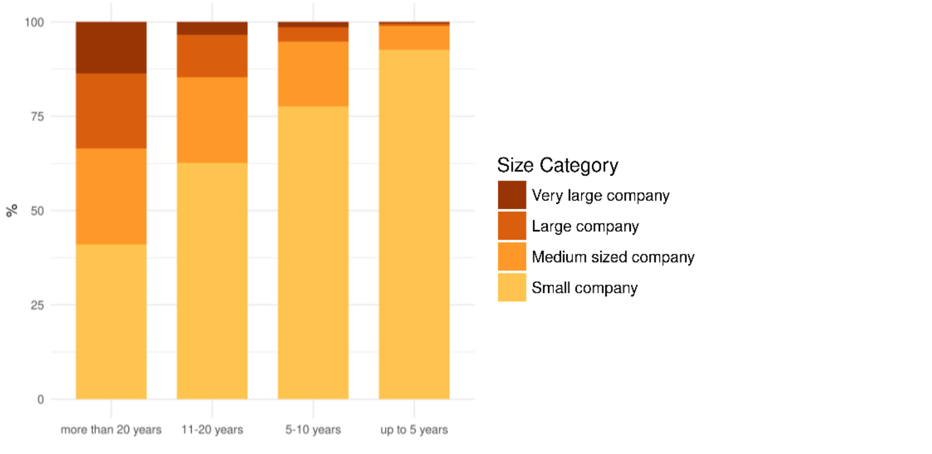


Note: Only Member States with more than 30 observations shown.

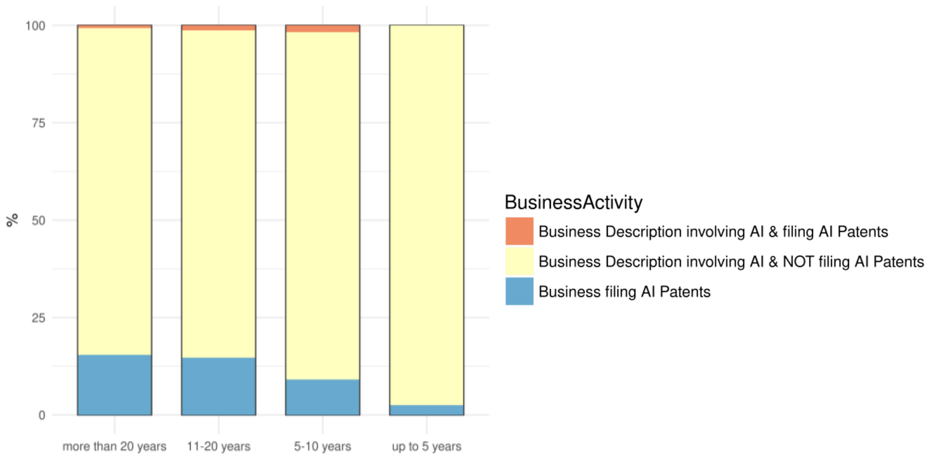
Source: JRC PREDICT- AI TES Dataset, 2019.

Figure 20 EU28 AI firms by economic sector, business activity and size class (%) in each age group, 2009-2018

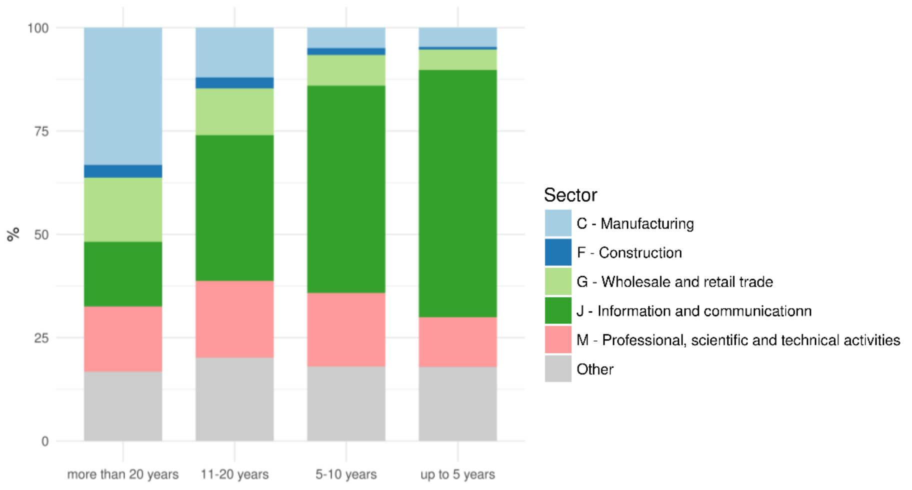
(a) Age group and size class (%)



(b) Age group and relation between core business and use of AI (%)



(c) Age group and economic sector (%)



Source: JRC PREDICT- AI TES Dataset, 2019.

6 Attractiveness & Excellence in AI research

This section provides a selection of indicators on attractiveness and excellence in research as addressed in the EC communications in 2018 (European Strategy on AI, Coordinated Plan on AI), and the Declaration of Cooperation on AI signed on 10 April 2018 by 24 Member States and Norway[26]. In the first part, we present the worldwide map of players involved in AI R&D processes. It should be noted that these players are not exclusively participating in R&D processes, as they can also be involved in industrial activities. Subsequently, certain indicators based on players' performance are developed so as to assess the countries' profile in R&D. As the EC Coordinated Plan on AI, and the Declaration of Cooperation point at fostering research cooperation, EU28 regions are further evaluated by analysing the network (and its subnetworks) of AI R&D collaborations. Finally, the most relevant regions are also studied concerning their AI thematic areas of specialisation.

When research activities are mentioned in this section, we target the participation of detected players in (i) patenting activities, and/or (ii) publications in AI conferences (see Annex 5). These two sub-types of activities, namely patenting and publications, are selected as the most representative R&D outcomes. With patenting activity we intend to capture industrial developments in the field. With publications we identify the most important theoretical advancements and other results of academic research. We also refer to this second sub-type of activities as "frontier research". As information regarding EU-funded projects is available, it is considered when analysing EU28 countries and the R&D network of EU regions. However, EU-funded projects are not used in the worldwide landscape analysis, as information on other national public policies is not available.

6.1 Location of Worldwide R&D players

The players involved in AI R&D processes are depicted in Figure 21, where their geographical density is represented from blue (low density) to red (high density). As already mentioned in other sections, **R&D players in AI are mainly located in China, the US and Europe**, with different distributions over the territory depending on the country. While the map can provide some confirmations, more detailed information at country level is represented in Figure 22(a), where R&D players' organisational type is considered, and Figure 22(b-d), where the geographical distribution of subgroups of R&D players is represented. As the R&D activities considered in this work are patents and frontier research publications, the presence of players by country is also analysed considering the two types of activities separately.

this is line with the fact that China is very active in this field (WIPO, 2019). This is also the main reason behind the large presence of Chinese players involved in AI R&D activities (Figure 22(a-b)). The presence of South Korea is moderate (7.04% of AI patenting players, in Figure 22(c)) while the US's is modest (12.33% of AI patenting players, in Figure 22(c)). However, as discussed in subsection 3.1.2, even if the number of US patenting players is not so large, their productivity is greater than any other country´s (3.2 patents per patenting player, in Figure 6(c)).

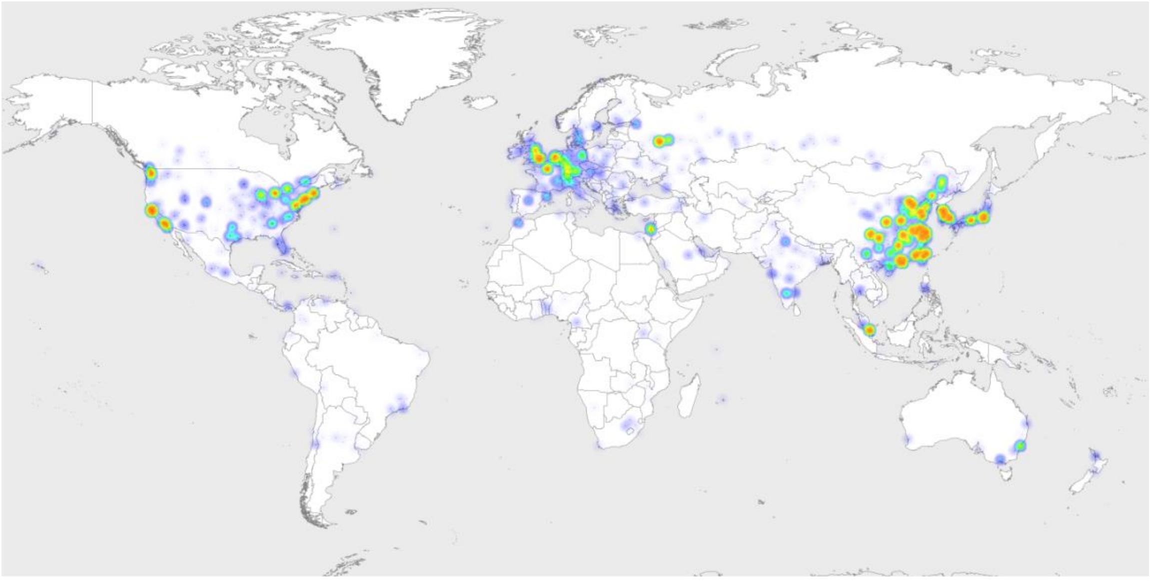
The landscape is different if we exclusively consider players with publications on AI related key areas. **EU28 is the most considerable area, with 34% of the total AI players** worldwide (Figure 22(d)). The United States follow closely, whereas China holds only 9.32%.

The presence of **firms participating in AI R&D processes is noticeable** (Figure 22) **in the US**. China and EU28 in fact both present a larger share of research institutions (Figure 22(a)). The position of South Korea is moderate, with 6.35% of worldwide AI R&D players (Figure 22(b)), mostly firms. The **percentage of Chinese patenting players** (over the overall number of patenting players) **is very large** (Figure 22(c)), and Worldwide countries' R&D score.

In this subsection, we compute and analyse a score based on the involvement of players in AI R&D activities, to evaluate the performance at country level. This indicator is based on the number of research activities (patenting and frontier research) in which the players of the corresponding geographic areas participate. The analysis is presented for worldwide countries.

26 The most relevant sentences are: "Europe is home to a world leading AI research community", "should strengthen its status as a powerhouse", "support AI technologies both in basic and industrial research". "Cooperate on reinforcing AI research centres and supporting their pan-European dimension". *"Europe will scale up national research capacities and reach a critical mass through tighter networks of European AI research excellence centres. The objective is to foster cooperation among the best research teams in Europe, joining forces to tackle more efficiently major scientific and technological challenges in AI and mobilise industry to be integrated in and find synergies with the research teams".*

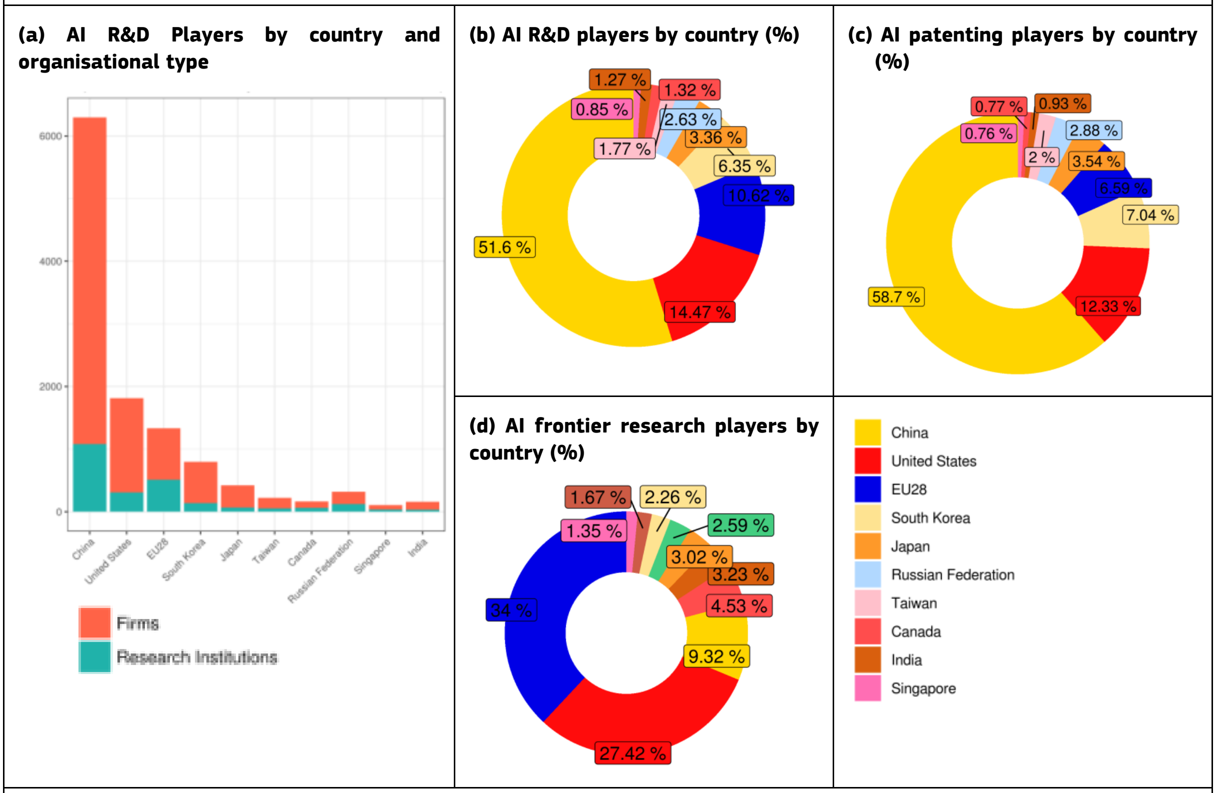
Figure 21. Density map of players involved in AI R&D activities, 2009-2018.



Note: Colours from blue to green to yellow to red progressively indicate a denser concentration of players involved in AI R&D processes.

Source: JRC PREDICT- AI TES Dataset, 2019.

Figure 22. Indicators on AI players in R&D activities, 2009-2018



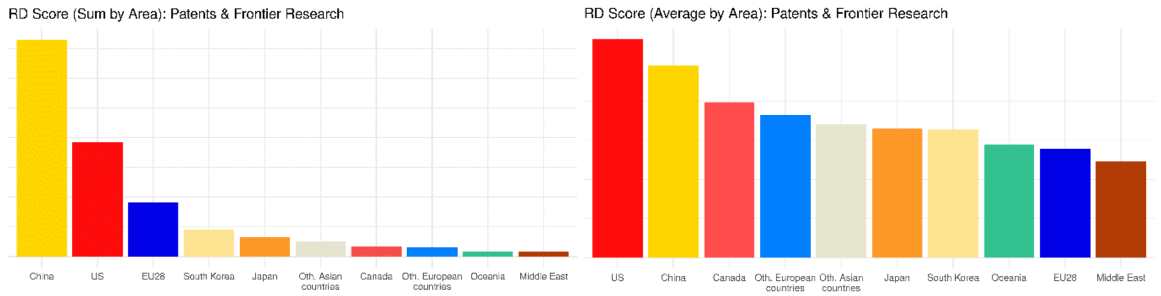
Source: JRC PREDICT- AI TES Dataset, 2019.

The R&D score proposed in this subsection is computed by considering the fractional count[27] of the number of activities in which each player participated (both shared activities and individual activities). As several types of activity are considered (patents, publications and, depending on the case, EU funded projects), presenting distributions with different scale and variance, the number of activities is normalised by type. Therefore, the fractional count of patent applications filed by each player is normalised in the interval [0,1] [28], and the same applies for the fractional count of publications and the fractional count of EU funded projects. This normalisation is a methodological procedure to allow different dimensions to have the same scale. Subsequently, for each player, the corresponding normalised values are summed. Finally, to compute the discussed indicators, the sum and the average of the players' R&D scores are calculated based on the geographical area where the players are located.

When the sum of players' R&D score (Figure 23(a)) is considered, three geographical areas emerge: China, the US and the EU28, holding very different scores. **The R&D score of China** is approximatively **twice the US' score** and almost **four times that of the EU28**. This indicates that the volume of Chinese AI R&D activities is considerably higher than other countries'. Nevertheless, this is likely due to **national dynamics**. On patenting, as described in subsection 3.1.2, Chinese AI patents prevail (WIPO, 2019). Regarding publications, although the number of Chinese articles is closely followed by the US, the research impact is reversely proportional according to their H-index of research publications (Ding 2018, McKinsey Global Institute, 2017a), although the impact is not captured by our R&D score.

Figure 23. Players' R&D score by geographical area: sum by area (a), and average by area (b), 2009-2018.

(a) Sum of R&D scores **(b) Average of R&D scores**



Source: JRC PREDICT- AI TES Dataset, 2019.

To further analyse the R&D performance, we compute the average R&D score by area (Figure 23(b)). This reveals that **players from the US are the most active, and that Canada has a considerable role.** The Canadian presence can be explained mainly by the AI experts concentration in the region of Ontario (Toronto) with high ranking universities and innovation hubs, and also by Ontario's position in the worldwide thematic profile (subsection 4.4) (China Institute for Science and Technology Policy at Tsinghua University, 2018). According to this indicator, China ranks second. **The overall activity of European players is significant**, but the **individual propensity to develop patents and/or publications is modest (EU28 ranks ninth for average R&D score)**. Not only the patenting players in the EU28 are less than those in the US and China, but on average they also tend to develop less patents (1.6 patent applications per player in the EU28, vs 3.3 patent applications per player in the US and 2.8 in China, as presented in Table 2).

27 If the activity is performed by more than one player, then the fractional count of this activity is equal to 1 divided by the number of participants. In case the players participating are 4, then for each of them the activity counts 0.25. If the activity is performed by a single player, then its fractional count is 1 as the player is the only one performing that activity.

28 By means of the min-max normalisation.

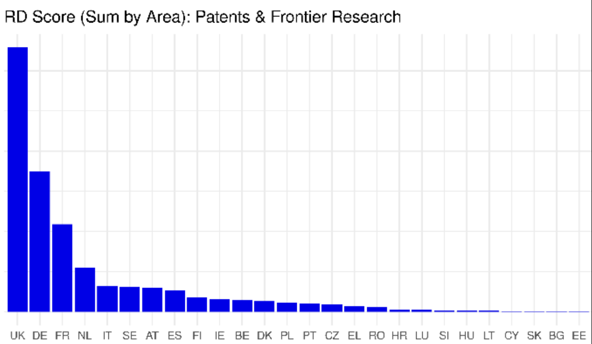
6.2 EU28: R&D excellence in patent, frontier research & EU funded projects

The objective of this subsection is to assess the distribution of the R&D score and relative positions of EU28 MS. The same R&D score as presented in subsection 6.1 is computed for the EU28 countries, with and without considering EU-funded projects (Figure 24 (a) and (b), respectively). For both of them, the sum and average of the R&D scores EU-funded are reported at country level. The first two plots (a1 and a2) illustrate the ranking based on the score of EU28 players with patenting and frontier research activities only. As this indicator is a sum of players' performances at country level, it is able to represent the overall research strength of countries when EU-funded projects are not considered. The second pair of plots (b1 and b2) include also participations in EU-funded projects as a third sub-type of research activity. By considering the two different distributions, also some considerations on the impact of EU-funded projects can be inferred.

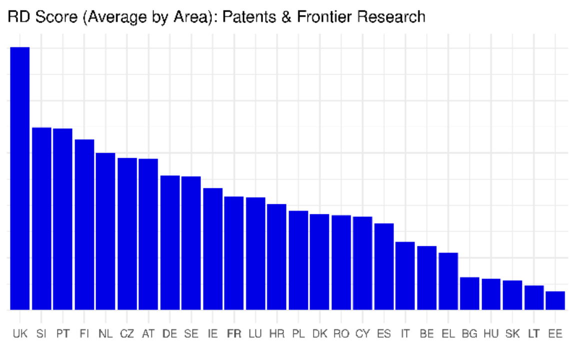
Figure 24. Players' R&D score in EU28 countries: sum and average. (a) Based on patenting and frontier research. (b) Based on patenting, frontier research and EU-funded projects, 2009-2018.

(a) Patenting activity and Frontier research

(a1) Sum of R&D scores

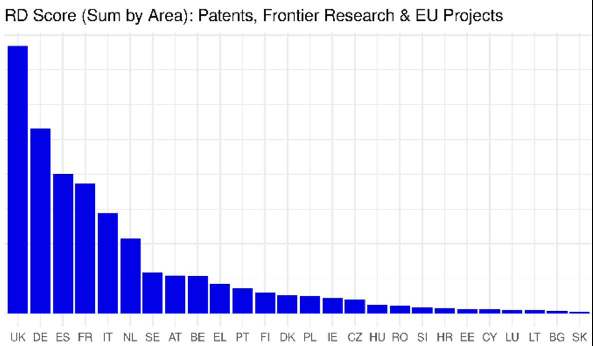


(a2) Average of R&D scores

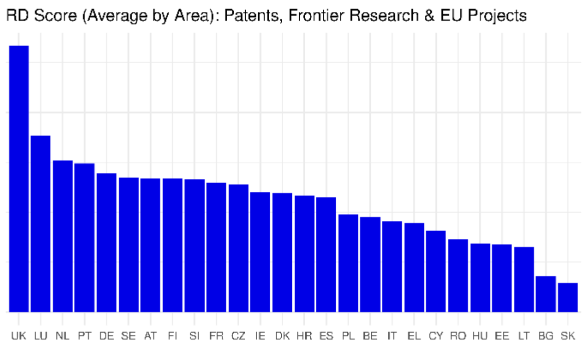


(b) Patenting activity, Frontier research and EU-funded projects

(b1) Sum of R&D scores



(b2) Average of R&D scores



Source: JRC PREDICT- AI TES Dataset, 2019.

When the sum of R&D score (Figure 24 (a1)) is considered based on the fractional count of participations in patents and frontier research, the **leading country is the UK**, followed by **Germany, France and Netherlands**. All the remaining countries are ranked with modest R&D scores, with their ranking marginally different from Table 3, where MS are sorted by number of players (all types). In particular, Spain, that is 5th in the ranking by number of AI players, is now 8th, with Italy, Sweden and Austria ranking higher. The distribution of the sum of R&D scores becomes less concentrated on the highest scoring countries when EU-funded projects are considered (b1). We can conclude that **EU projects assure a more balanced development of**

AI across Member States. Even though the ranking does not change considerably when including EU-funded projects, the difference of the sums of the R&D scores are smaller. **Large relative increases of sums of AI R&D scores** because of **EU projects** are **observed in Spain and Italy**.

Regarding the average R&D score of the players some interesting points emerge when comparing the average **activity of players of different countries**. When EU-funded projects are not considered (a2), UK still presents very active players. Slovenia and Portugal significantly improve their position, if comparing the rank of the average (a2) with the rank of the sum (a1). The position of the latter is likely to reflect Portugal's research funding schemes of the last decades[29]. When EU projects are considered (b2), UK is still leading.

6.3 The EU28 R&D network and its subnetworks: composition, most influential regions and degree of activity

In this subsection, we analyse the R&D activities from a network perspective. As many R&D processes generate collaborations among the involved players, and as one of the objectives of the Coordinated Plan on AI is to "*foster cooperation among the best research teams in Europe"*, we study the structure of relationships in AI R&D activities at regional level (NUTS2). Therefore, all the identified players involved in AI R&D are grouped depending on the region where they are located. This allows the evaluation of connections between regions, based on connections between players located in the corresponding regions. All three types of activities are used for the analysis: EU-funded projects, patents, and publications. These types are considered separately to generate distinct subnetworks (Table 6 (b)-(d)) and also collectively to compute the entire R&D network indicators (Table 6 (a)). This allows the evaluation of how EU players involved in AI R&D activities create synergies at regional level. Hence, this analysis provides insights on the structure of the interactions occurring in the European Union regarding the part of the AI ecosystem aimed at R&D. The results of the top 30 EU28 regions in terms of number of players involved in AI R&D processes are presented. The computed indicators are described as follows (in italic is indicated the name used in the tables):

- Number of players in a region over total number of AI R&D players detected in EU28, in percentage [*% of Players (over EU28)*]
- Percentage of research institutes in the region over the AI R&D players detected in the region [*% of Res. Institute (in the area)*]
- Percentage of firms in the region over the AI R&D players detected in the region [*% of Firms (in the area)*]
- Number of R&D activities in which regional players participated [*N. of Activities*]. We do not consider in this case the fractional count, as the statistic (as it is here computed) corresponds to the degree centrality of the areas
- Number of other regions with which at least one collaboration is detected [*N. of Collaborating Areas*]
- Total number of collaborations[30] in which players belonging to the corresponding region are involved [*N. of Collaborations*]
- Propensity to establish collaborations with other areas, computed as the number of areas with which there are active collaborations, divided by the number of activities developed by the players in the region [*Coll. Areas / N. of Activities*]
- Potential to influence the network, measured in terms of Weighted Betweenness Centrality [*WBC*] [Brandes, 2001], accounting fractional counting and normalized in interval [0-1]
- Average number of R&D activities in which regional players are involved [*N. of Activities / N. of Players*].

29 "The exponential growth of research and development in Portugal in the past 20–30 years is set to continue as the challenges of quality and international competitiveness are addressed." [Seabra, 2014].

30 The number of collaborations differs from the number of activities. While the activities are the initiatives in which players participate (e.g. a patent), the number of collaborations is determined by the number of players commonly participating in the same activity. If for example four players participate in the same patent, we infer the presence of six collaborations (under the hypothesis that they all establish a 1:1 collaboration with any of them).

The indicators are presented in Table 6(a-d). The variables regarding the **presence of research institutions** and **firms are blue**. The variables describing the potential to influence the network, namely the **level of influence to other regions' activities are red**. The variable describing the average **number of activities per player are green**. The different network/subnetworks are indicated by letters (a) to (d) in the subheading of the table. The R&D network and its subnetworks are analysed in this subsection.

We implement certain methodological steps to compute the statistics described in Figure 2. In the analysis proposed in the rest of the subsection, AI players involved in R&D activities are grouped according to the region where they belong, instead of the country as pictured in Figure 2. Nonetheless, the methodological steps for the analysis are the same.

Table 6. Network indicators of the top 30 EU28 regions, by number of AI R&D players: (a) Entire AI R&D network , (b) Subnetwork of EU-funded projects, (c) Subnetwork of patenting collaborations, (d) Subnetwork of frontier research collaborations, 2009-2018.

			(a)									(b)								
			All R&D Network									EU Project Network								
NUTS2 code	Region	Country	% of Players (over EU28)	% of Res. Institute (in the area)	% of Firms (in the area)	N. of Activities	N. of Collaborating Areas	N. of Collaborations	WBC	Coll. Areas / N. of Activities	N. of Activities / N. of Players	% of Players (over EU28)	% of Res. Institute (in the area)	% of Firms (in the area)	N. of Activities	N. of Collaborating Areas	N. of Collaborations	WBC	Coll. Areas / N. of Activities	N. of Activities / N. of Players
FR10	Île de France	FR	3.79%	62%	32%	469	278	4015	0.80	0.59	2.14	2.56%	78%	14%	284	253	3859	1.00	0.89	1.92
ES51	Cataluña	ES	2.53%	83%	12%	250	227	2122	0.16	0.91	1.71	2.16%	90%	5%	194	216	2082	0.33	1.11	1.55
ES30	Comunidad de Madrid	ES	1.99%	77%	17%	191	236	2152	0.10	1.24	1.66	1.68%	89%	3%	164	229	2131	0.18	1.40	1.69
UKI3	Inner London - West	UK	1.85%	49%	49%	343	223	1364	0.08	0.65	3.21	0.64%	84%	8%	143	193	1232	0.09	1.35	3.86
DE21	Oberbayern	DE	1.49%	65%	34%	326	227	2708	0.23	0.70	3.79	0.83%	92%	6%	217	219	2652	0.34	1.01	4.52
ITC4	Lombardia	IT	1.49%	76%	22%	127	213	1459	0.03	1.68	1.48	1.11%	92%	5%	88	198	1415	0.05	2.25	1.38
EL30	Attiki	EL	1.33%	86%	9%	120	209	1520	0.09	1.74	1.56	1.19%	88%	6%	111	207	1512	0.19	1.86	1.61
DE30	Berlin	DE	1.28%	69%	28%	116	188	962	0.02	1.62	1.57	0.67%	95%	0%	70	179	916	0.03	2.56	1.79
NL33	Zuid-Holland	NL	1.23%	56%	35%	128	188	1339	0.07	1.47	1.80	0.76%	84%	2%	79	177	1295	0.04	2.24	1.80
ITI4	Lazio	IT	1.19%	80%	14%	134	211	1760	0.04	1.57	1.94	1.00%	90%	3%	118	199	1738	0.10	1.69	2.03
BE10	Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest	BE	1.06%	54%	31%	95	201	1336	0.02	2.12	1.56	0.95%	51%	33%	83	197	1323	0.04	2.37	1.51
NL32	Noord-Holland	NL	1.04%	65%	35%	173	178	784	0.16	1.03	2.88	0.52%	97%	3%	75	147	698	0.05	1.96	2.50
FR71	Rhône-Alpes	FR	1.02%	66%	29%	100	157	649	0.02	1.57	1.69	0.64%	89%	3%	47	143	572	0.02	3.04	1.27
AT13	Wien	AT	0.99%	81%	14%	108	178	971	0.04	1.65	1.89	0.78%	87%	7%	78	173	946	0.02	2.22	1.73
IE02	Southern and Eastern	IE	0.99%	74%	25%	133	177	937	0.03	1.33	2.33	0.66%	95%	3%	67	168	910	0.08	2.51	1.76
SE23	Västsverige	SE	0.95%	56%	42%	138	174	1184	0.01	1.26	2.51	0.55%	91%	6%	67	168	1148	0.01	2.51	2.09
DE12	Karlsruhe	DE	0.88%	78%	22%	109	182	1005	0.01	1.67	2.14	0.55%	97%	3%	66	178	992	0.06	2.70	2.06
ES21	País Vasco	ES	0.87%	88%	12%	91	184	1583	0.00	2.02	1.82	0.76%	98%	2%	83	182	1574	0.01	2.19	1.89
SE11	Stockholm	SE	0.87%	70%	24%	113	173	995	0.02	1.53	2.26	0.55%	88%	3%	63	161	966	0.01	2.56	1.97
DEA2	Köln	DE	0.85%	76%	16%	125	193	1831	0.05	1.54	2.55	0.62%	83%	6%	99	185	1810	0.09	1.87	2.75
ITH5	Emilia-Romagna	IT	0.78%	80%	20%	62	151	879	0.01	2.44	1.38	0.59%	100%	0%	38	145	849	0.01	3.82	1.12
ITI1	Toscana	IT	0.76%	82%	11%	77	186	1309	0.06	2.42	1.75	0.64%	89%	3%	67	178	1293	0.10	2.66	1.81
NL41	Noord-Brabant	NL	0.76%	80%	16%	70	149	1176	-	2.13	1.59	0.62%	89%	6%	45	143	1163	0.01	3.18	1.25
ITC1	Piemonte	IT	0.74%	86%	12%	87	189	1621	0.02	2.17	2.02	0.66%	95%	3%	73	183	1604	0.05	2.51	1.92
DE71	Darmstadt	DE	0.71%	66%	34%	66	141	630	0.01	2.14	1.61	0.52%	83%	17%	46	136	612	0.01	2.96	1.53
UKJ1	Berkshire, Buckinghamshire and Oxfordshire	UK	0.71%	73%	24%	221	161	546	0.10	0.73	5.39	0.47%	89%	7%	69	133	445	0.03	1.93	2.56
FI1B	Helsinki-Uusimaa	FI	0.69%	70%	20%	93	174	1104	0.00	1.87	2.33	0.54%	77%	10%	57	162	1083	0.02	2.84	1.84
DK01	Hovedstaden	DK	0.66%	53%	32%	74	155	531	0.00	2.09	1.95	0.31%	67%	0%	41	149	514	0.01	3.63	2.28
DE11	Stuttgart	DE	0.64%	62%	35%	127	158	940	0.03	1.24	3.43	0.45%	85%	12%	51	144	900	0.00	2.82	1.96
HU10	Közép-Magyarország	HU	0.64%	70%	27%	36	102	301	0.00	2.83	0.97	0.40%	96%	0%	28	101	277	0.02	3.61	1.22

Source: JRC PREDICT- AI TES Dataset, 2019.

Table 6 (continuation).

NUTS2 code	Region	Country	(c) Patent Applications Network: % of Players (over EU28)	% of Res. Institute (in the area)	% of Firms (in the area)	N. of Activities	N. of Collaborating Areas	N. of Collaborations	WBC	Coll. Areas / N. of Activities	N. of Activities / N. of Players	(d) Frontier Research Network: % of Players (over EU28)	% of Res. Institute (in the area)	% of Firms (in the area)	N. of Activities	N. of Collaborating Areas	N. of Collaborations	WBC	Coll. Areas / N. of Activities	N. of Activities / N. of Players
FR10	Île de France	FR	0.67%	13%	87%	73	23	55	0.02	0.32	1.87	0.76%	57%	43%	112	42	101	0.10	0.38	2.55
ES51	Cataluña	ES	0.21%	25%	75%	16	7	13	0.01	0.44	1.33	0.31%	72%	28%	40	22	27	0.05	0.55	2.22
ES30	Comunidad de Madrid	ES	0.24%	21%	79%	17	3	6	-	0.18	1.21	0.16%	44%	56%	10	12	15	-	1.20	1.11
UKI3	Inner London - West	UK	0.61%	6%	94%	44	29	51	0.00	0.66	1.26	0.64%	54%	46%	156	46	81	0.07	0.29	4.22
DE21	Oberbayern	DE	0.50%	10%	90%	78	15	33	0.03	0.19	2.69	0.26%	87%	13%	31	16	23	0.01	0.52	2.07
ITC4	Lombardia	IT	0.24%	7%	93%	24	15	23	0.00	0.63	1.71	0.17%	70%	30%	15	12	21	0.01	0.80	1.50
EL30	Attiki	EL	0.05%	33%	67%	3	3	3	-	1.00	1.00	0.12%	71%	29%	6	4	5	-	0.67	0.86
DE30	Berlin	DE	0.21%	17%	83%	15	4	17	-	0.27	1.25	0.45%	58%	42%	31	20	29	0.02	0.65	1.19
NL33	Zuid-Holland	NL	0.36%	5%	95%	34	13	33	0.02	0.38	1.62	0.12%	43%	57%	15	7	11	0.00	0.47	2.14
ITI4	Lazio	IT	0.07%	0%	100%	4	9	9	0.00	2.25	1.00	0.14%	50%	50%	12	12	13	0.00	1.00	1.50
BE10	Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest	BE	0.03%	100%	0%	2	1	1	-	0.50	1.00	0.12%	86%	14%	10	11	12	0.02	1.10	1.43
NL32	Noord-Holland	NL	0.26%	0%	100%	15	9	13	-	0.60	1.00	0.28%	69%	31%	83	37	73	0.09	0.45	5.19
FR71	Rhône-Alpes	FR	0.21%	17%	83%	9	7	20	-	0.78	0.75	0.24%	57%	43%	44	25	57	0.03	0.57	3.14
AT13	Wien	AT	0.07%	25%	75%	4	3	5	-	0.75	1.00	0.16%	78%	22%	26	17	20	0.03	0.65	2.89
IE02	Southern and Eastern	IE	0.24%	14%	86%	56	14	23	0.01	0.25	4.00	0.12%	86%	14%	10	4	4	-	0.40	1.43
SE23	Västsverige	SE	0.36%	5%	95%	63	4	27	-	0.06	3.00	0.05%	67%	33%	8	9	9	0.00	1.13	2.67
DE12	Karlsruhe	DE	0.17%	30%	70%	19	1	5	-	0.05	1.90	0.21%	75%	25%	24	8	8	0.01	0.33	2.00
ES21	País Vasco	ES	0.16%	44%	56%	7	1	7	-	0.14	0.78	0.02%	100%	0%	1	1	2	-	1.00	1.00
SE11	Stockholm	SE	0.19%	18%	82%	41	11	20	0.03	0.27	3.73	0.12%	71%	29%	9	6	9	0.00	0.67	1.29
DEA2	Köln	DE	0.14%	25%	75%	8	4	5	0.00	0.50	1.00	0.12%	86%	14%	18	14	16	0.01	0.78	2.57
ITH5	Emilia-Romagna	IT	0.14%	13%	88%	9	8	11	0.02	0.89	1.13	0.07%	50%	50%	15	11	19	0.02	0.73	3.75
ITI1	Toscana	IT	0.07%	0%	100%	5	5	5	-	1.00	1.25	0.07%	100%	0%	5	10	11	0.01	2.00	1.25
NL41	Noord-Brabant	NL	0.10%	17%	83%	19	4	5	-	0.21	3.17	0.05%	67%	33%	6	5	8	-	0.83	2.00
ITC1	Piemonte	IT	0.09%	40%	60%	9	8	10	0.00	0.89	1.80	0.07%	75%	25%	5	7	7	-	1.40	1.25
DE71	Darmstadt	DE	0.19%	18%	82%	15	6	13	-	0.40	1.36	0.09%	80%	20%	5	4	5	-	0.80	1.00
UKJ1	Berkshire, Buckinghamshire and Oxfordshire	UK	0.14%	13%	88%	28	15	39	0.02	0.54	3.50	0.14%	75%	25%	124	41	62	0.06	0.33	15.50
FI1B	Helsinki-Uusimaa	FI	0.07%	25%	75%	18	2	2	-	0.11	4.50	0.12%	71%	29%	18	15	19	0.00	0.83	2.57
DK01	Hovedstaden	DK	0.14%	0%	100%	11	5	7	-	0.45	1.38	0.22%	69%	31%	22	7	10	-	0.32	1.69
DE11	Stuttgart	DE	0.17%	20%	80%	43	4	7	0.01	0.09	4.30	0.12%	43%	57%	33	22	33	0.05	0.67	4.71
HU10	Közép-Magyarország	HU	0.16%	0%	100%	5	4	20	0.01	0.80	0.56	0.09%	80%	20%	3	1	4	-	0.33	0.60

Source: JRC PREDICT- AI TES Dataset, 2019.

EU28 R&D regional network - Table 6 (a)

- Considering the type of the players involved in the most relevant R&D processes, in the majority of EU28 regions the involved players are mainly research institutions. A **higher presence of firms** is found in regions located in **North/West Europe**: Inner London–West (49% of firms), Västsverige (42% of firms), Zuid-Holland (35% of firms), Stuttgart (35% of firms), Noord-Holland (35% of firms), Darmstadt (34% of firms), Oberbayern (34% of firms), Île de France (32% of firms), Hovedstaden (32% of firms), Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest (31% of firms).
- The **most influential** region, given the structure of its relationships to the entire network (measured in terms of WBC), is **Île de France**. Second in terms of capacity to influence other regions is **Oberbayern.**. Other relevant regions are **Cataluña**, **Noord-Holland**, **Comunidad de Madrid** and **Berkshire, Buckinghamshire and Oxfordshire**.

Subnetwork of participations in EU-funded projects – Table 6 (b)

- As expected, this R&D subnetwork is dominated by the presence of research institutes. Two noticeable cases with higher firm presence are Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest and Île de France, with 33% and 14% of firms among the involved players, respectively.
- In terms of potential to influence the activity of other regions, three regions are prominent. In descending order, these are Île de France, Oberbayern and Cataluña.
- Regarding the involvement in EU-funded projects, the areas with the most active players on average are Oberbayern and Inner London–West.

Subnetwork of patenting collaborations – Table 6 (c)

- As expected, this is the only subnetwork where the **presence of firms** is more substantial than the presence of research institutions.
- The areas with the **highest potential to influence** other regions are Oberbayern and Stockholm.
- The **most productive regions** in terms of number of patents per player are in the **north-west of Europe**. These are Helsinki-Uusimaa (4.50 patents per player), Stuttgart (4.30), Southern and Eastern Ireland (4.00), Stockholm (3.73), Berkshire, Buckinghamshire and Oxfordshire (3.50), Noord-Brabant (3.17), and Västsverige (3.00).

Subnetwork of collaborations in frontier research – Table 6 (d)

- The subnetwork of publications in top AI conferences presents **numerous regions** with a **balanced number of firms and research institutes** developing research publications. These are Stuttgart (57% of firms), Zuid-Holland (57%), Comunidad de Madrid (56%), Lazio (50%), and Emilia-Romagna (50%).
- In terms of the **potential to influence the subnetwork of publications**, the most relevant areas for this subnetwork are Île de France, Noord-Holland and Inner London–West.
- The region presenting the **largest academic productivity is Berkshire, Buckinghamshire and Oxfordshire** (15.50 publications per player), with 25% of its players as firms. Other productive regions follow in decreasing order: Noord-Holland (5.19 publications per player and 31% of firms), Stuttgart (4.71 publications per player and 57% of firms), Inner London–West (4.22 publications per player and 46% of firms), Emilia-Romagna (3.75 publications per player and 50% of firms), and Rhône-Alpes (3.14 publications per player and 43% of firms).

6.4 R&D openness in patenting, frontier research and EU projects

In this subsection we evaluate the top 20 EU28 regions by number of players in AI-related R&D activities to study their peers' location and assess the openness and internationalisation of these regions. The location of the peers is considered in relative terms, i.e. *within the same region* (darkest colour in Figure 25), in a *different region of the same country* (mid-intensity colour), or in a *different country* (lightest colour). Moreover, we discuss and compare the structure of collaborations in patenting, in frontier research publications, and in EU-funded projects.

In all three R&D sub-networks (Figure 25(a-c)), the most common type of collaboration is with out-of-country players; however, each sub-network shows a different pattern of peers' location. The sub-network of EU-funded projects (a) activates almost exclusively **out-of-country collaborations**. The distribution of the number of collaborations over the different regions is more uniform than the respective distributions for patenting and frontier research. This suggests that EU funded projects are able to foster a homogeneous and balanced involvement in AI, as it is also noted in the analysis of R&D scores distribution among EU countries (see subsection 6.2).

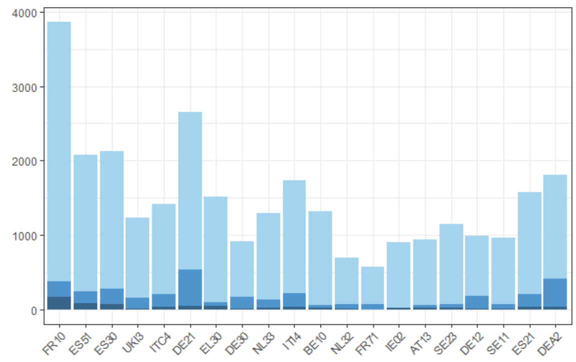
Regarding the collaborations in **patenting activities** Figure 25(b), the considered regions present a **higher propensity** to develop **collaborations within the same region**, in comparison with the other two subnetworks. This is the case of *Västsverige* (SE23) (85%), *Karlsruhe* (DE12) (80%), *Comunidad de Madrid* (ES30) (50%), *Rhône-Alpes* (FR71) (50%), and *País Vasco* (86%) (not plotted). Regarding **collaborations with other regions belonging to the same country**, the regions with the highest propensity to develop this kind of collaborations are all located in **Germany**: *Stuttgart* (DE11) (57%), *Köln* (DEA2) (40%), *Darmstadt* (DE71)

(38%), *Berlin* (DE30) (35%), and *Oberbayern* (DE21) (33%) (the three cited regions are not plotted in Figure 25). Nevertheless they are included in the top 30 ranking by number of players in R&D, as presented in Table 6. This confirms the intense collaborative structure among German players.

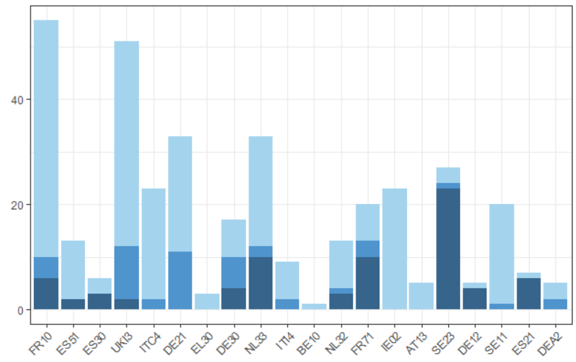
Regarding the collaborations in publications (c), we spot a relatively larger percentage of collaborations with players from the same country but not the same region (with respect to collaborations in other types of activities). Therefore, **academic research networks** of the largest European regions are **more structured within the country** than other types of R&D collaborations. In Figure 25(c), there are five regions with more than the 25% of collaborations with players from the same country but not the same region. These are *Île de France* (FR10) (27%), *Oberbayern* (DE21) (26%), *Berlin* (DE30) (28%), *Zuid-Holland* (NL33) (27%), and *Rhône-Alpes* (FR71) (32%). Given also the overall amount of activities developed by the players of these regions (namely 112, 31, 31, 15, and 44 respectively), these regions play a noticeable role within their country in terms of frontier research in AI.

Figure 25. Number of collaborations by location of the peers for top 20 EU28 Regions: (a) EU-funded projects, (b) Patenting activity, (c) Frontier research, 2009-2018

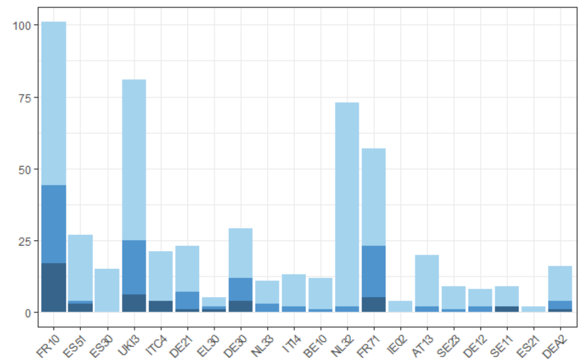
(a) EU-funded projects

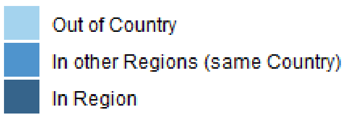


(b) Patenting activity



(c) Frontier research





The listed NUTS2 regions, from left to right in decreasing order by number of AI players in R&D process, are: FR10 - *Île de France*; ES51 – *Cataluña*; ES30 – *Comunidad de Madrid*; UKI3 – *Inner London–West*; DE21 – *Oberbayern*; ITC4 – *Lombardia*; EL30 – *Attica*; DE30 – *Berlin*; NL33 – *Zuid-Holland*; ITI4 – *Lazio*; BE10 – *Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest*; NL32 – *Noord-Holland*; FR71 – *Rhône-Alpes*; AT13 – *Wien*; IE02 – *Southern and Eastern*; SE23 – *Västsverige*; DE12 – *Karlsruhe*; ES21 – *País Vasco*; SE11 – *Stockholm*; DEA2 – *Köln*..

Notes:

Top 20 EU28 regions by number of players in AI-related R&D activities (in decreasing order by the sum of R&D players).

Indicators:

Out of Country: number of collaborations between regional players and players belonging to a different country.

In other Regions (same Country): number of collaborations between regional players and players belonging to the same country but not the same region.

In Region: number of collaborations between regional players and other regional players.

Collaborations in the three subnetworks are considered separately. Collaborations are identified depending on the location of the peers with respect to the considered region. For example, when considering region AT13, i.e. *Wien*, the "**In Region**" collaborations refer to collaborations between players located in *Wien*. The "**In other Regions (same Country)**" collaborations refer to collaborations with the first player in *Wien* and the second in a different Austrian region. The "**Out of Country**" collaborations refer to collaborations with the first player in *Wien* and the second in a different country. Note that figures have different scales.

Source: JRC PREDICT- AI TES Dataset, 2019.

7 Conclusions

This report studies the vast amount and great variety of activities in the landscape of AI in the last decade, 2009-2018. It presents a methodology well suited for the purpose of mapping the techno-economic segment of AI from a multidimensional perspective. The TES methodology identifies and analyses the firms, research institutes and governmental institutions playing an important role in the AI industrial and research landscape.

The quantitative evidence presented in this report demonstrates the dominating positions of the US, China and the EU, but also describes the interesting roles of some emerging AI economies such as India, Singapore or Israel. Both China and the US are leaders in different aspects of the AI world. US is the leader in industrial activity and in the AI thematic areas of AI Services, Robotics and Automation, and to an extent in Natural language processing. While China has a very strong governmental involvement and is the clear leader in Machine learning, Computer vision and Connected and Automated vehicles. Overall, the activity of both US and China indicates that they have an extensive presence in all AI areas. At regional level, the leadership of the two giants persists: out of the top-30 regions with highest activity level in AI, 15 are Chinese, 7 are from the US. California and Beijing are the top worldwide regions in AI in almost all thematic areas, along with Jiangsu and Guangdong.

The EU has a more targeted role in the AI landscape. The data presented in the report confirms EU's strong positioning in Robotics and Automation, and in AI Services, with a revealed comparative advantage in both areas. It additionally has an extensive, spread and collaborative research sector (sizeable even when EU-funded projects with AI content are not considered). Two EU regions are among the top 30 by level of AI activity worldwide: Inner London-West and Île de France. However, the EU compares less favourably regarding AI firms presence. A look into the EU firm structure reveals the comparatively lower propensity of AI firms to perform innovative activities in the AI field, submitting less AI related patents, and a large proportion of AI firms active in services sectors like Information and Communication or Professional, Scientific and Technical activities. These findings suggest a slower AI penetration in the remaining sectors relative to the US and China. The AI firm demographics of the EU show a young firm population, which on the one hand points to AI entrepreneurial activity, but on the other hand may mean limited activities of established firms in developing AI for their businesses. In the analysed period, UK played an undisputed role in the EU landscape. It is home to the highest number of AI players and takes the leadership in most AI thematic areas within the EU. Therefore, the withdrawal of the UK will inevitably lead to a less favourable positioning of the EU.

Other countries that outstand both in relative and absolute terms in one or more thematic areas are South Korea and Japan, which excel in Natural language processing, Computer vision and Connected and Automated vehicles, and India in AI Services and to an extent in Robotics and Automation.

Talent is a fundamental piece for the development of disruptive technologies and their subsequent adoption by the business world. Top world regions have the capacity to attract talent from all over the world, leading to a certain level of AI brain drain in other world regions. The EU's framework programmes have proven effective in building a research network of excellence throughout the Union. In order to keep and attract talent and reduce the risk of becoming mainly a consumer of AI services, the EU needs to play an active role to foster industrial capacity and research excellence and reinforce the academy-industry links. All of this in view of keeping to the fundaments of the European Union, following ethical rules, respecting human rights and striving for an inclusive society. Many challenges rest ahead in the AI landscape. It remains to be monitored how these different roles in the AI landscape will develop in the future.

Possible extensions of this project include the consideration of the ownership structure of the identified economic players, in order to analyse the internationalisation of AI enterprises. This additional information on. e.g. the location of the ultimate owner may offer an interesting alternative view of the global AI ecosystem. Furthermore, the analysis of the R&D excellence might benefit from additional indicators of R&D excellence that account for influential capacity, such as citation-related indices. This may also cause notable shifts in the research landscape. The investigation of the way governments support the development of AI could provide additional insights of the AI landscape. Future potential advancements could lie in the broadening of the AI conceptual framework, in order to acknowledge dimensions not explored in the current version of the landscape project. This could be to incorporate questions of ethics in AI, which covers concepts such as trust, fairness and accountability, and have become the focus of interest for academic and policy bodies.

References

Accenture (2018). Realising the economic and societal potential of responsible AI.

Agenzia per l'Italia Digitale (AGID), Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale (2018). Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino.

Aldous, D.J. (1985). Exchangeability and related topics. In: École d'Été de Probabilités de Saint-Flour XIII1983. Springer, pp. 1-198.

Atomico (2017). State of European Tech.

Atomico (2018). State of European Tech.

Bardt, H. (2017). Autonomous Driving - a Challenge for the Automotive Industry. Intereconomics, 52(3), 171-177.

BDI (2015). Industry 4.0 – an opportunity for Germany. Last accessed (22.05.2019) at: english.bdi.eu/topics/germany/industrie-40/

Blei, D.M., Laerty, J.D. (2009). Topic models. Text mining: classification, clustering, and applications. 10(71).

Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3, pp. 993-1022.

Brandes, U. (2001). A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25 (2), p. 163-177.

China Institute for Science and Technology Policy at Tsinghua University (2018). AI Development Report.

China's State Council (2017). Next Generation Artificial Intelligence Development Plan (AIDP). China

Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. In Proceedings of the international working conference on advanced visual interfaces (pp. 74-77). ACM.

Coad, A., Segarra, A., Teruel, M. (2016). Innovation and firm growth: Does firm age play a role? Research Policy 45(2), (pp. 387-400).

DG for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW). Regional Innovation Monitor Plus (RIM). Last accessed (13.05.2019) at: ec.europa.eu/growth/tools-databases/regional-innovation-monitor/content/regional-innovation-monitor-plus

Ding, J. (2018). Deciphering China's AI dream. Future of Humanity Institute Technical Report.

Etzkowitz, H. & Leydesdor, L.. (2000). The dynamics of innovation: from national systems and mode 2 to a triple helix of university{industry{government relations. Research policy 29(2) 109-123.

European Commission (2017). Europe's Digital Progress Report (EDPR) 2017 Country Profile Estonia.

European Commission (2018a). Artifical Intelligence for Europe, COM(2018) 237 final

European Commission (2018b). Coordinated Plan on Artificial Intelligence, COM(2018) 795 final

EU countries' representatives (2018). Declaration on Cooperation on Artificial Intelligence. April 10.

Federal Government (2018). Artificial Intelligence Strategy. Germany

HM Government (2017). Industrial Strategy: Building a Britain fit for the future Cm9528. United Kingdom

International Federation of Robotics. Last accessed (14.11.2018) at: eu-robotics.net/sparc/ about/robotics-in-europe/index.html.

Jurafsky D., Martin, J.H. (2014). Speech and language processing.

Kalemli-Ozcan, S., B. Sorensen, et al. (2015). How to construct nationally representative firm level data from the ORBIS global database, National Bureau of Economic Research.

Ke, D. L. H., & de Diego, E. (2019). Chinese innovation and entrepreneurship going abroad: from counterfeits and copycats to innovation exporters. In Handbook on China and Globalization (p. 43). Edward Elgar Publishing.

KPMG (2019). Autonomous vehicles readiness index.

Krovetz, R., 1993. Viewing Morphology as an Inference Process. s.l., s.n., pp. 191-203.

Lane, D., Pumain, D., Ernst van der Leeuw, S., and West, G. (2009). Complexity perspectives in innovation and social change, 7, Springer, Science & Business Media

Li, L. (2018). China's manufacturing locus in 2025: With a comparison of "Made-in-China 2025" and "Industry 4.0". Technological Forecasting and Social Change, 135, 66-74.

Margetts, H. and Naumann, A.(2016). Government as a platform: What can Estonia show the world? Oxford: Oxford Internet Institute.

Mas, M., Fernández De Guevara, J., Robledo, J.C., Righi, R., Cardona, M., Samoili, S., López Cobo, M. and De Prato, G. (2019): The 2019 PREDICT Key Facts Report. An Analysis of ICT R&D in the EU and Beyond, EUR 29770 EN, Publications Office of the European Union, Luxembourg

McKinsey Global Institute (2017a). Artificial Intelligence: Implications for China. Discussion Paper. In China Development Forum.

McKinsey Global Institute (2017b). Artificial Intelligence: The Next Digital Frontier. Discussion Paper.

Ministerio de Ciencia, Innovación y Universidades (2019). Estrategia Española De I+D+I En Inteligencia Artificial.

National Strategy for Artificial Intelligence #AIFORALL (2018). Discussion paper: National Strategy for Artificial Intelligence. India

OECD Directorate for Science, Technology and Innovation, Committee on Industry, Innovation and Entrepreneurship (2018). Identifying and Measuring Developments in Artificial Intelligence. DSTI/CIIE/ WPIA(2018)4

Pan-Canadian Artificial Intelligence Strategy (2017). Last accessed (13.05.2019) at: www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy

Papadimitriou, C.H., Raghavan, P., Tamaki, H., Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. Journal of Computer and System Sciences, 61(2), pp. 217-235.

Parliamentary Mission (Villani Mission): Villani C., Schoenauer M., Bonnet Y., Berthet C., Cornut A.-C., Levin F., Rondepierre B. (2018). For A Meaningful Artificial Intelligence Towards A French And European Strategy (Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne). France

Pinho, M. (2019). China and global structural change: past, present and future. In Handbook on China and Globalization (p. 212). Edward Elgar Publishing.

Plattform Industrie 4.0 (2015). Last accessed (22.05.2019) at: https://www.plattform-i40.de/PI40/Navigation/EN/Industrie40/WhatIsIndustrie40/what-is-industrie40.html

Porter, M., 1980. An algorithm for suffix stripping. Program, 14(3), pp. 130-137

Regulation No. 1415 (November, 2018). Conducting an Experiment in Testing the Use of Highly Automated Vehicles on Public Roads. Russian Federation.

Righi, R., Samoili S., Lopez-Cobo, M., De Prato G. (2017). International ICT Trade Dynamics 2004-2014: an explorative network analysis, DigiWorld Economic Journal, 107(3), pp 69-94.

Seabra M. (2014). Portugal's research funding is secure. Nature, 507 (306), doi.org/10.1038/507306b.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Sievert, C., Shirley, K. (2014). Ldavis: A method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp. 63-70.

Special Interest Group on Artificial Intelligence (2018). Dutch Artificial Intelligence Manifesto. Netherlands

Statista (2017). Artificial Intelligence.

Steering Group of the Artificial Intelligence Programme (appointed by: Ministry of Economic Affairs and Employment) (2017). Finland's Age of Artificial Intelligence.

Steyvers, M., Griffiths, T. (2007). Probabilistic topic models. Handbook of latent semantic analysis, 427(7), pp.424-440.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S. and Leyton-Brown, K. (2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA.

Straka, M., Strakova, J. (2017) Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Vancouver, Canada, Association for Computational Linguistics, 88-99.

Strategic Council for AI Technology (2017). Artificial Intelligence Technology Strategy. Japan

US Congressional Research Service (2019). Artificial Intelligence and National Security. United States of America

US Department of Defense, Govini (2018). Artificial intelligence, big data and cloud taxonomy. United States of America

Viereckl, R., Ahlemann, D., Koster, A., Hirsh, E., Kuhnert, F., Mohs, J., Fischer, M., Gerling, W., Gnanasekaran, K., Kusber, J. and Stephan, J. (2016). Connected car report 2016: Opportunities, risk, and turmoil on the road to autonomous vehicles. PwC. Last accessed (22.05. 2019) at: http://www.strategyand.pwc.com/reports/connected-car-2016-study.

Wartena, C., Brussee, R., & Slakhorst, W. (2010). Keyword extraction using word co-occurrence. In 2010 Workshops on Database and Expert Systems Applications (pp. 54-58). IEEE.

WIPO (2019). WIPO Technology Trends 2019: Artificial Intelligence. Geneva: World Intellectual Property Organization.

List of figures

List of tables

Annexes

Annex 1 Description of AI thematic areas

The topic N*atural language processing (NLP)* contains activities related to the process of machine identification of information from written and spoken human communications. This information can be retrieved, analysed and also generated as speech signals by a functional unit, with applications varying from speech-to-text to speech synthesis for every device and machine.

The topic C*omputer vision* refers to activities that identify human faces and objects in digital images, as part of object-class detection. In this task, the locations and sizes of the faces and objects are also spotted (e.g. pedestrians, vehicles etc.). Applications of this thematic topic are found in biometrics, human-computer interaction, surveillance, photography and other areas of the computer vision field.

In the Machine learning (ML) topic we find the theoretical concepts and libraries used in AI for both production and research, e.g. convolutional neural networks, stochastic reinforcement learning, TensorFlow (machine learning library), etc.

The thematic topic Robotics and Automation gathers activities related to application and research of the technological intelligent tools to assist or substitute human activity, or to enable actions that are not humanly possible (e.g. medical robots), to optimize technical limitations, labour or production costs. This topic involves all the phases of design, construction and operation of robotic systems. Communication between digital systems, performance of remote and local operations in the medical, technical, industrial, teaching, agricultural fields and other applications are included (e.g. robotic arms, drones, commercial robots as customer assistants, domestic robots etc.)

The Connected and Automated vehicles *(CAVs)* topic regards activities relative to the technologies of autonomous vehicles, connected vehicles and driver assistance systems, in all the phases from theoretical background, to design and construction and communication. All the automation levels are considered (no driver assistance in driving, to no human assistance in driving) and the different communication technologies (V2V, V2C, V2I, V2P, V2X).

The AI Services thematic field includes the activities providing databases, software, visualisation and other services allowing the deployment and maintenance of applications. These applications cover a variety of needs in a cloud, the web, or in local machines (e.g. for financial advice, travel planning, business decisions, cloud storage services, Virtual Private Network (VPN) clients etc.).

Annex 2 Worldwide regional thematic hotspots of key areas

Table A. 1. Normalised thematic hotspot indicator of top 30 regions worldwide, 2009-2018.

Region name (country code)	nTHI: Normalised Thematic Hotspot Indicator per thematic area					
	Natural Language Processing	Computer Vision	Machine Learning	Robotics & Automation	Connected & Automated Vehicles	AI Services
California (US)	0.92	0.30	0.56	1.00	0.38	1.00
Beijing (CN)	1.00	0.98	1.00	0.05	0.66	0.10
Jiangsu (CN)	0.26	0.97	0.67		1.00	0.02
Guangdong (CN)	0.59	1.00	0.52	0.07	0.74	0.05
New York (US)	0.71	0.09	0.19	0.22	0.07	0.25
Shanghai (CN)	0.32	0.49	0.32	0.37	0.49	0.03
Seoul (KR)	0.85	0.30	0.11	0.02	0.39	0.04
Zhejiang (CN)	0.14	0.37	0.36	0.05	0.40	0.02
Tokyo (JP)	0.68	0.31	0.09	0.03	0.16	0.06
Inner London-West (UK)	0.03	0.02	0.10	0.34	0.01	0.24
Sichuan (CN)	0.16	0.39	0.19		0.25	0.01
Massachusetts (US)	0.26	0.04	0.15	0.23	0.06	0.13
Washington (US)	0.47	0.09	0.11	0.10	0.06	0.08
Texas (US)	0.09	0.05	0.09	0.14	0.23	0.11
Shaanxi (CN)	0.04	0.24	0.20	0.01	0.19	0.00
Liaoning (CN)	0.04	0.22	0.16	0.17	0.22	0.00
Shandong (CN)	0.06	0.16	0.14		0.27	0.00
Fujian (CN)	0.08	0.24	0.14	0.04	0.13	0.00
Hubei (CN)	0.05	0.20	0.15	0.02	0.16	0.01
Ontario (CA)	0.06	0.02	0.06	0.15	0.01	0.12
Anhui (CN)	0.12	0.16	0.10		0.24	0.00
Michigan (US)	0.11	0.02	0.06	0.06	0.34	0.03
Central (SG)	0.07	0.07	0.08	0.17	0.05	0.07
Île de France (FR)	0.04	0.03	0.07	0.09	0.02	0.09
Pennsylvania (US)	0.02	0.01	0.17	0.08	0.04	0.05
Daejeon (KR)	0.26	0.11	0.03	0.01	0.11	0.00
Tianjin (CN)	0.04	0.14	0.08		0.16	0.00
Karnataka (IN)	0.02	0.01	0.02	0.05	0.01	0.10
Wan Chai (CN)	0.07	0.10	0.08	0.02	0.05	0.03
Hunan (CN)	0.11	0.11	0.08	0.01	0.11	0.00

Notes: Top-30 regions with highest overall number of AI activities. EU-funded activities are not included.

Source: JRC PREDICT- AI TES Dataset, 2019.

Annex 3 EU28: regional thematic hotspots of key areas

Table A. 2. Normalised thematic hotspot indicator of top 30 EU28 regions, 2009-2018.

Region name (country code)	nTHI: Normalised Thematic Hotspot Indicator per thematic area					
	Natural Language Processing	Computer Vision	Machine Learning	Robotics & Automation	Connected & Automated Vehicles	AI Services
Inner London-West (UK)	0.76	0.51	1.00	1.00	0.17	1.00
Île de France (FR)	1.00	0.72	0.71	0.25	0.32	0.39
Berlin (DE)	0.21	0.00	0.17	0.19	0.15	0.27
Berkshire, Buckinghamshire and Oxfordshire (UK)	0.30	0.13	0.71	0.14	0.08	0.10
Noord-Holland (NL)	0.13	0.10	0.43	0.06	0.05	0.16
Oberbayern (DE)	0.57	0.31	0.28	0.06	0.48	0.10
Southern and Eastern (IE)	0.58	1.00	0.08	0.06	0.03	0.12
East Anglia (UK)	0.27	0.01	0.37	0.17	0.04	0.09
Cataluña (ES)	0.08	0.10	0.21	0.14	0.07	0.12
Inner London-East (UK)	0.17	0.03	0.06	0.09	0.10	0.15
Stockholm (SE)	0.51	0.25	0.11	0.06	0.14	0.10
North Yorkshire (UK)	0.28	0.01	0.60	0.08	0.02	0.02
Comunidad de Madrid (ES)	0.19	0.06	0.04	0.10	0.14	0.11
Helsinki-Uusimaa (FI)	0.21	0.18	0.12	0.09	0.02	0.08
Västsverige (SE)	0.11	0.00	0.04	0.02	1.00	0.03
Lombardia (IT)	0.05	0.03	0.14	0.02	0.14	0.07
Zuid-Holland (NL)	0.15	0.37	0.09	0.06	0.16	0.05
Stuttgart (DE)	0.04	0.01	0.14	0.09	0.66	0.02
Gloucestershire, Wiltshire and Bristol/Bath area (UK)	0.43	0.04	0.12	0.16	0.12	0.05
Hovedstaden (DK)	0.19	0.03	0.14	0.04	0.05	0.06
Eastern Scotland (UK)	0.50	0.01	0.21	0.09	0.02	0.04
Rhône-Alpes (FR)	0.00	0.06	0.21	0.06	0.09	0.03
Surrey, East and West Sussex (UK)	0.10	0.07	0.02	0.09	0.01	0.06
Münster (DE)	0.15	0.01	0.00	0.00	0.87	0.01
Wien (AT)	0.04	0.04	0.14	0.06	0.03	0.04
Karlsruhe (DE)	0.32	0.04	0.17	0.04	0.07	0.02
Área Metropolitana de Lisboa (PT)	0.00	0.07	0.09	0.00	0.00	0.05
Greater Manchester (UK)	0.01	0.07	0.05	0.00	0.03	0.05
West Midlands (UK)	0.06	0.00	0.08	0.00	0.08	0.04
Mazowieckie (PL)	0.04	0.01	0.04	0.04	0.09	0.04

Notes: Top-30 region with highest overall number of activities. EU-funded activities are not included.
Source: JRC PREDICT- AI TES Dataset, 2019.

Annex 4 Methodological notes on thematic key areas

A technology is approximated by a number of collected activities, assumed to be representative of this technology's ecosystem. Each detected activity (patent, project, publication, firm description) is represented by an amount of textual information, considered as a document. Each document is represented by a set of semantically related terms. A mixture of semantically related terms creates a topic, regardless of the document it belongs to. A document and a term may be assigned to one or more topics with a certain membership. The non-exclusive assignment of a term to a topic is sustained by the observation that in a natural language words do not belong to one topic only.

A technology's topics are inferred among a number of documents that represent the R&D and industrial activities performed by a technology. In order to interpret the content of the corpus formed by these documents, in particular for real-world corpora analysis, expert assessment is found in relevant literature for technology mining methodologies, which then requires a further refinement of the analysis (Chuang, Manning, & Heer, 2012). To avoid this subjective approach, a generative three-level hierarchical Bayesian model is used to identify the topics generated based on the collected documents, namely the Latent Dirichlet Allocation (LDA). Therefore, the knowledge of a technological topic, which was only a property of experts, is approximated through this statistical unbiased method that extracts this information from the documents' content (Griffiths & Steyvers, 2004). The topics are probability distributions of terms, which occur from the underlying semantic structure of documents (Blei & Lafferty, 2009; Steyvers, 2007; Papadimitriou, Raghavan, Tamaki, & Vempala, 2000). So each document could be a combination of topics, which is a combination of terms from several documents. This bag-of-words assumption is common to natural language statistical models (Blei & Lafferty, 2006; Aldous, 1983). The optimal number of topics is found via a harmonic mean estimator and a post-evaluation. The computation of the probability of occurrence of a word given a topic, and the probability of occurrence of a topic given a document, involves the computation of the posterior probability over hidden variables (Dirichlet priors), which are used to sample the Dirichlet distribution for the topics and the topics' mixture. The computation of the *a posteriori* probabilities over the hidden variables is intractable for the LDA model (Blei et al., 2003), so the variational Bayesian approximation is used.

Annex 5 List of top AI conferences

The list of top 10 AI conferences is populated based on google.scholar.com top conferences on AI, and AI experts' suggestions. The considered conferences and short description of their aim follows.

International Conference on Learning Representations (ICLR, iclr.cc): ICLR is focused on learning representations, more commonly known as deep learning, as the aim of deep learning is to learn higher-level features, or else representations.

International Conference on Machine Learning (ICML, icml.cc): ICML is focused on the machine learning branch of AI.

Conference on Neural Information Processing Systems (NeurIPS, nips.cc): NeurIPS is focused, but not exclusively, on neural information systems in multiple aspects.

AAAI Conference on Artificial Intelligence (AAAI, aaai.org/Conferences/AAAI): AAAI is supported by the Association for the Advancement of Artificial Intelligence (AAAI), and focuses on theoretical and applied AI research.

Artificial Intelligence for Interactive Digital Entertainment Conference (AIIDE, aaai.org/Conferences/AIIDE): AIIDE is supported by the Association for the Advancement of Artificial Intelligence (AAAI), and is more applications-oriented to allow the fruitful interactions between AI entertainment software developers, academic and industrial AI researchers.

Innovative Applications of Artificial Intelligence Conferences (IAAI, aaai.org/Conferences/IAAI): IAAI is sponsored by the Association for the Advancement of Artificial Intelligence (AAAI), focusing on AI-based applications.

Computer Vision and Pattern Recognition (CVPR, cvprYEAR.thecvf.com): CVPR is supported by the IEEE Computer Society and intends to provide the latest insights for students, academics and industry researchers.

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD, ecmlpkdd2019.org): ECML PKDD is considered among the leading conferences on machine learning and knowledge discovery.

International Conference on Computer Vision (ICCV, iccvYEAR.thecvf.com): ICCV focuses on computer vision and applications in all its subdomains.

International Joint Conference on Artificial Intelligence (IJCAI, ijcai.org): IJCAI intends to assist in the dissemination of information on AI by including cutting.

Annex 6 Size categories of firms

For this analysis the size categorisation of the BvD Orbis database is adopted. Firms are categorised by BvD as

- Very large firms, when they match at least one of the following conditions:
    - Operating Revenue >= 100 million EUR (130 million USD)
    - Total assets >= 200 million EUR (260 million USD)
    - Employees >= 1,000
    - Listed
- Large firms, when they match at least one of the following conditions:
    - Operating Revenue >= 10 million EUR (13 million USD)
    - Total assets >= 20 million EUR (26 million USD)
    - Employees >= 150
    - Not Very Large
- Medium sized firms, when they match at least one of the following conditions:
    - Operating Revenue >= 1 million EUR (1.3 million USD)
    - Total assets >= 2 million EUR (2.6 million USD)
    - Employees >= 15
    - Not Very Large or Large
- Small firms, when they are not included in another category.

KJ-NA-30109-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union

doi:10.2760/85212

ISBN 978-92-76-16661-0