



# Trends in the use of AI in science

A bibliometric analysis



**R&I PAPER SERIES**  
WORKING PAPER 2023/04

David ARRANZ, Stefano BIANCHINI, Valentina DI  
GIROLAMO, Julien RAVET



## Trends in the use of AI in science

European Commission  
Directorate-General for Research and Innovation  
Directorate G — Common Policy Centre  
Unit G.1— Common R&I strategy and foresight service

Contact Alexandr HOBZA  
Email [RTD-ECONOMIC-ANALYSIS@ec.europa.eu](mailto:RTD-ECONOMIC-ANALYSIS@ec.europa.eu)  
[RTD-PUBLICATIONS@ec.europa.eu](mailto:RTD-PUBLICATIONS@ec.europa.eu)

European Commission  
B-1049 Brussels

Manuscript completed in March 2023.

This document has been prepared for the European Commission, however it reflects the views only of the authors, and the European Commission shall not be liable for any consequence stemming from the reuse.

More information on the European Union is available on the internet (<http://europa.eu>).

---

PDF	ISBN: 978-92-68-04289-2	doi: 10.2777/418191	KI-BD-23-003-EN-N
-----	-------------------------	---------------------	-------------------

---

Luxembourg: Publications Office of the European Union, 2023

© European Union, 2023



The reuse policy of European Commission documents is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

Reproduction is authorised provided the source is acknowledged.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightsholders. The European Union does not own the copyright in relation to the following elements:

Cover page : VectorMine #303873962 2020. Source : [stock.adobe.com](https://www.adobe.com/stock)

# Trends in the use of AI in science

*A bibliometric analysis*



Working paper

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>3</b>
1. Introduction .....	3
2. The potential benefits of AI in specific scientific domains. Measuring the use of AI in science. ....	4
<b>2.1 Quantifying the impact of AI</b>	<b>4</b>
<b>2.2 The use of AI in science</b>	<b>7</b>
<b>2.3 Approach taken in this paper</b>	<b>9</b>
3. AI in science: global trends and international benchmarking .....	10
<b>3.1 Global trends of AI in science</b>	<b>10</b>
<b>3.2 The EU's global position in AI scientific production</b>	<b>12</b>
4. The EU research landscape in AI in science .....	18
5. Conclusions.....	25
6. References.....	27
7. Annex: Methodological note .....	31
<b>7.1 Definitions of AI</b>	<b>31</b>
<b>7.2 Data collection methodology</b>	<b>31</b>

# ABSTRACT

*Artificial intelligence (AI) has the capability of boosting scientific productivity, augmenting human cognitive capabilities, and accelerating breakthroughs. Given the magnitude of this potential, efforts have increasingly been focused on tracking the diffusion and impact of AI in science and innovation. This paper provides new insights into the role of AI in scientific research across various domains of science. It is evident that AI as applied to science and research has been growing at a significant pace in recent years, with China leading the way, followed by the EU and the US. Indeed, the EU's performance is found to be highly heterogeneous across its Member States. If current trends persist, the probability of future scientific discoveries being driven primarily by AI applications and tools is set to increase significantly. Failure to keep pace with the development and uptake of AI in science poses important challenges for the EU, including the need to address concerns about strategic autonomy, trends in scientific work and the labour market and, more broadly, productivity and growth.*

## 1. Introduction

**Artificial intelligence (AI) constitutes a ground-breaking suite of technologies with the potential to have a profound effect on all aspects of human existence.**

Its importance is escalating rapidly, in terms not only of the cutting-edge innovations being made but also of the prominence it has gained in public debates. The development of large language models (LLMs) – in particular, since the launch of ChatGPT at the end of 2022 – has acted as a catalyst for both innovation and competition among the main AI players, positioning AI technology as a veritable “game changer” in scientific research. Moreover, given its general-purpose nature, AI has the ability to have a transversal impact on all sectors and is expected to play a critical role in the Twin – green and digital – Transition (EIB, 2021) and in achieving the Sustainable Development Goals (SDGs) (Vinuesa et al., 2020; Bianchini et al., 2023).

**AI is rapidly becoming an indispensable tool in the scientific process.** It already serves as a robust research mechanism for many scientists, and its adoption is expected to undergo rapid growth in the immediate future. Its capabilities underpin the process of scientific discovery (Bianchini et al., 2022; Krenn, 2022) and innovation (Rammer et al., 2022), playing, for example, a central role in processing large-scale scientific data and generating patterns, predictions, and models. Moreover, it facilitates understanding of scientific output by means of information retrieval, natural language processing, and recommender systems in extensive repositories of scientific papers. Indeed, according to the ‘AI Index Report 2023’ (Institute for Human-Centered AI, 2023), in 2022, AI played a significant role in areas as diverse as plasma control in nuclear fusion, algorithm optimizations for matrix multiplications (with important implications for efficient computing) and, thanks to generative AI models, the accelerated discovery of antibodies.

**AI has the potential to serve as a catalyst for scientific productivity, resulting in more efficient outcomes and pushing back scientific boundaries.** AI

increases human cognitive capacities, solving problems and expanding the limits of human achievement by tackling complex processes, such as data analysis, and producing research that would be unattainable using traditional tools (e.g., the protein folding problem<sup>1</sup>). Although its overall impact on scientific productivity remains unconfirmed, AI has the potential to shorten the typical timeframe needed for scientific discoveries to be taken up and, thus, to contribute to addressing pressing global challenges, especially at a time when scientific productivity seems to be stagnating and new ideas are “getting harder to find” (Bloom et al., 2020).

**In recent times, there has been an increasing effort to track the dissemination of AI within science and innovation, albeit with less focus on quantifying its impact.** Having high-quality data and metrics to monitor the development and adoption of this emerging technology remains critical from both a policy and strategic perspective. The European Commission places the accelerated adoption of AI technologies at the heart of its strategy to establish EU global technological leadership, and the 2021 Coordinated Plan on Artificial Intelligence lays out the actions to be undertaken by EU Member States to accelerate AI investments and reduce fragmentation within the EU Single Market (European Commission, 2021). Nevertheless, the process of mapping AI research faces **two critical challenges**: first, establishing a precise definition of AI, and second, distinguishing between fundamental AI research and its applications across various fields.

The aim of this paper is to add on the existing evidence on current trends of AI applications, looking at the **global trends of AI in science**, with a focus on the EU’s performance vis-à-vis that of other big global innovators, and **mapping the diffusion of AI technology in science and its application in different domains within the EU’s research landscape**.

The remainder of the paper is structured as follows. Section 2 describes the methodological approach adopted in the paper, building on the work of other studies in the relevant literature. Section 3 and 4 provide an overview of the main findings of the analysis. Section 5 concludes the paper.

## 2. The potential benefits of AI in specific scientific domains. Measuring the use of AI in science.

### 2.1 Quantifying the impact of AI

**AI lacks a universally accepted definition.** Since its inception, the field has been associated with ever-shifting definitions (see Nilsson, 2009; and Russell and Norvig, 2021, for comprehensive accounts of AI history): some place greater emphasis on the operational characteristics of intelligent machines; others prefer to focus on the objectives of AI research. Although the most recent definitions aim to be technically accurate, technology-neutral, and applicable to both short- and long-term horizons,

---

<sup>1</sup> See Jumper et al. (2021). See also <https://www.science.org/doi/10.1126/science.370.6521.1144>

AI spans multiple evolving, overlapping domains, making it difficult to identify just what does and does not qualify as AI.<sup>2</sup>

**The diffusion of AI knowledge is conducted via a mix of different channels.**

Most AI mapping studies rely on scientific publications and patents for their data sources, although AI development also occurs extensively in cloud-based repositories (GitHub, Hugging Face, Apache Allura, etc.) as well as on blog platforms (New Scientist, The Conversation, WIRED, etc.).

**Bibliometric analyses of publications face specific constraints and difficulties.**

Here, we focus solely on *scientific publications* which, despite their limitations, provide extremely valuable information about the advances occurring within and across all fields of science. In general, because of the absence of an accepted AI ontology, three main approaches have emerged in the identification of AI-related scholarly activity, namely: (i) a *predefined keyword* approach; (ii) *machine learning* approach(es); and (iii) a *combination of the two*, which allows words to be added to an initial list.

- i. The first approach is relatively straightforward. It involves searching for a set of ad hoc and predefined terms in a publication's title, abstract, keywords, and, where feasible, the entire body of the document (e.g., Cockburn et al., 2018; Tran et al., 2019; Guo et al., 2020; De la Vega Hernández et al., 2023). Relying on a list of search terms for document retrieval is common practice in research on emerging technologies and sciences. However, because of the challenge of defining just what AI is exactly, there is no definitive, "ready-to-use" list of terms that might be deemed authoritative for this purpose. In selecting terms for inclusion on the list, a researcher can either adopt a highly parsimonious approach and consider only a limited number of essential terms, or opt for a broader approach, resulting in a more extensive list (e.g., Gargiulo et al., 2022). Nevertheless, the essential terms selected typically include general keywords such as 'Artificial Intelligence', 'Machine Learning', and 'Deep Learning', among others. It should be stressed that these terms capture the main bulk of scientific activity – as shown below in this study (see Table 1A in Annex). Moreover, it is not uncommon to encounter various terms that refer to specific machine learning techniques, such as 'Random Forest', 'Decision Tree', and 'Boosting'.
- ii. The second approach – often referred to as "using AI to define AI" – leverages certain machine learning (ML) techniques to analyse publication abstracts (e.g., Klinger et al., 2021; Bianchini et al., 2022). In this way, search terms are not defined ex ante, but, instead, are "learned" directly from the data. A common strategy here is to use embedding algorithms, such as Word2Vec and its variants, which map the corpus from a high- into a lower-dimensional space, while preserving semantic relationships, thus making it easier to cluster words of similar meaning or from similar contexts. Once the

---

<sup>2</sup> See the methodological annex at the end of this paper for the most recent definitions of AI.

model has been trained on a large sample of publications (e.g., arXiv or PubMed), it can be queried to obtain a set of words that fall within the cluster that includes the term ‘Artificial Intelligence’ or the like. Thus, a list of search terms can be generated and used to identify AI-related documents.

While this approach is appealing, it is often more pragmatic than scientific, inasmuch as it requires the making of many, often iterative, decisions, based on trial and error. These decisions include, for instance, text input cleaning (e.g., tokenization, stop-word removal, stemming, and lemmatization), setting hyperparameters (e.g., embedding size, learning rate), and several post-processing operations (e.g., assembling word embeddings, clustering). Interestingly, previous research using natural language processing-based approaches typically ended up with virtually the same search terms as those defined beforehand by experts.

- iii. The third approach – i.e., keywords *plus* ML – consists generally of two steps. First, keywords present in publications across all journals are tagged as AI-related (e.g., the “All Science Journal Classification” in Scopus and “Subject Categories” in Web of Science), and, second, the set is then augmented and refined using text mining techniques and by expert validation. Using this method, Baruffaldi et al. (2020), for example, were able to compile a list of 193 AI-related keywords and use them to identify over 2.4 million documents in Scopus spanning the period 1996 to 2016.

The same approach can be employed in reverse, that is, starting with a large list of keywords and concepts identified from various sources – including AI textbooks, news items, online blogs, and others – and then using ML tools to streamline and refine the list to a more manageable size. By adopting this method, for example, a team at Elsevier (2020) was able to identify more than 600,000 AI documents for the period 1998 to 2017.

Regardless of the approach, various caveats should be borne in mind:

- First, some terms may be overly general and open to contention. The term ‘Robot’ is an example in point: while some robots may incorporate AI technology so as to enhance their functionality, others may contain no AI components whatsoever (Russell and Norvig, 2021 – Ch.26). The term ‘Big Data’ is another example: although AI can be a powerful tool for extracting insights and patterns from big data, it is by no means a prerequisite for working with big data (De Mauro et al., 2015).
- Second, special attention needs to be paid to certain terms that, while significant in the field of AI, may not refer to it exclusively. For instance, the term ‘Neural Network’ is potentially problematic inasmuch as some publications using this term may not necessarily be discussing artificial intelligence, but rather human intelligence and the biological brain. Studies have shown, however, that this problem is mainly confined to the field of neuroscience and that it has minimal impact in other scientific disciplines (Bianchini et al., 2022 – Section 5).



- Third, as AI technology becomes increasingly integrated into scientific research and permeates the scientific system, references to AI techniques and tools in the titles, abstracts, and keywords of scientific publications may no longer be a reliable method for identifying AI-related scientific activity. This means novel approaches, based on the comprehensive content analysis of full texts, will need to be developed.

## 2.2 The use of AI in science

A major challenge in mapping the diffusion of AI in the sciences is differentiating between research that *develops* AI technology – i.e., fundamental or basic AI – and research that *uses* AI to solve field-specific scientific problems – i.e., applied AI. To illustrate this, consider two papers: the first aims to train more efficiently a neural network architecture, while the second applies the same architecture to detect, say, cancer from an MRI image. While the former would be categorized as basic AI research, the latter is clearly an example of applied AI research in the medical field.

However, the line between basic AI and its applications can, at times, be quite blurred, given just how interconnected theoretical and applied studies can be. On the one hand, advances in basic AI research can lead to the development of more effective and more efficient AI tools that can be applied to a range of scientific problems, while, on the other, feedback from the application of AI in specific scientific settings can help identify areas where basic AI research can be improved or expanded.

Over time, the focus of basic AI research has shifted from an initial emphasis on neuroscience and philosophy to a more computer science-oriented and – albeit to a lesser extent – mathematical approach. Bibliographic studies of the period 1950 to 2017 confirm that AI has transitioned towards computational research in recent decades, especially with the emergence of deep learning techniques (Frank et al., 2019). Thus, earlier studies tend to consider that AI publications in all areas other than computer science represent applications of AI techniques designed to address field-specific research problems (e.g., Cockburn et al., 2018; Bianchini et al., 2022). In what follows, we also opt to adhere to this approach (see Table 2A in Annex).

### Box 1: Applications of AI in the sciences<sup>3</sup>

AI has a vast array of potential applications that span a continuum between the two extremes of *search* and *discovery*. At the search end of the spectrum, AI can support access to knowledge and information, especially at times characterized by the explosion of data and information; at the discovery end of the spectrum, often as the end-result of a research project, AI can be employed to identify patterns in data in an open-ended manner, leading to new discoveries and insights (Agrawal et al., 2018; Raghu and Schmidt, 2020; Xu et al., 2021,

---

<sup>3</sup> Note that the examples considered constitute a non-exhaustive list of potential applications of AI in science.

Bianchini et al., 2022). Below, we consider some of the most common applications of AI in the scientific pipeline.

The most common use of AI in science is to address complex prediction problems – that is, mapping inputs to predicted outputs. The problems can be of any kind, as can the type of methodological approach adopted. For instance, convolutional neural networks (CNNs) can be used to process MRI images and to predict the possible presence of cancer. Examples of the many computer vision tasks include semantic segmentation, where the goal is to categorize pixels according to the high-level group to which they belong, and pose estimation, where the goal is to predict and track the location of a person or object. Other techniques, such as recurrent neural networks (RNNs), are common in scientific applications involving the prediction of sequential structures, such as in genomics and proteomics, but also in finance.

A second common application of AI is to perform transformations of input data, including dimensionality reduction, clustering, data augmentation, and image super-resolution, to name but a few. Dimensionality reduction and clustering are simple but effective methods for revealing hidden properties in data and are often the first step in exploring and visualizing data, before any other prediction tasks are undertaken. Image super-resolution and data compression are other common applications that can facilitate data analysis and enable the researcher to save and optimize space.

A third application is the optimal parameterization of complex systems. Here, techniques such as reinforcement learning can be used to search for the optimal set of parameters that maximize or minimize a specific objective function or produce a desired outcome. A recent example is the configuration of tokamaks (for nuclear fusion) with deep reinforcement learning, which has enabled scientists to model and maintain a high-temperature plasma within the tokamak vessel, a problem that had hitherto proved impossible to solve (Delgrave et al., 2022).

Another valuable application of AI is automating (or partially automating) the literature review process, which can be facilitated by powerful search engines based on LLMs. Platforms like Elicit and Perplexity work through a chatbot-style interface, enabling researchers to interact dynamically with the machine. The researcher can initiate a conversation to search for information about past research in a certain area and receive a summary of key information about that field. The newest tools can even remember the conversational context, improving the quality of the exchange between user and machine. Most of these AI-powered platforms offer other functionalities, such as assisting researchers in brainstorming research questions and directions – i.e., rephrasing their research questions and suggesting potential research directions based on the current state of the art – and providing suggestions on how to improve prose writing and editing.

Still within the context of academic literature reviews, an interesting application is literature-based discovery, where AI can uncover implicit, hidden associations from existing studies, resulting in interesting, surprising, non-trivial hypotheses

that are worth studying. Machine reading comprehension systems are particularly useful in this context, as they can identify gaps in the literature and propose variations on existing experiments.

Finally, AI or, specifically, simple robotics can be used to *automate tedious, routine laboratory tasks* such as media and buffer preparation or pipetting. These tasks require a high degree of accuracy but have relatively low value-added.

### **2.3 Approach taken in this paper**

**The publication data used in this paper are drawn from the Web of Science (WoS) Core Collection.**<sup>4</sup> We are specifically concerned with scholarly activity related to AI in the period 2000 to 2021 (data for 2022 were incomplete and not fully comparable with those of previous years at the time of extraction). Relevant papers are identified using a list of AI-related keywords and these studies are classified into different fields in line with established practices in the field, albeit with a number of specific adjustments.

**Publications are assigned to a country and to a specific domain of science.** The link between a paper and a country is established based on the affiliations of the authors, taking a “weighted-by-author” approach, to avoid double counting issues (i.e., assigning one paper to different countries). Papers are originally classified in more than 250 different WoS categories. This granular categorization is consolidated in 21 macro-categories (domains). A paper can be associated with more than one domain, which means there could be some level of double-counting in the analysis per scientific domain.

**When analysing the applications of AI in science, we exclude papers that belong exclusively to the domain of ‘Computer Science’.** In the classification per domain, we distinguish between the field of ‘Computer Science’ and all other fields. When a paper is exclusively assigned to areas within the computer sciences, we consider this paper as belonging to the development of AI and close applications, which we refer to as “Core AI”. Papers that are classified in at least one domain other than that of computer science are considered as belonging to applications of AI in science.

*More details on the methodology are provided in the methodological annex at the end of this paper.*

---

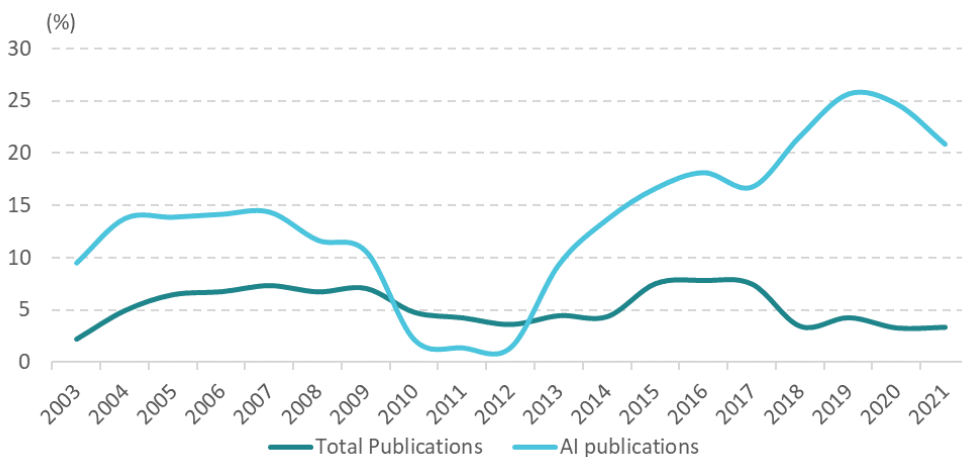
<sup>4</sup> <https://www.webofscience.com/wos/woscc/basic-search>

### 3. AI in science: global trends and international benchmarking

#### 3.1 Global trends of AI in science

**The field of AI is growing at a faster rate than that of scientific production as a whole.** In general, global scientific activity has grown at around 5% per year between 2004 and 2021. In the same period, the annual growth rate of AI-related publications has been around or above 15%, except for the years between 2010 and 2012, when scientific production in the field of AI stagnated (Figure 1), presumably because of the reorientation in research priorities and funding linked to the onset of the financial crisis. The slowdown observed since 2019 is presumably attributable to the effects of the COVID-19 pandemic.

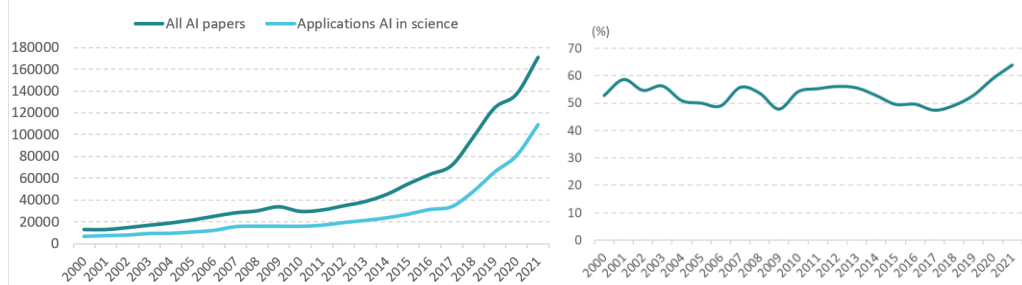
**Figure 1. Growth in scientific activity (3-year average yearly growth)**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data. Annual growth calculated as a 3-year rolling average.

**Scientific activity related to AI applications accounts for a significant share of total publications in the field of AI.** Between 2000 and 2021, the evolution undergone by the total number of AI publications and publications related exclusively to AI applications in science followed a similar pattern of growth (Figure 2 – left panel), with the latter accounting for around 50% of total AI publications up to 2018. Since that date, there has been a significant increase in this share, indicating a decoupling of AI applications in science from the overall growth in the field of AI. This indicates that **AI applications in science are growing faster than the AI field as a whole** (Figure 2 – right panel).

**Figure 2. Number of publications (left) and share of publications on AI applications in science (right)**

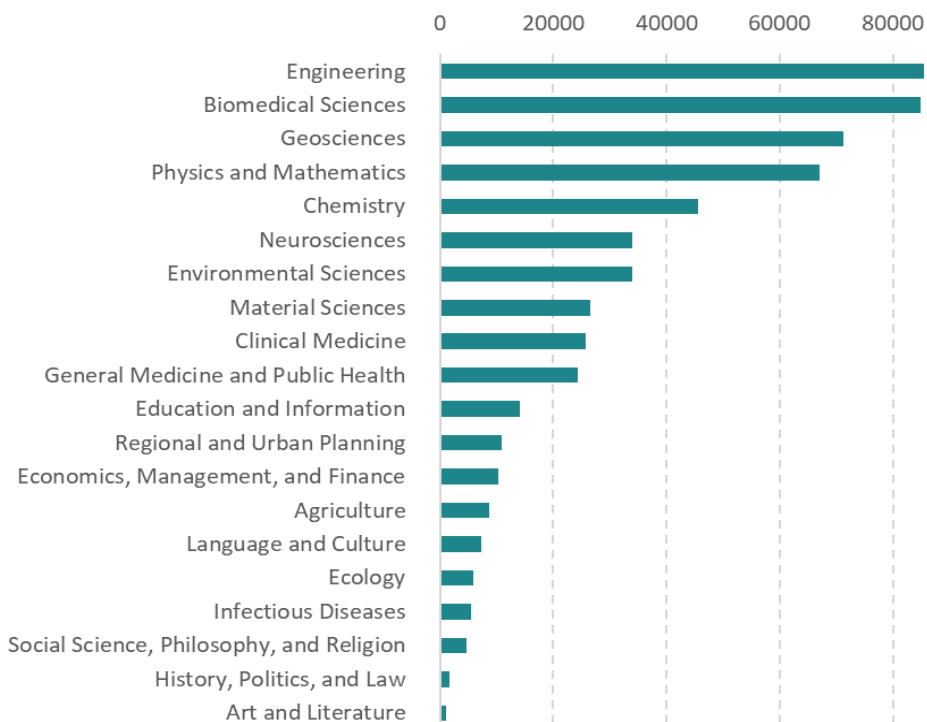


Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

**The impact of AI could vary considerably across different scientific domains.**

Over the last 5 years, applied sciences (such as, engineering and materials), as well as the natural and life sciences are the fields that have reported the highest number of publications on AI applications. The social sciences, including economics and the humanities, account for a lower share of publications in which AI is used as a tool (Figure 3). In terms of growth, the material sciences is the discipline with the highest growth rate (almost 50% on a yearly basis), followed by medicine (clinical and general – around 45%). Surprisingly, the neurosciences, one of the disciplines that served as a reference for the development of AI, presents one of the lowest growth rates (22%), while the lowest is found in the discipline of art and literature (5%).

**Figure 3. Number of publications (2017–2021) on AI per scientific domain**

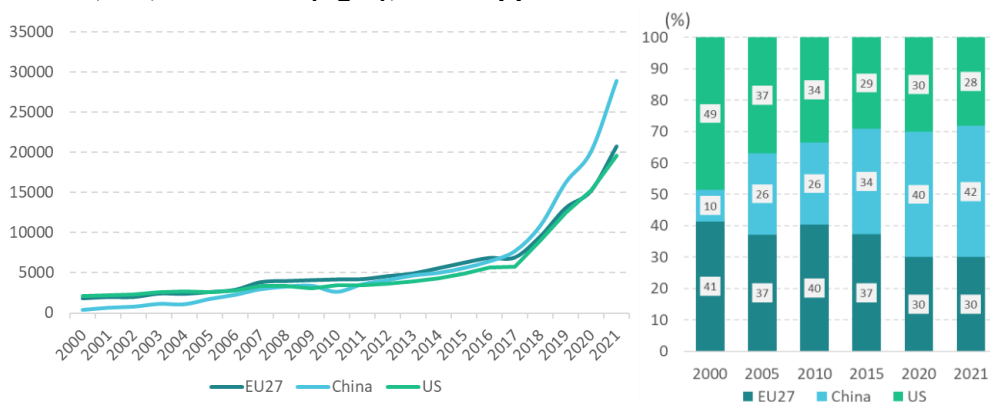


Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

### 3.2 The EU's global position in AI scientific production

**China, the US and the EU are global leaders in AI scientific production and in publications related to AI applications in science.** More specifically, the EU and the US have reported similar levels of AI publications over the last few decades, with the EU showing a modest advantage up to 2017, before the trend was slightly reversed in the US's favour. Particularly impressive has been the performance of China, which entered the 21st century lagging behind its main competitors, but it has been able to catch up quickly and, since 2017, has outperformed both the US and the EU (Figure 4 – left panel). A similar trend is also observed if we focus solely on the share of publications related to AI applications, with China reporting a remarkable growth at the expense of the other two economies (Figure 4 – right panel).

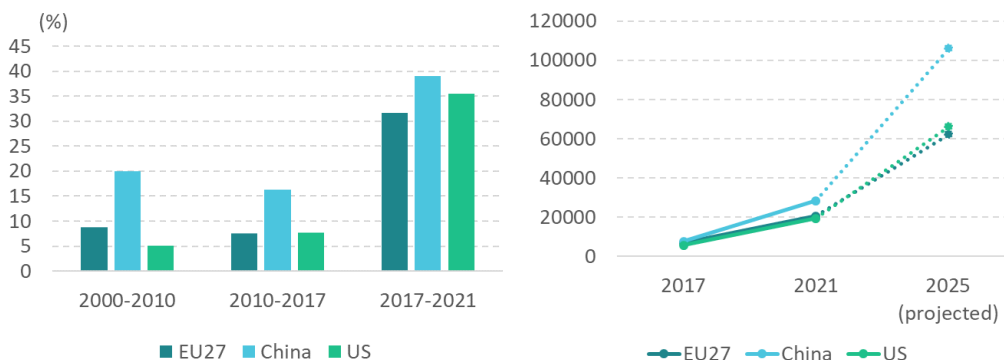
**Figure 4. Number of publications (left), and relative share of publications of the EU, US, and China (right), on AI applications in science**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

**The number of publications dedicated to AI applications in science has grown at a faster rate since 2017.** The period between 2017 and 2021 has witnessed a marked acceleration in the pace of publications. This is particularly noteworthy in the case of China, which reported an average yearly growth rate of 39%, followed by the US (36%) and the EU (32%) (Figure 5 – left panel). If these growth rates remain unchanged over the next four years, the gap between China and the EU will widen further (Figure 5 – right panel), and EU's publications will represent less than 60% of Chinese production.

**Figure 5. Average yearly growth by period (left), and projected number of publications (right), on AI applications in science**



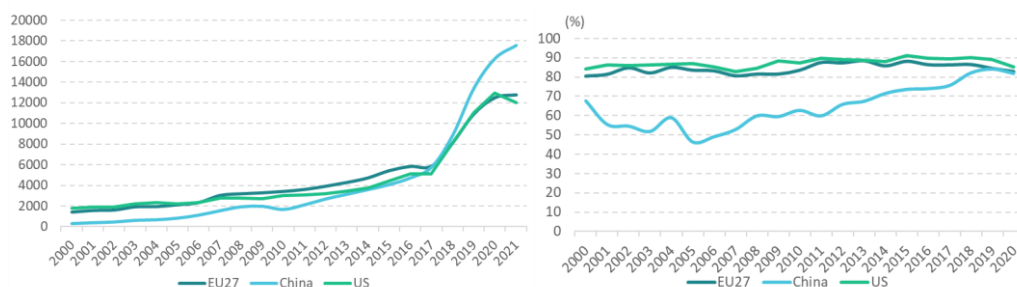
Note: the projections for 2021 are calculated applying the yearly growth rate from 2017–2021 for each of the countries/region.

Source: European Commission, DG Research & Innovation, calculations based on Web of Science data.

**However, the widening gap between China and its counterparts is narrowed when accounting for the quality of publications.** Figure 6 shows the trend in the number of publications on AI applications in science, with at least one citation (Figure 6 – left panel). When the quality of publications is accounted for, the overall performance of the three main innovators decreases compared to the overall trend reported in Figure 5. However, this drop in performance is more significant in the case of China than it is for the EU and the US, as the gap between the Chinese performance and that of its counterparts becomes narrower.

**Yet, the share of Chinese publications with zero citations has fallen in recent years.** Indeed, the share of Chinese publications reporting at least one citation has converged with that of both the US and the EU (with the former leading the way)<sup>5</sup> (Figure 6 – right panel), signalling that the quality of Chinese scientific production in AI applications has improved.

**Figure 6. Number of publications (left), and share of publications (right) with at least one citation, on AI applications in science**

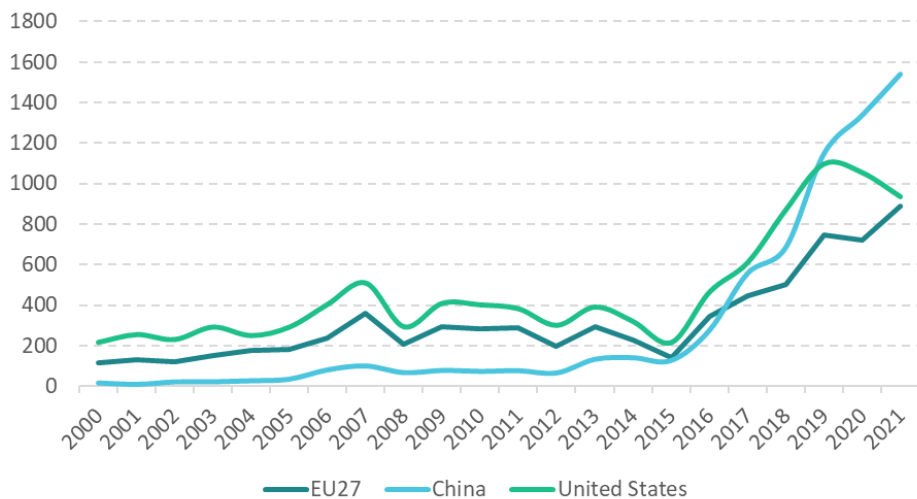


Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

<sup>5</sup> Data for 2021 have been excluded as the share of papers without citations increases as we approach the data extraction cut-off date.

**The distance between the three main players narrows further when focusing solely on top publications** (Figure 7). If we consider the top 10% of publications on AI in science by number of citations<sup>6</sup>, China only reached the head of the rankings in 2019, replacing the US in a role it had held for the previous two decades. Interestingly, the gap in high quality scientific production between the latter and the EU has also been reduced over time, with the performance of the EU progressively converging towards that of the US.

**Figure 7. Number of top publications (top 10% of publications by number of citations) on AI applications in science**



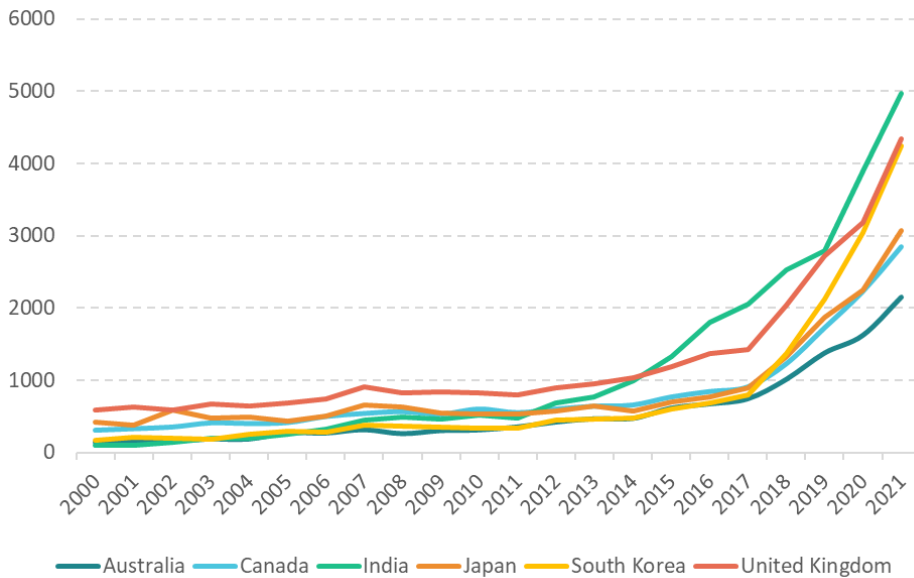
Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

Beyond the three main players, **the major performers in scientific activity related to AI applications include India, the UK, South Korea, and Japan**. The publication performance of India underwent an acceleration around 2011–2012, followed by that of the UK and South Korea. Growth in the latter became significant in 2017, with an annual growth rate of 53% for the period 2017–2021 (a much higher rate than that reported by the rest of the advanced economies analysed), followed by Japan with a yearly growth of 37%. Meanwhile, India’s publication performance has progressively slowed since 2017, reporting a yearly growth rate of 25% thereafter.

<sup>6</sup> We also considered weighted citation, but found no significant differences in the trend.



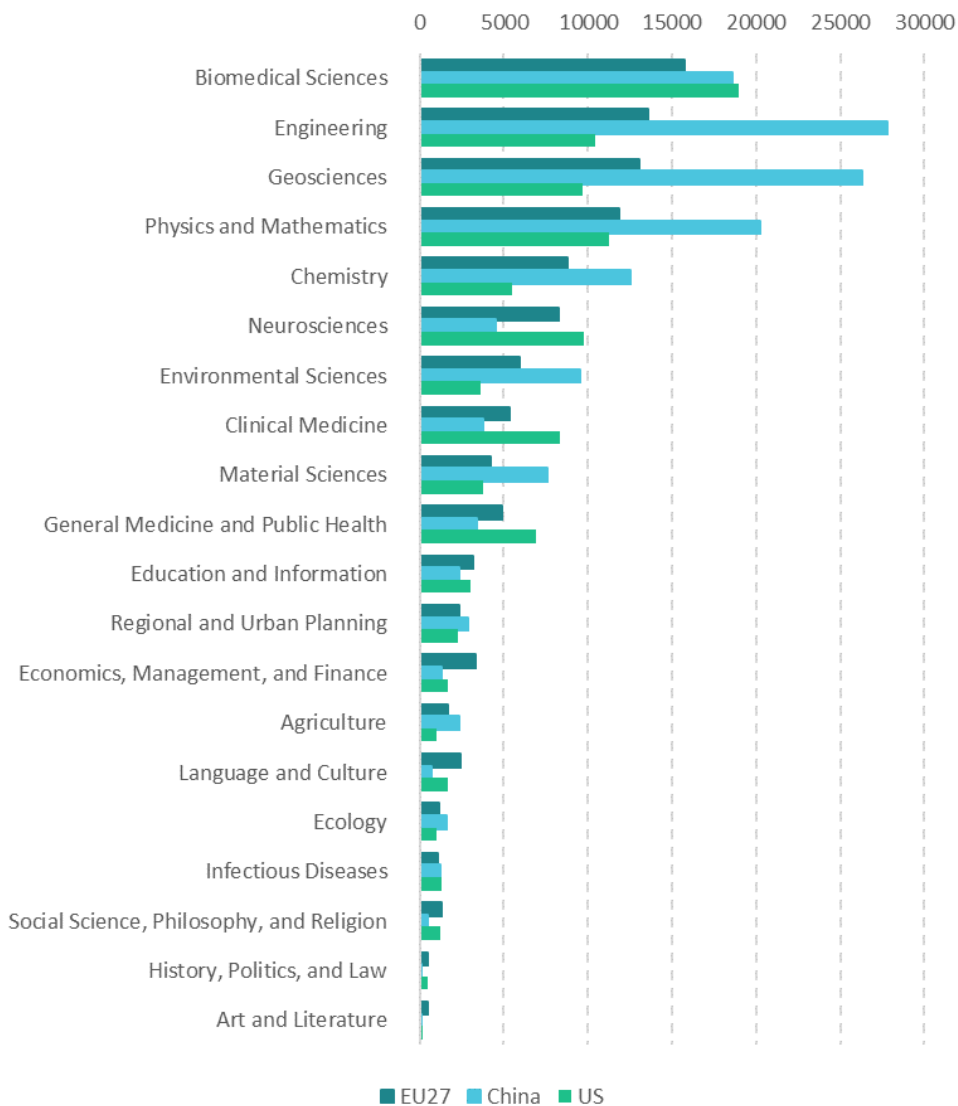
**Figure 8. Number of publications, for a selection of countries, on AI applications in science**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

**Much of the gap in scientific production between China and the EU is attributable to China’s pre-eminence in the scientific domains in which AI plays a more prominent role.** These include engineering, geosciences, and physics and mathematics where, between 2017 and 2021, China recorded the highest number of publications (Figure 9). The US heads the rankings in health-related domains, while the EU leads the field in social sciences and humanities, but lags behind in all the other major scientific domains.

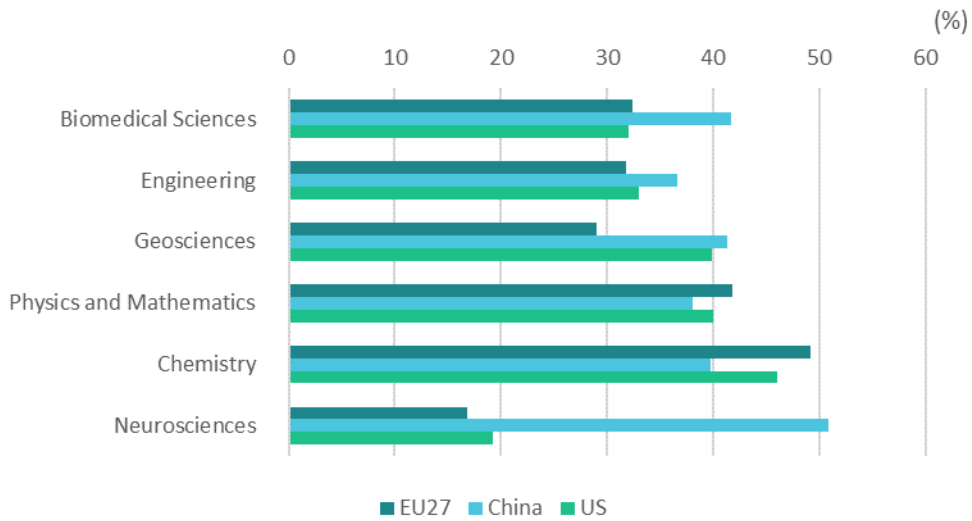
**Figure 9. Number of AI publications (2017-2021) per domain of science**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

Additionally, **Chinese production has grown significantly in all three of the main areas of AI applications** (i.e., biomedicine, engineering, and geosciences) **over the last 5 years, further consolidating its leadership in these domains.** However, it is in the neurosciences that China has shown strongest growth, pulling further and further away from both the EU and the US. The EU, meanwhile, has undergone fastest growth in the domains of physics, mathematics, and chemistry (Figure 10).

**Figure 10. Average yearly growth rate (2017–2021) in the use of AI per domain of science (for a selection of domains)**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

In terms of quality, **all the main scientific domains are characterized by shares of publications with citations of between 70 and 80%**, with the same incidence across China, the US, and the EU. Only the social sciences and humanities report a lower number of publications with citations (around 60%), with China presenting a broader gap with respect to this domain than that presented by either the US or the EU.

**Box 2: An examination of the high-impact sectors flagged up in the ‘AI Coordinated Plan’**

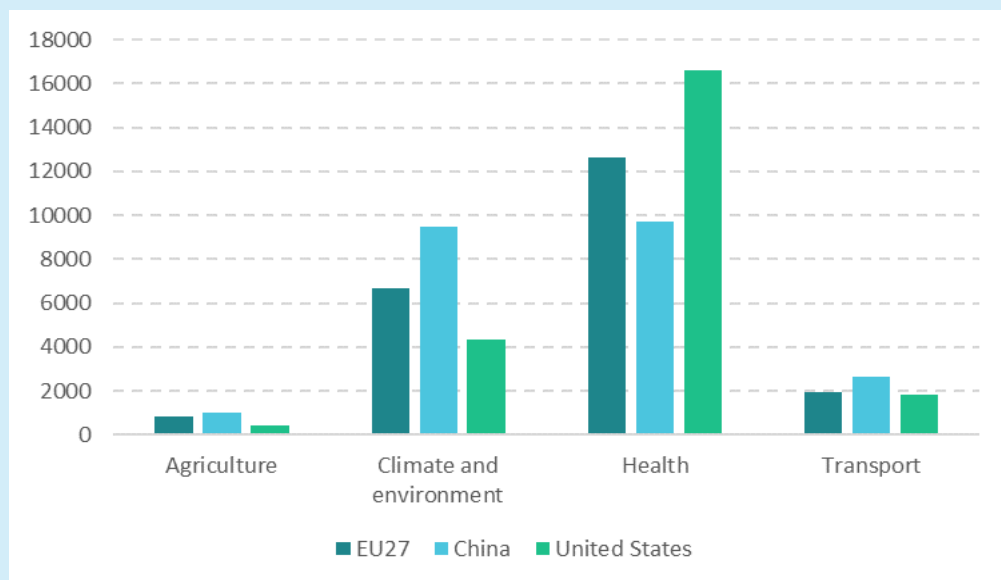
The Coordinated Plan (European Commission, 2021) identifies a set of sectors in which the EU should be striving to build strategic leadership.<sup>7</sup> Among these areas of action, there are four whose publications can be usefully analysed from a bibliometric perspective: Agriculture, Environment, Health, and Transport. In this way, a picture can be approximated of the state of play in these high-impact sectors.

To undertake such an analysis, we opted to examine the level of classification of publications in the Web of Science at a more granular level and to establish a number of synthetic categories that might proxy these sectors. The sectors, as shown in Figure 2.1, differ markedly in their adoption of AI. Thus, it is evident that the Health and Environment sectors make much greater use of AI than the other

<sup>7</sup> <https://digital-strategy.ec.europa.eu/en/policies/build-leadership-ai>

two and that, among the three main players, the EU ranks second in each, the US leading the way in Health and China in the other three sectors.

**Figure 2.1. AI publications in selected sectors (2017-2021)**

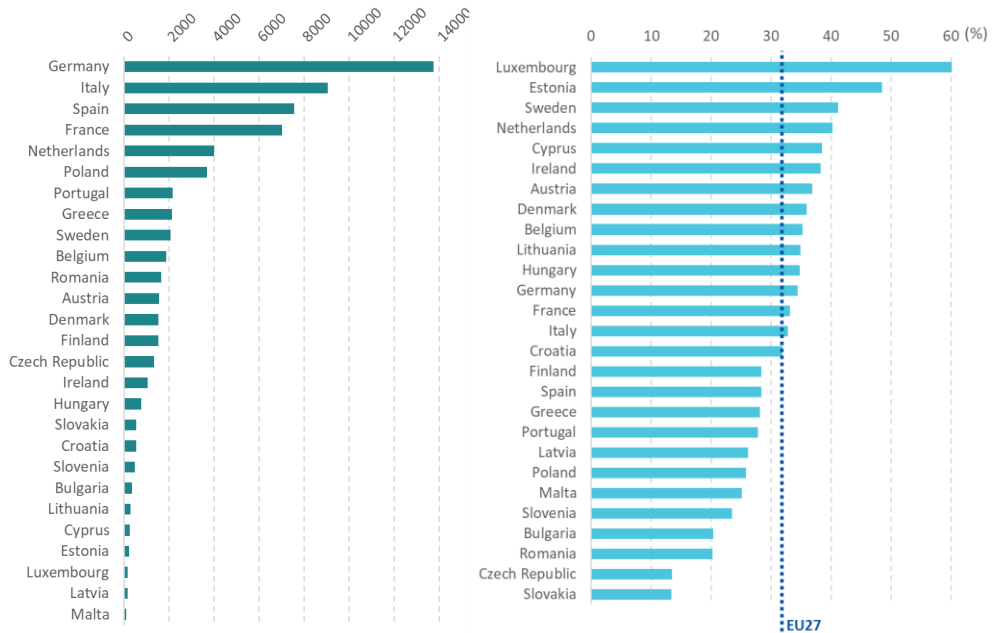


Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

#### 4. The EU research landscape in AI in science

**Among the EU Member States, Germany, Italy, Spain, and France are the leading producers of publications related to AI applied to science.** In absolute numbers, these four countries are also the leading AI-publishing nations within the EU (Figure 11 – left panel). All four, with the exception of Spain, also report a yearly growth rate slightly above the EU average (Figure 11 – right panel). Among the rest of the Member States, Sweden and the Netherlands also report both a high absolute number of publications and a high growth rate, albeit that Estonia and Luxembourg report the highest growth, although these two countries both start from a significantly lower baseline in terms of the absolute number of publications.

**Figure 11. Number of publications (left), and average yearly growth in the number of publications (right) for EU countries, 2017–2021, on AI applications in science**

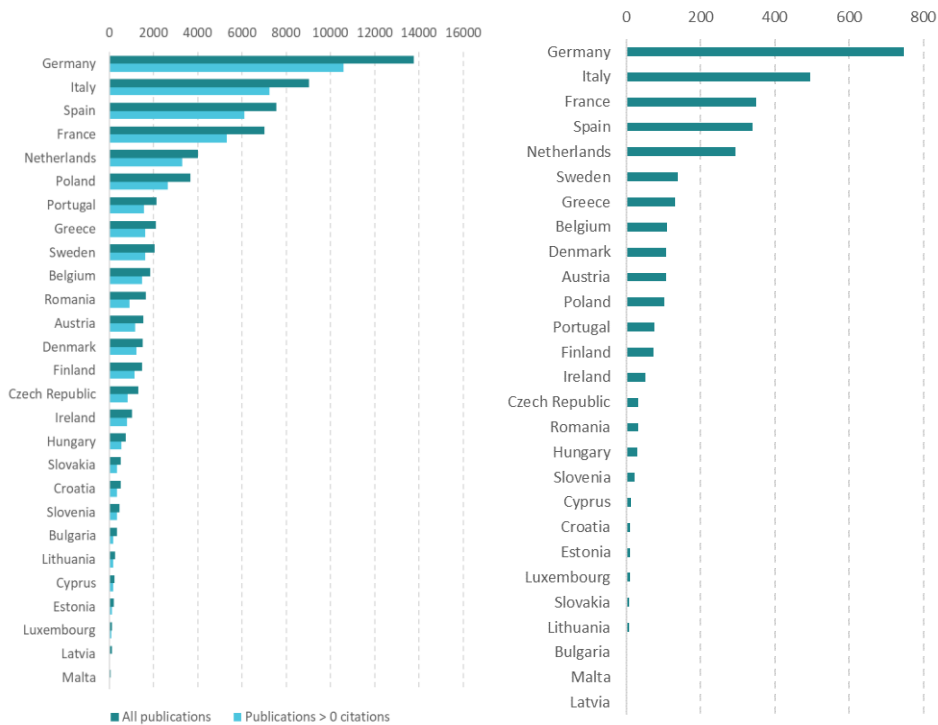


Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

**The quality of publications on AI applications is quite heterogeneous across the Member States.** Overall, the EU research landscape is characterized by a high incidence of publications on AI in science with no citations (Figure 12 – left panel),<sup>8</sup> standing at between 20 and 30% for the majority of countries. However, for some Eastern European countries, such as Romania and Czechia, the share is even higher, with more than 40% of their AI-related publications in science failing to receive a single citation. If we focus on the top 10% of the most cited papers (Figure 12 – right panel), Germany again leads the rankings, followed by Italy. France is third, just above Spain, with the Netherlands ranked just below them. Sweden follows them in 6th position by number of publications in the top 10%, and 9th by total number of publications.

<sup>8</sup> Note that this incidence is especially high for 2021, the most recent data point in the dataset.

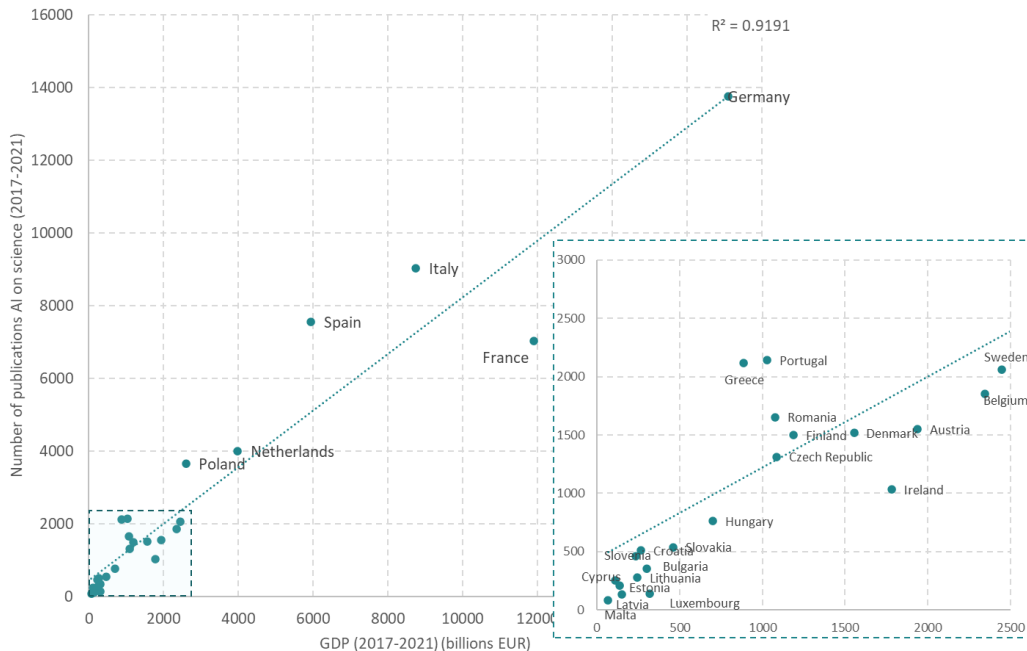
**Figure 12. Total number of publications, and number of publications with at least 1 citation (left), and number of publications among the world's top 10% cited papers (right) for EU countries, 2017-2021, on AI applications in science**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

**The number of AI publications in science produced by a country strongly correlates with its GDP.** While this correlation might be expected, GDP is surprisingly a better predictor than the expenditure on R&D (which might be assumed to be a more similar concept). As such, this correlation (shown in Figure 13) can be used to detect countries with a performance above (the case of Italy and Spain) or below (the case of France) that expected from their GDP.

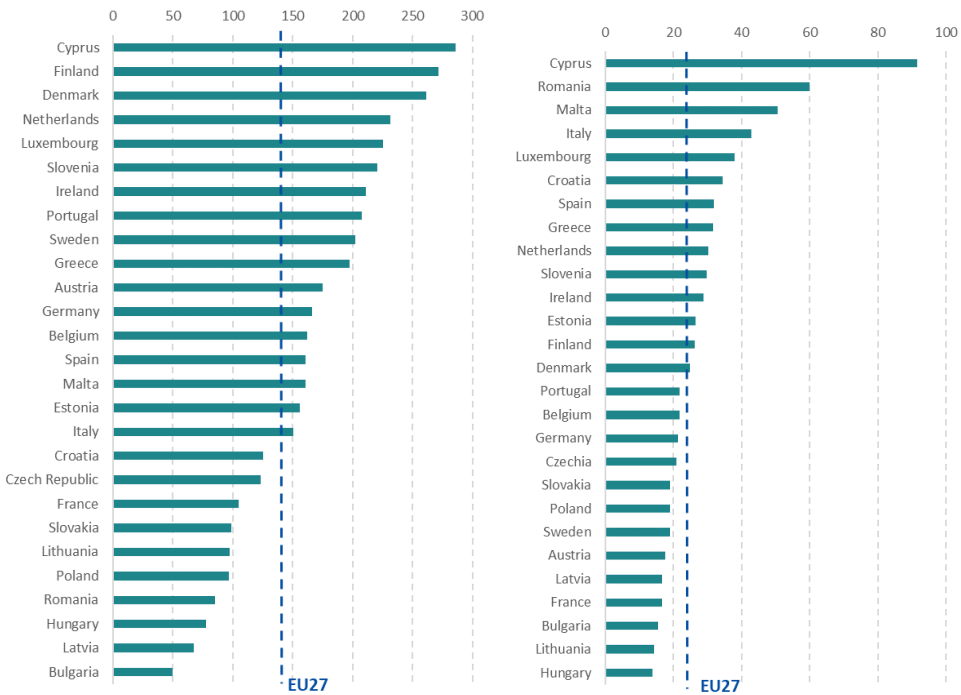
**Figure 13. GDP (2017–2021) and total number of publications of EU Member States, 2017–2021, on AI applications in science**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data and Eurostat [nama\_10\_gdp]

**Publication intensity provides a better metric for comparing the performance of the Member States.** While absolute numbers of publications may not be readily interpretable due to size differences between the EU Member States, relative measures – i.e., the number of publications by population and by number of researchers – can be more insightful (see Figure 14). These two measures complement each other, given that the number of researchers can also differ notably. When applying both indicators, Cyprus is found to perform well. Among the largest Member States, scientific output by the number of researchers in Italy and Spain is exceptional, which explains in part their respective positions in relation to GDP.

**Figure 14. Number of publications on AI applications in science per million inhabitants (left), and per thousand researchers (right), for EU countries, 2017–2021**



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data and Eurostat [rd\_p\_persage]

**Publications on AI by scientific domain follow similar patterns to that presented by the overall trend.** The countries leading the publication rankings per domain coincide, in most cases, with those reporting the highest number of publications (Table 1), with Germany leading the way in most domains (i.e., the biomedical sciences, engineering, geosciences, and also the neurosciences). While yearly growth rates are high for most countries and most domains, a few fields present very high growth rates (Table 2). Typically, the highest growth rates are found in domains with a medium to low number of overall publications, while the largest domains present growth rates more similar to those of the overall growth presented by publications on AI applications. The different growth rates per domain recorded in some countries may be indicative of the development of certain degrees of specialization in specific domains.



**Table 1. Ranking of publications per scientific domain, EU countries, 2017–2021**

	Ranking				
	1	2	3	4	5
<b>Biomedical Sciences</b>	Germany	Italy	France	Spain	Netherlands
<b>Engineering</b>	Germany	Italy	Spain	France	Poland
<b>Geosciences</b>	Germany	Italy	France	Spain	Netherlands
<b>Physics and Mathematics</b>	Germany	Italy	Spain	France	Poland
<b>Chemistry</b>	Spain	Germany	Italy	Poland	France
<b>Neurosciences</b>	Germany	Italy	Netherlands	France	Spain
<b>Environmental Sciences</b>	Spain	Italy	Germany	France	Poland
<b>Clinical Medicine</b>	Germany	France	Netherlands	Italy	Spain
<b>General Medicine and Public Health</b>	Germany	Italy	Spain	France	Netherlands
<b>Material Sciences</b>	Germany	Spain	Italy	Poland	France
<b>Economics, Management, and Finance</b>	Germany	Italy	Romania	Czech Republic	Spain
<b>Education and Information</b>	Spain	Germany	Italy	Romania	Portugal
<b>Language and Culture</b>	Germany	Spain	France	Italy	Netherlands
<b>Regional and Urban Planning</b>	Germany	Italy	Spain	France	Netherlands
<b>Agriculture</b>	Spain	Germany	Italy	France	Netherlands
<b>Social Science, Philosophy, and Religion</b>	Germany	Italy	Netherlands	Spain	France
<b>Ecology</b>	Germany	Italy	France	Spain	Portugal
<b>Infectious Diseases</b>	Germany	France	Italy	Spain	Netherlands
<b>History, Politics, and Law</b>	Germany	Italy	Netherlands	Spain	Belgium
<b>Art and Literature</b>	Italy	Germany	Spain	Denmark	France

Note: Domains are ordered by total number of publications in the EU (2017–2021).

Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

**Table 2. Yearly growth rates per scientific domain, in selected EU countries, 2017–2021**

	Germany	Italy	France	Spain	Netherlands	Poland
Biomedical Sciences	35.8	33.4	31.4	28.0	32.4	22.1
Engineering	33.2	32.5	33.3	30.1	48.6	22.5
Physics and Mathematics	23.7	25.7	21.2	35.2	19.5	50.7
Geosciences	30.1	32.8	45.8	29.7	25.5	62.0
Chemistry	29.3	36.2	12.8	47.7	20.2	30.9
Neurosciences	58.6	37.7	44.4	8.0	87.5	5.6
Environmental Sciences	52.5	58.9	97.8	38.7	58.2	69.5
Clinical Medicine	61.8	45.6	65.5	10.6	66.0	43.7
General Medicine and Public Health	55.0	33.7	54.5	35.9	73.9	5.0
Material Sciences	28.1	59.0	14.6	60.8	-3.4	82.5
Economics, Management, and Finance	29.5	32.9	37.2	48.2	65.7	14.5
Education and Information	21.3	6.1	17.7	20.6	36.0	12.3
Regional and Urban Planning	41.3	8.6	26.7	14.7	38.9	-27.0
Agriculture	31.5	48.7	26.9	18.1	23.2	67.5
Language and Culture	32.7	6.6	31.5	22.3	49.5	-11.9
Ecology	31.3	16.0	51.9	65.8	13.2	14.2
Social Science, Philosophy, and Religion	32.7	56.4	1.7	33.9	68.8	12.2
Infectious Diseases	47.6	52.8	30.3	68.3	66.7	64.8
History, Politics, and Law	42.0	12.6	-17.1	9.8	42.8	2.8
Art and Literature	38.5	7.5	38.7	46.7	-18.5	

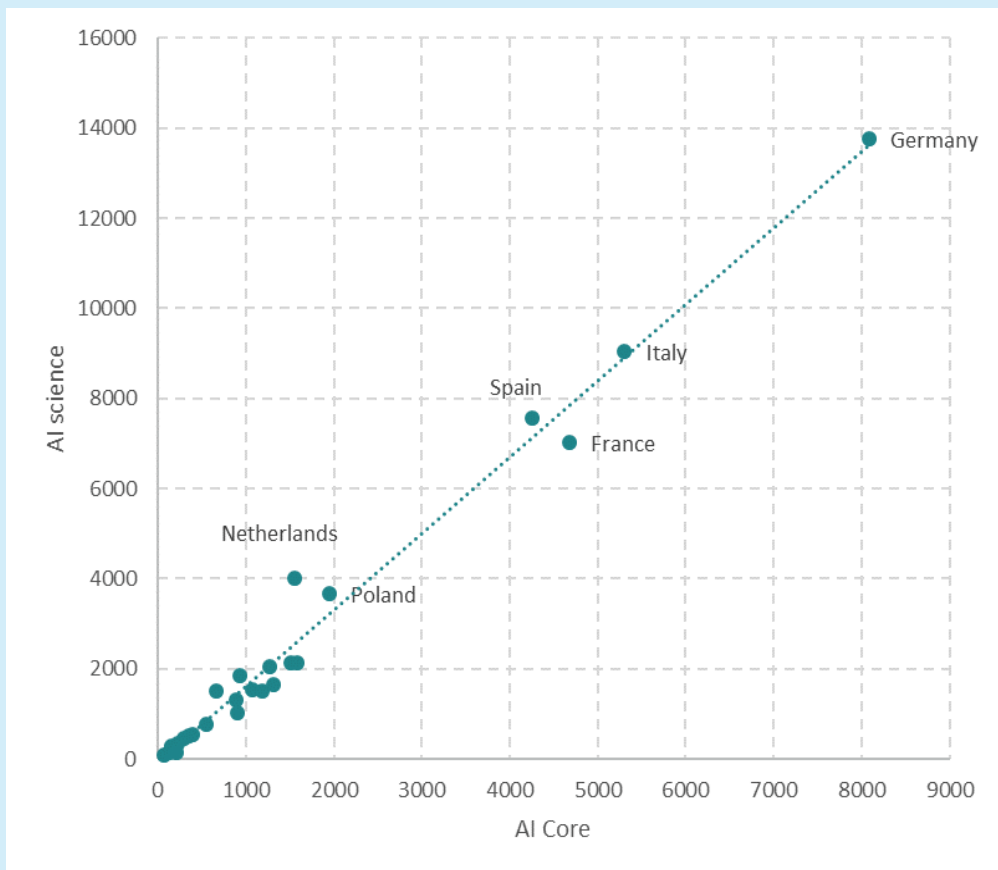
Note: Domains are ordered by total number of publications in the EU (2017–2021). Highlighted in green the top growth rate among the selected countries.

Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

### Box 3: Applications of AI in science vs “Core AI”

Technology development is fundamental in the development of applications of that technology. In the case of AI, the core of its development, and many of its applications, lie in the domain of Computer Science. For this reason, we would expect to find a strong correlation between the publications exclusive to the domain of Computer Science and those that fall within other scientific domains. However, some traits might manifest themselves at the national level, with some countries focusing their efforts more on developing “core” technology and its applications, while other countries are more active in developing AI applications in science.

Figure 3.1: Number of publications by EU countries, 2017–2021



Source: European Commission, DG Research & Innovation, calculations based on Web of Science data

This strong correlation is more than apparent in the case of the EU Member States (Figure 3.1). Yet this mapping does not allow us to draw any inferences about the direction of the causal link between the two. That is, does the development of the technology facilitate its applications, or does the need to apply it in different fields

act as a stimulus for the development of the technology? Answering this question remains critical for policymakers in their efforts to understand the dynamics underpinning this field of science, and there is an obvious need for further research on this topic.

## 5. Conclusions

**AI has an enormous potential to further advances in science and technology.** Indeed, AI constitutes a powerful tool that is capable of inducing positive change across a broad spectrum of fields, thanks to its ability to allow humans to achieve more, and at a faster pace, by enhancing existing skills. This means that AI technologies are set to play a fundamental role in increasing the efficiency of scientific and innovation actions aimed at solving complex global challenges.

**The applications of AI in science and research have grown at a significant pace in recent years.** At the global level, China is the most productive, in terms of both the absolute number of publications and the growth rate of this output, followed by other key players, most notably, the EU and the US. Although the Chinese advantage is mitigated, in part, when considering indicators of research quality, its leadership in the development of AI and in its applications in science remains significant.

Based on the evidence presented in this paper, **the gap between the EU and China in this regard can be expected to increase in the future.** The EU led the way in the application of AI in science up to 2017, when it was overtaken by China. EU production is currently at a similar level to that of the US but it presents a somewhat slower growth rate. If this trend is confirmed, the window available for the EU to catch up with its competitors is likely to shrink further in the future.

**Current trends highlight the need for actions that can strengthen the EU's position in the application of AI in research and scientific activities.** Given its multiple applications across a range of fields, AI is one of the digital technologies with the greatest potential to boost EU productivity, and to revitalize the green transition and, in this way, increase EU competitiveness. Moreover, if future scientific discoveries are likely to be driven in the main by AI applications and tools, lagging behind in the development and uptake of AI in science poses major challenges for the EU's strategic autonomy, increasing the risk of developing dependencies in strategic scientific fields.

**From a policy perspective, creating the right conditions to facilitate the uptake of AI across all scientific fields is no easy task.** Based on the results of this paper, strengthening both the EU's R&I ecosystems, as well as those of its Member States, to facilitate the further adoption of AI in science, is a key issue. Moreover, AI technologies introduce a series of broader challenges that policymakers have to address. Here, one of the most significant impacts to take into consideration is the impact of AI on jobs (including those in the domains of science). All changes in technology invariably lead to disruptions and threaten to exacerbate existing

disparities. AI is no exception and particular attention will have to be paid to boost the uptake of AI technologies, while respecting human rights and preserving the value of human endeavour and intellect, in line with a human-centric, transparent approach that promotes public trust in this field.

**Improving understanding of this technology at every stakeholder level is also critical.** To achieve this, current tools and systems need to be reimagined to help researchers, companies, and policymakers exploit the potential of AI to the full. This, likewise, calls for a better understanding of the current state of knowledge in the field, so as to enable an effective prioritization of research at the boundaries of knowledge.

**The increasing use of AI as a tool to advance science is accompanied by new challenges and dilemmas.** Being a novel technology, AI is the cause of new concerns among scientists. The swift adoption curve, as described in this paper, will serve merely to intensify these fears. While some of these problems are known or, at least, expected, new dilemmas are likely to present themselves. A prime example of the challenges faced is the way in which generative models, especially LLMs, might affect different aspects of the scientific process (Birhane et al, 2023; van Dis et al., 2023). The perils of misuse also exist, most notably in potentially hazardous areas such as drug discovery (Shankar and Zare, 2022). At a broader scale, AI could impact the scientific process by introducing bias (for instance, in literature reviews) or creating obstacles for the reproducibility and interpretability of results.

**It is, therefore, critical that steps be taken to improve the quality of available evidence when monitoring future trends and developments.** In this regard, the relevance of the business sector is especially important. The interest shown in the application of AI in science by the tech giants, having large technological and financial power, (such as, Google, Microsoft, and Meta), and by their related start-ups, could be a major game changer in this field. The role of corporations in this regard (albeit an issue not specifically examined herein) could be of great relevance too. For example, in the UK, a huge share (70%) of leading publications on AI are generated solely by DeepMind.<sup>9</sup> Likewise, open science might well be impacted as the result of increased competition between companies, since this would have serious ramifications for openness in the uptake of ideas by the research community.

**Finally, the impact of AI on scientific productivity clearly requires further analysis and monitoring.** Quantifying the current impact of AI on the productivity of science is no easy task. The fact that the number of scientific publications using AI is growing faster today than the overall number of publications can be interpreted as an initial indication of the potential of AI to boost research productivity. The anecdotal evidence of projects such as AlphaFold<sup>10</sup> moreover points in this same direction. Results from this specific project highlighted enormous productivity gains

---

<sup>9</sup> The UK's share of citations in the top 100 recent AI papers stands at 7.8% but once DeepMind's contributions are removed, it falls to 1.9%: <https://www.ft.com/content/470e9848-b2dd-4ad5-94cb-65e95c226545>

<sup>10</sup> AlphaFold, with an interdisciplinary team of around 20 scientists, was able to predict the structure of 200 million proteins in a period of between 5 to 6 years. Before this, predicting just one structure could be the work of a significant part of a PhD thesis.

in the scientific process, and opened up, at the same time, multiple additional lines of research in medicine and drug discovery.<sup>11</sup> This is an excellent example of the potential of AI tools. However, further research is needed to quantify the actual impact of AI and its evolution, taking into account other criteria beyond that of scientific production (e.g., novelty, originality, and disruptiveness).

## 6. References

Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., De Prato, G., Feijoo, C., Fernandez Macias, E., Gomez Gutierrez, E., Iglesias Portela, M., Junklewitz, H., Lopez Cobo, M., Martens, B., Figueiredo Do Nascimento, S., Nativi, S., Polvora, A., Sanchez Martin, J.I., Tolan, S., Tuomi, I. and Vesnic Alujevic, L. (2018) Artificial Intelligence: A European Perspective, Craglia, M. editor(s), EUR 29425 EN, Publications Office of the European Union, Luxembourg, <https://publications.jrc.ec.europa.eu/repository/handle/JRC113826>

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D., & Squicciarini, M. (2020), *Identifying and measuring developments in artificial intelligence: Making the impossible possible*, OECD Science, Technology and Industry Working Papers, No. 2020/05, OECD Publishing, Paris. <https://doi.org/10.1787/5f65ff7e-en>

Bianchini, S., Müller, M., & Pelletier, P. (2022), 'Artificial intelligence in science: An emerging general method of invention', *Research Policy*, 51(10), 104604. <https://doi.org/10.1016/j.respol.2022.104604>

Bianchini, S., Bottero, P., Colagrossi, M., Damioli, G., Ghisetti, C. and Michoud, K. (2023), 'Sustainable Development Goals and Digital Technologies: Mapping Scientific Research', *European Commission*, JRC133609. <https://publications.jrc.ec.europa.eu/repository/handle/JRC133609>

Birhane, A., Kasirzadeh, A., Leslie, D. et al. (2023), 'Science in the age of large language models', *Nat Rev Phys* 5, 277–280. <https://doi.org/10.1038/s42254-023-00581-4>

Bloom, N. et al. (2020), 'Are ideas getting harder to find?', *American Economic Review*, Vol. 110/4, pp. 1104-1144. <https://doi.org/10.1257/aer.20180338>

Chen, C. (2017), 'Science mapping: a systematic review of the literature', *Journal of Data and information Science*, 2(2), 1-40. <https://doi.org/10.1515/jdis-2017-0006>

Cockburn, I. M., Henderson, R., & Stern, S. (2018), The impact of artificial intelligence on innovation: An exploratory analysis, In *The economics of artificial*

---

<sup>11</sup> The AlphaFold study was the most cited AI paper in 2022: <https://www.zeta-alpha.com/post/must-read-the-100-most-cited-ai-papers-in-2022>

intelligence: An agenda (pp. 115-146), University of Chicago Press, <https://www.nber.org/papers/w24449>

De la Vega Hernández, I. M., Urdaneta, A. S., & Carayannis, E. (2023), 'Global bibliometric mapping of the frontier of knowledge in the field of artificial intelligence for the period 1990–2019', *Artificial Intelligence Review*, 56(2), 1699-1729, <https://doi.org/10.1007/s10462-022-10206-4>

De Mauro, A., Greco, M., & Grimaldi, M. (2015), What is big data? A consensual definition and a review of key research topics, AIP conference proceedings (Vol. 1644, No. 1, pp. 97-104), American Institute of Physics. <https://doi.org/10.1063/1.4907823>

EIB (2021), Artificial intelligence, blockchain and the future of Europe: How disruptive technologies create opportunities for a green and digital economy. <https://www.eib.org/en/publications/artificial-intelligence-blockchain-and-the-future-of-europe-report>

Elsevier (2020), Artificial intelligence: How knowledge is created, transferred, and used. <https://www.elsevier.com/connect/resource-center/artificial-intelligence>

Frank, M. R., Wang, D., Cebrian, M., & Rahwan, I. (2019), 'The evolution of citation graphs in artificial intelligence research', *Nature Machine Intelligence*, 1(2), 79-85. <https://doi.org/10.1038/s42256-019-0024-5>

European Commission (2021), Coordinated Plan on Artificial Intelligence 2021 Review <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

European Commission (2021b), Proposal for a Regulation on Artificial Intelligence (2021). <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

Gargiulo, F., Fontaine, S., Dubois, M., & Tubaro, P. (2022), A meso-scale cartography of the AI ecosystem, arXiv preprint arXiv:2212.12263. <https://arxiv.org/pdf/2212.12263>

Guo, Y., Hao, Z., Zhao, S., Gong, J., & Yang, F. (2020), 'Artificial intelligence in health care: Bibliometric analysis', *Journal of Medical Internet Research*, 22(7), [10.2196/18228](https://doi.org/10.2196/18228)

Institute for Human-Centered AI (2023), The AI Index 2023 Annual Report, AI Index Steering Committee, Stanford University, Stanford, CA, [https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI\\_AI-Index-Report\\_2023.pdf](https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf)

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021), 'Highly accurate protein structure prediction with

AlphaFold', *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>

Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2021), 'Deep learning, deep change? Mapping the evolution and geography of a general-purpose technology', *Scientometrics*, 126, 5589-5621. <https://doi.org/10.1007/s11192-021-03936-9>

Krenn, M., Pollice, R., Guo, S.Y. et al (2022), 'On scientific understanding with artificial Intelligence', *Nature*, <https://www.nature.com/articles/s42254-022-00518-3>

Leydesdorff, L., & Rafols, I. (2009), 'A global map of science based on the ISI subject categories', *Journal of the American Society for Information Science and Technology*, 60(2), 348-362, <https://doi.org/10.1002/asi.20967>

Nilsson, N. J. (2009), *The quest for artificial intelligence*, Cambridge University Press. <https://doi.org/10.1017/CBO9780511819346>

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris. <https://doi.org/10.1787/eedfee77-en>

Rammer, C., Fernández, G. P., & Czarnitzki, D. (2022), 'Artificial intelligence and industrial innovation: Evidence from German firm-level data', *Research Policy*, 51(7), 104555, <https://doi.org/10.1016/j.respol.2022.104555>

Russell, S., & Norvig, P. (2021), *Artificial Intelligence: A Modern Approach*, Global Edition 4th. Foundations, <https://aima.cs.berkeley.edu>

Shankar, S., Zare, R.N. (2022), 'The perils of machine learning in designing new chemicals and materials', *Nature Machine Intelligence* 4, 314–315, <https://doi.org/10.1038/s42256-022-00481-9>

Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q. H., Ho, M. T., Vuong, T. T., ... & Ho, R. C. (2019), 'Global evolution of research in artificial intelligence in health and medicine: A bibliometric study', *Journal of Clinical Medicine*, 8(3), 360. <https://doi.org/10.3390/jcm8030360>

van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023), 'ChatGPT: five priorities for research', *Nature*, 614(7947), 224-226.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020), 'The role of artificial intelligence in achieving the Sustainable Development Goals', *Nature Communications*, 11(1), 233, <https://doi.org/10.1038/s41467-019-14108-y>

WIPO (2019), *WIPO Technology Trends 2019: Artificial Intelligence*, Geneva: World Intellectual Property Organization, <https://www.wipo.int/publications/en/details.jsp?id=4386>

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021), 'Artificial intelligence: A powerful paradigm for scientific research', *The Innovation*, 2(4), 100179. <https://doi.org/10.1016/j.xinn.2021.100179>



## 7. Annex: Methodological note

### 7.1 Definitions of AI

Today, most definitions of AI centre on the concept of *emulating intelligent behaviour through machines*, where intelligence refers to the capacity to perform complex tasks in real-world environments and learn by experience:

- *Machines or agents that are capable of observing their environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions.* (EU Commission, 2018, p.19 – Annoni et al.)
- *An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments.* (OECD, 2019, p.23)
- *Learning systems, that is, machines that can become better at a task typically performed by humans with limited or no human intervention.* (WIPO, 2019, p.19)
- *AI is a term commonly used to describe machines performing human-like cognitive functions (e.g., learning, understanding, reasoning and interacting).* (OECD, 2020, p.10 – Baruffaldi et al.)
- *‘Artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I<sup>12</sup> and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.* (AI Act, European Commission 2021b)
- *A system that is capable of acquiring, analyzing, and utilizing data to learn and make decisions based on that learning.* (International Organization for Standardization (ISO), 2021)

Taken together these definitions converge in their focus on machines or AI systems that (i) possess learning capabilities, (ii) can make intelligent decisions, (iii) influence the environment, (iv) improve tasks autonomously, and (v) exhibit human-like cognitive functions.

### 7.2 Data collection methodology

The data used in this work are taken from the *Web of Science* (WoS) Core Collection. The analysis considers all scientific articles published in peer-reviewed journals and conference proceedings, and some residual categories, including books, series, etc.

---

<sup>12</sup> Annex I of the AI Act proposal

We employ a comprehensive list of 40+ search terms designed to encompass not only developments in AI technologies per se, but also AI-related developments and applications:

Artificial intelligig\*; Automated reasoning; Backpropagation; Computer vision; Data mining; Data science; Deep learning; Expert system\*; Face detection; Feature extraction; Generative adversarial net\*; Gesture recognition; Image classification; Image recognition; Image segmentation; Information retrieval; Intelligent machine\*; Kernel machine\*; Knowledge representation; Machine intelligence; Machine learning; Machine translation; Meta-learning; Multilayer perceptron\*; Natural language processing; Neural net\*; Object detection; Object identification; Object recognition; Pattern recognition; Pose estimation; Reinforcement learning; Semantic search; Semi-supervised learning; Sentiment analysis; Speech recognition; Statistical learning; Supervised learning; Text classification; Transfer learning; Transformer net\*; Unsupervised learning; Voice recognition.

The query returned all 1,275,134 publications from 2000–2022 containing at least one search term in the title, abstract, or keywords. Our approach ensures a lower-bound estimate of AI-related scientific activity.

**Table 3. AI-related keywords sorted by relevance**

Keyword	Count	Share [%]	Keyword	Count	Share [%]
Neural net*	642,003	33.58	Unsupervised learning	13,813	0.72
Machine learning	260,256	13.61	Pose estimation	11,643	0.61
Deep learning	156,046	8.16	Knowledge representation	10,673	0.56
Artificial intellig*	92,893	4.86	Image recognition	10,536	0.55
Feature extraction	92,107	4.82	Backpropagation	8,266	0.43
Data mining	75,707	3.96	Gesture recognition	7,724	0.4
Pattern recognition	65,070	3.4	Semi-supervised learning	7,524	0.39
Computer vision	57,757	3.02	Machine translation	7,470	0.39
Image segmentation	39,934	2.09	Text classification	7,431	0.39
Reinforcement learning	37,731	1.97	Data science	7,409	0.39
Object detection	29,841	1.56	Face detection	6,850	0.36
Natural language processing	28,639	1.5	Statistical learning	6,506	0.34
Information retrieval	28,605	1.5	Meta-learning	2,794	0.15
Speech recognition	28,333	1.48	Object identification	2,139	0.11
Image classification	26,773	1.4	Voice recognition	1,865	0.1
Object recognition	26,276	1.37	Semantic search	1,676	0.09
Supervised learning	22,324	1.17	Machine intelligence	1,381	0.07
Expert system*	19,182	1	Automated reasoning	1,324	0.07
Transfer learning	18,214	0.95	Intelligent machine*	1,073	0.06
Generative adversarial net*	16,346	0.85	Transformer net*	961	0.05
Sentiment analysis	14,225	0.74	Kernel machine*	862	0.05
Multilayer perceptron*	13,813	0.72			

We decompose the scientific literature into subdisciplines, aggregating 251 subject categories from the Web of Science into 21 macro-categories. This classification is inspired and enriched by the taxonomies developed in previous scientometric research (see, e.g., Leydesdorff and Rafols, 2009; Chen, 2017). A paper can be classified in multiple categories. A paper that falls into a category other than that of Computer Science is considered an application of AI *within* a scientific domain. All papers co-classified as Computer Science are also included as AI applications. The trends discussed in this report are robust to the use of a more restrictive criterion, that is, considering as AI applications only those papers that do not fall within the Computer Science category. Although any attempt at suggesting a clear disciplinary structure of science is open to criticism, we believe that our approach is sufficiently comprehensive to capture the dynamics of science of interest to us in the context of the diffusion of AI applications within different domains.

To assign a paper to a country, we apply a weighted fractional count (WFC) based on author affiliation: the total FC available per paper is 1, which is shared between authors affiliated to different countries, assuming that each author contributed equally to the research – i.e., a paper with two authors from two different countries (A and B) will be counted as 0.5 for country A and 0.5 for country B. The main trends discussed in this paper are robust to other aggregation strategies (e.g., aggregating certain residual categories and full vs. weighted counts, among others).

## GETTING IN TOUCH WITH THE EU

### In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online ([european-union.europa.eu/contact-eu/meet-us\\_en](https://european-union.europa.eu/contact-eu/meet-us_en)).

### On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: [european-union.europa.eu/contact-eu/write-us\\_en](https://european-union.europa.eu/contact-eu/write-us_en).

## FINDING INFORMATION ABOUT THE EU

### Online

Information about the European Union in all the official languages of the EU is available on the Europa website ([european-union.europa.eu](https://european-union.europa.eu)).

### EU publications

You can view or order EU publications at [op.europa.eu/en/publications](https://op.europa.eu/en/publications). Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre ([european-union.europa.eu/contact-eu/meet-us\\_en](https://european-union.europa.eu/contact-eu/meet-us_en)).

### EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex ([eur-lex.europa.eu](https://eur-lex.europa.eu)).

### EU open data

The portal [data.europa.eu](https://data.europa.eu) provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

This paper provides new insights into the role of Artificial Intelligence in scientific research across various domains of science. It is evident that AI as applied to science and research has been growing at a significant pace in recent years, with China leading the way, followed by the EU and the US. Indeed, the EU's performance is found to be highly heterogeneous across its Member States. If current trends persist, the probability of future scientific discoveries being driven primarily by AI applications and tools is set to increase significantly. Failure to keep pace with the development and uptake of AI in science poses important challenges for the EU, including the need to address concerns about strategic autonomy, trends in scientific work and the labour market and, more broadly, productivity and growth.

*Studies and reports*

