Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

AI Technological Trajectories in Patent Data: General Purpose Technology and Concentration of Actors

Kerstin Hötte, Taheya Tarannum,
Vilhelm Verendel & Lauren Bennett

30 May 2023

INET Oxford Working Paper No. 2023-09

AI Technological Trajectories in Patent Data

General Purpose Technology and Concentration of Actors

Kerstin Hötte,[1,2,3] Taheya Tarannum,[4] Vilhelm Verendel,[2,5] Lauren Bennett[4,6]

[1] *The Alan Turing Institute, London, UK*
[2] *Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK*
[3] *Faculty of Business Administration and Economics, University of Bielefeld, DE*
[4] *Oxford Martin Programme on Technological and Economic Change, University of Oxford, UK*
[5] *Department of Physics, Chalmers University, SE*
[6] *The Leverhulme Doctoral Training Programme for the Ecological Study of the Brain, University College London, UK*

May 30, 2023

It is often claimed that Artificial Intelligence (AI) is the next general purpose technology (GPT) with profound economic and societal impacts. However, without a consensus definition of AI and its empirical measurement, there are wide discrepancies in beliefs about its trajectory, diffusion, and ownership. In this study, we compare four AI patent classification approaches reflecting different technological trajectories, namely (1) short-range, (2) academic, (3) technical, and (4) broad interpretations of AI. We use US patents granted between 1990-2019 to assess the extent to which each approach qualifies AI as a GPT, and study patterns of its concentration and agency. Strikingly, the four trajectories overlap on only 1.36% of patents and vary in scale, accounting for shares of 3-17% of all US patents. Despite capturing the smallest set of AI patents, the short-range trajectory identified by the latest AI keywords demonstrates the strongest GPT characteristics of high intrinsic growth and generality. All trajectories agree, however, that AI inventions are highly concentrated within a few firms and this has consequences for competition policy and market regulation. Our study highlights how various methods of defining AI can lead to contrasting as well as similar conclusions about its impact.

JEL Classifications: O31, O33, O34
Keywords: Patent, Artificial Intelligence, AI, Classification, General Purpose Technology, Concentration, Inventions, Innovation

Highlights

- We compare 4 different technological trajectories of AI patents (short-range, academic, technical, broad view).
- The 4 trajectories differ by growth and scale (60k-600k patents), and overlap in only 1.36%.
- All trajectories exhibit GPT characteristics (growth, generality, complementarity).
- The keyword-based short-range trajectory performs best to reproduce GPT characteristics.
- Concentration of a few leading innovative firms is consistent across the different trajectories.
- Conclusions about the growth and impact of AI may be sensitive to choice of trajectory.

Contents

1. Introduction

Artificial intelligence (AI) is often claimed to be the next general purpose technology (GPT) that will shape the technological and economic evolution of the 21st century (Agrawal et al., 2019; Cockburn et al., 2019; Brynjolfsson et al., 2021). AI is expected to affect a wide range of industries through its ability to augment and complement products and processes (Brynjolfsson and T. Mitchell, 2017; Cockburn et al., 2018), as epitomised by the release of ChatGPT in November 2022. Researchers predict radical impacts on labour markets producing winners and losers (Trajtenberg, 2018), with some estimating that the technology may render 47% of existing jobs in the US obsolete (Frey and Osborne, 2017). On the other hand, AI optimists envision that, if firms up-skill workers and adapt their business to exploit AI, rising productivity will expand output, transform occupational task structures and stimulate long-run employment that will help overcome some of the 21st century's greatest challenges (Gries and Naude, 2018; Acemoglu et al., 2020).

Research on the impact of AI requires robust measures of AI diffusion (Valdes and Rudyk, 2017; Webb, 2019; Alderucci et al., 2020). However, the ex-ante scientific measurement of radical novelty is one of the greatest challenges for scholars of technological change (Schumpeter, 2005). AI development and diffusion is widely believed to still be at an early stage (Brynjolfsson et al., 2021), and it is not yet clear how the technology will impact long-run economic variables. Moreover, a consensus about a clear technological trajectory of AI has not yet emerged (Jacobides et al., 2021).

Patents provide a quantitative and qualitative track-record of inventions and are often used to analyse innovation: patent classifications that associate patents with technological fields provide insights into the trajectories of technological change (Jaffe and De Rassenfosse, 2019). As AI inventions spread through the economy, existing classification schemes may be unfit to capture complex technological networks with many mutual dependencies (Hall and Trajtenberg, 2006; Petralia, 2020). Furthermore, AI's definition is itself a product of the scale and scope of its use within the current political economy, as it is delineated by those developing, using, and investing in it. Measuring AI's diffusion is important for more than investment, as barriers to entry and economies of scale raise concerns about market and political power if AI invention and innovation become increasingly concentrated within a few key actors (Petit, 2017; Jacobides et al., 2021; Fanti et al., 2022).

In this paper, we use four distinct methods of classifying AI inventions to assemble and compare four sets of patents corresponding to different AI trajectories. Specifically, we compare methods based on:

1. keywords focusing on recent trends in neural networks, robotics, and natural language processing (NLP);

2. scientific citations reflecting the academic origins of AI;

3. the World Intellectual Property Organization (WIPO) classification method accounting for both the hardware and software dimensions of AI; and

4. the United States Patent and Trademark Office (USPTO) approach capturing the widespread use of AI in other inventions.

By these trajectories, we identify a total of 736k AI patents from 1990 to 2019, although this varies greatly: 67k patents are captured by keywords, 178k by science citations, 159k by the WIPO, and 595k by the USPTO approach. Strikingly, all four trajectories agree on only 1.36% of AI patents with pairwise overlaps of 20% or less throughout the entire period. Additionally, the four approaches also reflect disparate trends of AI inventions over time, with the Science and USPTO approaches demonstrating a slow-down in recent years that contrasts the continued growth observed in Keyword and WIPO patents.

We then evaluate each approach according to the three GPT characteristics proposed by Bresnahan and Trajtenberg (1995), Hall and Trajtenberg (2006), and Petralia (2020):

1. GPTs are engines of growth with a wide scope for technical improvement (Petralia, 2020). For each of our four AI pathways, we measure this feature by calculating the growth rates of patents, as well as the growth rates of patents that rely on these AI inventions (indicated by citation links). GPTs would show positive or accelerating growth.

2. GPTs can be used across a wide range of products and processes. To measure this, we examine the technological diversity of patent citations. Historically, GPTs experience delays of up to forty years before becoming widely utilised throughout the economy (Comin and Mestieri, 2018). Following Hall and Trajtenberg (2006), we therefore measure the citation lag between AI and subsequent inventions.

3. GPTs complement extant and novel technologies by transforming myriad production processes (Petralia, 2020). We measure this by studying the diversity of co-classifications of AI patents across different technology groups. A high diversity suggests that AI complements a wide range of technological fields.

Furthermore, we also analyse how each approach frames the agency dimension of AI invention by comparing the contribution of key actors across industries, research centres and the public sector.

Given the heterogeneity of these trajectories and the lack of a clear definition of AI (Krafft et al., 2020), we do not aim to identify the best method of measuring AI. Instead, we highlight how the choice of trajectory alters the projections of GPT characteristics of AI, and its industrial concentration, both of which are central aspects to investment and to the current political and academic debate.

The rest of our paper is structured as follows: Section 2 describes previous attempts to measure GPTs and AI in patent data, while Section 3 introduces our own approach. Our results are presented in Section 4, followed by a discussion in Section 5. Section 6 concludes. Appendix A provides details on the data and methods, and Appendix B describes the four approaches to tracking AI's trajectory.

2. Background

Here, we review the arguments for AI being a GPT, discuss their limitations and introduce the use of patent data to study AI technological trajectories.

2.1. What is a GPT, and does it matter for AI?

Bresnahan and Trajtenberg (1995) introduce the concept of a GPT as a technology which pervades the economy and spurs inventions, both endogenously and through complementarities. Well-established GPTs, such as electricity and Information and Communication Technologies (ICT), are considered as drivers of growth and technological progress. Much research has focused on their identification in quantitative data. GPTs are said to share three characteristics: pervasiveness throughout the economy, capacity for rapid intrinsic improvement, and the ability to spawn spillover productivity across sectors though complementarities (Bresnahan and Trajtenberg, 1995; Lipsey et al., 2005).

The first of these criteria refers to GPTs' ability to engender new methods of production or innovation. Due to their pervasiveness, GPTs inspire a wave of technological inventions as they embody new universal tools for production and research. Many impactful technologies, such as nuclear power and fMRI, lack the generality required to pervade a significant number of sectors (Agrawal et al., 2018; Brynjolfsson et al., 2019).

The second criterion refers to GPTs' inherent capacity to rapidly improve. The final criterion captures how GPTs spawn productivity spillovers across a range of industries through their ability to augment and complement extant products and processes. Indeed, many GPTs act as agents of creative destruction by restructuring established processes throughout the economy (Lipsey et al., 2005).

Altogether, these criteria necessitate a longer time period for GPTs to evolve, spread and unleash their full economic impact (Lipsey et al., 2005). However, once they have sufficiently evolved, GPTs rely on supporting infrastructure and secondary inventions in order to restructure production processes throughout the economy.

Recent work has raised potential issues with conceptualising AI as a GPT from a policy perspective. Bresnahan and Trajtenberg (1995) argue that GPTs could create market failures because they are useful everywhere (public good characteristics), which may generate lost social gains due to underinvestment. However, it is unclear whether AI represents a single GPT or is an amalgamation of many technologies that fulfill different functions, including data provision, prediction, classification, soft- and hardware, and edge applications (Jacobides et al., 2021). As such, the study of GPTs may need to be broadened to include key actors that shape technological evolution.

The concentration of AI development within a few key players (Klinger et al., 2022; Babina et al., 2023) has raised concerns about the power of Big Tech (Jacobides et al., 2021) and hubs (Klinger et al., 2022). Indeed, this concentration has raised doubts about the need for AI to be publicly funded, (as proclaimed by the GPT perspective (Jacobides et al., 2021)), as this may entrench existing asymmetries in the absence of further intervention from competition policy and market regulation (Petit, 2017; Hennemann, 2020).

In this paper, we analyse whether these concerns are consistent with four different theoretical perspectives on AI. We begin by introducing these classification approaches and then analyse the trajectories of AI innovations they define. Comparing multiple definitions of AI in this way serves many purposes: firstly, different definitions of AI impact the timescale it is viewed on and the perceived history of its trajectory. Secondly, theoretical approaches will differ in how they identify key actors shaping AI innovation and the impact on market power. Thirdly, the choice of definition and technological trajectory may spur differing guidance on the direction of funding and investment to AI technologies.

2.2. AI history and different technological trajectories

Broadly speaking, AI is concerned with systems and technologies associated with intelligence. As there is no consensus around a precise definition of AI, research evaluating AI as a GPT and its impact has this as an open and often unaddressed question.

The conceptual roots of AI stem back to Alan Turing (Turing, 1950) in the 1950s, when a series of inventions in the following decades led to breakthroughs in the ability of computers to perform reasoning and solve complex problems. High expectations coupled with limited computing power and funding withdrawal saw a slowdown in AI development during both the late 1970s and again in the 1990s, forming periods now referred to as 'AI winters' (Stuart and Norvig, 2003).

Modern AI (at latest, since the mid 2000s) is to a large degree based on machine learning, or computational approaches to detecting and modeling patterns in various data sources (T. M. Mitchell et al., 2007). Machine learning builds upon computer science, statistics, probability and optimisation. As machine learning methods are often paired with sensors, actuators, and controls to create intelligent systems, the mix of machine learning software with computer hardware has been considered as one way to build AI systems. From a historical perspective, AI has been associated with several technological trends occurring broadly on three different time-scales.

First, from a *long-range* perspective (decades to centuries), AI can be viewed as being only the latest step in a long trend towards technological automation (Aghion et al., 2018). Historically, technologies such as the thermostat and the microprocessor enabled breakthroughs in automation based on physics and general digital control, creating the possibility to perform automated tasks, e.g., counting. Modern AI can therefore be considered as the outcome of the long-term co-evolution of hardware, software, and networking technologies including the internet (Bonaccorsi and Moggi, 2020).

Second, from a *mid-range* perspective (decades), AI is associated with many more modern forms of computing. This AI includes systems based on digital sensors and actuators that are paired with computational methods to perform tasks previously associated with human-like intelligence. Such a view relates AI to technologies beyond machine learning. As a specific example, the modern standard textbook for AI describes technological systems *that receive percepts from the environment and take actions that affect that environment* (Stuart and Norvig, 2003). The notion that AI is broader than the latest trends in machine learning is also referred to as "the good old-fashioned AI".

Third, from a *short-range* perspective (years to a decade) AI is believed to consist of the latest developments in machine learning, involving natural language processing, computer vision, and deep learning. Recent work Bianchini et al. (2020), Klinger et al. (2022), Jurowetzki et al. (2021), and Whittaker (2021) contends that modern AI research is being rapidly privatised and narrowed towards deep learning (DL) at the expense of other relatively unexplored domains. However, it remains to be seen whether this narrowing will impact on the ability of this modern definition to capture the full scope of future AI inventions that may include sub-fields that are, at present, relatively unexplored.

Given the complexity of defining AI, which view should a policy-oriented analysis take? Clearly, there is some arbitrariness of choice here. To review a recent example, the proposed EU AI Act (European Commission, 2021) illustrates the balance between including the latest developments in machine learning, with well-established methods from other disciplines in AI's definition. In the proposed EU AI Act, AI is defined as,

> software that is developed with one or more of the **techniques and approaches** listed in Annex I 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

and these techniques and approaches are further specified (Annex I, European Commission, 2021) as,

> (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
> (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
> (c) Statistical approaches, Bayesian estimation, search and optimisation methods.

Clearly, the scope of AI here includes mathematics, statistics, optimisation and computer science. In other words, it appears as though AI policy in the EU may be guided by the mid-range definition of the technology.

2.2.1. Technological trajectories of AI

These differing perspectives on AI are ultimatley associated with different *technological trajectories* (Dosi, 1988). Trajectories are innovative activities characterised by being,

- *selective*: focusing on some specific means (e.g., means to fulfil a human purpose);
- *precise*: narrow in their directions;
- *cumulative*: providing increased problem-solving capabilities.

In other words, technological trajectories serve as uncertainty-reducing representations for communities studying them (Bonaccorsi and Moggi, 2020). Although AI has previously been associated with technological trajectories, (Fanti et al., 2022; Iori et al., 2021), we aim to show how four technical trajectories of AI differ in their conclusions regarding the projected impact and spread of AI, as well as what role it will play in current political and economic debates, investment and projections.

2.2.2. Patents as a data source

In our study, we use patent data to investigate overlaps and divergences in the characteristics and ownership between four AI trajectories. Patents are detailed track records of inventions and require the disclosure of the technological knowledge embodied in the patented invention. Patent offices hold stock of millions of patents assigned to different technological fields using hierarchical patent classification systems. One patent can be classified into a variety of technology fields, which describe the nature of the patented technology (Jaffe and De Rassenfosse, 2019). Patents have been previously been used to study AI inventions for technological trajectories on the level of individual patent citation networks (Iori et al., 2021), as well as AI inventions in the aggregate (WIPO, 2019b; Verendel, 2023).

Moreover, patent data is well-suited to the study of technological trajectories. Patents are typically organised around protected claims (selective in functionality), which must be highly specific (being precise) and novel (cumulative in problem-solving capabilities). Additionally, patents are assigned to one or more codes from hierarchical patent classification systems, such as the Cooperative Patent Classification system developed by the European and US patent office.[1] These codes help describe the qualitative nature of a patent, for example whether it is related to specific computing techniques or hardware for data transmission. Patent documents also include citation links to other patents and, in some cases, academic articles. This helps distinguish the patented invention from previous inventions and inventors are required to establish the novelty of their invention compared to existing technologies. These citations are frequently used by innovation scholars as they reveal the cumulative dependencies between technologies, and may be interpreted as an indicator for the extent to which an invention builds on previous technological knowledge (Jaffe and De Rassenfosse, 2019). Previous discussions of AI as a GPT have also used patent data (Cockburn et al., 2018) and firm-level non-patent data (Klinger et al., 2018), both of which we detail below.

3. Approach

In our analysis, we compare the GPT characteristics of four sets of AI patents identified by distinct classification techniques, reflecting different technological trajectories associated with AI. More specifically, we contrast the estimated growth, generality and complementarity of AI patents classified by either (1) keywords, (2) science citations,

[1] https://worldwide.espacenet.com/patent/cpc-browser#

(3) the WIPO, and the (4) USPTO method from all USPTO patents granted from 1990-2019.

First, the Keyword method is replicated from Cockburn et al. (2018) and relies on text search terms (NLP, robotics, neural networks). This approach reflects short-term perspectives on AI and attributes it's progress to themes that became dominant over the past decade.

Second, using scientific citations to AI research (including grey literature and conference proceedings) Marx and Fuegi (2020) conceptualises AI technologies as an outcome of science and academic research (Arthur, 2009; Jee and Sohn, 2023a). As seen below, this approach is subject to limitations as citation practices differ across technological fields.

Third, the WIPO method focuses on the technical underpinnings of AI, including AI functions, techniques, ML, and various areas of applied computing, by combining a set of keywords with computer-specific technological classification codes (WIPO, 2019a).

Fourth, the method by the USPTO relies on the broadest understanding of AI. This technique uses a trained ML classifier on patent text and citations to identify innovations related to knowledge processing, speech, hardware, evolutionary computation, NLP, ML, computer vision, planning and control (Giczy et al., 2021). This broad conceptualisation of AI is also reflected in the volume of AI patents identified through this method, including a large share of downstream AI applications.

In the literature, the term *technological trajectory* has been sometimes used to specifically refer to trajectories of patent citations, and otherwise in a broader, more conceptual sense. In this article, we use the term trajectory to refer to one of the four sets of patents identified through the different classification methods. For further background on these technological trajectories and detail on their implementation, we refer the reader to supplementary information Sections A and B.

4. Results

In the following analysis, we compare the four trajectories according to:

1. General trends in the ownership and industrial affiliation of AI inventions;
2. The extent to which they demonstrate the three GPT characteristics of *growth*, *generality* and *complementarity* over time; and
3. The market concentration of key actors generating AI.

4.1. General Characteristics of AI trajectories

In this section, we explore the actors driving AI inventions in each trajectory. Across all trajectories, we observe that AI inventions became increasingly diversified across industries and countries over the past three decades. Table 1 gives an overview of the trajectories, showing their patent counts, growth rates, inventor types, industry affiliations, countries of origin, public support, and main technology classes.

The trajectories agree that AI inventions are disproportionately borne from commercial enterprises and concentrated in the fields of computer manufacturing and machinery. Despite classifying numbers of patents that differ on the order of magnitude, all approaches show the expected dominance of the US in AI patents, although our use of US patents over-represents US inventors.

At the points of difference, the growth rates for the Science and the USPTO approaches suggest that the rate of AI inventions has, at least for now, passed its peak. As may be expected, the Science approach captures the most non-commercial AI inventions from individuals, non-profit organisations and universities, with a notably higher concentration of patenting ascribed to the pharmaceutical industry, compared to machinery and manufacturing for both WIPO and Keyword. The high share of pharmaceutical patents in the Science trajectory partly arises from the fact that patents in biotechnology are more frequently filed by academic inventors who generally include more scientific citations. Geographically, the Keyword and WIPO trajectories include more AI patents of foreign inventors, especially from Japan.

Table 1: AI trajectories by four classification approaches

	Keyword	Science	WIPO	USPTO
A. Volume				
Number of patents	67,187	178,004	158,652	595,047
B. Growth				
Growth rate (1990-99)	3.57	6.69	2.38	3.87
Growth rate (2000-09)	0.87	1.32	1.21	1.16
Growth rate (2010-19)	4.31	0.91	3.02	1.07
C. Inventor type (patent assignee)				
% Commercial	0.89	0.85	0.90	0.91
% Individual	0.05	0.31	0.05	0.05
% Non-profit	0.07	0.13	0.06	0.05
% University	0.05	0.10	0.04	0.03
D. Industry affiliation (patent assignee)				
% Pharmaceuticals (manufacturing)	0.01	0.17	0.01	0.01
% Computer (manufacturing)	0.74	0.76	0.79	0.84
% Machinery & equipment (manufacturing)	0.43	0.28	0.54	0.22
% Other manufacturing	0.13	0.10	0.09	0.09
% Computer programming (service)	0.08	0.07	0.09	0.12
E. Country of origin (patent applicant)				
% USA	0.64	0.75	0.65	0.72
% Japan	0.13	0.08	0.15	0.10
% S. Korea	0.04	0.02	0.03	0.02
% Germany	0.04	0.03	0.03	0.03
% China	0.02	0.01	0.02	0.01
% Canada	0.02	0.02	0.02	0.02
F. Public support				
% Public support (1990-99)	0.25	0.39	0.29	0.21
% Public support (2000-09)	0.31	0.48	0.31	0.23
% Public support (2010-15)	0.30	0.52	0.28	0.22
G. CPC 1-digit codes				
% Human necessities (A)	0.13	0.18	0.08	0.08
% Performing operations (B)	0.21	0.05	0.11	0.04
% Chemistry; Metallurgy (C)	0.02	0.16	0.01	0.01
% Physics (G)	0.70	0.75	0.95	0.79
% Electricity (H)	0.32	0.26	0.27	0.33
% General/Cross-sectional (Y)	0.05	0.04	0.03	0.04

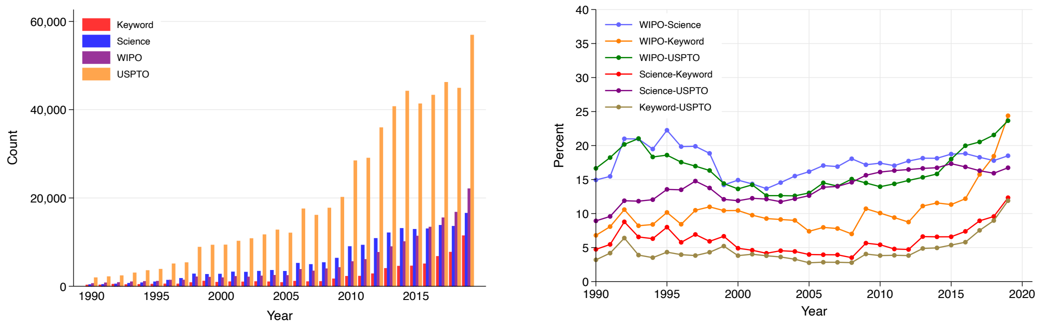
Notes: This table compares scale and scope of AI invention identified by each trajectory, with respect to inventor type (commercial, individual, non-profit, university), industry affiliation (based on NACE Rev. 2 classification), country of origin and technological classification. The Keyword approach identifies the smallest sample, whereas the USPTO sample is almost nine times bigger. Keywords and WIPO estimates a larger growth rate over the last decade (4.31% and 3.02%, respectively). The science approach identifies larger shares of non-commercial patents, pharmaceutical patents, and patents receiving public support. Since we use the USPTO data, the largest country share comes from the US. The distribution across technology classes as shown by CPC 1-digit codes look largely similar. The notable difference is that WIPO identifies a larger share of patents coming from technology class G (Physics).

Each trajectory paints a different picture with respect to trends in government support for AI innovations. By the USPTO trajectory, only one-fifth of AI patents received public support, however for the Science trajectory, this jumps to half in the last decade.

Comparing how the inventions are classified by CPC 1-digit codes, WIPO patents are most notably clustered in the 'Physics' category, while 'Chemistry' patents are most clearly identified by the Science approach. Patents identified by the Keyword and Science approaches produce the most diverse range of AI patents.

Figure 1 shows the pace of AI invention in each trajectory. Represented as a share of all granted US patents, the USPTO approach identifies roughly 16.6% of all patents as AI in 2019 (see Figure E.10 in Appendix E.3.1). For all approaches, we find that this share increased over time from 1-2% in the 1990s to 3-17% in 2019.

Figure 1: AI Patents by Year (1990-2019)



(a) Number of AI patents

(b) Jaccard Similarities

Notes: The right panel shows the number of AI patents over time as identified by the four approaches. The left panel shows the evolution of the Jaccard similarities computed for each year in our dataset.

To quantify the overlap and differences between the approaches, we compute Jaccard similarities and illustrate their evolution over time.[2] In Figure 1b, we plot the evolution of the Jaccard similarity for pairwise comparisons of the AI trajectories. All of them show a relatively low pairwise overlap, ranging at or below 20%. The WIPO approach shows the highest agreement with the others. The WIPO-Keyword pair and the Science-USPTO pair experienced the strongest increase in similarity over time. The Keyword approach demonstrates the lowest Jaccard similarity to the other classification approaches.

Altogether, we find that only 10,859 or 1.36% of all unique granted patents are uniformly identified as AI patents by all four approaches, highlighting how the choice of classification greatly impacts the perceived scale and reach of AI.

[2] The Jaccard similarity for two sets of patents is given by

$$J(A, B) = \frac{|\text{patents in both A and B}|}{|\text{patents in union of A and B}|} = \frac{|A \cap B|}{|A \cup B|}$$

with $J(A, B) \in [0, 1]$ where $J(A, B) = 0$ if both sets do not overlap and $J(A, B) = 1$ if both sets are identical. In other terms, the overlap between the sets can range from 0% to 100%.

4.2. GPT Characteristics of AI

In this section, we compare the growth, generality, and complementarity of the trajectories. Table 2 summarises the key characteristics.

Table 2: Measure of GPT Characteristics of AI

	Keyword	Science	WIPO	USPTO
	A. Growth			
Avg. growth rate	0.16	0.15	0.14	0.13
	B. Generality			
Avg. generality index (1 digit)	0.76	0.73	0.72	0.68
Avg. generality index (3 digit)	0.91	0.87	0.87	0.84
Avg. generality index (4 digit)	0.96	0.95	0.94	0.93
	C. Complimentarity			
Avg. number of CPC (1 digit)	1.39	1.40	1.36	1.27
Avg. number of CPC (3 digit)	1.61	1.64	1.64	1.43
Avg. number of CPC (4 digit)	1.88	1.97	2.05	1.64

Notes: This table gives a comparison of the GPT-like characteristics of AI inventions classified by each distinct technique. Note that the generality index is defined as share of citations to patents in different CPC classes at different aggregation levels. Citations within the same class are excluded. levels

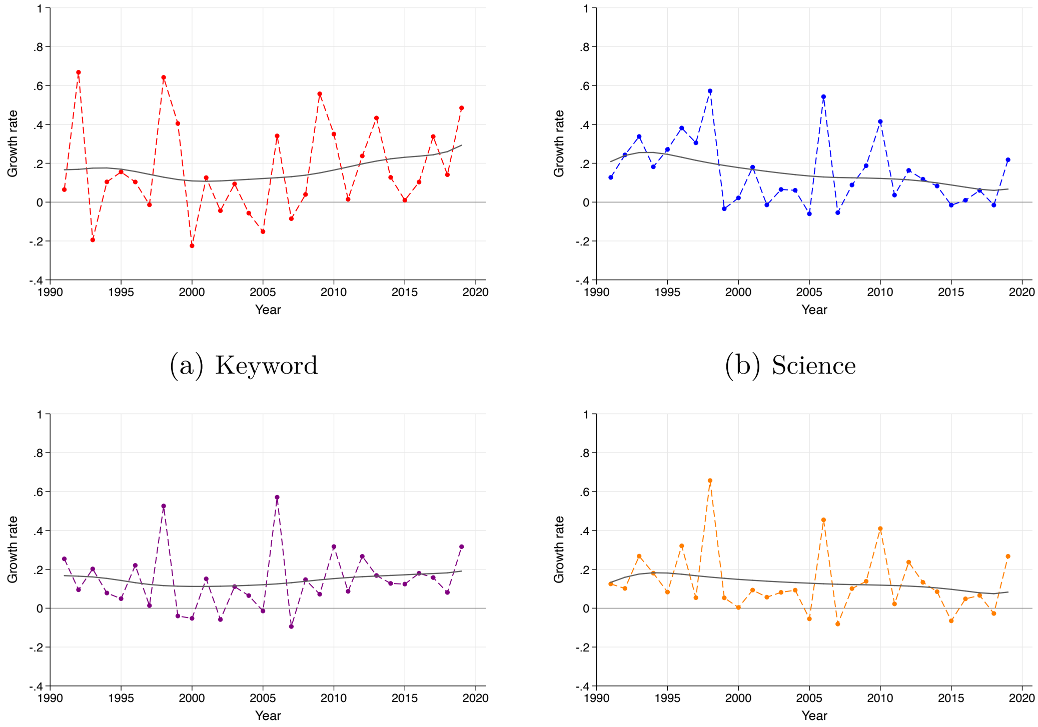
4.2.1. Growth

We study growth as the rate of change of patents between consecutive years. Figure 2 shows the growth rates, with 1 corresponding to 100%. We include lowess (local regression) smoother fits (Cleveland, 1979) in each plot to show the overall pattern.

We observe the following: first, each trajectory produces positive growth rates in most years, as anticipated for GPTs. Second, each trajectory demonstrates a dip in growth during the early 2000s before taking off again in recent years. However, a comparison with benchmark groups shows that patenting decreased across many sectors during this period (see Figure E.4 in Appendix E.1.1). Third, two of the smaller groups produced by the Keyword and WIPO approaches show an accelerating growth rate in the last few years, in contrast to the USPTO and Science pool. Taken together, we see positive growth rates and differences in time trends between the approaches.

In the Supplementary Information (Figures E.4), we study benchmarks that have been discussed in the literature as GPTs. Figure E.4 shows that the AI approaches exhibit high, or higher, growth rates than most other technologies that have been put forward as GPTs. Additionally, we make a series of significance test using a Wilcoxon signed-rank test for the significance of the observed differences in average growth rates (Appendix E.2.1). We find that the growth patterns of the AI samples cannot be statistically distinguished from our GPT benchmark candidates, although we observe the patterns of growth to be significantly higher than those of average patents.[3] Generally, we find that the differences between the average growth rates are not significant across different AI samples, except for the USPTO approach which shows the lowest growth.

Figure 2: Growth of Patents by Year



(a) Keyword

(b) Science

(c) WIPO

(d) USPTO

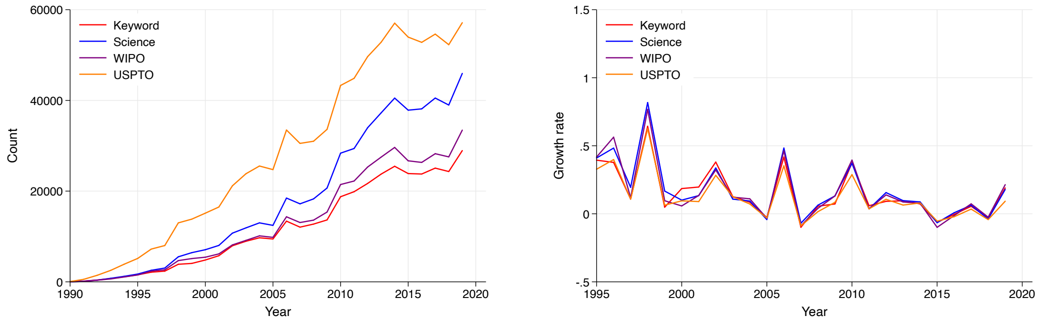
Notes: The four AI approaches have different growth patterns over time. The averages for all are positive, but the Keyword and WIPO approach both have increasing growth rates.

To investigate whether the different AI trajectories are associated with spillovers to follow-up inventions in non-AI sectors, we investigate patterns of growth in patent citations. Figure 3 shows that the patents citing AI (excluding those that are themselves AI by the respective approach) have similar profiles over time. The size ranking also corresponds with the counts and shares in previous plots, with growth rates that have, on average, been positive. This confirms that each approach generates a growing number of invention spillovers to non-AI sectors, which is characteristic for a GPT. The significance tests in E.9 indicate that the differences between the triple Keyword-Science-WIPO cannot be statistically distinguished, except for Keyword showing the slowest uptake in the 1990s, but to take off thereafter. These three approaches, however consistently score significantly higher than the USPTO group. Figure 3b suggests similarity between the

[3]Note that these tests rely on a very small number of observations and growth patterns of the four AI samples that fluctuate and differ across the three decades.

four time profiles and growth series, thus implying that the trajectories capture different portions of a larger group of similar technologies.

Figure 3: Patents Citing AI



(a) Counts

(b) Growth rate (1=100% growth)

Notes: Panel (a) shows actual number of AI citing patents. Panel (b) shows growth rates plotted from 1995 and onwards.

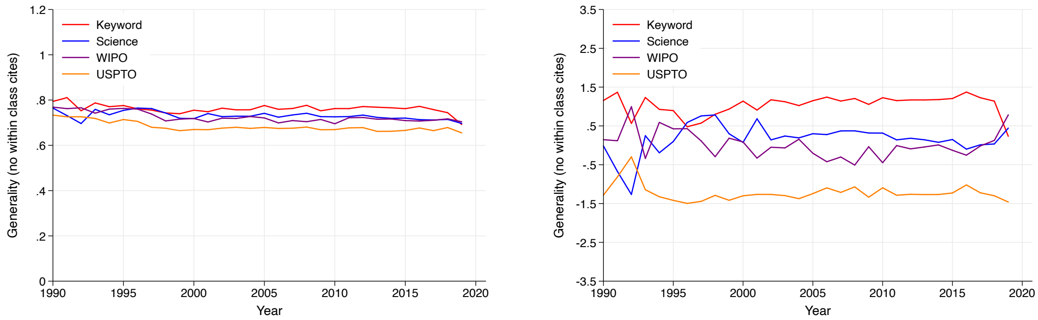
4.2.2. Generality

We use two indicators to evaluate the generality of AI inventions, i.e. the extent to which these inventions are used in diverse technology fields. First, we assess the generality index, which is defined in the Appendix A.2. Additional results for other aggregation levels are provided in Appendix E.3.2. The Keyword trajectory shows the highest level of generality across all levels of aggregation. Hence, citations from other technology fields to Keyword AI patents are most equally distributed across different technology fields. The WIPO and Science produce similar generality indices, with a slightly higher value for Science. The USPTO group produces the lowest generality index across all CPC levels.

Next, we present the second GPT measure of wide usefulness, based on the mean annual average number of unique citing classes citing to an AI patent. This accounts for the different number of annual patents and avoids an over-representation of the generality of patents in more recent years, as the number of patents is steeply increasing over time. Again, we compile the metric for different levels of aggregation. At all levels of aggregation, the Keyword patents are the most general (see Table 2). At the 1-digit level, Science scores second, closely followed by WIPO. WIPO scores higher at the 3- and 4-digit level. The USPTO patents demonstrate the lowest number of unique 1-, 3-, and 4-digit patent citations.

In Figure 4a and 4b we show the evolution of the generality index over time, with 4b showing the z-scaled version. The figures confirm that the Keyword patents are consistently the most general during the whole period, with the Science and WIPO approach producing similar, but lower, results. Again, the USPTO patents show the lowest generality. Towards the end of the time period, the decline in generality of the Keyword and USPTO methods needs to be considered with caution, as the absolute number of citations of recently granted patents is lower when the time window to be cited is small.

Figure 4: Generality Index at the 1-digit CPC-section level



(a) Generality index

(b) Z-score scaled generality

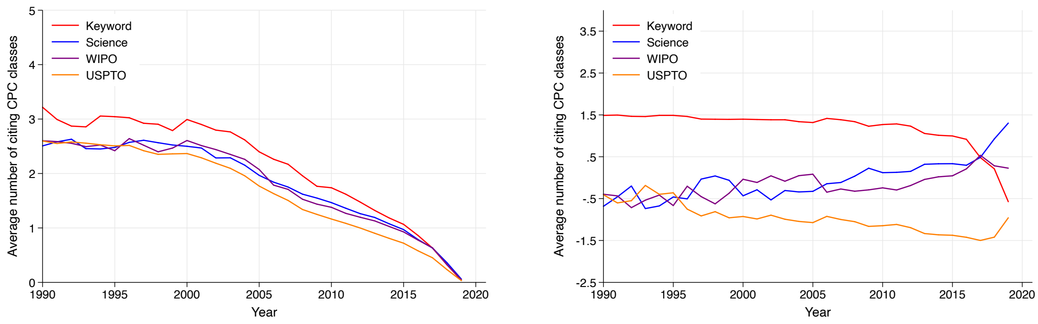
Notes: The z-scored value equals the level of the generality index minus its average across the four approaches divided by the standard deviation for each year.

We also calculate the generality index for all patents and our set of benchmark GPTs. Table E.2 shows that the generality index across all patents is higher than AI categories, which is a natural feature confirming the usefulness of the generality index to capture the widespread use of patents. Among other groups, biochemistry/genetic engineering, nanotechnology, and climate inventions related patents have high generality score. Figure E.5b shows that the time-series pattern of the generality score is quite stable, for some end-of-the-period fluctuations. For our second measure of generality, we find that AI patents show more generality than the entire patent universe at any level of CPC codes (Table E.3). Biochemistry/genetic engineering and climate-related patents show high patent-level generality. These numbers show that our generality measures work well, at least for some known GPT candidates.

In Appendix E.2.2 we show the results of a series of significance tests for the differences between the time series. These tests confirm that the highest score of the Keyword and the lowest of the USPTO method are statistically significant, while the differences between generality scores of the science and WIPO approaches are negligible at the 1-digit level. In Appendix E.3.2, we also report the number of unique citing technology classes for the set of AI patents that are cited at least once (see Table E.39). Interpreting backward citations as a proxy for the technological value of a patent (Kogan et al., 2017), this alternative measure focuses only on high-value patents. Again, the Keyword patents show the highest generality across all levels of aggregation. Science scores slightly higher than WIPO at the 1-digit level and vice versa at the 3- and 4-digit level. The USPTO approach shows the least generality.

We provide additional results on the generality of the 'descendants' of AI patents, i.e. those patents that cite AI patents, but are not themselves AI (Appendix E.4). These patents produce similar results, although the differences between the four AI trajectories are smaller.

Figure 5: Average Number of Classes Citing AI



(a) All patents

(b) All patents (z-score scaled)

Notes: The z-scored value equals the level of the generality index minus its average across the four approaches divided by the standard deviation for each year.

Figure 5 and the significance tests shown in Appendix E.2.2 confirm that the Keyword sample of AI patents shows a significantly larger number of different citing CPC classes compared to the other three methods. Notably, time series of citation data are sensitive to truncation towards the end of the period, as more recently granted patents have a shorter period of being cited. Looking at the z-score scaled data (Fig. 5b), we find that the science approach is the least sensitive to this truncation. Appendix E.4 provides additional results for the patents cited at least once.

Table 3: Average Citation Lags by Approach

Period	Keyword	Science	WIPO	USPTO
all periods	9.83	8.90	9.63	9.80
1990-1999	13.72	13.26	13.77	13.64
2000-2009	9.84	9.08	9.38	9.34
2010-2019	4.26	4.15	4.19	4.33

Notes: This table shows the average number of years taken until a patent in the sample is cited. The average number of years is lower during the more recent decade as the maximal time lag is truncated when our data ends in 2019.

An analysis of the average citation lags are shown in Table 3. The lags are given by the average number of years between the grant year of a patent and the grant year of the patents citing the patent. Hall and Trajtenberg (2006) argue that long citation lags are characteristic for GPTs, when their early 'learning and destruction' phase relies on the accretion of complementary inventions and structural and organisational changes throughout the economy (Crafts, 2021). Keyword patents show the longest citation lag, i.e. patented inventions captured by this approach take more time to be taken up in subsequent patents.

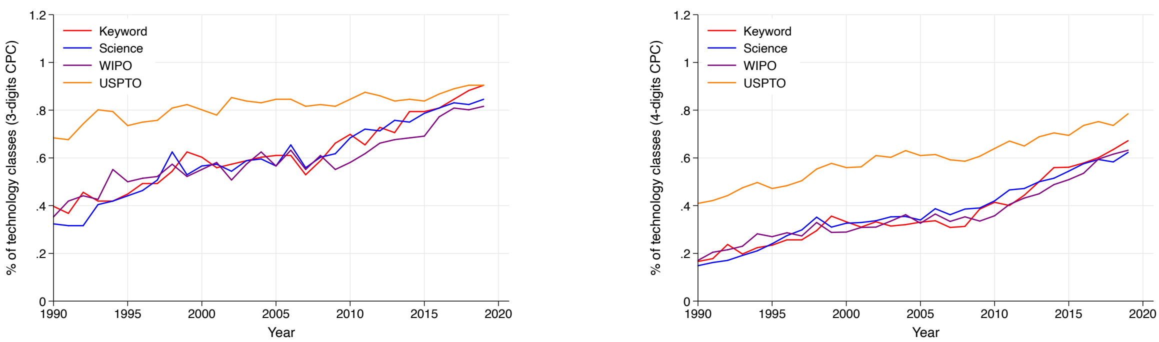
Altogether, the AI trajectory identified by the Keyword approach produces a set of AI patents with the highest level of generality across different metrics.

4.2.3. Complementarity

To understand how strongly AI inventions complement existing or novel products and processes, we examine the co-occurrence of the technology classes ascribed to AI patents. If AI can be combined with many other technologies, AI patents would be co-classified across diverse technological classes.

Figure 6 shows the percentage of 3- and 4-digit CPC classes associated with each pool of AI patents. In panel (a), USPTO patents span the most diverse pool of technology classes, with CPC classes ranging from 70-90% of all possible 3-digit codes. This result, however, could be caused by the large number of AI patents identified by the USPTO approach. The other three approaches identify a substantially smaller pool of AI patents (see Table 1), yet these patents cover a large share (80-90%) of 3-digit codes by 2019. Panel (b) in the same figure shows the share of 4-digit CPC classes. Starting around 2010, the share of technology classes embedded in Keyword, Science, and WIPO patents rapidly increased. Although USPTO patents represent the most diverse portfolio — at 78% of all technology classes in 2019 — the gap to the other three approaches has narrowed in recent years. Our significance tests show that the differences between the triple Keyword-Science-WIPO are statistically insignificant, but all score significantly lower than USPTO patents (see Table E.29 and E.31).

Figure 6: Diversity of Patent Pool – Share of Technology Classes



(a) % of all 3-digit CPC

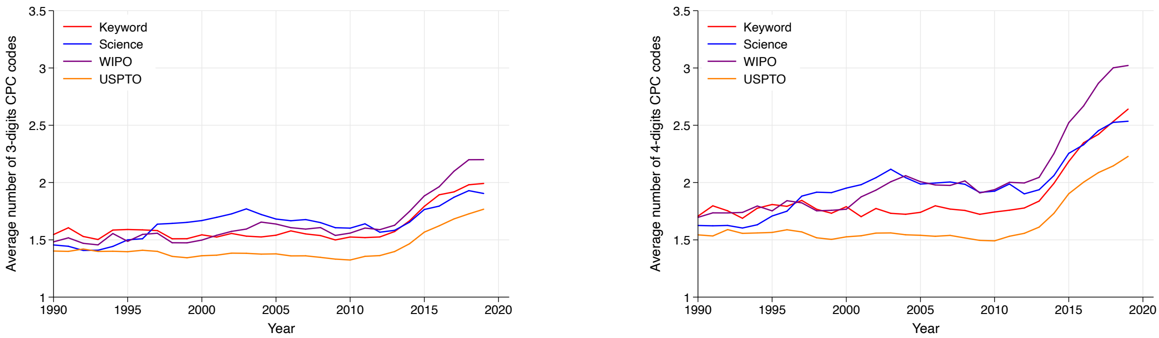
(b) % of all 4-digit CPC

Note: Panel (a) shows the percentage of 3-digit CPC codes and panel (b) shows the percentage of 4-digit CPC as a share of all codes in the respective category. Note that the total number of 3-digit and 4-digit CPC codes are 136 and 674, respectively according to the February 2022 version.

To understand whether AI patents became more multidisciplinary, we calculate the annual averages of the number of 1-, 3- and 4-digit CPC codes per patent. As Table E.42 in Appendix E.3.2 shows, an average WIPO patent is associated with 1.64 3-digit classes and 2.05 4-digit classes across all years. An average Keyword or Science patent is associated with slightly fewer classes, whereas the USPTO patents appear to be the least multidisciplinary by this metric. The highest score of the WIPO method at the 4-digit level and the weak diversity performance of the USPTO approach are statistically significant.

At the 1-digit level, an average Keyword and Science patent show both similarly high numbers of co-classifications (1.39-1.40) and cannot be statistically distinguished. The other approaches rank significantly lower, with the USPTO approach showing the lowest, despite its scale. In Figure 7, we check how the average number of technology classes per patent varies over time for different aggregation levels. All panels demonstrate a rising number of technology classes per patent towards the second half of the last decade, with the most profound increase for WIPO patents. The significance tests suggest that the Keyword and Science methods cannot be statistically distinguished at the 1- and 3-digit level (see Appendix E.3.3). At other aggregation levels, the differences diminish, except for the low scoring of the USPTO method and the highest score of the WIPO method at the 4-digit level.

Figure 7: Patent-Level Diversity - Average Technology Classes



(a) Average number of 3-digit CPC

(b) Average number of 4-digit CPC

In summation, although the USPTO patents dominate by scale, the other three approaches produce the dynamics of increasingly diverse patent pools over time. It is remarkable that the Keyword approach with only 67,186 AI patents cover almost 90% of all possible 3-digit CPC codes and 67% of all possible 4-digit CPC codes.

Examining the diversity of average patents, we find widespread recent increases in the diversity of technology classes per AI patent. In recent years, the WIPO approach seems to capture the most multidisciplinary patents, which could be due to the more explicit incorporation of computer hardware compared to other approaches (see Section B.4).

In Appendix E.1.1, we also use these measures for a set of benchmark GPT candidates to see how well they capture technological diversity. We find that our measures are performing equally well in capturing the increasing diversity of other GPTs suggested in the literature.

4.3. Market Power of AI-producing Firms

As we saw above, there could be a need to look beyond AI as a GPT, as new and pervasive technologies can have profound impacts on the distribution of market power. We now turn to the key inventors and examine concentration of inventive activities across the four technological trajectories.

Table 4: Top AI-Producing Firms

Keyword		Science		WIPO		USPTO	
Company name	share	Company name	share	Company name	share	Company name	share
IBM Corp	0.04	IBM Corp	0.07	IBM Corp	0.07	IBM Corp	0.07
Microsoft	0.03	Microsoft	0.06	Microsoft	0.04	Microsoft	0.04
Samsung	0.02	Google	0.03	Google	0.03	Google	0.03
Google	0.02	Apple	0.02	Canon	0.02	Intel	0.01
Amazon	0.01	Sony	0.01	Samsung	0.02	Samsung	0.02
Siemens	0.01	Siemens	0.01	Sony	0.02	Sony	0.02
Intel	0.01	Hewlett Packard	0.01	Intel	0.01	Amazon	0.01
Canon	0.01	Intel	0.01	Amazon	0.01	Canon	0.02
Sony	0.01	AT&T	0.01	Siemens	0.01	Siemens	0.01
AT&T	0.01	Canon	0.01	Fujitsu	0.01	Fujitsu	0.01

Notes: This table reports the top-ten AI producing firms within each trajectory. IBM, Microsoft, and Google are among the top-five AI patenting firms across all four groups. The column share reports the share of commercial patents accounted by a firm within the each trajectory. For example, IBM accounts for 4-7% of all AI patents produced by commercial firms.

Table 4 presents the top AI-patenting firms across all four trajectories. All approaches identify that AI patenting is dominated by technology and communication companies, with WIPO and USPTO even identifying perturbed lists of actors.

At first look, the top-ten AI producing firms suggests that the trajectories largely agree on the market concentration of the largest firms. However, because the table does not represent the full distribution, we calculate the Herfindahl–Hirschman Index (HHI) which in our context denotes the sum of squares of the share of AI patents produced by each firm.

Table 5 summarises the concentration of the top 10 patenting firms identified by each classification approach. Clearly, firms identified via the Keyword approach are the least concentrated and with the lowest HHI: a result consistent with findings in Section 4.2 that this approach best captures other characteristics of AI as a GPT.

Contrary to some of the concerns raised, we do not observe that AI inventions became clearly more concentrated in firms during the past three decades. We observe minor fluctuations, but cannot observe any clear trend for any of the four trajectories. These findings need to be combined with the insight that our data only reflects trends in the distribution of patented AI inventions across firms: Market concentration in AI can also refer to market shares and control over data, computational resources, and platforms.

Table 5: Concentration of Firms innovating in AI

	Keyword	Science	WIPO	USPTO
Concentration Ratio (CR)				
Four-firm CR	0.11	0.17	0.16	0.18
Eight-firm CR	0.16	0.22	0.22	0.23
Herfindahl–Hirschman Index (HHI)				
HHI (overall)	.006	.012	.011	.014
HHI (1990-1999)	.006	.012	.012	.014
HHI (2000-2009)	.005	.013	.015	.016
HHI (2010-2019)	.007	.013	.011	.014

Notes: This table shows the measures of concentration of AI-producing commercial firms. HHI is calculated as the sum of squares of shares of patent produced by each firm within each of the four trajectories. Concentration ratios measure the market share of top-four (or top-eight) firms.

5. Discussion

Our preceding analysis highlights how sensitive the estimation AI's scale and scope is to its classification approach. All four trajectories document the rise of AI inventions over the past three decades, as captured by patent counts and share of AI patents. All also document a slow-down in AI patenting from 1990 to the early 2000s (a so-called 'AI Winter') followed by a period of increased patenting since the 2010s (Stuart and Norvig, 2003; Klinger et al., 2022). To validate our measures, we evaluated the same three GPT characteristics for other technologies that were proposed as GPT candidates in the post-1990s and find similar trends to those seen in AI patents. All of them outperform average US patents, indicating their potential GPT nature. These supports the often made claim of AI having the technical characteristics of a GPT (Trajtenberg, 2018).

All four trajectories exhibit key GPT characteristics, but to different extents. Despite capturing the smallest number of patents, the Keyword sample shows the highest average growth rate in the post 2010s and most strikingly demonstrates AI's wide-ranging usefulness across several technology fields. AI patents identified by the USPTO approach most strongly show the feature of complementarity from 1990-2010, while the other three samples show more heterogeneous trends. This is not too surprising, as the USPTO applies a very broad understanding of what inventions could be considered as AI-related, leading to a wide spread of AI patents across technological classes. When accounting for the differences in patent counts, complementarity is most clearly demonstrated in the WIPO patents, closely followed by Science and Keywords. The difference in results when using an aggregate versus patent-level measure of complementarity suggest that narrow definitions may perform better to capture the core properties of an invention, while the wide AI conception, as encoded in the USPTO method, might be better suited to capture AI diffusion.

Another noteworthy result is that we do not find any empirical support for an increasing concentration in AI as often argued in the literature (Petit, 2017). This is reflected in stagnating concentration measures. Further, when using the Keyword method, which has a focus on very recent trends and captures best the GPT of AI, we find much lower levels of concentration than for any of the other trajectories. Of course, our results are limited to concentration in inventive activity, and we cannot draw definite conclusions about the concentration of power in the market and control over essential resources such as data, computing facilities, social networks, and platforms. While we confirm that the powerful position of few Big Tech companies is also reflected in our data, we cannot confirm that concentration became stronger over time or can be attributed to the GPT characteristics of AI that raise great expectations about its technological potential.

The results suggest that barriers to entry in AI invention did not recently become more difficult to climb. This may also be due to the nature of digital inventions, which are associated with a relatively low capital intensity. An alternative hypothesis is that the modularisation of the AI ecosystem facilitates the entry of novel players, as it allows computational resources and basic hardware to be externally sourced at affordable prices. Those interested in regulation and competition policy, as well as economists studying the impact of AI, should therefore be specific about which segments of the AI ecosystem are considered, and note the limitations of their approach (cf. Jacobides et al., 2021).

Our comparison of four AI trajectories have implications for policy and research, as they underscore how the definition used to identify a novel technology can impact the outcome of research in economics and innovation studies. For example, when investigating the role of public R&D funding on AI innovation outcomes, the Science trajectory would suggest a much higher importance of funding than the other approaches. While the USPTO and Science trajectory report a slow-down in AI inventions, this result would not be supported when applying the short-range view of the Keywords method or the technical perspective by the WIPO. This may lead to different conclusions and candidate explanations, for example, whether the decline or rise can be attributed to a narrowing and corporatisation of AI research (Klinger et al., 2022; Jee and Sohn, 2023b).

As an example, in 2021 the UK government launched a package of economic, policy, and regulatory reforms, dubbed the National Artificial Intelligence Strategy, to make the UK a 'global AI superpower' and plan for the long-term needs of AI technologies (Artificial Intelligence, 2021). Given that the Science and USPTO trajectories suggest that AI inventions are plateauing, while others (Keywords and WIPO) show that it is increasingly taking off, we recommend considering multiple techniques to counteract the dependence of policy conclusions on the choice of the method.

Our study indicates that the simple Keyword approach generates key GPT-features most strikingly. The other approaches are conceptually and computationally more complex: the USPTO approach relies on sophisticated machine learning (ML) methods, the Science approach leverages additional data sources, and the WIPO method relies on the symbolic combination of technology codes and keywords. The narrowing of AI research into deep learning, as highlighted above, may partially explain why a method based on human-selected and therefore topical keywords related to deep learning produces the characteristics so strikingly.

5.1. General Purpose Technology indicators for Technological Trajectories

As summarised in Table 6, the Keyword approach produces the patent pool with the highest growth rates, generality and second-highest level of complementarity. The WIPO and Science pools produce similar levels of each characteristics, but WIPO being stronger when it comes to complementarity. The USPTO patents least profoundly demonstrate each of the three GPT characteristics.

Table 6: Summary of Findings

	Keyword	USPTO	WIPO	Science	Metric	Based on
Growth					Growth rate	Counts
Generality					Generality index	Citations
Complementarity					Avg. # tech. classes	CPC codes

This brief summary of our results shows which patent group generates the strongest average estimate of each General Purpose Technology characteristic over the last 10 years. Red (yellow) colour indicates the strongest (weakest) performance.

Remarkably, the Keyword approach generates the most profound evidence of each of the three GPT characteristics using a much smaller (67k) set of patents than the other approaches (159-595k). The WIPO and science-based patent groups both demonstrate increasing generality, complementarities, and growth of AI invention over time, yet these methods rely on more complex classification and face limitations regarding patent coverage.

Despite relying on the most computationally-intensive ML approach and covering the largest group of patents, the USPTO patents are the least diverse, have the lowest growth rates, and exhibit the fewest complementarities.

Notably, each approach reproduces key time-trends in AI inventions, including a substantial uptick in the number of patents citing an increasingly diverse spread of AI patents after 2010. All methods show a high generality of average AI patents and decreasing number of distinct citing technological classes since 2000.

5.2. Concentration of AI inventions and political implications

Our comparison of top AI-producing firms across all trajectories clearly demonstrates that top technology firms dominate AI inventions. Whichever approach is used, it is clear that IBM, Microsoft, and Google among the top-five AI patenting firms. IBM accounts for 7% of the patents in Science, WIPO, and USPTO trajectories, yet only 4% in Keyword patents, which attributes a higher share to Samsung. Keyword patents are the least concentrated, with the top-four firm accounting for 11% of patents, compared to 16-18% in other approaches. The Keyword trajectory also appears to be least concentrated in terms of HHI. These results, along with our measure of higher generality index for Keywords, suggest that the Keyword trajectory represents variety of patents represented by more variation among the inventing agents. On the other hand, the other approaches were shown to capture highly concentrated AI invention activities, thereby identifying less of the general purpose AI and more special purpose patents.

It should also be noted that while the levels of industrial concentration are consistent, this did not clearly increase during the past decades. Thus, arguments about an increasing concentration of AI development can not be directly based on shares of patented AI inventions. Clearly, these could be based on other qualitative aspects, such as an increasing utility of AI inventions.

Taken together, our results demonstrate that introducing an agency dimension consistently reproduces a high concentration among a few powerful actors in AI research and development. This can be viewed as consistent support for arguments that have argued the possible need to consider regulating AI and look beyond the question of GPT (Jacobides et al., 2021).

5.3. Results in a Wider Context

Our results highlight that the extent to which AI constitutes a GPT, as well as the scale and scope of its present and future impact, is significantly sensitive to the classification technique used. Moreover, our results call into question the literature that takes as given that AI is a GPT that is projected to expand in the ensuing years (Cockburn et al., 2019; Goldfarb et al., 2019; Brynjolfsson et al., 2021).

Relying on the USPTO classification method, it can be said that average patents now show the lowest generality index and there has been no growth in the number of 3-digit CPC codes assigned an average AI patent since 2000. Moreover, by this method, the growth of AI patenting has plateaued after 2010 —a result that may generate questions over new, large-scale initiatives designed to bolster AI research.

Recent breakthroughs in the technology, data, and resources underpinning AI and its applications have spurred an international race between countries to become the global AI leader through hefty investments and significant policy decisions. In this context, the UK has recently launched the National Artificial Intelligence Strategy, to bolster the 'long-term needs of AI technologies' to ready the economy for the 'Age of AI' (Artificial Intelligence, 2021). Yet by the USPTO's own classification methods, it can be said that AI inventions as reflected in patents began to slow back in 2010.

From a research efficiency perspective, our results raise the question whether human reasoning used to define relevant keywords outperforms more computationally complex classification methods. The small, simple and tractable Keyword approach may be desirable for researchers aiming to catch emerging GPTs. Small patent groups may indicate a clearer distinction from other patents and, because such technologies are 'new', there is greater potential for future growth, compared to the USPTO approach which suggests that 16.6% of all US patents today are based on AI. This being said, other classification approaches may be better suited to other questions surrounding the knowledge base of inventions (Science) or the trajectory of inventions through the universe of technologies (USPTO).

During our analysis, we also discovered that the Keyword method reproduced from Cockburn et al. (2018) may be further simplified. We found that the majority of patents can be identified using a narrower set of four terms (machine learning, neural network, robot, pattern recognition), rather than the original list of over forty words.

Finally, the phenomena of 'AI narrowing', as discussed in Klinger et al. (2022), suggests that AI innovation has become significantly less diverse during the time period of our analysis. The authors suggest that the concentration of AI research within a small group of private firms is leading to the premature narrowing of its GPT potential. We argue that the intensive focus on deep learning techniques since 2010, combined with path dependency, has reduced incentives for such leaders to explore alternative methods. We argue that this 'locking in' of AI to a subset of deep learning techniques is driving under-investment in other applications, including ethical and environmentally conscious AI. Our investigation of arXiv AI research reveals a sharp turn towards computer vision and language AI dominated by deep neural networks as well as a stagnation of various metrics of AI research diversity.

The Keyword and WIPO approaches both include terms closely related to machine and deep learning, and may be the most sensitive to this premature narrowing. However, this is not visible in our results, which instead show stabilised generality and increasingly diverse co-classifications since 2010. Another possible explanation is that a certain hype around keywords unfolds with a time lag, and that incentives support using these keywords. On the other hand, this is made more difficult by the fact that approaches using keywords match key sections in patents that are carefully reviewed for inventiveness. A recent focus on certain techniques may simply represent a technological trend, whereby existing keywords may soon be supplanted by new inventions.

5.4. Limitations of Our Approach

Our evaluation is subject to two major caveats. First, the performance of the different approaches to capture the GPT-like characteristics of AI may be dependent on the time period chosen. Given that our aim is to inform researchers who are interested in the impact of AI, we have chosen to investigate the period from 1990-2019, when AI began to grow and disseminate more widely (Cockburn et al., 2019). Because that the performance of the keyword approach may reflect the popularity of a narrow set of technological buzzwords, our results may differ in other time periods.

Second, we emphasise that the evaluation of different classification methods depends on the intended purpose. Here, we aim to provide guidance for researchers interested in studying the economic impact of AI inventions that they take to be a GPT. Other methods may be more appropriate to study the dissemination of AI-related applications or the knowledge origins of AI. For example, the USPTO approach tells an interesting story, suggesting that 14% of all patented inventions in the US today already rely on AI; the Science approach in these contexts would provide interesting extrapolation to the early origins of AI research in the 1950s.

Ultimately, our results suggest that the extent to which AI can be seen as a GPT, as well as its future scale and scope, is sensitive to the chosen classification method. This underscores the importance of balancing multiple classification techniques when considering how political and economic measures could affect AI's projected future impact. Furthermore, we suggest that those wishing to show that AI is indeed a strong GPT projected to accelerate into the future could do so by employing a keyword-based classification approach.

6. Concluding Remarks

In this paper, we have performed a systematic analysis of four separate approaches to identifying patented AI inventions. The evidence suggests that these are distinct and only partially overlapping groups of patents that can represent different potential technological trajectories associated with AI. Ultimately, we demonstrate how sensitive each key GPT-like feature of AI is to the classification method chosen, and investigate levels of market concentration.

First, our results provide overall guidance to policy-makers and innovation scholars on how to identify patented AI inventions. For researchers interested in studying the GPT-like characteristics of AI, a simple keyword-based approach may be more successful at producing a narrowly defined set of patented technologies that demonstrates the canonical GPT features of intrinsic growth, wide usefulness, and complementarity. Different trajectories are associated with different levels of GPT-like characteristics.

We find consistent support for the need go beyond thinking of AI as a GPT and study (Jacobides et al., 2021), e.g., by considering the agency dimension related to AI. Among the strongest results that are most consistent across the different technological trajectories, we have found similar levels of market concentration. That a few firms strongly lead AI inventions is something which appears valid across methods to find AI inventions in data, which could be relevant for competition policy and market regulation (Petit, 2017; Hennemann, 2020).

Our work therefore provides both robust findings as well as important caveats to those making policy and funding decisions for the future pace of AI inventions, such as the recently released UK National Artificial Intelligence Strategy (Artificial Intelligence, 2021). Taking an approach of AI as a General Purpose Technology is reasonable to do for different technological trajectories, but there is clearly a need to observe that an agency dimension suggests strong concentration of market power. This illustrates the usefulness to further track AI inventions using patent data. To acknowledge other respects where the findings differ, we also underscore the need for multiple AI classification techniques in order to counteract the dependence of significant policy conclusions on the choice of classification approach.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo (2020). *AI and jobs: Evidence from online vacancies*. Working Paper. National Bureau of Economic Research. DOI: 10.3386/w28257.

Aghion, Philippe, Benjamin F Jones, and Charles I Jones (2018). "Artificial intelligence and economic growth". In: *The economics of artificial intelligence: An agenda*. University of Chicago Press, pp. 237–282. DOI: 10.7208/chicago/9780226613475.001.0001.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2019). "Economic policy for artificial intelligence". In: *Innovation Policy and the Economy* 19.1, pp. 139–159. DOI: 10.1086/699935.

Agrawal, Ajay, John McHale, and Alexander Oettl (2018). "Finding needles in haystacks: Artificial intelligence and recombinant growth". In: *The economics of artificial intelligence: An agenda*. University of Chicago Press, pp. 149–174. DOI: 10.7208/chicago/9780226613475.003.0005.

Ahmadpoor, Mohammad and Benjamin F Jones (2017). "The dual frontier: Patented inventions and prior scientific advance". In: *Science* 357.6351, pp. 583–587. DOI: 10.1126/science.aam9527.

Alderucci, Dean, Lee Branstetter, Eduard Hovy, Andrew Runge, and Nikolas Zolas (2020). "Quantifying the impact of AI on productivity and labor demand: Evidence from US census microdata". In: *Allied Social Science Associations - ASSA 2020 Annual Meeting*.

Arthur, W Brian (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.

Artificial Intelligence, Office for (2021). *National AI Strategy*. Tech. rep. UK Government. URL: https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version.

Babina, Tania, Alex X He, Anastassia Fedyk, and James Hodson (2023). "Artificial Intelligence, Firm Growth, and Product Innovation". In: *Journal of Financial Economics (Forthcoming)*.

Bianchini, Stefano, Moritz Müller, and Pierre Pelletier (2020). "Deep learning in science". In: *arXiv preprint arXiv:2009.01575*. DOI: 10.48550/arXiv.2009.01575.

Bonaccorsi, Pietro and Massimo Moggi (2020). *The long wave of the Internet*. Tech. rep. LEM Working Paper Series.

Bresnahan, Timothy F and Manuel Trajtenberg (1995). "General purpose technologies 'Engines of growth'?" In: *Journal of econometrics* 65.1, pp. 83–108. DOI: 10.1016/0304-4076(94)01598-T.

Brynjolfsson, Erik and Tom Mitchell (2017). "What can machine learning do? Workforce implications". In: *Science* 358.6370, pp. 1530–1534. DOI: 10.1126/science.aap8062.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson (2019). "Artificial intelligence and the modern productivity paradox". In: *The economics of artificial intelligence: An agenda* 23. DOI: 10.3386/w24001.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson (2021). "The productivity J-curve: How intangibles complement general purpose technologies". In: *American Economic Journal: Macroeconomics* 13.1, pp. 333–72. DOI: 10.1257/mac.20180386.

Cleveland, William S (1979). "Robust locally weighted regression and smoothing scatterplots". In: *Journal of the American statistical association* 74.368, pp. 829–836.

Cockburn, Iain M, Rebecca Henderson, and Scott Stern (2018). *The impact of artificial intelligence on innovation*. Working Paper. National bureau of economic research. DOI: 10.3386/w24449.

– (2019). "The Impact of Artificial Intelligence on Innovation: an Exploratory Analysis". In: *The Economics of Artificial Intelligence*. University of Chicago Press, pp. 115–148. DOI: 10.7208/9780226613475-006.

Comin, Diego and Marti Mestieri (2018). "If technology has arrived everywhere, why has income diverged?" In: *American Economic Journal: Macroeconomics* 10.3, pp. 137–78. DOI: 10.1257/mac.20150175.

Crafts, Nicholas (2021). "Artificial intelligence as a general-purpose technology: an historical perspective". In: *Oxford Review of Economic Policy* 37.3, pp. 521–536. DOI: 10.1093/oxrep/grab012.

Dosi, Giovanni (1988). "Sources, procedures, and microeconomic effects of innovation". In: *Journal of economic literature*, pp. 1120–1171.

EPO (2021). *PATSTAT online (Version 2.7.2), Patent information for experts [Data set]*. Global. European Patent Office. URL: https://data.epo.org/expert-services/index.html.

European Commission (2021). *Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*.

Fanti, Lucrezia, Dario Guarascio, and Massimo Moggi (2022). "From Heron of Alexandria to Amazon's Alexa: a stylized history of AI and its impact on business models, organization and work". In: *Journal of Industrial and Business Economics* 49.3, pp. 409–440.

Frey, Carl Benedikt and Michael A Osborne (2017). "The future of employment: How susceptible are jobs to computerisation?" In: *Technological forecasting and social change* 114, pp. 254–280. DOI: 10.1016/j.techfore.2016.08.019.

Giczy, Alexander V, Nicholas A Pairolero, and Andrew Toole (2021). "Identifying artificial intelligence (AI) invention: A novel AI patent dataset. USPTO Economic Working Paper No. 2021-2". In: *The Journal of Technology Transfer*. DOI: 10.2139/ssrn.3866793.

Goldfarb, Avi, Bledi Taska, Florenta Teodoridis, et al. (2019). "Could machine learning be a general-purpose technology? Evidence from online job postings". In: *SSRN*. DOI: 10.2139/ssrn.3468822.

Gries, Thomas and Wim Naude (2018). *Artificial intelligence, jobs, inequality and productivity: Does aggregate demand matter?* MERIT Working Papers 2018-047. United Nations University - Maastricht Economic, Social Research Institute on Innovation, and Technology (MERIT). URL: https://ideas.repec.org/p/unm/unumer/2018047.html.

Hall, Bronwyn H and Manuel Trajtenberg (2006). "Chapter 14: Uncovering General Purpose Technologies with Patent Data". In: *New Frontiers in the Economics of Innovation and New Technology Essays in Honour of Paul A. David*. Ed. by Cristiano Antonelli, Dominique Foray, Bronwyn H. Hall, and W. Edward Steinmueller. Cheltenham, UK: Edward Elgar Publishing. DOI: 10.4337/9781845427924.00022.

Hennemann, Moritz (2020). "Artificial Intelligence and Competition Law". In: *Regulating Artificial Intelligence*, pp. 361–388.

Hötte, Kerstin, Anton Pichler, and François Lafond (2021). "The rise of science in low-carbon energy technologies". In: *Renewable and Sustainable Energy Reviews* 139, p. 110654. DOI: 10.1016/j.rser.2020.110654.

Iori, Martina, Arianna Martinelli, and Andrea Mina (2021). *The direction of technical change in AI and the trajectory effects of government funding*. Tech. rep. LEM Working Paper Series.

Jacobides, Michael G, Stefano Brusoni, and Francois Candelon (2021). "The evolutionary dynamics of the artificial intelligence ecosystem". In: *Strategy Science* 6.4, pp. 412–435.

Jaffe, Adam B and Gaetan De Rassenfosse (2019). "Patent citation data in social science research: Overview and best practices". In: *Research handbook on the economics of intellectual property law*. DOI: 10.4337/9781789903997.00043.

Jee, Su Jung and So Young Sohn (2023a). "A firm's creation of proprietary knowledge linked to the knowledge spilled over from its research publications: the case of artificial intelligence". In: *Industrial and Corporate Change*, dtad002. DOI: 10.1093/icc/dtad002.

– (2023b). "Firms' influence on the evolution of published knowledge when a science-related technology emerges: the case of artificial intelligence". In: *Journal of Evolutionary Economics* 33.1, pp. 209–247. DOI: 10.1007/s00191-022-00804-4.

Jovanovic, Boyan and Peter L Rousseau (2005). *General purpose technologies*. Working Paper. National Bureau of Economic Research. DOI: 10.3386/w11093.

Jurowetzki, Roman, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos (2021). "The Privatization of AI Research (-ers): Causes and Potential Consequences–From university-industry interaction to public research brain-drain?" In: *arXiv preprint*. DOI: 10.48550/ARXIV.2102.01648.

Klinger, Joel, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos (2022). "A narrowing of AI research?" In: *CoRR* abs/2009.10385. arXiv: 2009.10385. URL: https://arxiv.org/abs/2009.10385.

Klinger, Joel, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos (2018). "Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology". In: *Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018)*.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017). "Technological innovation, resource allocation, and growth". In: *The Quarterly Journal of Economics* 132.2, pp. 665–712. DOI: 10.1093/qje/qjw040.

Krafft, PM, Meg Young, Michael Katell, Karen Huang, and Ghislain Bugingo (2020). "Defining AI in policy versus practice". In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 72–78. DOI: 10.1145/3375627.3375835.

Lipsey, Richard G, Kenneth I Carlaw, and Clifford T Bekar (2005). *Economic transformations: general purpose technologies and long-term economic growth*. Oxford University Press, Oxford. URL: https://EconPapers.repec.org/RePEc:oxp:obooks:9780199290895.

Marx, Matt and Aaron Fuegi (2019). *Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles*. Boston University Questrom School of Business Research Paper No. 3331686. November 15, 2019. DOI: 10.2139/ssrn.3331686.

– (2020). "Reliance on science: Worldwide front-page patent citations to scientific articles". In: *Strategic Management Journal* 41.9, pp. 1572–1594. DOI: 10.1002/smj.3145.

– (2021). *Reliance on Science in Patenting (Version v30) [Data set]*. Zenodo. DOI: 10.5281/zenodo.4850763.

Meyer, Martin (2000). "Does science push technology? Patents citing scientific literature". In: *Research Policy* 29.3, pp. 409–434. DOI: 10.1016/S0048-7333(99)00040-2.

Mitchell, Tom Michael et al. (2007). *Machine learning*. Vol. 1. McGraw-hill New York.

Moser, Petra and Tom Nicholas (2004). "Was electricity a general purpose technology? Evidence from historical patent citations". In: *American Economic Review* 94.2, pp. 388–394. DOI: 10.1257/0002828041301407.

Narin, Francis, Kimberly S Hamilton, and Dominic Olivastro (1995). "Linkage between agency-supported research and patented industrial technology". In: *Research Evaluation* 5.3, pp. 183–187. DOI: 10.1093/rev/5.3.183.

Petit, Nicolas (2017). "Antitrust and artificial intelligence: a research agenda". In: *Journal of European Competition Law & Practice* 8.6, pp. 361–362. DOI: 10.1093/jeclap/lpx033.

Petralia, Sergio (2020). "Mapping general purpose technologies with patent data". In: *Research Policy* 49.7, p. 104013. DOI: 10.1016/j.respol.2020.104013.

Schumpeter, Joseph A (2005). "Development". In: *Journal of Economic Literature* 43 (1). post-hum published and introduced by Knudsen, Thorbjørn, pp. 108–120. DOI: 10.1257/0022051053737825.

Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang (2015). *An Overview of Microsoft Academic Service (MAS) and Applications*. In Proceedings of the 24th International Conference on World Wide Web (WWW 15 Companion). ACM, New York, NY, USA, 243-246. DOI: 10.1145/2740908.2742839.

Stuart, Russell and Peter Norvig (2003). *Artificial intelligence: a modern approach*. Upper Saddle River, NJ, USA: Prentice Hall.

Tijssen, Robert JW (2001). "Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows". In: *Research Policy* 30.1, pp. 35–54. DOI: 10.1016/S0048-7333(99)00080-3.

Trajtenberg, Manuel (2018). *AI as the next GPT: a Political-Economy Perspective*. Tech. rep. National Bureau of Economic Research. DOI: 10.3386/w24245.

Trajtenberg, Manuel, Rebecca Henderson, and Adam Jaffe (1997). "University versus corporate patents: A window on the basicness of invention". In: *Economics of Innovation and New Technology* 5.1, pp. 19–50. DOI: 10.1080/10438599700000006.

Turing, Alan (1950). "Computing machinery and intelligence". In: *Mind* 59.236, p. 433.

Valdes, YM and Ilja J Rudyk (2017). *Patents and the fourth industrial revolution*. Tech. rep. European Patent Office, Handelsblatt Research Institute. URL: https://www.epo.org/service-support/publications.html?pubid=163#tab3.

Verendel, Vilhelm (2023). "Tracking artificial intelligence in climate inventions with patent data". In: *Nature Climate Change*, pp. 1–8.

Webb, Michael (2019). *The impact of artificial intelligence on the labor market*. Available at SSRN 3482150. Stanford University. DOI: 10.2139/ssrn.3482150.

Whittaker, Meredith (2021). "The steep cost of capture". In: *Interactions* 28.6, pp. 50–55. DOI: 10.1145/3488666.

WIPO (2019a). "Technology Trends 2019: Artificial Intelligence: Data collection method and clustering scheme: Background paper". In: *Geneva: World Intellectual Property Organization*. URL: https://www.wipo.int/export/sites/www/tech_trends/en/artificial_intelligence/docs/techtrends_ai_methodology.pdf.

– (2019b). "WIPO Technology Trends 2019: Artificial Intelligence". In: URL: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

A. Measuring GPTs

Currently, there exists a number of alternative metrics to capture GPT characteristics. Given the lack of consensus, many believe GPTs should be better identified as sophisticated networks of technologies sharing 'underlying principles and mutual dependencies' (Petralia, 2020).

Historically, patent growth rates have been used to capture the endogenous elaboration of technologies similar to GPTs (Moser and Nicholas, 2004; Jovanovic and Rousseau, 2005; Petralia, 2020). Petralia (2020) uses patent growth rates, co-classifications, and a text-mining algorithm to successfully reproduce the canonical GPTs contained within the broad USPTO categories of electricity and computer communication. However, the author finds great heterogeneity within these pools of patents, which contain both dynamic and stagnant inventions. Moreover, the author notes that the identification of more diffuse and diverse GPTs, such as AI, may require 'bottom-up' classification approaches using lower levels of aggregation that can scan multiple technological classes for common principles.

Hall and Trajtenberg (2006) attempt to capture GPTs by measuring the patent growth rates and unbiased generality measures for the most-cited US patents and the patents which cite them. The authors also find great heterogeneity between patents, which underscores the need for multiple metrics to satisfactorily capture GPTs.

In the next section, we motivate our selection of patent measures for GPT characteristics and connect each with empirical facts about AI's dissemination and the three canonical GPT features.

A.1. Growth

For more than a decade, AI methods have become more powerful and complex as a result of new technical methods, increased data availability, and improved hardware. Consequently, AI invention has shifted away from specific application-based methods to more generalised learning-orientated systems (Cockburn et al., 2019). With this refinement, the performance of many sub-fields of AI, such as image and text recognition, have seen remarkable improvements in performance (Brynjolfsson et al., 2021). This is reflected in the exponential growth of patenting activity referencing terms such as machine learning and deep learning (see Supplementary Information, Figure D.2).

Based on these observations, we measure improvements in AI via the growth rates of each group of patents and changes to their share of all patents, from 1990 to 2020 (Hall and Trajtenberg, 2006; Petralia, 2020). We also look at the growth of the patents that cite such technologies: the 'GPT hypothesis' in previous work has been that inventions that build on GPT-like technologies should spawn more new inventions (Hall and Trajtenberg, 2006).

Let $N_{i,t}$ denote the number of patents in a group $i \in \{\text{keyword}, \text{science}, \text{WIPO}, \text{USPTO}\}$ at time t, indexed by year. We compute the growth rate as

$$\frac{N_{i,t} - N_{i,t-1}}{N_{i,t-1}}. \tag{A.0}$$

A.2. Generality

AI has already begun to pervade a myriad of industries as it expands beyond computer science into such diverse fields as structural biology, transport, and imaging (Cockburn et al., 2019). In the early 1990s, AI methods remained largely confined to computer science. However, over the past decade, the majority of patents referencing these technologies have appeared in secondary domains (Cockburn et al., 2019). Based on the work of Trajtenberg et al. (1997), we capture this stylised fact through the 'generality' of patents, measuring the dissemination of AI across different technology fields.

To do so, we build on patent citation data and assume that a forward citation link entails information about the use of a patent in a subsequent invention (Jaffe and De Rassenfosse, 2019). To operationalise wide usefulness, we rely on a modified version of the generality metric by Trajtenberg et al. (1997) and Hall and Trajtenberg (2006) given by

$$1 - \sum_{j} \left(\frac{\#cites_{ij,t}}{\sum_{j=1}^{N_j} \#cites_{ij,t}} \right)^2 \tag{A.0}$$

where $\#cites_{ij}$ is the sum of citations to patents labelled as AI by classification approach i from technology class j, whereby we use the CPC 1-digit level as class. The number of citations $\#cites_{ij}$ excludes citations within the same class: N_i is the number of patents in approach i and N_j is the number of different CPC classes. Our approach differs to that of Trajtenberg et al. (1997) as we apply the method to each group of AI patents belonging to a variety of CPC sections. For the main analysis, we focus on 1-digit CPC sections, as these are more technologically distant than 3-digit or 4-digit classes and subclasses, whose results we also report.

Our generality measure is calculated for the entire group of patents in i with N_i unique patents. To address concerns that this metric may be affected by differences between group sizes, we additionally calculate patent-level metrics given by the average number of citing classes, i.e.

$$\frac{1}{N_{i,t}} \sum_{p=1}^{N_{i,t}} \sum_{j=1}^{N_d} \mathbb{1}(\#ncites_{p,j,t} \geq 0) \tag{A.0}$$

where $\mathbb{1}(\#ncites_{p,j,t} \geq 0) = 1$ if patent p in i is cited by at least one patent in technology class j out of the total number of classes N_d at level with $d \subset \{1, 3, 4\}$ s in the code. Again, we exclude within-class citations and present results at both the 1-digit CPC section level ($d = 1$) and higher orders of disaggregation ($d = 3$ or $d = 4$).

A.3. Complementarity

Thirdly, GPTs augment existing products and processes in a range of novel contexts to generate productive complementarities throughout the economy (Bresnahan and Trajtenberg, 1995; Petralia, 2020). AI technologies have been shown to complement and rely on secondary inventions, related to areas such as cloud computing and big data, which increase access to larger and more affordable data-sets (Brynjolfsson et al., 2019). Furthermore, because diverse AI systems share similar underlying structures and can share information, advances in one application of ML, such as machine vision, can spur inventions in other fields, such as autonomous vehicles.

Following the approach of Petralia (2020), we measure the extent to which AI patents enhance and supplement other inventions through the diversity of their technology class co-occurrences. For our analysis, we calculate the share of 3- and 4-digit CPC codes ($d = 3, 4$) assigned to the patents in each group of AI patents. Specifically, we calculate the following *diversity measure* over time is

$$\frac{\#CPCs_{i,d,t}}{N_d} \tag{A.0}$$

where i denotes each of the four patent classification approaches, d is the classification level and t is year. N_d refers to the number of CPC codes found in use for a particular group of patents, where the codes include d digits. Note that there are 136 and 674 CPC codes, respectively at the 3- and 4-digit level (according to the February 2022 version of CPC codes).

As the above measure could be biased by patent volume, we also calculate the average number of distinct 1-, 3-, and 4-digit CPC codes per patent per year. The *diversity per patent* over time is

$$\frac{1}{N_{i,t}} \sum_p \#CPCs_{p,i,d,t} \tag{A.0}$$

where d represent the technology class represented by 1-, 3-, or 4-digit CPC codes. The time series graphs for the latter measures depict how an average patent's complementarity across technology sections evolves over time.

B. Measuring AI

In our analysis, we compare four methodologically and conceptually distinct approaches to identifying AI inventions in patents based on (1) keyword search, (2) science citations, (3) the WIPO, and the (4) USPTO method. Here, we introduce these classification approaches in detail.

B.1. Data Source

We apply our methods to all patents granted by the USPTO from 1990-2019. For the analysis, we create four groups of AI patents for each classification method and complement each with supplementary information.

From PATSTAT (Spring 2021 edition, EPO (2021)) we sourced patent grant dates and from the USPTO we downloaded the Master classification file (April 2021 version) which contains CPC classifications of patents.[4] We added further data on patent-to-patent citations and patent titles from GooglePats obtained in an earlier project (Hötte et al., 2021). For our analysis, we supplemented the citation data with citation year and the technology classes of both the citing and cited patent. In doing so, we obtained networks which represent citations from technology fields at different levels of aggregation to our four sets of AI patents. We also made use of the Reliance on Science database (Marx and Fuegi, 2020) for citation data between patents and science.

B.2. Keyword search

Our first classification technique is a straight-forward approach based on keyword search, in which researchers use their discretion to develop a set of terms that reflect the most recent developments in AI. In this paper, we use the set of keywords provided in the appendix of Cockburn et al., 2018.[5] The keywords used in this paper focus on three sub-fields of AI: symbolic systems, learning algorithms, and robotics (see Table C.1 for the full list of keywords). According to the authors, the symbolic systems represent 'complex concepts through logical manipulation of symbolic representations' and include 'natural language processing' and 'pattern recognition'. Learning algorithms include core analytic techniques such as neural networks, deep learning, and machine learning. The last category, robotics, is related to automation or applications of AI (e.g. computer vision and sensory networks).

We search for these keywords in patent titles, abstracts, claims, and descriptions using USPTO data. We match the resulting list with patents granted by the USPTO between 1990 and 2019. The main advantage of the keyword approach is its simplicity and ease of implementation. Moreover, carefully chosen keywords can capture recent changes in the AI field. However, the success of this approach depends on the judgement and familiarity of the researcher to the field of AI. Missing important keywords could lead to under-representation of a sub-field. Our approach yields 67,187 patents.

[4] `https://bulkdata.uspto.gov/data/patent/classification/cpc/`

[5] While we use the keywords from Cockburn et al., 2018, we do not fully replicate their approach. They use two subsets of patents: (1) patents classified by the USPC code 706 (Artificial Intelligence) and 901 (Robots); and (2) patents identified by searching titles for the selected keywords. Here we use patents identified by keyword search only, but we extend our search to match keywords also from abstract, claims, and description. We do not use the USPC classification codes since the WIPO method takes a more comprehensive approach combining keywords with IPC or CPC classifications. Also, with our extensive keyword search, we miss only a few patents which are in the first group (i.e., USPC 706 and 901) but not in the second group of Cockburn et al., 2018.

B.3. Science Citations

This classification approach harnesses the scientific basis of patents. In particular, we classify a patent as an AI patent if it makes at least one citation to a scientific paper in the scientific field of 'Computer Science; Artificial Intelligence' (short, AI paper) as categorised by the Web of Science (WoS). Scientific citations are added to patent documents for multiple reasons such as describing the technological content of the invention or distinguishing the legal claim from other publicly available knowledge (see Narin et al., 1995; Meyer, 2000; Tijssen, 2001; Ahmadpoor and Jones, 2017; Marx and Fuegi, 2019). A citation link to an AI paper indicates that the patent is technologically related to AI because it builds on scientific advancements in this field. A limitation of this approach is that it only identifies AI patents within the subset of patents that make citations to science.

For this method, we use data from the Reliance on Science (RoS) database (Marx and Fuegi, 2019; Marx and Fuegi, 2020; Marx and Fuegi, 2021) which comprises a mapping from patents to scientific articles indexed in Microsoft Academic Graph (MAG) (Sinha et al., 2015). Scientific articles are tagged by the WoS fields indicating the field of science into which an article is grouped.[6]

The citation links in the RoS database cover citations made by both the patent applicant and examiner, as well as citations indicated at both the front page and body of the patent document. Marx and Fuegi (2019) identified citations through a sequential probabilistic text recognition technique. Each citation link is tagged with a confidence score indicating the reliability of the matching approach. In the RoS data, roughly one third (34%) of all US patents granted in 2019 can be attributed with at least one citation to science.

In our study, we identified AI papers by their WoS categories and extracted all patents with at least one citation link to an AI paper. We kept only citation links with a reliability score greater than three, which corresponds to a precision rate of 99.5% and a recall of 93%. This approach yields 178,004 AI patents.

B.4. World Intellectual Property Organisation (WIPO) Method

The WIPO methodology for classifying AI patents was published in 2019 and validated by a team of patent experts (WIPO, 2019a; WIPO, 2019b). The aim behind the methodology is to capture three aspects of AI invention: (1) core AI techniques (deep learning, other learning methods, various type of logic, clustering, etc.); (2) functional applications of AI that can be used to simulate human-like cognitive capacities (such as vision, language, or decision-making); and (3) end-user application fields (such as automation in business, health, or military).

This methodology is based on both a keyword search of patent texts and the use of patent classification codes (CPC and IPC). In this technique, some patents are classified

[6]Note that this assignment was made at the paper level using a probabilistic mapping which is different from the journal-based categorisation of Clarivate Analytics (Web of Science).

based on only a subset of the technological codes, or keywords, whilst others are identified by a combination of both.

The list of keywords used in this approach covers core AI methods as well as computing and mathematical concepts used in such technologies. These keywords are matched to the text in the patent titles, abstracts, and claims.

This approach identifies 158,652 patents.

B.5. United States Patent and Trademark Office (USPTO) Classification

The USPTO approach uses a supervised machine learning (ML) classifier to identify AI patents (see Giczy et al., 2021). This ML model is trained to classify eight components of AI technologies, namely: machine learning, evolutionary computation, natural language processing, speech, vision, knowledge processing, planning/control, and AI hardware. The ML model is trained on the abstracts and claims of a seed (positive set) and an anti-seed (negative set). The seeds are chosen carefully for each respective component by taking an intersection of CPC, IPC, and USPC codes, as well as Derwent's World Patents Index™. The seeds are expanded based on patent families, CPC codes, and citations to identify all patents linked to the seed set. The anti-seed set is selected randomly from all remaining patents. For training, each text is pre-processed and embedded via the Word2Vec algorithm. The ML models also encode backward and forward citations in a citation vector. The predictions from the ML model are further validated using a small subset of patents that are manually examined.

Published in August 2021, the resulting dataset contains 13.2 million USPTO patents and pre-grant publications issued or published between 1976 and 2020. For consistency with our other approaches, we only consider patents granted between 1990 and 2019 and exclude pre-grant publications. The remaining data yields 595,047 patents.

C. List of Keywords Used in Keyword Approach

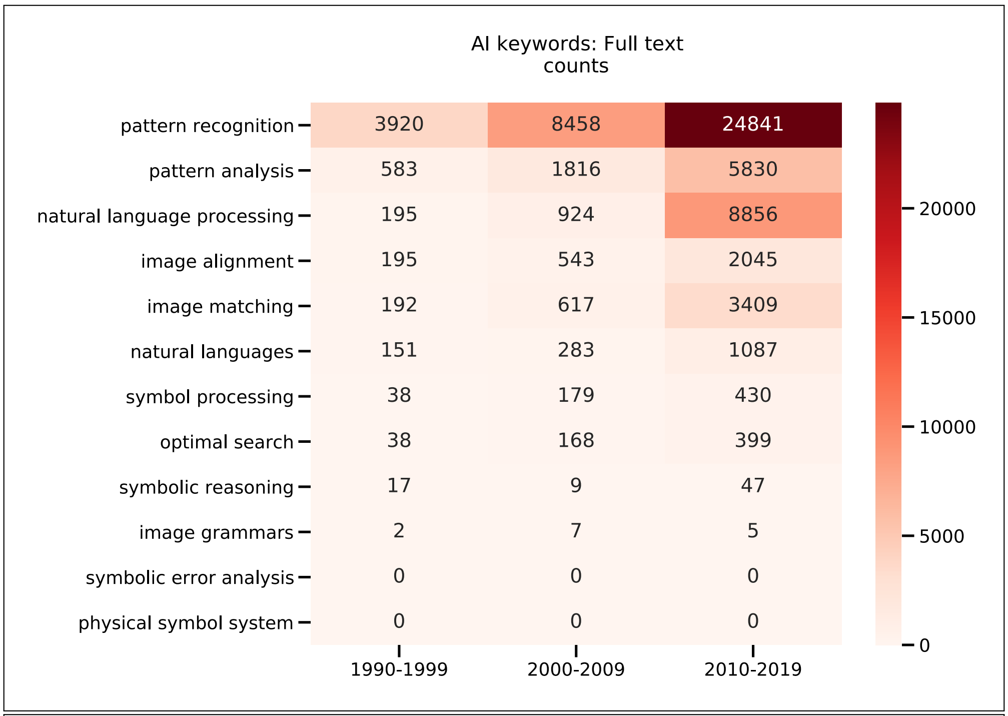
Table C.1: List of Keywords from Cockburn et al., 2018

Symbols	Learning	Robotics
natural language processing	machine learning	computer vision
image grammars	neural networks	robot
pattern recognition	reinforcement learning	robots
image matching	logic theorist	robot systems
symbolic reasoning	bayesian belief networks	robotics
symbolic error analysis	unsupervised learning	robotic
pattern analysis	deep learning	collaborative systems
symbol processing	knowledge representation and reasoning	humanoid robotics
physical symbol system	crowdsourcing and human computation	sensor network
natural languages	neuromorphic computing	sensor networks
pattern analysis	decision making	sensor data fusion
image alignment	machine intelligence	systems and control theory
optimal search	neural network	layered control systems
symbolic reasoning		
symbolic error analysis		

D. A Study of AI Keywords in Patent Texts

We split all the patent texts into three time periods (1990-1999, 2000-2009, 2010-2019) and search through the texts for keywords. Then, in each period (and for each category) we counts the unique number of matching documents and what percentages of the AI patents match according to this keyword. Figures D.1, D.2, and D.3 below illustrate both counts and shares.

Figure D.1: Symbolic Keywords: in Full Texts



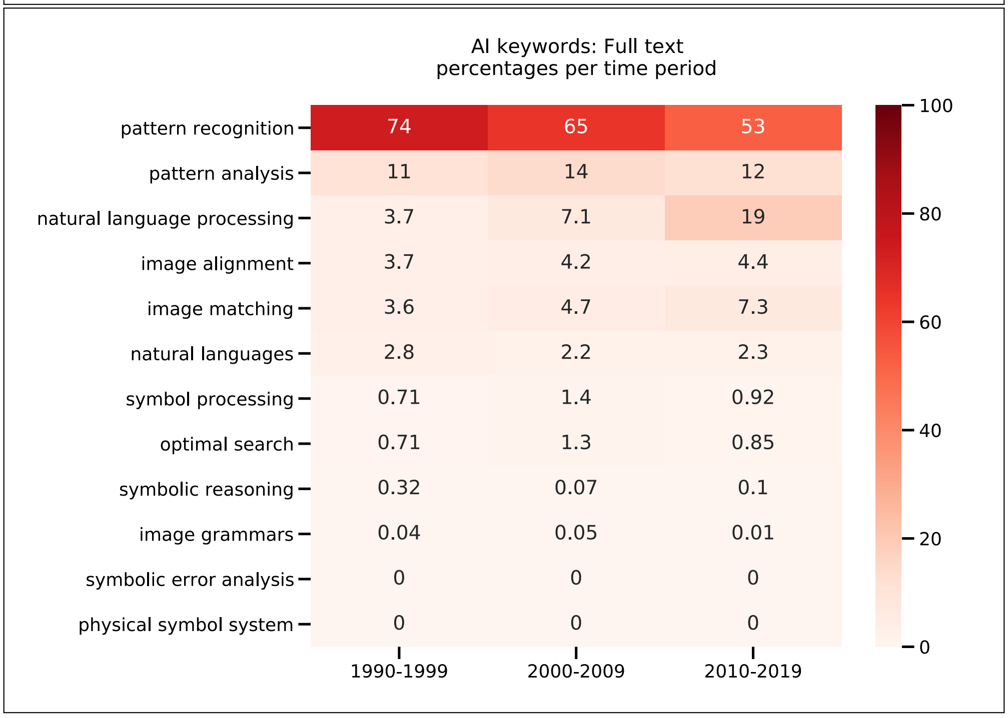
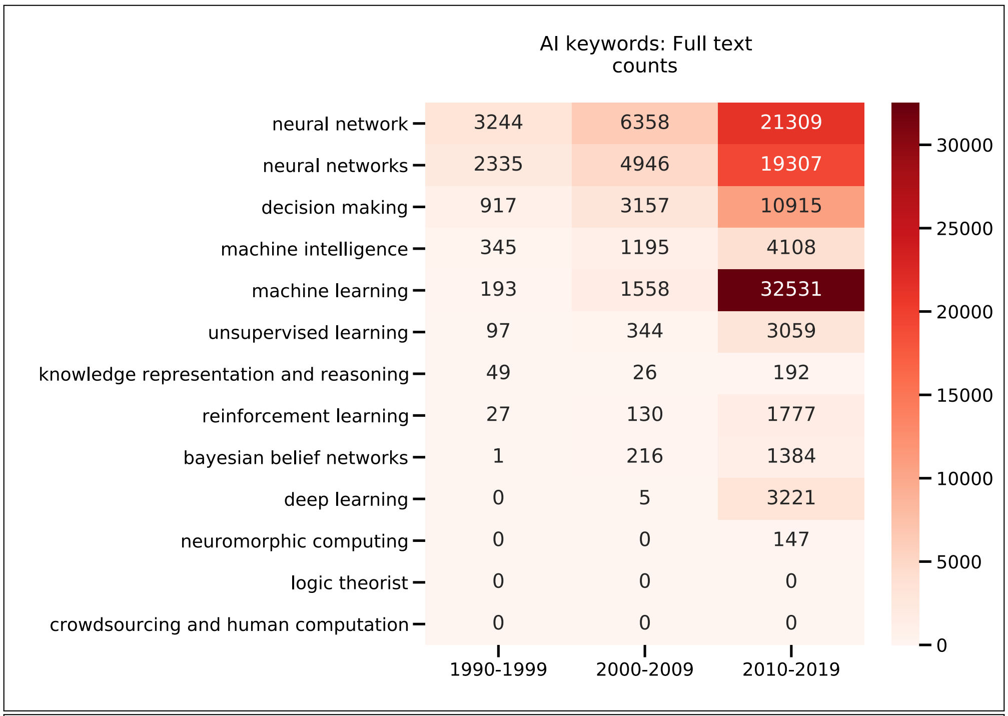


Figure D.2: Learning Keywords: in Full Texts



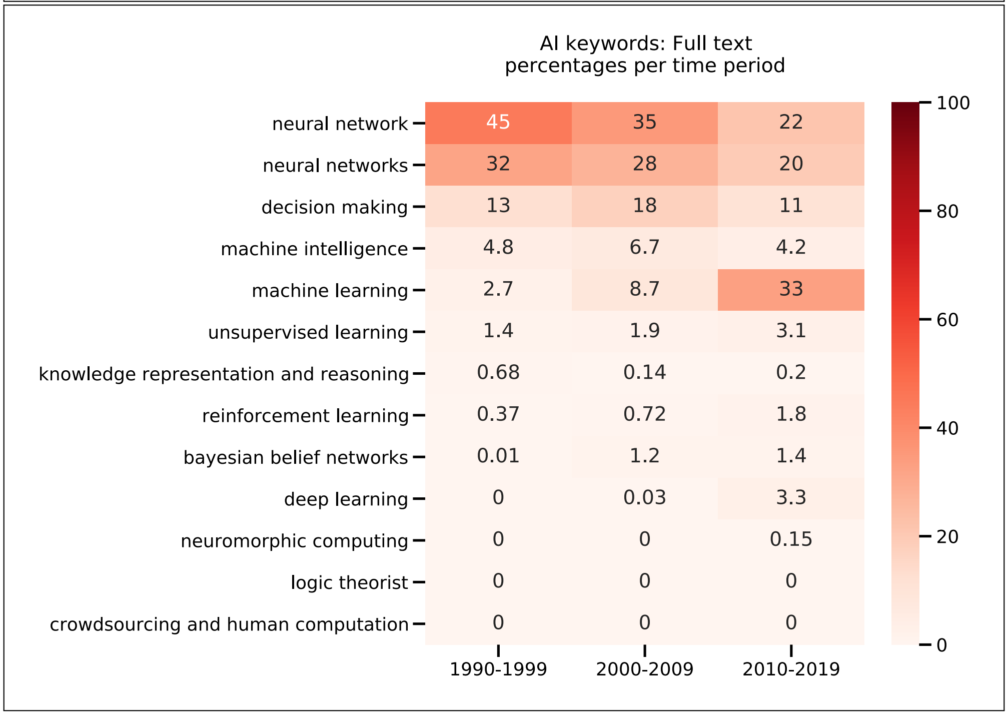
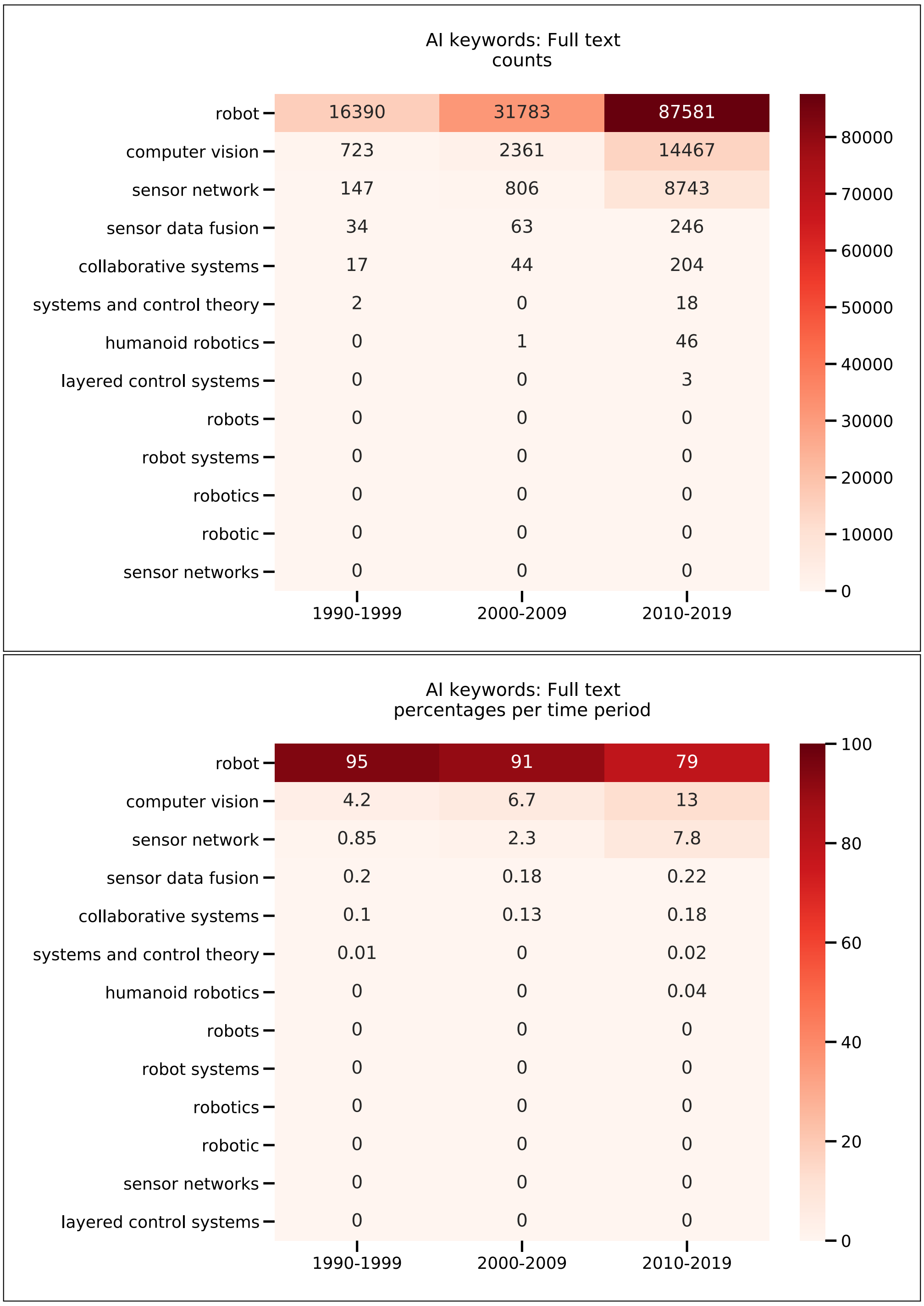


Figure D.3: Robotics Keywords: in Full Texts



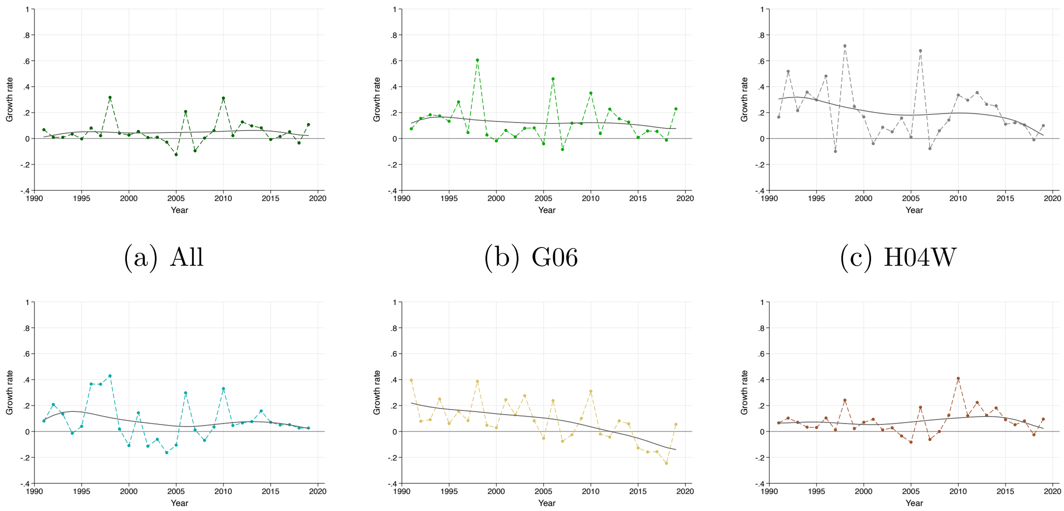
E. Additional Results

E.1. Comparison to Benchmarks

The following figures reproduce time series of growth rates, counts, and shares for additional groups of patents. The benchmarks were identified in previous discussions of GPT technologies in the literature (nanotechnology, biochemistry, green technologies, computing). Climate patents were also included as a group of technologies where one can expect wide diversity, as climate inventions can be expected to cover many sectors of the economy.

E.1.1. Growth

Figure E.4: Growth Rates of Benchmark Patents by Year



(a) All (b) G06 (c) H04W

(d) C12 (e) B82 (f) Y02

Note: ‘All’ refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

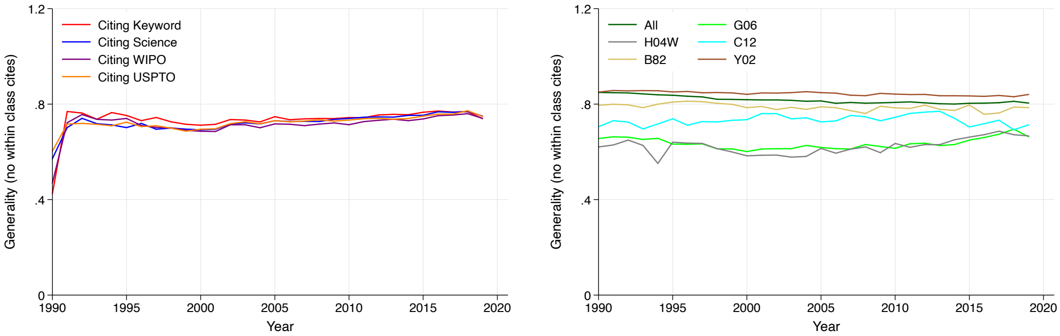
E.1.2. Generality

Table E.2: Average Generality Index (1990-2019): Benchmark Categories

	All	G06	H04W	C12	B82	Y02
1 digit	0.82	0.62	0.62	0.74	0.79	0.85
3 digit	0.95	0.82	0.82	0.85	0.92	0.95
4 digit	0.82	0.62	0.62	0.74	0.79	0.85

Notes: The generality index is defined as share of citations to patents in different CPC classes at different aggregation levels (see A.2). Citations within the same class are excluded. 'All' refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

Figure E.5: Generality Index at the 1-digit CPC-section Level



(a) AI citing patents

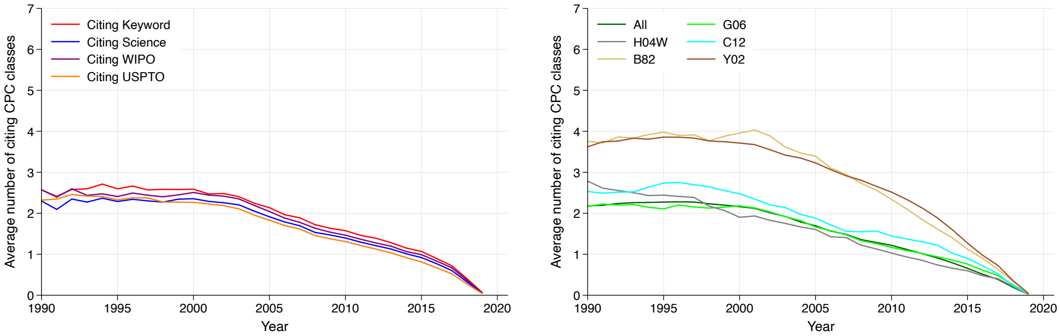
(b) Benchmark

Table E.3: Average CPC Classes Making Citations: Benchmark Categories

	All	G06	H04W	C12	B82	Y02
1 digit	1.27	1.00	0.68	1.46	2.36	1.99
3 digit	2.48	1.99	1.19	2.52	4.39	3.32
4 digit	3.97	3.31	2.80	4.18	6.42	5.10

Notes: The table shows number of different CPC classes making a citation to an average patent of the respective group. Citations within the same class are excluded. 'All' refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

Figure E.6: Average Number of CPC Classes Citing AI



(a) AI citing patents (b) Benchmark

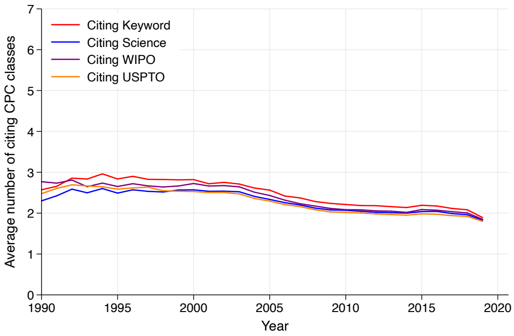
Note: ‘All’ refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

Table E.4: Average Number of Citing CPC Classes – Cited Patents

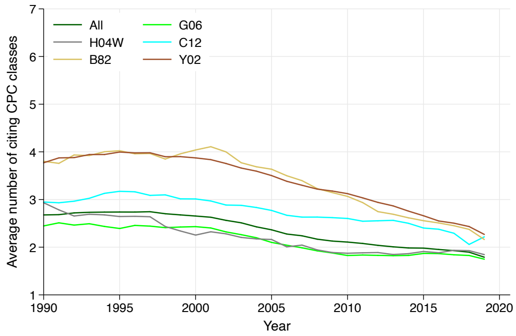
	All	G06	H04W	C12	B82	Y02
1 digit	2.39	2.03	2.00	2.78	3.35	3.29
3 digit	4.28	3.78	3.35	4.60	6.22	5.48
4 digit	6.44	5.96	5.63	7.47	9.08	8.42

Notes: The table reports numbers of different CPC classes making a citation to an average patent of the respective group that receives at least one citation. Citations within the same class are excluded. ‘All’ refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

Figure E.7: Average Number of CPC Classes Citing AI: Subset of Cited Patents



(a) AI citing patents



(b) Benchmark

Note: 'All' refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

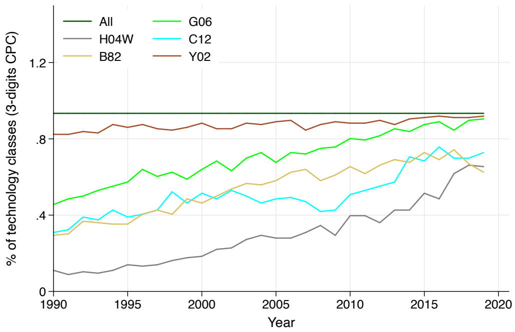
Table E.5: Average Citation Lags by Patents in Benchmark Categories

Period	All	G06	H04W	C12	B82	Y02
1990-1999	13.57	12.78	12.47	14.58	12.14	13.49
2000-2009	9.19	9.00	8.75	10.15	8.58	8.92
2010-2019	4.29	4.19	3.83	4.29	4.33	4.08

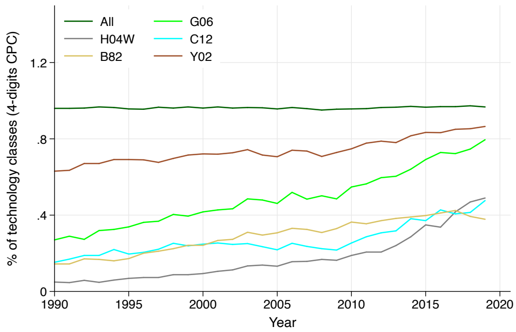
Notes: This table shows the average number of years it takes until a patent in the sample is cited. The average number of years is lower during the more recent decade as the maximal time lag is truncated since our data ends in 2019. 'All' refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

E.1.3. Complementarity

Figure E.8: Share of Technology Classes: Diversity of Benchmark Categories



(a) % of all 3-digit CPC



(b) % of all 4-digit CPC

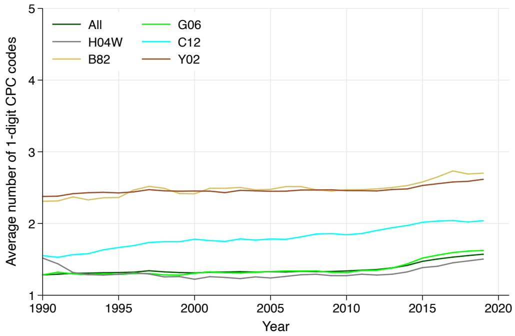
Note: Panel (a) shows the percentage of 3-digit CPC codes and panel (b) shows the percentage of 4-digit CPC as a share of all codes in the respective category. Note that the total number of 3-digit and 4-digit CPC codes are 136 and 674, respectively. 'All' refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

Table E.6: Yearly Average Number of 3- and 4-digits CPC Codes per Patent

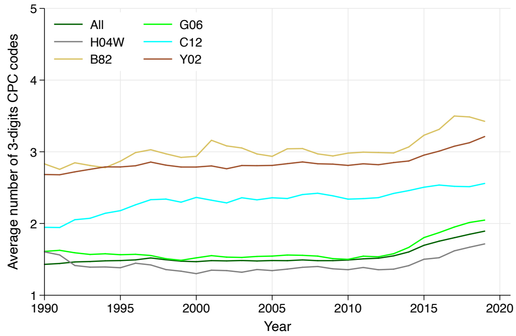
	All	G06	H04W	C12	B82	Y02
1 digit	1.36	1.36	1.32	1.80	2.48	2.47
3 digit	1.54	1.62	1.43	2.32	3.03	2.85
4 digit	1.80	1.81	2.26	2.93	3.50	3.39

Note: The table shows the average of annual average number of technology classes by 1-, 3- or 4-digit CPC per patent. 'All' refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

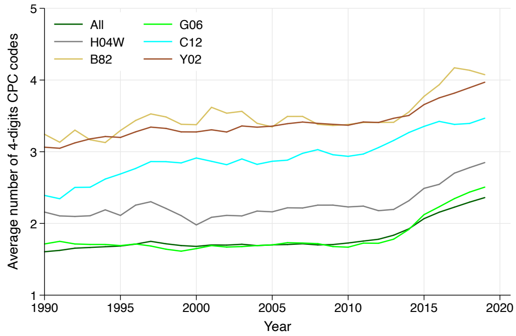
Figure E.9: Average Technology Classes: Patent-Level Diversity of Benchmark Categories



(a) Average Number of 1-digit CPC



(b) Average Number of 3-digit CPC



(c) Average Number of 4-digit CPC

Note: ‘All’ refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate-related patents, respectively.

E.2. Significance Tests

Here, we provide the results of a series of pair-wise Wilcoxon signed rank tests showing whether the differences between the means reported in Table 2, E.40, E.41, E.39, E.42 are significant.

E.2.1. Growth

Table E.7: Growth Rates

period	pair	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	1.00								
1990-2019	WIPO	1.00	1.00							
1990-2019	USPTO	1.00	1.00	1.00						
1990-2019	All	0.07	0.00	0.00	0.00					
1990-2019	G06	1.00	1.00	1.00	1.00	0.00				
1990-2019	H04W	1.00	0.64	0.77	0.17	0.00	0.04			
1990-2019	C12	0.91	0.08	0.41	0.79	1.00	0.78	0.01		
1990-2019	B82	1.00	0.29	1.00	0.71	1.00	0.90	0.05	1.00	
1990-2019	Y02	1.00	0.63	0.33	0.70	0.03	0.91	0.00	1.00	1.00
1990-1999	Science	1.00								
1990-1999	WIPO	1.00	0.96							
1990-1999	USPTO	1.00	1.00	1.00						
1990-1999	All	1.00	0.47	1.00	0.18					
1990-1999	G06	1.00	1.00	1.00	1.00	0.47				
1990-1999	H04W	1.00	1.00	1.00	1.00	0.47	0.96			
1990-1999	C12	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
1990-1999	B82	1.00	1.00	1.00	1.00	0.18	1.00	1.00	1.00	
1990-1999	Y02	1.00	0.33	1.00	0.33	1.00	0.18	0.47	1.00	0.47
2000-2009	Science	1.00								
2000-2009	WIPO	1.00	1.00							
2000-2009	USPTO	1.00	1.00	1.00						
2000-2009	All	1.00	0.43	1.00	0.26					
2000-2009	G06	1.00	1.00	1.00	1.00	0.56				
2000-2009	H04W	1.00	1.00	1.00	1.00	1.00	1.00			
2000-2009	C12	1.00	0.43	1.00	0.78	1.00	1.00	1.00		
2000-2009	B82	1.00	1.00	1.00	1.00	0.43	1.00	1.00	1.00	
2000-2009	Y02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2010-2019	Science	0.55								
2010-2019	WIPO	1.00	0.74							
2010-2019	USPTO	0.89	1.00	0.89						
2010-2019	All	0.14	0.74	0.09	0.98					
2010-2019	G06	0.89	1.00	0.30	1.00	0.09				
2010-2019	H04W	1.00	0.98	1.00	1.00	0.14	1.00			
2010-2019	C12	1.00	1.00	0.30	1.00	1.00	1.00	0.19		
2010-2019	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.40	
2010-2019	Y02	1.00	1.00	1.00	1.00	0.19	1.00	0.89	0.74	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed rank test for the hypothesis that the compared pair ranks equal.

Table E.8: Summary Statistics for Growth Rates

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.16	0.15	0.14	0.13	0.05	0.13	0.21	0.08	0.08	0.08
Median 1990-2019	0.10	0.12	0.12	0.09	0.03	0.08	0.17	0.05	0.08	0.07
St.dev. 1990-2019	0.24	0.17	0.16	0.16	0.10	0.15	0.20	0.15	0.16	0.10
Mean 1990-1999	0.21	0.26	0.16	0.20	0.06	0.19	0.32	0.18	0.17	0.08
Median 1990-1999	0.10	0.27	0.09	0.12	0.03	0.16	0.30	0.13	0.09	0.07
St.dev. 1990-1999	0.29	0.17	0.17	0.19	0.10	0.18	0.23	0.17	0.14	0.07
Mean 2000-2009	0.06	0.10	0.09	0.09	0.01	0.08	0.12	-0.01	0.09	0.03
Median 2000-2009	-0.00	0.06	0.07	0.09	0.01	0.07	0.07	-0.06	0.09	0.02
St.dev. 2000-2009	0.24	0.18	0.19	0.15	0.09	0.15	0.21	0.14	0.13	0.09
Mean 2010-2019	0.22	0.11	0.18	0.12	0.08	0.12	0.19	0.09	-0.02	0.14
Median 2010-2019	0.19	0.07	0.16	0.08	0.07	0.09	0.19	0.06	-0.03	0.11
St.dev. 2010-2019	0.17	0.13	0.09	0.15	0.10	0.12	0.12	0.09	0.16	0.12

Table E.9: Growth Rates (Citing AI)

period	pair	Keyword	Science	WIPO	USPTO	G06	H04W	C12	B82
1990-2019	Science	0.65							
1990-2019	WIPO	1.00	1.00						
1990-2019	USPTO	0.67	0.00	0.01					
1990-2019	G06	0.09	0.00	0.00	1.00				
1990-2019	H04W	0.17	0.85	1.00	0.01	0.00			
1990-2019	C12	1.00	0.65	1.00	1.00	1.00	0.11		
1990-2019	B82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
1990-2019	Y02	1.00	0.12	1.00	1.00	1.00	0.05	1.00	1.00
1990-1999	Science	0.14							
1990-1999	WIPO	0.14	1.00						
1990-1999	USPTO	1.00	0.14	0.24					
1990-1999	G06	1.00	1.00	0.14	1.00				
1990-1999	H04W	0.24	1.00	1.00	1.00	0.14			
1990-1999	C12	1.00	1.00	1.00	1.00	1.00	0.79		
1990-1999	B82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
1990-1999	Y02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2000-2009	Science	1.00							
2000-2009	WIPO	1.00	1.00						
2000-2009	USPTO	1.00	0.70	1.00					
2000-2009	G06	1.00	1.00	1.00	1.00				
2000-2009	H04W	1.00	1.00	1.00	1.00	0.96			
2000-2009	C12	1.00	1.00	1.00	1.00	1.00	1.00		
2000-2009	B82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
2000-2009	Y02	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2010-2019	Science	1.00							
2010-2019	WIPO	1.00	1.00						
2010-2019	USPTO	1.00	0.66	1.00					
2010-2019	G06	1.00	0.21	0.66	1.00				
2010-2019	H04W	1.00	1.00	1.00	1.00	0.48			
2010-2019	C12	1.00	1.00	1.00	1.00	1.00	1.00		
2010-2019	B82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
2010-2019	Y02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Table excludes those patents that themselves are AI by the respective classification approach. Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.10: Summary Statistics for Growth Rates (Citing AI)

	Keyword	Science	WIPO	USPTO	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.30	0.75	1.74	0.53	0.96	1.29	0.70	1.67	0.70
Median 1990-2019	0.11	0.13	0.12	0.09	0.11	0.19	0.08	0.12	0.10
St.dev. 1990-2019	0.48	2.56	8.01	1.71	4.08	5.34	2.58	7.77	2.65
Mean 1990-1999	0.73	2.15	5.38	1.52	2.91	3.89	2.07	5.14	2.07
Median 1990-1999	0.54	0.58	0.68	0.54	0.44	0.73	0.43	0.39	0.40
St.dev. 1990-1999	0.68	4.43	14.24	2.93	7.21	9.43	4.50	13.83	4.63
Mean 2000-2009	0.14	0.13	0.13	0.10	0.11	0.17	0.11	0.16	0.10
Median 2000-2009	0.11	0.10	0.12	0.09	0.09	0.12	0.04	0.13	0.09
St.dev. 2000-2009	0.16	0.16	0.16	0.13	0.13	0.19	0.22	0.17	0.15
Mean 2010-2019	0.08	0.09	0.09	0.06	0.05	0.09	0.07	0.07	0.07
Median 2010-2019	0.07	0.07	0.08	0.05	0.05	0.09	0.06	0.06	0.07
St.dev. 2010-2019	0.12	0.13	0.14	0.10	0.10	0.13	0.15	0.13	0.10

Notes: Table excludes those patents that themselves are AI by the respective classification approach.

E.2.2. Generality

Table E.11: Generality Index at 1-Digit Level.

Period	1-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.05							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.00	0.03			
1990-2019	C12	0.00	0.47	0.47	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.14								
1990-1999	WIPO	0.11	0.77							
1990-1999	USPTO	0.09	0.09	0.09						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.09	0.25			
1990-1999	C12	0.09	0.20	0.14	0.32	0.09	0.09	0.09		
1990-1999	B82	0.14	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.09								
2010-2019	WIPO	0.09	0.22							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	0.22	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.22	0.09	0.49			
2010-2019	C12	0.22	0.22	0.19	0.09	0.09	0.09	0.09		
2010-2019	B82	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed rank test for the hypothesis that the compared pair ranks equal.

Table E.12: Summary Statistics for Generality Index at 1-Digit Level.

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.76	0.73	0.72	0.68	0.82	0.64	0.62	0.73	0.79	0.84
Median 1990-2019	0.76	0.73	0.72	0.68	0.81	0.63	0.62	0.73	0.79	0.85
St.dev. 1990-2019	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.01	0.01
Mean 1990-1999	0.77	0.74	0.75	0.7	0.84	0.64	0.62	0.72	0.8	0.85
Median 1990-1999	0.77	0.75	0.76	0.71	0.84	0.64	0.63	0.73	0.8	0.85
St.dev. 1990-1999	0.02	0.02	0.02	0.02	0.01	0.02	0.03	0.01	0.01	0
Mean 2000-2009	0.76	0.73	0.71	0.67	0.81	0.62	0.6	0.74	0.78	0.84
Median 2000-2009	0.76	0.73	0.72	0.68	0.81	0.61	0.59	0.74	0.79	0.85
St.dev. 2000-2009	0.01	0.01	0.01	0	0.01	0.01	0.02	0.01	0.01	0.01
Mean 2010-2019	0.76	0.72	0.71	0.67	0.81	0.65	0.65	0.73	0.78	0.84
Median 2010-2019	0.76	0.72	0.71	0.67	0.8	0.64	0.66	0.74	0.78	0.84
St.dev. 2010-2019	0.02	0.01	0.01	0.01	0	0.02	0.02	0.03	0.01	0

Table E.13: Generality Index at 3-Digit Level

Period	3-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.90							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.23	0.00	0.90			
1990-2019	C12	0.00	0.03	0.01	0.14	0.00	0.00	0.00		
1990-2019	B82	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.00	0.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	1.00							
1990-1999	USPTO	0.09	0.10	0.09						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.09	1.00			
1990-1999	C12	0.09	0.09	0.09	0.22	0.09	1.00	1.00		
1990-1999	B82	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.19							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	1.00			
2000-2009	C12	0.09	0.09	0.10	0.09	0.09	0.09	0.09		
2000-2009	B82	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	1.00	0.09	0.09	0.09	0.09
2010-2019	Science	0.09								
2010-2019	WIPO	0.09	0.59							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	1.00	0.09				
2010-2019	H04W	0.09	0.97	0.09	0.59	0.09	0.16			
2010-2019	C12	0.09	0.09	1.00	0.09	0.09	0.09	0.12		
2010-2019	B82	0.97	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.97	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.14: Summary Statistics for Generality Index at 3-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.9	0.87	0.87	0.84	0.95	0.82	0.82	0.85	0.91	0.94
Median 1990-2019	0.9	0.86	0.86	0.83	0.95	0.82	0.82	0.85	0.92	0.94
St.dev. 1990-2019	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01
Mean 1990-1999	0.92	0.88	0.89	0.86	0.96	0.84	0.83	0.84	0.92	0.95
Median 1990-1999	0.92	0.88	0.89	0.87	0.96	0.85	0.83	0.84	0.92	0.95
St.dev. 1990-1999	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0
Mean 2000-2009	0.9	0.87	0.86	0.83	0.94	0.81	0.81	0.85	0.91	0.95
Median 2000-2009	0.9	0.87	0.86	0.83	0.95	0.81	0.81	0.85	0.92	0.94
St.dev. 2000-2009	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0.01	0
Mean 2010-2019	0.89	0.85	0.86	0.81	0.94	0.82	0.83	0.86	0.9	0.94
Median 2010-2019	0.9	0.85	0.86	0.81	0.94	0.82	0.84	0.86	0.9	0.94
St.dev. 2010-2019	0.02	0.01	0.01	0.01	0	0.01	0.02	0.01	0.02	0.01

Table E.15: Generality Index at 4-Digit Level

Period	4-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.00							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.00	0.00			
1990-2019	C12	0.00	0.00	0.42	0.01	0.00	0.00	0.00		
1990-2019	B82	0.04	0.04	0.01	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	0.42							
1990-1999	USPTO	0.09	0.97	1.00						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
1990-1999	C12	0.09	0.09	0.10	0.39	0.09	0.09	0.09		
1990-1999	B82	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.64	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.10	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.23	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	1.00								
2010-2019	WIPO	0.67	0.58							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	1.00	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2010-2019	C12	0.09	0.49	1.00	0.09	0.09	0.09	0.09		
2010-2019	B82	0.21	1.00	1.00	0.18	0.09	0.18	0.09	1.00	
2010-2019	Y02	0.09	0.09	0.09	0.09	1.00	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.16: Summary Statistics for Generality Index at 4-Digit Level.

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.96	0.95	0.94	0.93	0.98	0.92	0.89	0.94	0.95	0.98
Median 1990-2019	0.96	0.95	0.94	0.93	0.98	0.92	0.88	0.94	0.95	0.98
St.dev. 1990-2019	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0.02	0.01
Mean 1990-1999	0.96	0.95	0.95	0.95	0.99	0.93	0.89	0.94	0.96	0.99
Median 1990-1999	0.97	0.95	0.95	0.95	0.98	0.93	0.89	0.94	0.97	0.99
St.dev. 1990-1999	0.01	0.01	0.01	0.01	0	0.01	0.01	0	0	0
Mean 2000-2009	0.96	0.94	0.94	0.93	0.98	0.92	0.88	0.94	0.95	0.98
Median 2000-2009	0.96	0.95	0.94	0.93	0.98	0.92	0.88	0.94	0.95	0.98
St.dev. 2000-2009	0	0	0	0	0	0	0	0	0.01	0
Mean 2010-2019	0.95	0.94	0.94	0.92	0.98	0.92	0.89	0.93	0.94	0.98
Median 2010-2019	0.95	0.94	0.94	0.92	0.98	0.92	0.89	0.94	0.94	0.98
St.dev. 2010-2019	0.01	0	0.01	0	0	0	0.01	0.01	0.02	0.01

Table E.17: Average Number of Citing Classes (All) at 1-Digit Level

Period	1-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	1.00							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.01	1.00				
1990-2019	H04W	0.00	0.00	0.00	0.00	1.00	1.00			
1990-2019	C12	0.00	1.00	1.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	1.00							
1990-1999	USPTO	0.09	1.00	1.00						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	1.00	1.00	1.00	0.21	0.09			
1990-1999	C12	0.09	1.00	0.84	0.84	0.09	0.09	0.84		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.12
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.75							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.92				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.11
2010-2019	Science	0.22								
2010-2019	WIPO	0.22	0.09							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.75					
2010-2019	G06	0.09	0.09	0.09	0.09	0.22				
2010-2019	H04W	0.09	0.09	0.09	0.09	0.22	0.09			
2010-2019	C12	0.09	0.25	1.00	0.09	0.09	0.09	0.09		
2010-2019	B82	0.09	0.22	0.10	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.18: Summary Statistics for Average Number of Citing Classes (All) at 1-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	2.15	1.83	1.82	1.68	1.55	1.54	1.54	1.81	2.82	2.8
Median 1990-2019	2.51	2.06	2.17	1.86	1.74	1.74	1.64	1.93	3.43	3.29
St.dev. 1990-2019	0.93	0.76	0.79	0.82	0.72	0.68	0.81	0.8	1.27	1.17
Mean 1990-1999	2.97	2.54	2.52	2.5	2.24	2.17	2.44	2.61	3.85	3.78
Median 1990-1999	2.96	2.54	2.52	2.52	2.25	2.16	2.44	2.59	3.87	3.79
St.dev. 1990-1999	0.12	0.06	0.08	0.09	0.04	0.04	0.21	0.1	0.08	0.07
Mean 2000-2009	2.46	2.04	2.07	1.84	1.74	1.74	1.59	1.94	3.37	3.24
Median 2000-2009	2.51	2.06	2.17	1.86	1.74	1.74	1.64	1.93	3.43	3.29
St.dev. 2000-2009	0.42	0.35	0.43	0.4	0.31	0.33	0.28	0.34	0.52	0.37
Mean 2010-2019	1.03	0.92	0.87	0.7	0.68	0.72	0.59	0.88	1.23	1.38
Median 2010-2019	1.12	1.02	0.98	0.76	0.72	0.81	0.63	0.96	1.26	1.43
St.dev. 2010-2019	0.56	0.45	0.43	0.37	0.4	0.37	0.32	0.49	0.76	0.85

Table E.19: Average Number of Citing Classes (All) at 3-Digit Level

Period	3-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.00							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.82	0.04			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.01	0.44		
1990-2019	B82	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	0.34							
1990-1999	USPTO	0.09	0.70	0.32						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.09	0.58			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.16		
1990-1999	B82	0.64	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.13	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.13	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.25								
2010-2019	WIPO	0.27	0.67							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	1.00	0.25				
2010-2019	H04W	0.09	0.09	0.09	0.09	0.25	0.09			
2010-2019	C12	0.09	0.09	0.09	0.19	0.09	0.67	0.15		
2010-2019	B82	0.58	1.00	1.00	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	1.00	0.67	0.67	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.20: Summary Statistics for Average Number of Citing Classes (All) at 3-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	5.18	4.14	4.39	3.91	3.1	3.47	3.23	3.24	5.51	4.9
Median 1990-2019	6.01	4.54	5.12	4.15	3.43	3.74	3.25	3.37	6.78	5.73
St.dev. 1990-2019	2.59	2.07	2.22	2.27	1.61	1.87	2.05	1.66	2.89	2.31
Mean 1990-1999	7.61	6.28	6.48	6.32	4.71	5.39	5.62	5.02	7.86	7.01
Median 1990-1999	7.65	6.16	6.48	6.38	4.73	5.42	5.73	4.95	7.94	7.09
St.dev. 1990-1999	0.37	0.39	0.27	0.26	0.1	0.17	0.68	0.25	0.38	0.23
Mean 2000-2009	5.91	4.46	4.99	4.18	3.44	3.79	3.13	3.4	6.79	5.64
Median 2000-2009	6.01	4.54	5.12	4.15	3.43	3.74	3.25	3.37	6.78	5.73
St.dev. 2000-2009	1.28	0.92	1.3	1.22	0.77	1.02	0.77	0.74	1.61	0.88
Mean 2010-2019	2	1.67	1.7	1.23	1.16	1.24	0.93	1.31	1.89	2.06
Median 2010-2019	2.11	1.79	1.85	1.28	1.19	1.35	0.96	1.39	1.82	2.05
St.dev. 2010-2019	1.22	0.9	0.93	0.72	0.72	0.69	0.53	0.78	1.28	1.34

Table E.21: Average Number of Citing Classes (All) at 4-Digit Level

Period	4-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.00							
1990-2019	USPTO	0.00	0.16	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.82	0.28	0.06	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.04	0.00		
1990-2019	B82	0.00	0.09	0.57	0.00	0.00	0.00	0.16	0.00	
1990-2019	Y02	0.00	0.16	0.16	0.08	0.00	0.00	0.74	0.00	0.09
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	0.95							
1990-1999	USPTO	0.09	1.00	1.00						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	1.00	0.09	0.77	0.09	0.09	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	1.00	0.09		
1990-1999	B82	0.09	0.95	1.00	0.77	0.09	0.09	1.00	0.09	
1990-1999	Y02	0.09	1.00	0.09	0.84	0.09	0.09	0.25	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.80	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	1.00	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	1.00	0.09	0.09		
2000-2009	B82	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.93	0.09	0.09	0.09	0.09	0.09	0.12
2010-2019	Science	0.18								
2010-2019	WIPO	0.19	0.18							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	0.12	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2010-2019	C12	0.09	0.09	0.09	0.18	0.09	1.00	0.09		
2010-2019	B82	0.09	0.09	0.09	0.18	0.09	0.14	0.77	0.09	
2010-2019	Y02	0.09	1.00	1.00	0.09	0.09	0.09	0.12	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.22: Summary Statistics for Average Number of Citing Classes (All) at 4-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	8.9	7.41	8.04	7.13	4.96	5.83	7.52	5.48	8.28	7.66
Median 1990-2019	10.49	8.12	9.44	7.58	5.49	6.31	7.81	5.47	10.06	9.03
St.dev. 1990-2019	4.48	3.76	4.12	4.21	2.59	3.2	4.63	3.01	4.65	3.76
Mean 1990-1999	13.02	11.31	11.81	11.54	7.53	9.06	12.82	8.82	12.19	11.06
Median 1990-1999	13.11	11.08	11.85	11.7	7.57	9.06	13.16	8.76	12.21	11.22
St.dev. 1990-1999	0.51	0.67	0.41	0.31	0.23	0.16	1.28	0.51	0.86	0.51
Mean 2000-2009	10.31	8	9.28	7.7	5.54	6.43	7.53	5.58	10.14	8.9
Median 2000-2009	10.49	8.12	9.44	7.58	5.49	6.31	7.81	5.47	10.06	9.03
St.dev. 2000-2009	2.26	1.69	2.51	2.37	1.28	1.83	1.82	1.34	2.84	1.56
Mean 2010-2019	3.39	2.93	3.04	2.14	1.83	2.01	2.22	2.05	2.51	3.01
Median 2010-2019	3.51	3.09	3.28	2.22	1.87	2.16	2.26	2.14	2.39	2.92
St.dev. 2010-2019	2.13	1.64	1.74	1.32	1.17	1.18	1.34	1.25	1.75	2.04

Table E.23: Average Number of Citing Classes (Cited) at 1-Digit Level

Period	1-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	1.00							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.05	0.00					
1990-2019	G06	0.00	0.00	0.00	0.02	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.76	0.00	0.02			
1990-2019	C12	1.00	0.00	0.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	1.00							
1990-1999	USPTO	0.09	1.00	0.79						
1990-1999	All	0.09	0.14	0.22	1.00					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.26	0.14	0.11	0.97	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	1.00
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.97							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.34	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.14	0.09	0.97			
2000-2009	C12	0.97	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.09								
2010-2019	WIPO	0.09	1.00							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	1.00	1.00	0.09					
2010-2019	G06	0.09	0.09	0.09	0.09	0.09				
2010-2019	H04W	0.09	0.14	0.09	0.09	0.14	0.09			
2010-2019	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2010-2019	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.24: Summary Statistics for Average Number of Citing Classes (Cited) at 1-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	2.71	2.42	2.41	2.26	2.36	2.14	2.22	2.75	3.39	3.39
Median 1990-2019	2.87	2.52	2.55	2.25	2.4	2.15	2.17	2.8	3.66	3.55
St.dev. 1990-2019	0.49	0.36	0.38	0.44	0.33	0.28	0.35	0.3	0.62	0.55
Mean 1990-1999	3.18	2.78	2.8	2.75	2.71	2.45	2.64	3.05	3.92	3.92
Median 1990-1999	3.2	2.79	2.82	2.77	2.73	2.44	2.65	3.06	3.95	3.92
St.dev. 1990-1999	0.13	0.06	0.08	0.09	0.03	0.04	0.16	0.09	0.09	0.07
Mean 2000-2009	2.82	2.48	2.46	2.24	2.4	2.15	2.13	2.79	3.65	3.53
Median 2000-2009	2.87	2.52	2.55	2.25	2.4	2.15	2.17	2.8	3.66	3.55
St.dev. 2000-2009	0.29	0.23	0.29	0.27	0.19	0.2	0.15	0.15	0.34	0.25
Mean 2010-2019	2.13	1.98	1.98	1.77	1.98	1.83	1.88	2.41	2.61	2.71
Median 2010-2019	2.18	1.99	2	1.78	1.98	1.83	1.88	2.45	2.59	2.71
St.dev. 2010-2019	0.19	0.08	0.07	0.05	0.09	0.03	0.03	0.18	0.27	0.28

Table E.25: Average Number of Citing Classes (Cited) at 3-Digit Level

Period	3-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.00							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.03					
1990-2019	G06	0.00	0.00	0.00	0.00	0.50				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.79	0.79			
1990-2019	C12	0.00	0.00	0.00	0.50	0.00	0.04	0.05		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	0.92							
1990-1999	USPTO	0.09	0.97	0.96						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.16	0.97			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.97	0.97		
1990-1999	B82	0.97	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.14	0.14	0.11	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.10					
2000-2009	G06	0.09	0.09	0.09	0.09	1.00				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	1.00	0.09	0.26	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.26	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.21								
2010-2019	WIPO	0.33	0.33							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	0.93	0.09				
2010-2019	H04W	0.09	0.09	0.09	1.00	0.53	1.00			
2010-2019	C12	0.27	0.34	1.00	0.09	0.09	0.09	0.09		
2010-2019	B82	0.16	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.50

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.26: Summary Statistics for Average Number of Citing Classes (Cited) at 3-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	5.99	5	5.23	4.74	4.21	4.31	4.25	4.55	6.38	5.77
Median 1990-2019	6.54	5.26	5.63	4.73	4.34	4.33	4.11	4.66	7.21	6.18
St.dev. 1990-2019	1.93	1.48	1.59	1.74	1.09	1.33	1.42	1.01	2.05	1.51
Mean 1990-1999	7.87	6.59	6.8	6.68	5.35	5.75	5.97	5.63	7.99	7.26
Median 1990-1999	7.91	6.5	6.84	6.73	5.37	5.77	6.11	5.57	8.07	7.32
St.dev. 1990-1999	0.36	0.35	0.25	0.25	0.09	0.14	0.62	0.24	0.39	0.23
Mean 2000-2009	6.45	5.16	5.54	4.78	4.36	4.39	3.99	4.65	7.31	6.14
Median 2000-2009	6.54	5.26	5.63	4.73	4.34	4.33	4.11	4.66	7.23	6.18
St.dev. 2000-2009	1.07	0.77	1.07	1.02	0.6	0.83	0.58	0.46	1.31	0.71
Mean 2010-2019	3.65	3.25	3.35	2.75	2.92	2.78	2.78	3.37	3.83	3.92
Median 2010-2019	3.76	3.24	3.43	2.75	2.92	2.83	2.8	3.41	3.72	3.86
St.dev. 2010-2019	0.72	0.35	0.38	0.32	0.35	0.24	0.13	0.44	0.76	0.64

Table E.27: Average Number of Citing Classes (Cited) at 4-Digit Level

Period	4-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.00							
1990-2019	USPTO	0.00	0.12	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	1.00	0.18	1.00	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.11	0.00	0.01	0.18		
1990-2019	B82	0.01	0.01	0.15	0.00	0.00	0.00	0.04	0.00	
1990-2019	Y02	0.00	0.03	1.00	0.04	0.00	0.00	0.29	0.00	0.12
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	1.00							
1990-1999	USPTO	0.09	0.59	1.00						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	1.00	0.09	1.00	0.48	0.09	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	1.00	0.09		
1990-1999	B82	0.09	1.00	1.00	1.00	0.09	0.09	1.00	0.09	
1990-1999	Y02	0.09	1.00	0.29	0.71	0.09	0.09	0.38	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.32	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.10				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.22	0.09	1.00	0.32		
2000-2009	B82	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	1.00	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.45								
2010-2019	WIPO	1.00	0.41							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	1.00	1.00				
2010-2019	H04W	0.09	0.09	0.09	1.00	0.41	0.16			
2010-2019	C12	0.45	1.00	0.84	0.09	0.09	0.09	0.09		
2010-2019	B82	0.10	1.00	0.84	0.09	0.09	0.09	0.09	1.00	
2010-2019	Y02	1.00	0.09	0.41	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.28: Summary Statistics for Average Number of Citing Classes (Cited) at 4-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	9.92	8.54	9.05	8.21	6.36	6.95	8.35	7.41	9.43	8.92
Median 1990-2019	11.14	9.13	9.98	8.32	6.65	7.11	8.16	7.4	10.7	9.74
St.dev. 1990-2019	3.61	2.94	3.2	3.48	1.9	2.45	3.81	2.07	3.61	2.68
Mean 1990-1999	13.3	11.63	12.06	11.99	8.29	9.52	12.9	9.69	12.39	11.45
Median 1990-1999	13.34	11.45	12.07	12.09	8.35	9.51	13.21	9.61	12.47	11.65
St.dev. 1990-1999	0.55	0.6	0.43	0.28	0.24	0.16	1.26	0.49	0.88	0.53
Mean 2000-2009	10.94	8.95	9.87	8.47	6.7	7.24	7.92	7.47	10.9	9.68
Median 2000-2009	11.14	9.13	9.98	8.32	6.65	7.11	8.16	7.4	10.73	9.74
St.dev. 2000-2009	2.04	1.53	2.23	2.11	1.06	1.59	1.62	0.92	2.44	1.3
Mean 2010-2019	5.51	5.04	5.22	4.18	4.08	4.1	4.22	5.06	5	5.62
Median 2010-2019	5.66	4.99	5.34	4.21	4.11	4.21	4.25	5.12	4.87	5.51
St.dev. 2010-2019	1.42	0.83	0.89	0.74	0.68	0.57	0.54	0.87	1.15	1.18

E.2.3. Complementarity

Table E.29: Share of CPC Classes at 3-Digit Level

Period	3-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.29								
1990-2019	WIPO	0.15	0.29							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.29	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.32								
1990-1999	WIPO	0.32	0.14							
1990-1999	USPTO	0.09	0.09	0.13						
1990-1999	All	0.13	0.13	0.13	0.09					
1990-1999	G06	0.10	0.13	0.13	0.13	0.09				
1990-1999	H04W	0.13	0.09	0.09	0.09	0.13	0.09			
1990-1999	C12	0.10	0.32	0.09	0.09	0.13	0.09	0.13		
1990-1999	B82	0.13	0.13	0.09	0.09	0.13	0.09	0.09	0.17	
1990-1999	Y02	0.09	0.09	0.13	0.13	0.13	0.09	0.09	0.13	0.09
2000-2009	Science	1.00								
2000-2009	WIPO	1.00	0.77							
2000-2009	USPTO	0.18	0.09	0.18						
2000-2009	All	0.18	0.18	0.18	0.18					
2000-2009	G06	0.18	0.18	0.18	0.18	0.18				
2000-2009	H04W	0.18	0.09	0.09	0.18	0.18	0.09			
2000-2009	C12	0.09	0.09	0.12	0.09	0.18	0.09	0.09		
2000-2009	B82	0.63	0.63	1.00	0.09	0.18	0.18	0.18	0.18	
2000-2009	Y02	0.18	0.09	0.09	0.18	0.18	0.18	0.18	0.09	0.18
2010-2019	Science	0.88								
2010-2019	WIPO	0.09	0.13							
2010-2019	USPTO	0.13	0.09	0.09						
2010-2019	All	0.13	0.09	0.09	0.13					
2010-2019	G06	0.13	0.13	0.13	0.88	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.09	0.13	0.09			
2010-2019	C12	0.09	0.09	0.13	0.09	0.13	0.13	0.09		
2010-2019	B82	0.09	0.09	0.88	0.09	0.13	0.09	0.13	0.88	
2010-2019	Y02	0.09	0.09	0.09	0.13	0.13	0.13	0.13	0.13	0.13

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.30: Summary Statistics for Share of CPC Classes at 3-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.61	0.6	0.59	0.82	0.93	0.7	0.3	0.51	0.54	0.87
Median 1990-2019	0.6	0.59	0.57	0.83	0.93	0.71	0.28	0.49	0.57	0.88
St.dev. 1990-2019	0.15	0.16	0.12	0.06	0	0.13	0.17	0.12	0.14	0.03
Mean 1990-1999	0.47	0.43	0.48	0.76	0.93	0.56	0.13	0.4	0.37	0.85
Median 1990-1999	0.45	0.43	0.51	0.75	0.93	0.56	0.12	0.4	0.36	0.85
St.dev. 1990-1999	0.08	0.1	0.07	0.05	0	0.06	0.03	0.06	0.06	0.02
Mean 2000-2009	0.59	0.59	0.57	0.82	0.93	0.7	0.27	0.48	0.57	0.87
Median 2000-2009	0.6	0.58	0.57	0.83	0.93	0.71	0.28	0.49	0.57	0.88
St.dev. 2000-2009	0.04	0.03	0.04	0.02	0	0.04	0.05	0.04	0.05	0.02
Mean 2010-2019	0.78	0.77	0.71	0.87	0.93	0.85	0.49	0.64	0.68	0.9
Median 2010-2019	0.79	0.77	0.69	0.86	0.93	0.85	0.46	0.69	0.67	0.91
St.dev. 2010-2019	0.08	0.06	0.08	0.03	0	0.04	0.11	0.09	0.04	0.02

Table E.31: Share of CPC Classes at 4-Digit Level

Period	4-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.83								
1990-2019	WIPO	0.83	0.33							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.77								
1990-1999	WIPO	0.48	0.42							
1990-1999	USPTO	0.09	0.09	0.09						
1990-1999	All	0.09	0.09	0.09	0.09					
1990-1999	G06	0.09	0.09	0.09	0.09	0.09				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
1990-1999	C12	0.09	0.48	0.09	0.09	0.09	0.09	0.09		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.92	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.48								
2010-2019	WIPO	0.09	0.22							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	0.22	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2010-2019	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.22		
2010-2019	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.25	0.46	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.32: Summary Statistics for the Share of CPC Classes at 4-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.37	0.37	0.36	0.6	0.96	0.49	0.17	0.27	0.29	0.74
Median 1990-2019	0.33	0.35	0.33	0.6	0.96	0.48	0.14	0.25	0.31	0.72
St.dev. 1990-2019	0.14	0.13	0.12	0.1	0.01	0.15	0.13	0.08	0.09	0.06
Mean 1990-1999	0.24	0.24	0.25	0.48	0.96	0.33	0.06	0.2	0.18	0.68
Median 1990-1999	0.24	0.23	0.27	0.48	0.96	0.33	0.06	0.2	0.17	0.68
St.dev. 1990-1999	0.06	0.07	0.05	0.05	0	0.05	0.02	0.03	0.03	0.03
Mean 2000-2009	0.33	0.36	0.33	0.6	0.96	0.47	0.14	0.24	0.3	0.72
Median 2000-2009	0.33	0.35	0.33	0.6	0.96	0.48	0.14	0.24	0.31	0.72
St.dev. 2000-2009	0.02	0.02	0.02	0.02	0	0.03	0.03	0.01	0.03	0.01
Mean 2010-2019	0.54	0.53	0.5	0.71	0.97	0.66	0.32	0.36	0.39	0.81
Median 2010-2019	0.56	0.53	0.5	0.7	0.97	0.67	0.31	0.38	0.39	0.82
St.dev. 2010-2019	0.09	0.07	0.09	0.05	0.01	0.08	0.11	0.07	0.02	0.04

Table E.33: Average Diversity at 1-Digit Level

Period	1-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	1.00								
1990-2019	WIPO	0.02	0.05							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.02	1.00	0.00					
1990-2019	G06	0.00	0.06	1.00	0.00	1.00				
1990-2019	H04W	0.00	0.01	0.03	0.00	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.82
1990-1999	Science	1.00								
1990-1999	WIPO	0.09	1.00							
1990-1999	USPTO	0.09	0.84	0.15						
1990-1999	All	0.25	1.00	0.09	0.09					
1990-1999	G06	0.09	1.00	0.15	0.09	0.45				
1990-1999	H04W	1.00	1.00	0.84	0.09	1.00	1.00			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.71
2000-2009	Science	0.09								
2000-2009	WIPO	0.29	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.83	0.09					
2000-2009	G06	0.09	0.09	0.83	0.09	0.83				
2000-2009	H04W	0.09	0.09	0.09	0.29	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.29
2010-2019	Science	1.00								
2010-2019	WIPO	1.00	1.00							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.58	0.11	0.84	0.09					
2010-2019	G06	1.00	1.00	1.00	0.09	0.84				
2010-2019	H04W	0.09	0.09	0.09	0.27	0.09	0.09			
2010-2019	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2010-2019	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.34: Summary Statistics for Average Diversity at 1-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	1.39	1.4	1.36	1.27	1.36	1.36	1.32	1.8	2.48	2.47
Median 1990-2019	1.36	1.42	1.33	1.24	1.33	1.32	1.29	1.78	2.48	2.46
St.dev. 1990-2019	0.09	0.1	0.12	0.08	0.08	0.11	0.08	0.15	0.11	0.05
Mean 1990-1999	1.35	1.3	1.27	1.24	1.31	1.29	1.33	1.64	2.39	2.43
Median 1990-1999	1.35	1.28	1.27	1.24	1.32	1.29	1.29	1.65	2.37	2.43
St.dev. 1990-1999	0.03	0.08	0.02	0.01	0.02	0.01	0.08	0.08	0.08	0.03
Mean 2000-2009	1.35	1.43	1.33	1.24	1.33	1.32	1.26	1.79	2.48	2.46
Median 2000-2009	1.36	1.43	1.33	1.24	1.33	1.32	1.26	1.78	2.48	2.45
St.dev. 2000-2009	0.02	0.02	0.03	0.01	0.01	0.01	0.02	0.04	0.03	0.01
Mean 2010-2019	1.47	1.47	1.48	1.35	1.45	1.47	1.37	1.97	2.58	2.52
Median 2010-2019	1.48	1.47	1.47	1.35	1.44	1.48	1.36	1.99	2.55	2.51
St.dev. 2010-2019	0.11	0.08	0.14	0.11	0.09	0.12	0.09	0.08	0.1	0.06

Table E.35: Average Diversity at 3-Digit Level

Period	3-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.65								
1990-2019	WIPO	0.38	0.87							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.79	0.79	0.46	0.00	0.00				
1990-2019	H04W	0.00	0.00	0.00	0.79	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	1.00								
1990-1999	WIPO	0.09	1.00							
1990-1999	USPTO	0.09	0.09	0.09						
1990-1999	All	0.09	1.00	0.76	0.09					
1990-1999	G06	1.00	1.00	0.09	0.09	0.09				
1990-1999	H04W	0.09	1.00	0.49	1.00	1.00	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.92	0.09	0.09	0.09	0.09				
2000-2009	H04W	0.09	0.09	0.09	0.86	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	1.00								
2010-2019	WIPO	0.09	0.10							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.32	1.00	0.09	0.09	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.10	0.09	0.09			
2010-2019	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2010-2019	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

Table E.36: Summary Statistics for Average Diversity at 3-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	1.61	1.64	1.64	1.43	1.54	1.62	1.43	2.32	3.03	2.85
Median 1990-2019	1.55	1.65	1.58	1.39	1.48	1.56	1.39	2.35	2.99	2.81
St.dev. 1990-2019	0.15	0.14	0.21	0.12	0.12	0.15	0.11	0.16	0.19	0.12
Mean 1990-1999	1.55	1.51	1.5	1.39	1.48	1.57	1.43	2.16	2.88	2.77
Median 1990-1999	1.56	1.48	1.48	1.4	1.48	1.57	1.4	2.16	2.86	2.79
St.dev. 1990-1999	0.04	0.1	0.04	0.02	0.03	0.04	0.09	0.15	0.09	0.06
Mean 2000-2009	1.54	1.69	1.58	1.36	1.48	1.54	1.35	2.36	3.01	2.81
Median 2000-2009	1.54	1.68	1.59	1.36	1.48	1.54	1.35	2.36	3.01	2.81
St.dev. 2000-2009	0.02	0.05	0.05	0.02	0.01	0.02	0.03	0.04	0.07	0.03
Mean 2010-2019	1.74	1.73	1.85	1.53	1.67	1.75	1.49	2.46	3.2	2.96
Median 2010-2019	1.73	1.71	1.82	1.52	1.65	1.74	1.46	2.48	3.15	2.91
St.dev. 2010-2019	0.2	0.14	0.26	0.17	0.15	0.21	0.14	0.08	0.22	0.14

Table E.37: Average Diversity at 4-Digit Level

Period	4-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.01								
1990-2019	WIPO	0.00	0.10							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.00					
1990-2019	G06	0.00	0.00	0.00	0.00	0.52				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	1.00								
1990-1999	WIPO	1.00	1.00							
1990-1999	USPTO	0.09	0.09	0.09						
1990-1999	All	0.09	1.00	0.09	0.09					
1990-1999	G06	0.09	1.00	0.09	0.09	1.00				
1990-1999	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.26							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.28				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.25								
2010-2019	WIPO	0.09	0.09							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	0.09	0.64				
2010-2019	H04W	0.09	0.09	0.70	0.09	0.09	0.09			
2010-2019	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2010-2019	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.15

Notes: Entries show the p-value of a two-sided paired Wilcoxon signed-rank test for the hypothesis that the compared pair ranks equal.

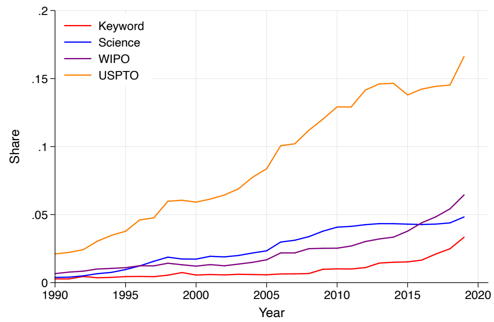
Table E.38: Summary Statistics for Average Diversity at 4-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	1.88	1.97	2.05	1.64	1.8	1.81	2.26	2.93	3.5	3.39
Median 1990-2019	1.77	1.97	1.96	1.56	1.71	1.71	2.2	2.89	3.41	3.36
St.dev. 1990-2019	0.26	0.25	0.38	0.21	0.21	0.25	0.21	0.3	0.27	0.23
Mean 1990-1999	1.77	1.73	1.76	1.55	1.68	1.69	2.16	2.64	3.31	3.2
Median 1990-1999	1.77	1.67	1.75	1.56	1.68	1.71	2.14	2.65	3.3	3.21
St.dev. 1990-1999	0.05	0.13	0.04	0.03	0.04	0.04	0.07	0.2	0.14	0.1
Mean 2000-2009	1.75	2	1.95	1.53	1.7	1.69	2.16	2.9	3.46	3.35
Median 2000-2009	1.75	1.99	1.98	1.54	1.7	1.69	2.17	2.89	3.44	3.36
St.dev. 2000-2009	0.03	0.06	0.09	0.02	0.01	0.03	0.09	0.07	0.1	0.05
Mean 2010-2019	2.12	2.19	2.43	1.83	2.01	2.05	2.45	3.24	3.72	3.63
Median 2010-2019	2.09	2.16	2.39	1.82	2	2.02	2.4	3.31	3.66	3.58
St.dev. 2010-2019	0.35	0.26	0.44	0.28	0.24	0.32	0.26	0.2	0.33	0.22

E.3. Extra Tables Moved Down

E.3.1. Volume and Time Trends

Figure E.10: AI Patents by Year (1990-2019)



(a) Share of AI patents

Notes: This figure shows the evolution of AI patents over time as identified by the four different approaches, as a share of all US patents granted in the same year.

E.3.2. Generality

Table E.39: Average Number of Citing CPC Classes (1990-2019): Cited Patents

	Keyword	Science	WIPO	USPTO
1 digit	2.71	2.42	2.41	2.26
3 digit	5.99	5.00	5.23	4.74
4 digit	9.92	8.54	9.05	8.21

Notes: This table shows the numbers of different CPC classes making a citation to an average patent of the respective group conditional on the patent being cited at least once. Citations within the same class are excluded.

Table E.40: Average Generality Index (1990-2019)

	Keyword	Science	WIPO	USPTO
1 digit	0.76	0.73	0.72	0.68
3 digit	0.91	0.87	0.87	0.84
4 digit	0.96	0.95	0.94	0.93

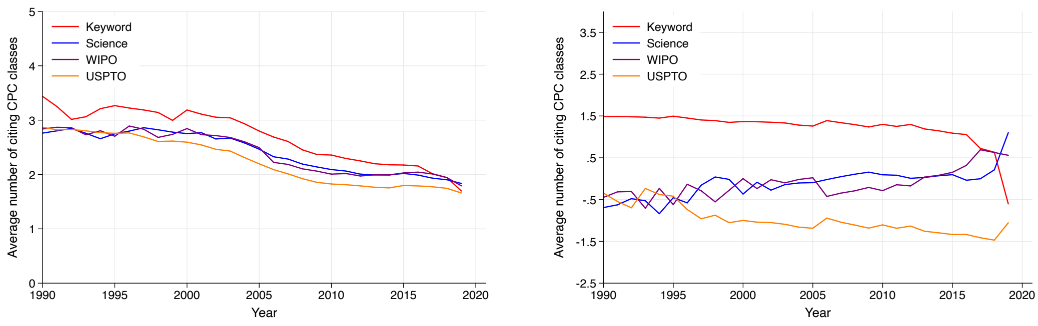
Notes: The generality index is defined as share of citations to patents in different CPC classes at different aggregation levels (see A.2). Citations within the same class are excluded.

Table E.41: Average Number of Citing CPC Classes (1990-2019)

	Keyword	Science	WIPO	USPTO
1 digit	2.15	1.83	1.82	1.68
3 digit	5.18	4.14	4.39	3.91
4 digit	8.90	7.41	8.04	7.13

Notes: This table shows the numbers of different CPC classes making a citation to an average patent of the respective group. Citations within the same class are excluded.

Figure E.11: Average Number of Classes Citing AI



(a) Subset of cited patents (b) Subset of cited patents (z-score scaled)

Notes: The z-scored value equals the level of the generality index minus its average across the four approaches divided by the standard deviation for each year.

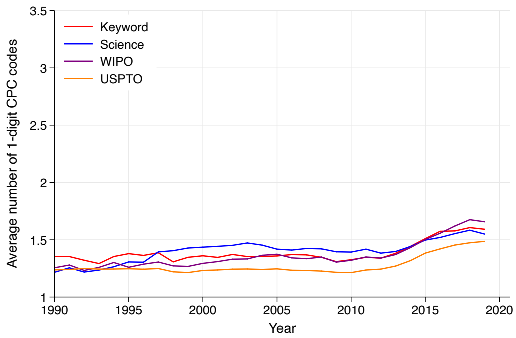
E.3.3. Complementarity

Table E.42: Average Number of 1-, 3- and 4-digit CPC Classes per Patent

	Keyword	Science	WIPO	USPTO
1 digit	1.39	1.40	1.36	1.27
3 digit	1.61	1.64	1.64	1.43
4 digit	1.88	1.97	2.05	1.64

Note: The table shows the average of annual average number of technology classes by 1-, 3- or 4-digit CPC per patent.

Figure E.12: Patent-Level Diversity - Average Technology Classes



(a) Average number of 1-digit CPC

E.4. Generality of AI Descendants

In this section, we report additional results for the wide-ranging usefulness of technological descendants of AI, i.e. those patents that cite an AI patent but are not AI themselves. This serves as an additional indicator of the wide spread of AI in a range of different products and processes. The results confirm the persistence of the ranking, indicating the highest generality of keyword patents across different indicators.

Table E.43: Average Generality Index: AI Descendants

	Citing Keyword	Citing Science	Citing WIPO	Citing USPTO
1 digit	0.74	0.73	0.72	0.72
3 digit	0.89	0.87	0.87	0.88
4 digit	0.96	0.95	0.95	0.95

Notes: Generality is measured as $G = 1 - \sum(s^2)$ with s as share of citations to patents in different CPC classes at different aggregation levels. Citations within the same class are excluded.

Table E.44: Average Number of Citing CPC Classes: AI Descendants

	Citing Keyword	Citing Science	Citing WIPO	Citing USPTO
1 digit	1.32	1.15	1.23	1.16
3 digit	2.75	2.27	2.57	2.33
4 digit	4.82	3.99	4.57	4.06

Notes: This table shows the number of different CPC classes making a citation to an average patent of the respective group. Citations within the same class are excluded.

Table E.45: Average Number of Citing CPC Classes (Cited): AI Descendants

	Citing Keyword	Citing Science	Citing WIPO	Citing USPTO
1 digit	2.35	2.17	2.24	2.17
3 digit	4.60	4.04	4.39	4.10
4 digit	7.51	6.60	7.29	6.63

Notes: The table shows the numbers of different CPC classes making a citation to an average patent of the respective group that receives at least one citation. Citations within the same class are excluded.

Table E.46: Average Citation Lags by Group of AI Citing Patents

Period	Citing Keyword	Citing Science	Citing WIPO	Citing USPTO
1990-1999	12.79	12.59	12.66	12.47
2000-2009	8.95	9.03	9.01	8.84
2010-2019	4.30	4.29	4.22	4.28

Notes: This table shows the average number of years it takes until a patent in the sample is cited. The average number of years is lower during the more recent decade as the maximal time lag is truncated since our data ends in 2019. Note that the group of AI citing includes all patents that cite AI but are not identified as AI by the respective approach.