
**Working Paper No.
2024-02**

INET Oxford Working Paper Series
18 March 2024



Measuring artificial intelligence: A systematic assessment and implications for governance

Kerstin Hötte, Taheya Tarannum, Wilhelm Verendel and
Lauren Bennett



Measuring artificial intelligence: a systematic assessment and implications for governance

Kerstin Hötte,^{1,2} Taheya Tarannum,³ Wilhelm Verendel,^{2,4}
Lauren Bennett⁵

¹ *The Alan Turing Institute, London, UK*

² *Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK*

³ *Oxford Martin Programme on Technological and Economic Change, University of Oxford, UK*

⁴ *Department of Physics, Chalmers University, SE*

⁵ *The Leverhulme Doctoral Training Programme for the Ecological Study of the Brain, University College London, UK*

March 18, 2024

Governing artificial intelligence (AI) is high on the political agenda, but it is still not clear how to define and measure it. We compare four approaches to identifying AI patented inventions that reflect different ways of understanding AI with divergent definitions. Using US patents from 1990-2019, we assess the extent to which each approach qualifies AI as a general purpose technology (GPT) and study patterns of concentration, which both are criteria relevant for regulation. The four approaches overlap on only 1.37% of patents and vary in scale, accounting for shares that range from 3-17% of all US patents in 2019. The smallest set of AI patents in our sample, identified by the latest AI keywords, is most GPT-like with high levels of growth and generality. All four approaches show AI inventions to be concentrated in few firms, confirming worries about competition. Our results suggest that regulation may not be straightforward, as the identification of AI inventions ultimately depends on how AI is defined.

JEL Classifications: O31, O33, O34

Keywords: Artificial Intelligence, Governance, General Purpose Technology, Concentration, Patent, Classification

Highlights

- Attempts to AI regulation and impact evaluation require a clear definition and measurement of AI.
- We systematically compare 4 approaches to AI classification using patents, and assess criteria relevant for policy.
- Across these 4 definitions of AI, we find significant differences by growth rates and scope (54k-600k patents) and little overlap (1.37%).
- All definitions classify AI as a GPT (by growth, generality, complementarity), but to different extents.
- Concentration of a few top-ranked AI inventing firms is consistent, but lower ranks differ.

1. Introduction

Artificial intelligence (AI) governance is high on the agenda of national and international policy, reflecting the view that future AI could and should be shaped by policy to mitigate risks and make it beneficial for all (Jelinek et al., 2021; Mazzucato et al., 2022; Schmitt, 2022). Governance requires a clear identification of AI technologies, but so far a consensual definition is lacking (Krafft et al., 2020), which hampers research that aims to assess the impact of AI and the effectiveness of regulatory policy (Dafoe, 2018). In this paper, we use patents as a quantitative and qualitative record of AI inventions. Comparing the differences across four different AI classification methods, we show how empirical conclusions about the characteristics and concentration in AI development can be sensitive to the chosen classification method.

AI is often claimed to be a general purpose technology (GPT) shaping the future technological and economic evolution, with wide-ranging impacts on production processes, labour markets, and technological and economic leadership at the national and international level (Agrawal et al., 2019; Cockburn et al., 2019; Brynjolfsson et al., 2021; Cockburn et al., 2018; Valdes and Rudyk, 2017; Webb, 2019; Alderucci et al., 2020). Testing whether such predictions hold requires a robust empirical measure of AI, but the ex-ante measurement of radical technological novelty is challenging (Schumpeter, 2005). AI is believed to be still at an early stage (Brynjolfsson et al., 2021), and its long-run effects have not yet manifested and may be affected by decisions undertaken today (Jacobides et al., 2021; Petit, 2017; Fanti et al., 2022).

In this paper, we do not claim to identify the best definition, but we explore the empirical implications when using different measurement methods of AI inventions and discuss how this matters for AI policy and research. We do so by comparing four samples of AI patents, each reflecting different ways to understand AI, as outlined in the literature. The samples are identified by:

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.

In the union of all samples, we identify 732k AI patents from 1990 to 2019. However, the different samples vary greatly by scale: 54k patents are captured by the Keyword, 178k by the Science, 159k by the WIPO, and 595k by the USPTO approach. Strikingly, all four methods agree on only 1.37% of AI patents, with pairwise overlaps of 20% or less throughout the entire period. Additionally, the four approaches reflect disparate trends of AI inventions over time, with the Science and USPTO sample showing an AI

slowdown in recent years, while Keyword and WIPO patents tell a story of continuous growth up until 2020.

We evaluate whether the four samples of AI patents reflect the beliefs of AI being a GPT, by assessing three criteria Bresnahan and Trajtenberg (1995), Hall and Trajtenberg (2006), and Petralia (2020):

1. GPTs are engines of growth with continued technological improvements (Petralia, 2020). We measure this feature by the growth rates of each AI sample, and those sets of patents that rely on them, as indicated by citations.
2. GPTs can be used across a wide range of products and processes. We examine the technological diversity of patent citations. As GPTs often experience long delays before being widely taken up (Comin and Mestieri, 2018), we measure citation lags between AI and subsequent inventions (Hall and Trajtenberg, 2006).
3. GPTs complement technologies in many fields (Petralia, 2020), being reflected in a high technological diversity. We quantify the diversity of co-classifications of AI patents across technology groups.

Evaluating the “GPTness” of AI is pertinent because it can be a desirable feature that promises continued growth in all areas of the economy (Lipsey et al., 2005). GPTs bring large social benefits in the long run, but they are also characterised by public good features, which can be a justification for public support (Bresnahan and Trajtenberg, 1995). The realisation of GPT benefits for all does not necessarily come about on its own: commercial actors may maximise the private instead of the social value of an invention, leading to a pre-mature lock-in in an inferior pathway of AI development (Klinger et al., 2022; Bresnahan, 2023). This consequently motivates an additional analysis of the agency dimension of AI patenting, identifying the key actors in its development, and evaluating the concentration of inventive activities.

Our results indicate that all methods classify AI as a GPT, yet the AI inventions identified by Keywords and the WIPO method show the highest level of GPTness. Furthermore, inventive activities by firms appear least concentrated in those technology areas identified by Keywords. Relying on GPTness and diversity in the market of AI development, one may conclude that Keywords, compared to the other methods, suggest larger possibilities for public benefits. This could give policy guidance, aimed at supporting inventions for the public good.

It should be noted that our approach which is based on patent data only describes a small niche of the entire AI ecosystem (Jacobides et al., 2021) and a narrow set of evaluation criteria that may be relevant for policy. Attempts to empirically assess the impact of AI and policy on pathways of AI development should therefore rely on a variety of methods of identifying AI and evaluation criteria, in line with the goals set by policy. Our systematic comparison of the four classification methods has research implications for patent-based studies, showing that researchers need to be aware of the implications of choosing a particular method and definition when analysing AI.

Measuring AI innovation remains a challenge, but our comparison informs discussions on its definitions by showing consensual features of AI and quantifying nuanced

differences in the types of industries, technology fields, and key players involved in AI development.

The rest of our paper is structured as follows: Section 2 introduces GPTs and AI in patent data; Section 3 describes the methods, followed by the results (Section 4). In Section 5, we discuss the findings, and Section 7 concludes.

2. Background

Here, we review the arguments for AI being a GPT, and discuss their limitations and the use of patent data to explore different definitions of AI.

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 technology 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. In research, 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 through 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. If the technology is sufficiently mature to be valuable for many uses, this should be reflected in relatively high growth rates.

The third 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 longer or varying periods for GPTs to evolve, spread and unleash their full economic impact (Lipsey et al., 2005). The idea is that once the inventions have sufficiently evolved and spread, GPTs rely on supporting infrastructure and secondary inventions to restructure organisation and production processes throughout the economy.

While much work has discussed AI as a GPT (see above), it has also been raised that the GPT framework may be a simplification in the context of innovation policy. Recent work (Jacobides et al., 2021) 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), and be associated with lost social gains due to under-investment. However, it is unclear whether AI represents a single GPT or is an amalgamation of many technologies that fulfil 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. A possible concentration of AI development to include only 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). 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).

Below, we analyse whether this view is consistent with the four different approaches to defining AI inventions in patent data. Comparing multiple definitions of AI serves many purposes. First, timescale and history: different definitions of AI impact the timescale it is viewed on and the perceived history of its development. Second, key actors: theoretical approaches will differ in how they identify key actors shaping AI innovation and the impact on market power. Third, concentration and GPT characteristics: the choice of definition may spur differing guidance on the direction of funding and investment in AI technologies.

2.2. AI history and different definitions

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 will need to make choices with impact on definition and scope, which is often handled on a case-to-case basis.

The conceptual roots of AI stem back to Alan Turing (Turing, 1950) around 1950, and a series of research results and inventions in the following decades led to breakthroughs in the ability of computers to be useful in computational problem-solving. 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 methods from machine learning, typically computational approaches to detect and model patterns in various data sources (Mitchell et al., 2007). Machine learning builds upon computer science, statistics, probability, and optimisation. Machine learning methods software are often paired with hardware sensors, actuators, and controls to create intelligent systems: Such a mix of machine learning software with computer hardware has been considered as one way to build AI systems.

Recent work (Bianchini et al., 2020; Klinger et al., 2022; Jurowetzki et al., 2021; Whittaker, 2021) contends that modern AI research is being rapidly privatised and narrowed towards deep learning (a particular form of machine learning) at the expense

of other relatively unexplored domains. Moreover, a dependence appears to exist on relatively speaking narrow and deep digital transformation in firms (Bresnahan, 2023). It remains to be seen whether this narrowing will impact what definitions will be used to capture the full scope of future AI inventions: These may also include new sub-fields that are, at present, relatively unexplored.

Taken together, we can see both general longer-time trends and the recent concentration of inventions that often use few particular technologies. There are important research- and policy-relevant choices regarding how large the scope of an AI definition can and should be. We can also see current attempts to make AI definitions more technology-neutral, e.g., illustrated in the forthcoming EU legislation that regulates AI. The EU AI Act is one of the first international AI-specific regulations and defines an *AI system*

a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments

and the specific connection to machine learning is further emphasized in the text (in line with the description above) as

AI systems often have machine learning capacities that allow them to adapt and perform new tasks autonomously. Machine learning refers to the computational process of optimizing the parameters of a model from data, which is a mathematical construct generating an output based on input data. Machine learning approaches include, for instance, supervised, unsupervised and reinforcement learning, using a variety of methods including deep learning with neural networks. [...] AI systems can be used as stand-alone software systems, integrated into a physical product (embedded), used to serve the functionality of a physical product without being integrated therein (non-embedded) or used as an AI component of a larger system. If this larger system would not function without the AI component in question, then the entire larger system should be considered as one single AI system under this Regulation.

There is a potential for abstracting away from specific technologies and a possible challenge to be sufficiently specific to draw clear system boundaries. Stated more generally, there is an arbitrariness of choice. It is the different implications that follow from the choice of AI perspective (definition) that we explore here.

2.2.1. Patents as a data source and other options

Here, we use patent data to investigate overlap and differences in the characteristics between four AI definitions. 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. It is also possible to track citations. One patent can be classified into a variety of technology fields, which describe the nature of the patented technology (Jaffe and De Rassenfosse, 2017). Patents have been previously used to study AI inventions (Cockburn et al., 2018; WIPO, 2019b; USPTO, 2020; Verendel, 2023).

Patents are assigned to one or more codes from hierarchical patent classification systems, such as the Cooperative Patent Classification (CPC) system developed by the European and US patent offices.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, or related to functions in another sector of the economy.

Patent citation links to other patents and, in some cases, academic articles, help distinguish the patented invention from previous inventions. Inventors are required to establish the novelty of their invention compared to existing technologies in patent applications. These citations are frequently used by innovation scholars as they reveal the cumulative dependencies between technologies, and may be interpreted as an indicator of the extent to which an invention builds on previous technological knowledge (Jaffe and De Rassenfosse, 2017). Previous discussions of AI as a GPT have also used both patent data (Cockburn et al., 2018); it should clearly be noted that various alternatives have been used including industry-, firm-, and technology-level non-patent quantitative and qualitative data (Klinger et al., 2018; Bresnahan, 2023; Trajtenberg, 2018; Goldfarb et al., 2023).

3. The Four Different Approaches to AI Patents

In our analysis, we sample four sets of AI patents identified by applying distinct AI definitions to patent data, and compare the four sets by GPT and other characteristics. More specifically, for GPT characteristics we contrast the estimated growth, generality and complementarity of AI patents classified by either (1) keywords, (2) science citations, (3) the WIPO, and the (4) USPTO method from all USPTO patents granted from 1990-2019. For analysis of the concentration of inventive activity, we use metrics for concentration (details below). The four different definitions differ in how they are computationally operationalised as follows.

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 its 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) conceptualise AI technologies as an outcome of science and academic research (Arthur, 2009; Jee and Sohn, 2023). As seen below, this approach is subject to limitations because citation practices differ across technological fields.

¹<https://worldwide.espacenet.com/patent/cpc-browser>

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 implemented 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 significant volume of AI patents identified through this method, including a large share of downstream AI applications.

For further background on these definitions and details on their implementation, we refer the reader to the Appendices A-B.

4. Results

Here we compare the four methods by their general characteristics (4.1), GPT characteristics (4.2), and concentration (4.3).

4.1. General characteristics of AI

Table 1 shows, for each method, growth rates (panel A), inventor types (panel B), main industries (panel C), countries of origin (panel D), public support (panel E), and main technology classes (panel F).

For all methods, AI inventions have become increasingly diversified across industries and countries over the past three decades, suggesting that AI has become increasingly disseminated throughout the economy. Further, all methods show that AI inventions are disproportionately borne from commercial enterprises (panel B) and associated with the computer and machinery manufacturing industries (panel D). Despite classifying numbers of patents that differ on the order of magnitude, all approaches show the expected dominance of the US in AI patents. However, our use of US patents over-represents inventors from this country.

We find differences in the growth rates of the Science and the USPTO approaches suggesting that these two groups have not grown as clearly as the other categories during the 2010s.

Table 1: AI patents by four classification approaches

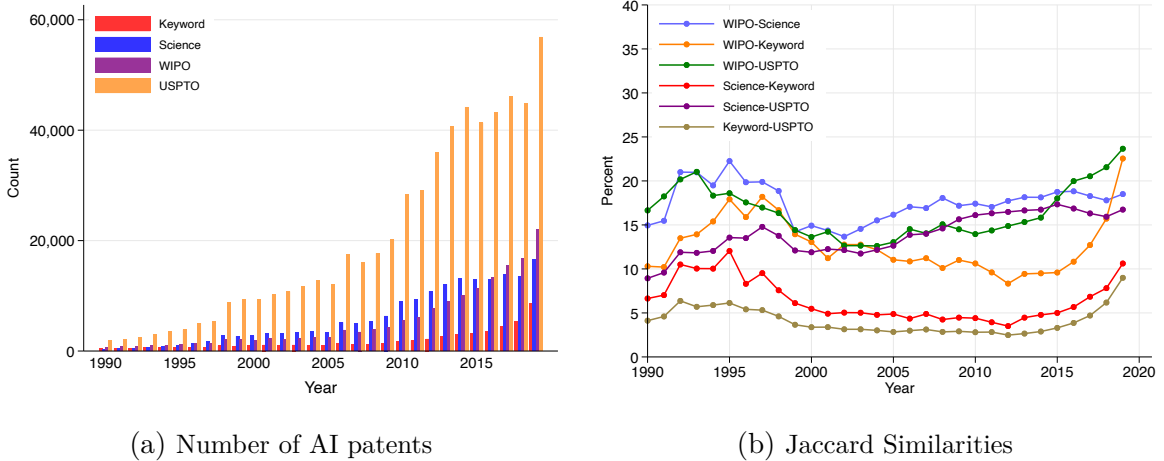
	Keyword	Science	WIPO	USPTO
<i>A. Growth</i>				
Growth rate (1990-99)	0.98	6.69	2.38	3.87
Growth rate (2000-09)	0.39	1.32	1.21	1.16
Growth rate (2010-19)	3.77	0.91	3.02	1.07
<i>B. Inventor type (patent assignee)</i>				
% Commercial	0.86	0.85	0.90	0.91
% Individual	0.06	0.03	0.05	0.05
% University	0.05	0.10	0.04	0.03
% Other non-profit	0.03	0.04	0.02	0.02
<i>C. Industry affiliation (patent assignee)</i>				
% Pharmaceuticals (manufacturing)	0.01	0.17	0.01	0.01
% Computer (manufacturing)	0.68	0.76	0.79	0.84
% Machinery & equipment (manufacturing)	0.49	0.28	0.54	0.22
% Other manufacturing	0.14	0.10	0.09	0.09
% Computer programming (service)	0.06	0.07	0.09	0.12
<i>D. Country of origin (patent applicant)</i>				
% USA	0.62	0.75	0.65	0.72
% Japan	0.14	0.08	0.15	0.10
% S. Korea	0.04	0.02	0.03	0.02
% Germany	0.05	0.03	0.03	0.03
% China	0.01	0.01	0.02	0.01
% Canada	0.02	0.02	0.02	0.02
<i>E. Public support</i>				
% Public support (1990-99)	0.31	0.39	0.29	0.21
% Public support (2000-09)	0.34	0.48	0.31	0.23
% Public support (2010-17)	0.33	0.52	0.28	0.22
<i>F. CPC 1-digit codes</i>				
% Human necessities (A)	0.15	0.18	0.08	0.08
% Performing operations (B)	0.31	0.05	0.11	0.04
% Chemistry; Metallurgy (C)	0.03	0.16	0.01	0.01
% Physics (G)	0.66	0.75	0.95	0.79
% Electricity (H)	0.25	0.26	0.27	0.33
% General/Cross-sectional (Y)	0.05	0.04	0.03	0.04
Number of patents	54,145	178,004	158,652	595,047

Notes: This table compares the scale and scope of AI invention identified by each definition, concerning inventor types (commercial, individual, non-profit, university), industry affiliation (based on NACE Rev. 2 classification), country of origin, reliance on public R&D support, and technological classification. Note that the data on public support ends in 2017.

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. The high concentration of WIPO in ‘Physics’ is caused by the data construction method: the WIPO method filters by design for a high share of patents classified into CPC-section ‘G - Physics’ (WIPO, 2019a).

Figure 1 shows the pace of AI invention. 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 D.10 in Appendix D.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)



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 between the different AI definitions (the degree to which they extract the same sets of patents or not), we compute the pairwise Jaccard similarities and their evolution over time, shown in Figure 1b.² All pairs of AI samples show low overlaps, ranging at or below 20%. The WIPO approach shows the highest agreement with the other groups, and the similarity of the WIPO-Keyword pair and the Science-USPTO increased most strongly over time. Keywords demonstrate the lowest Jaccard similarity to the other classification approaches, which may be due to the small number of patents in this sample.

Overall, only 10,062 or 1.37% of all unique granted patents are uniformly identified as AI patents by all four methods, indicating that quantitative conclusions about the scale and reach of AI are very sensitive to the chosen classification method. In the next section, we adopt a more qualitative perspective, and discuss how these differences impact on the extent to which AI qualifies as a GPT.

²The 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

We study the GPTness of AI by comparing the four AI samples by their growth, generality, and complementarity, summarised in Table 2.

Table 2: Measure of GPT characteristics of AI

	Keyword	Science	WIPO	USPTO
<i>A. Growth</i>				
Avg. growth rate	0.12	0.15	0.14	0.13
<i>B. Generality</i>				
Avg. generality index (1 digit)	0.81	0.77	0.76	0.73
Avg. generality index (3 digit)	0.94	0.90	0.90	0.87
Avg. generality index (4 digit)	0.97	0.96	0.95	0.95
<i>C. Complementarity</i>				
Avg. number of CPC (1 digit)	1.43	1.40	1.36	1.27
Avg. number of CPC (3 digit)	1.67	1.64	1.64	1.43
Avg. number of CPC (4 digit)	1.92	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

Panel A of Table 2 shows the average growth rate between two consecutive years, and Figure 2 shows how these rates evolved over time. We include lowess (local regression) smoother fits (Cleveland, 1979) in each plot to show the overall pattern.

We observe the following: first, each sample shows positive growth rates in most years, as anticipated for GPTs. Second, each AI definition demonstrates a dip in growth during the early 2000s before taking off again in recent years.³ Third, the smaller AI samples produced by the Keyword and WIPO methods 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.

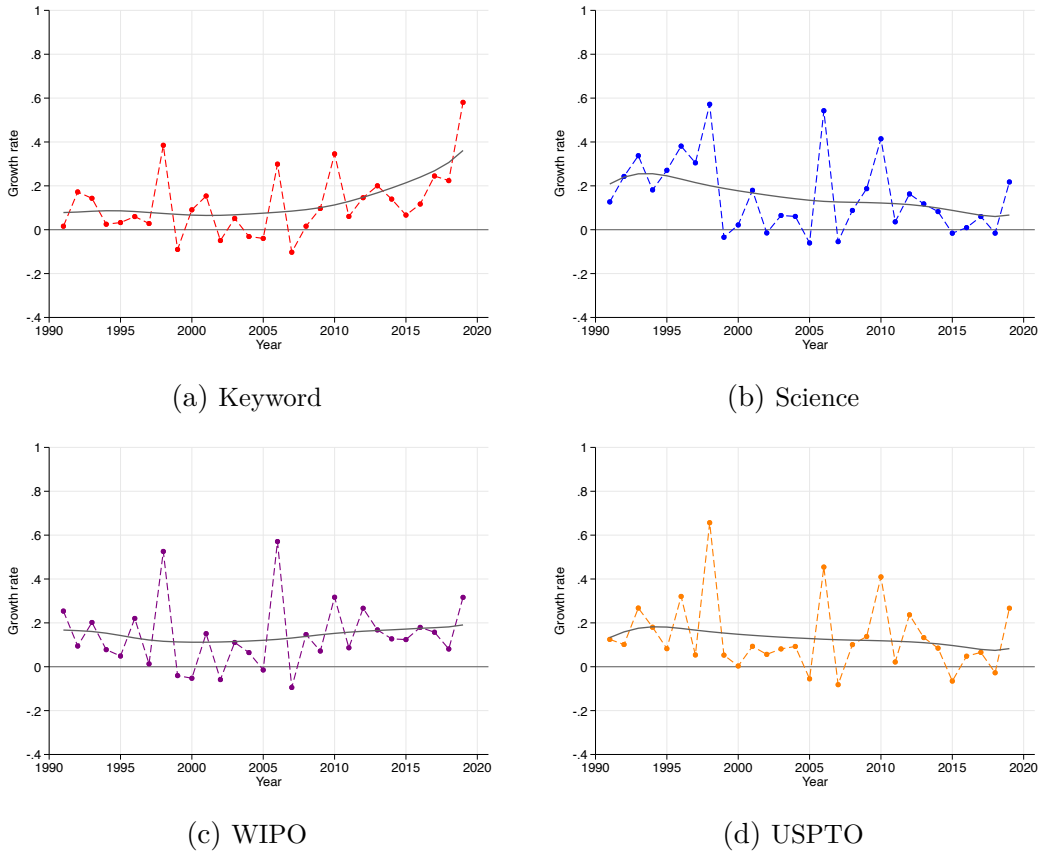
In the Appendix (Figures D.4), we study benchmarks that have been discussed in the literature as GPTs. We find that the AI samples show higher growth rates compared to most other GPT-candidates, but performing a Wilcoxon signed-rank test, we find that most of these differences are not significant (see Appendix D.2.1), although they are significantly higher than those of average patents.⁴ Across our AI samples, the

³However, a comparison with benchmark groups shows that patenting decreased across many sectors during this period (see Figure D.4 in Appendix D.1.1).

⁴Note 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.

differences between the four classification methods are not significant, except for the USPTO approach, which shows significantly lower growth rates compared to the others.

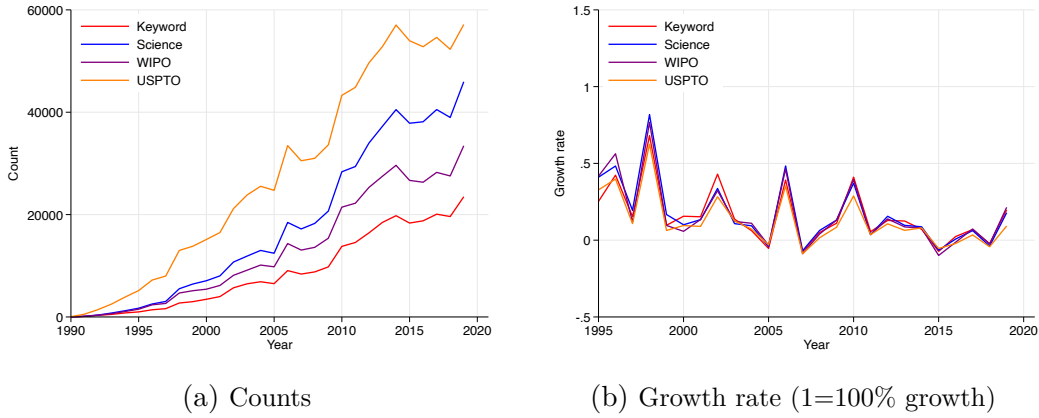
Figure 2: Growth of Patents by Year



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

If AI was a GPT, it should be an enabler of follow-up inventions in non-AI sectors, as reflected in the growth of patent citations by non-AI patents. 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 is consistent with the counts and shares shown above (Figure 2). The positive downstream growth suggests that each approach generates a growing number of invention spillovers to non-AI sectors. Significance tests in D.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 and taking off thereafter. These three approaches score significantly higher than the USPTO sample. Figure 3b suggests that all four methods capture different portions of a larger group of similar technologies.

Figure 3: Patents Citing AI



Notes: Panel (a) shows the 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; describing the extent to which AI inventions are cited in diverse technology fields.

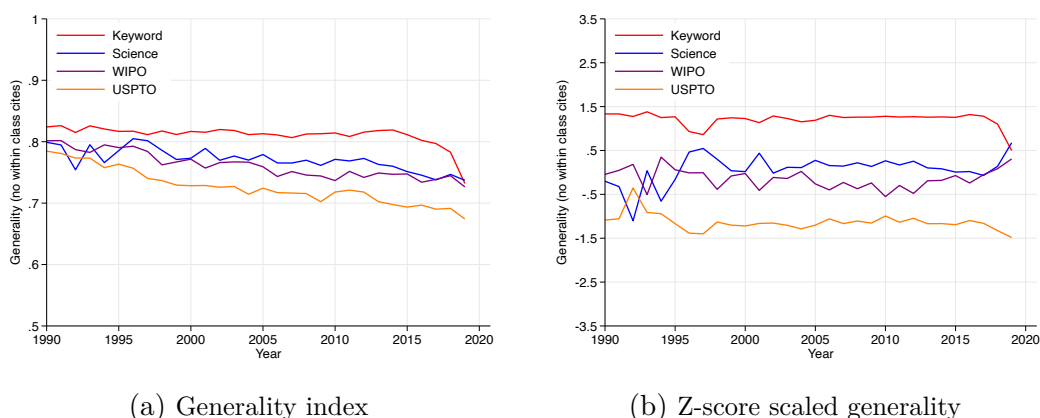
First, we assess the generality index (Trajtenberg et al., 1997; Hall and Trajtenberg, 2006), given by the inverse of the concentration of the 1-digit CPC-codes of AI-citing patents, similar to a Hirschmann-Herfindahl Index (HHI).⁵ Technical detail and additional results are provided in Appendices A.2 and D.3.2

⁵The formula is given by

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

where $\#cites_{ij}$ is the number of citations to patents labeled as AI by method i from CPC class j , using CPC codes at the 1-digit, 3-digit, and 4-digit level; $\#cites_{ij}$ excludes citations within the same class. N_j is the number of different CPC classes.

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

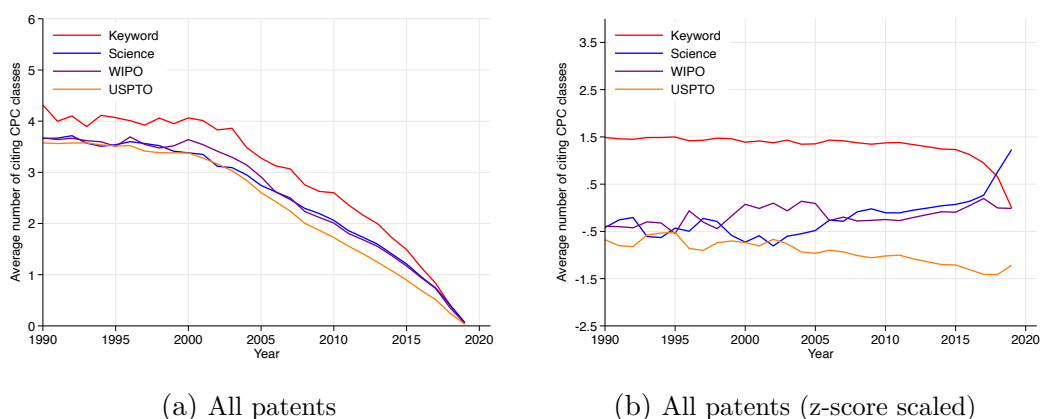


Notes: The z-scored value equals the level of the generality index minus its average across the four approaches divided by the standard deviation for each year.

Figure 4 shows the evolution of the generality index over time. The Keyword sample shows the highest level of generality, which is consistent over the three decades (Figure 4b). Hence, citations from other technology fields to Keyword AI patents are most equally distributed across different fields. The WIPO and Science samples score at similar levels, with a slightly higher generality score given to Science. The USPTO sample shows the lowest generality across all CPC levels. Towards the end of the time period, the decline in the generality of the Keyword and USPTO methods needs to be considered with caution, as the number of citations of recently granted patents is lower given the time needed for innovations to be cited.

Second, we assess generality by the mean annual average number of unique citing classes citing to an AI patent, accounting for the increase in annual patenting rates and avoiding an over-representation of the generality of patents in more recent years. Again, we compile this metric for different CPC digit levels.

Figure 5: Average number of classes citing AI



The z-scored value equals the level of the generality index calculated at the 1-digit level minus its average across the four approaches divided by the standard deviation for each year.

At all levels of aggregation, the Keyword patents are the most general (Table 2 and Figure 5). At the 1-digit level, Science scores second, closely followed by WIPO. WIPO ranks higher at the 3- and 4-digit level. The USPTO patents demonstrate the lowest number of unique 1-, 3-, and 4-digit patent citations. Again, the decline in recent years can be explained by the time lags in receiving citations.

Significance tests shown in Appendix D.2.2 confirm that the highest generality scores of the Keyword and the lowest of the USPTO method are statistically significant. The differences between the science and WIPO approaches are negligible at the 1-digit level. In Appendix D.3.2, we also report the number of unique citing technology classes for AI patents that are cited at least once (see Table D.39). Interpreting backward citations as an indicator of the value of a patent (Kogan et al., 2017), this measure focuses only on high-value patents. Again, the Keyword patents show the highest generality across all CPC digit levels. 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 again the lowest generality.

As before, we benchmark our AI samples against all patents and other GPT candidates discussed in the literature. Table D.2 shows that the generality index of all patents is higher than our AI samples, which is a natural feature, confirming the usefulness of the generality index in capturing the widespread use of patents. Among other groups, biochemistry/genetic engineering, nanotechnology, and climate inventions related patents have high generality scores. Figure D.5b shows that the time-series pattern of the generality score is quite stable, except 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 D.3). Biochemistry/genetic engineering and climate-related patents show high patent-level generality. These numbers show that our generality measures are appropriate, at least for some known GPT candidates.

Again, we provide additional results on the generality of AI-citing patents that are not AI themselves. These are qualitatively consistent, although the differences between the AI samples are smaller (see Appendix D.4).

Table 3: Average Citation Lags by Approach

Period	Keyword	Science	WIPO	USPTO
all periods	10.16	8.90	9.63	9.80
1990-1999	14.17	13.26	13.77	13.64
2000-2009	9.92	9.08	9.38	9.34
2010-2019	4.33	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 in recent years, because the data ends in 2019 causing a truncation of the maximal time lag.

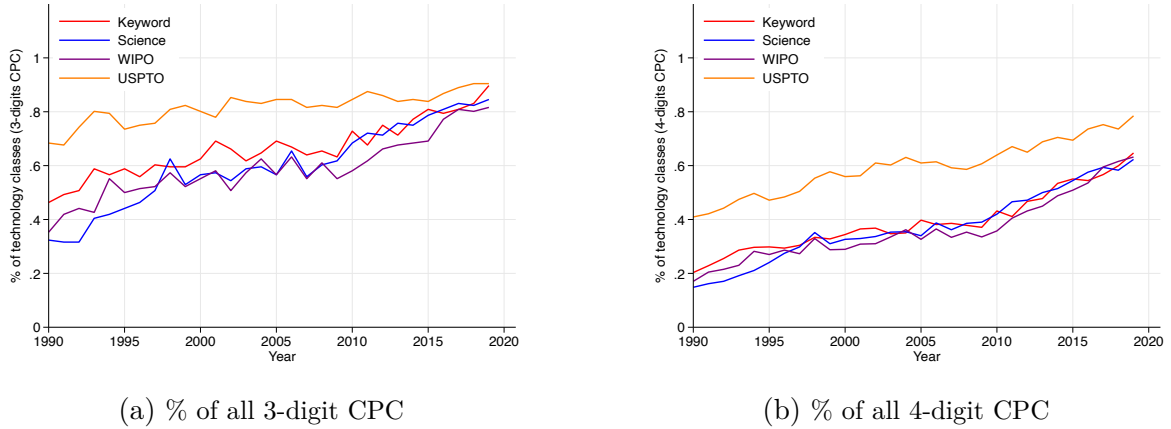
Patent citations to GPTs are said to occur with long lags (Hall and Trajtenberg, 2006), as organisational and complementary inventions are required during an early phase of ‘learning and destruction’ before a GPT can spread through the economy (Crafts, 2021; Bresnahan, 2023). In Table 3, we show the average citation lags of our AI samples, given by the average number of years between the grant year of a patent and the patents citing the patent. Keyword patents again rank top, by showing the longest average lags.

Altogether, the AI sample identified by the Keywords shows consistently the highest level of generality across different metrics.

4.2.3. Complementarity

As a measure of complementarity, we examine the co-classification of AI by multiple CPC codes. If AI complements a wide range of other technologies, AI patents would be co-classified across diverse technological fields.

Figure 6: Diversity of AI – Share of technology classes



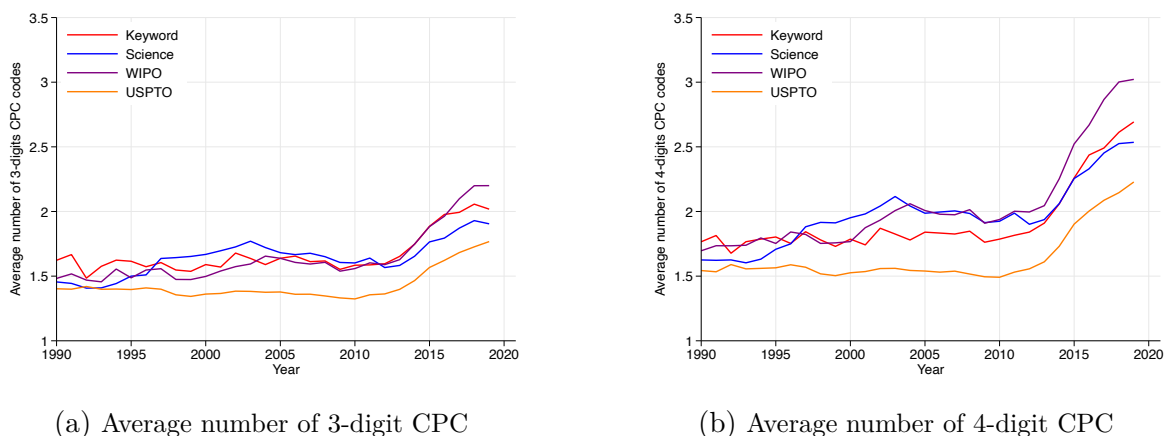
Note: Panel (a) shows the percentage of 3-digit CPC codes and panel (b) shows the percentage of 4-digit CPC as a share of all codes in the respective category. Note that the total number of 3-digit and 4-digit CPC codes are 136 and 674, respectively according to the February 2022 version.

Figure 6 shows the percentage of 3- and 4-digit CPC classes associated with each set of AI patents. USPTO patents span the most diverse pool of technology classes, with CPC classes ranging from 70-90% of all possible codes. This can be explained by the large number of AI patents in the USPTO approach, compared to the others (Table 1).

However, the smaller AI samples took off over time: starting around 2010, the share of codes associated with Keyword, Science, and WIPO patents rapidly increased. Significance tests show that the differences between the triple Keyword-Science-WIPO are statistically insignificant, but all score significantly lower than USPTO patents (Table D.29 and D.31).

To take account of the differences in sample sizes, we calculate the annual averages of the number of 1-, 3- and 4-digit CPC codes per patent, shown in Table 2, D.42 and D.3.2. An average WIPO patent is associated with 1.64 3-digit classes and 2.05 4-digit classes

Figure 7: Patent-level diversity - Average technology classes



across all years. Keyword or Science patents are on average 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, Keyword and Science AI patents show similar values (1.39-1.40) and cannot be distinguished statistically. The other approaches rank significantly lower, with the USPTO approach showing the least complementarity. In Figure 7, we show the average number of co-classifications over time at the 3- and 4-digit level. All panels show a rising diversity towards the second half of the last decade, with the strongest increase for WIPO patents.

In Appendix D.1.1, we provide results for the benchmark GPT candidates, confirming that our measures are performing equally well in capturing the increasing diversity of other GPTs suggested in the literature.⁶

Summing up, all AI samples show an increasing technological diversity. Accounting for differences in sample size, we find that the WIPO method captures the highest diversity at the more disaggregate level, while Science and Keyword AI patents are more diverse at the 1-digit section level. The high diversity of WIPO patents at more granular levels may be explained by the method’s design which focuses on patents classified into various definitions of the computing and hardware components of AI (see Section B.4). In contrast, Keywords and Science patents are not constrained by the CPC classes included.

⁶Given the heterogeneity of CPC classes, it is difficult to evaluate the compare the different candidates by their diversity as our selection of benchmarks is entirely based on CPC codes at the 3- and 4-digit level. For example, it is almost a natural feature of the Y02-codes used to identify climate technology span a wide range of technologies.

4.3. Concentration in AI

Many policy-related discussions on AI center around the concerns about an increasing concentration in the market for AI technologies, undermining fair competition and equal participation in the benefits of AI. Here, we also investigate the empirical basis of these concerns for different conceptions of AI, as they are embodied in our four samples of AI inventions. The Science approach captures the most non-commercial AI inventions by 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. This may be explained by the high share of biotech-related patents in the Science sample, which are more often filed by academic inventors whose inventions originate from their research. Geographically (Panel E), the Keyword and WIPO patents include more AI patents of foreign inventors, especially from Japan. Chinese inventors are not present in the top-list, despite China’s vital role in AI development in recent years (Jacobides et al., 2021). This can be partially explained by the time period covered, as China’s position at the global frontier in high-quality patenting only became apparent during the preceding decade. However, this omission may also be caused by a potential bias of relying only on US patents in this study.⁷

Moreover, each of the four classification methods draws a different picture of the reliance of AI inventions on public R&D support (Table see 1 Panel F). In the USPTO sample, only one-fifth of AI patents received public support, while in the Science sample this jumps to half in the last decade. This resonates with the inventor types (Table 1 Panel C): AI as classified by the USPTO appears to be mostly commercially driven, while academic AI inventions take a higher share in the Science sample, as public R&D support is often channelled through academic research institutions.

In Table 4, we show the top-ten firms ranked by their share in AI inventions for each of our AI samples. Consistently, we find that AI inventions are dominated by a few key technology and communication companies. We also uncover a broad consistency among the firms, with IBM, Microsoft, and Google occupying the top ranks.

To analyse the full distribution beyond the top-ten firms, we calculate the concentration ratio (CR) and HHI, shown in Table 5. The four-firm (eight-firm) CR is given by the share of patents filed by the top-four (top-eight) firms inventing AI, and the HHI is given by the sum of squares of the share of AI patents produced by each firm. Consistently, across measures of concentration, inventions in AI identified by Keywords are the least concentrated with the lowest HHI and CRs.

Contrary to concerns raised in the literature (Petit, 2017), we do not observe that AI inventions have become more concentrated in fewer firms over the past three decades. We observe minor fluctuations, but cannot observe any clear trend for any of the four AI samples. Note that our data only reflects trends in the distribution of patented AI inventions across firms: market concentration in AI can also refer to shares in the market for goods, and control over data, computational resources, and platforms (Jacobides et al., 2021).

⁷A discussion of the implications of this bias can be found below (Section 6).

Table 4: Top AI-producing firms

Keyword		Science		WIPO		USPTO	
Company name	%	Company name	%	Company name	%	Company name	%
IBM Corp	0.05	IBM Corp	0.07	IBM Corp	0.07	IBM Corp	0.09
Microsoft	0.02	Microsoft	0.06	Microsoft	0.04	Microsoft	0.04
Samsung	0.02	Google	0.03	Google	0.03	Google	0.02
Fanuc Corp	0.02	Apple	0.02	Canon	0.02	Intel	0.02
Google	0.02	Sony	0.01	Samsung	0.02	Samsung	0.01
Siemens	0.02	Siemens	0.01	Sony	0.02	Hewlett Packard	0.01
Honda Motor	0.01	Hewlett Packard	0.01	Intel	0.01	AT&T	0.01
Amazon	0.01	Intel	0.01	Amazon	0.01	Sony	0.01
Intel	0.01	AT&T	0.01	Siemens	0.01	Amazon	0.01
Sony	0.01	Canon	0.01	Fujitsu	0.01	Canon	0.01

Notes: This table reports the top-ten AI producing firms for each AI definition. 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 each AI definition. For example, IBM accounts for 4-7% of all AI patents produced by commercial firms.

Table 5: Concentration of Firms innovating in AI

	Keyword	Science	WIPO	USPTO
<i>A. Concentration Ratio (CR)</i>				
Four-firm CR	0.117	0.174	0.162	0.177
Eight-firm CR	0.175	0.219	0.221	0.228
<i>B. Herfindahl–Hirschman Index (HHI)</i>				
HHI (overall)	0.007	0.012	0.011	0.014
HHI (1990-1999)	0.008	0.012	0.013	0.014
HHI (2000-2009)	0.006	0.013	0.015	0.016
HHI (2010-2019)	0.009	0.013	0.011	0.014

Notes: This table shows the measures of concentration of AI-producing commercial firms. The concentration ratio (CR) measures the market share of top-four (or top-eight) firms. The Herfindahl–Hirschman Index (HHI) is calculated as the sum of squares of shares of patent produced by each firm within each of the four AI definition patent samples.

5. Discussion

Defining AI is a challenging task, as with every novel technology that has not yet unleashed its full impact and scope (Krafft et al., 2020; Schumpeter, 2005), which can be shaped by the socio-technical and regulatory environment (Geels, 2005). Shaping the future of AI has gained increasing attention in current discussions on AI governance, aimed at mitigating risks and maximising benefits for all (Mazzucato et al., 2022; Bresnahan, 2023). However, AI policy requires a clear identification of the area of AI where action may be required (Krafft et al., 2020).

Our analysis of the GPTness and concentration contributes to two dimensions of the AI governance: (1) policy could prioritise AI inventions that are most beneficial for broader society with the greatest public benefits (Mazzucato et al., 2022; Bresnahan, 2023), which may be indicated by high levels of GPTness (Lipsey et al., 2005). (2) AI policy could aim to avoid risks of power concentration when the core resources to develop and use AI are unequally distributed. Preventing concentration also reduces the risk that AI development pathways are prematurely narrowed in those areas that promise the greatest short-term private benefits, whilst other areas remain underdeveloped (Klinger et al., 2022; Jacobides et al., 2021; Bresnahan, 2023), although they would be beneficial for a more diverse range of actors.

5.1. AI as a General Purpose Technology

Here, we studied four different approaches to identifying patented AI inventions. Our results show a qualitative consistency across the methods: all classification techniques show low levels of AI inventions during the 1990s (also known as “AI winter”) followed by a period of increased patenting since the 2010s (Stuart and Norvig, 2003; Klinger et al., 2022); all confirm that AI can be viewed as a GPT; all associate AI with similar industries and core technology fields; and all show patterns of concentration amongst similar groups of key firms being the drivers between AI inventions. Several characteristics are therefore similar: this may inform innovation scholars and regulators in search of a robust definition of AI, as a basis for impact evaluation and regulatory action (Krafft et al., 2020).

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 GPT characteristic over the last 10 years. Red (yellow) colour indicates the strongest (weakest) performance.

Quantitatively, our four classification methods show some disagreement: AI as identified by Keywords and WIPO took off over the past decade, while the other two suggest

a relative innovation slowdown. Both identify patents that are most GPT-like, and capture more from machinery and equipment manufacturing, which is also reflected in the key firms being identified as leading inventors of AI. The Keyword method captures only a very narrow set of inventions (54k), but these patents show the highest levels of GPTness (see Table 6) – and also lowest levels of concentration in inventive activity.

5.2. Concentration in AI patenting

Concentration and GPTness interact, as GPTs have public good characteristics, with high innovation spillovers and lagged private returns for diverse actors, making the short-term profit maximisation based on these technologies more unlikely (Bresnahan and Trajtenberg, 1995). This may be reflected in lower levels of the concentration of AI inventions that are most GPT-like. In our results, the low concentration in AI as captured by Keywords resonates with its high ranking by GPTness (Section 4.2), supporting the claim that AI inventions as captured by Keywords are most GPT-like. The Keyword method identifies many patents centred around machine learning, which was also by Goldfarb et al. (2023) identified as a technology with a high GPT potential.

The realisation of GPT benefits for all does not necessarily come by itself: commercial actors may maximise the private instead of the social value of an invention, leading to a pre-mature lock-in in an inferior pathway of AI development (Klinger et al., 2022; Mazzucato et al., 2022; Bresnahan, 2023). Our identification of core inventions and technological areas, where AI exhibits strongest GPTness and lowest levels of concentration, as identified by Keywords, may be indicative of promising areas of AI where public support can be eventually justified. However, any such decision should rely on methodologically pluralist insights, as every method including ours has limitations (see Section 6).

Our analysis has consistently demonstrated, across all methods, that concentration in AI inventions did not increase over the past three decades. This may suggest that barriers to entry to the market of AI innovation did not recently become higher, and contrasts with other research raising this to be a major concern (Tambe et al., 2020; Agrawal et al., 2019; Mazzucato et al., 2022). This may be due to the nature of digital inventions, which are associated with a relatively low capital intensity and high levels of modularity (Bresnahan, 2023). The modularisation of the AI ecosystem facilitates the entry of novel players (Jacobides et al., 2021; Bresnahan, 2023) as it allows computational resources and basic hardware to be externally sourced at affordable prices. However, our restriction to the level of AI patents does not allow us to deduce trends in vertical integration across all levels of AI systems, or control over network resources, both of which may be areas for competition policy (Petit, 2017; Ducuing, 2020).

5.3. Choosing methodology to measure AI patents

Our systematic comparison of the four classification methods has research implications for patent-based studies (Fujii and Managi, 2018) as we show that empirical results may differ across methods. For example, the USPTO and Science methods suggest a

slowdown in AI inventions in recent years, which is not found when using Keywords or 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, 2023). We also find minor differences in the top-ranking of firms being among leaders in AI, and their nationalities. For example, the Science method qualifies Apple as a top-ten inventor of AI, but it excludes Samsung, which is among the top-five for all other methods. Such differences can have material impacts, as investors and banks base their funding decisions and loan conditions on prospective economic performance indicators, which may include patents in emergent technologies such as AI.

Our comparison may guide patent-based AI research on the choice of method, which may be dependent on the intended research goals. Our results suggest that the impact of *AI as a GPT* can be best studied by relying on the Keyword or WIPO methods. Other methods may be more appropriate for studying the wide-spread diffusion of AI, which seem to be well captured in the USPTO sample, or the knowledge origins of AI encoded in Science citations.

Measuring AI remains a challenge, but our comparison can inform discussions on AI definitions by showing consensual features of AI and quantifying nuanced differences in the types of industries, technology fields, and key players being involved in its development. Often, definitions of technologies are determined ex-post, after research, development, and diffusion have been accomplished (Schumpeter, 2005). Here, we quantify the consensus and differences in the nuances of AI that is developed by commercial inventors filing patents. We can only speculate, but the consensus can be indicative of the type of AI is most likely to be pursued further.

The simple Keyword approach generating a narrow set of key patents appears desirable for researchers aiming to catch emerging GPTs. Small patent groups may indicate a clearer distinction from other patents (Kovács et al., 2021). There is also a greater potential for future growth, compared to the USPTO approach that suggests that 16.6% of all US patents today are already based on AI.

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. However, we cannot rule out that the high levels of GPTness of the Keyword method reflect only a short term trend of the uptake of machine learning in commercially valuable AI inventions, as captured by patents, while the full scope of AI remains unleashed (Klinger et al., 2022).

6. Limitations

Patent data comes with various well-documented limitations and biases that may have an impact on the interpretation of our measures of AI, GPTs, and their implications for AI policy.

First, patents are used heterogeneously across industries, firms, and technologies, and some inventions cannot be protected by patents. Other means of IP protection such as trade secrets, trademarks, copyrights, and lead time may dominate in particular sectors, causing a bias (Granstrand, 2009; Jung et al., 2023; Verendel, 2023; Hötte and Jee, 2022). Further, the number of co-classifications, and patent and science citations differs across technology fields (Jaffe and De Rassenfosse, 2017; Hötte et al., 2021), which may affect our measures of GPTness.

Second, patents measure technical inventions, but not diffusion and the emergence of AI-related services (Goldfarb et al., 2023). Data and various other relevant parts of AI ecosystems (Jacobides et al., 2021) are not captured, as there are no patents in these areas. Diffusion may be reflected in changing production technology and input requirements of firms (Bresnahan, 2023). However, our results are consistent with (Goldfarb et al., 2023) who use data on skill requirements in jobs and confirm that machine learning is likely to be a GPT.

Third, we are restricted to patents filed at the US patent office. While US patents can serve as a good proxy of the global technological frontier, this may not be universally true for AI. AI specialisation strategies are heterogeneous across countries, including China (Fujii and Managi, 2018; Jacobides et al., 2021). Foreign, and especially Chinese, inventors are not generally excluded from filing patents in the US, but they may face barriers and/or have reduced incentives compared to filing domestic patents. Also, examination practices at patent offices in different jurisdictions are heterogeneous, which can affect patent citations or co-classifications (Jaffe and De Rassenfosse, 2017).

Fourth, the performance of the different approaches to capture the GPTness of AI may be specific for the time period (1990-2019), while the future of AI may be influenced by the decisions made today (Bresnahan, 2023; Goldfarb et al., 2023). Human-selected inputs like keywords, technology codes, or definitions of AI-science may also be chosen differently in the future (Schumpeter, 2005). The high GPTness of the Keyword and WIPO approaches could reflect the recent popularity of a narrow set of technology buzzwords (especially machine learning), but our results may differ when applied to other time periods in the future. This limitation can be related to the discussion on “AI narrowing” Klinger et al., 2022, suggesting that AI research became significantly less diverse in recent time, concentrated in a small group of private firms and focused on a narrow set of deep learning techniques.

Our analysis is restricted to GPTness and concentration as evaluation criteria relevant for AI policy, while being silent about other objectives such as safety, societal impact, and privacy (Roche et al., 2023; Krafft et al., 2020). Attempts to empirically assess the impact of policy on the evolution of AI should rely on a variety of methods for identifying AI and evaluation criteria, in line with the goals set by policy.

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.

7. Conclusions

We performed a systematic analysis of four separate approaches to identifying patented AI inventions that are distinct and only partially overlapping. We demonstrate how sensitive each key GPT-like feature of AI is to the classification method selected. Furthermore, we also investigate levels of market concentration and firm-level ownership of AI innovations.

Our results provide overall guidance to policy-makers and innovation scholars on how to identify patented AI inventions. For example, researchers interested in studying the GPT-like characteristics of AI, a simple keyword-based approach produces a narrowly defined set of patented technologies that demonstrates the canonical GPT features of intrinsic growth, wide usefulness, and complementarity.

In our analysis of AI patent ownership, we find consistent support for the need to go beyond thinking of AI as a GPT (Jacobides et al., 2021) and consider its agency dimension. Some striking results from our analysis of the four classification techniques relate to similarly high levels of market concentration, which may prove relevant for competition policy and market regulation (Petit, 2017; Hennemann, 2020).

Our work provides both robust findings of AI as a GPT using multiple classification strategies, and important caveats to those making policy and funding decisions for the future of AI inventions, including the recently released UK National Artificial Intelligence Strategy (Artificial Intelligence, 2021). Altogether, our results illustrate the utility of tracking AI inventions using patent data, and underscore the need for using multiple AI classification techniques to counteract the dependence of significant policy conclusions on the choice of classification approach.

References

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.48550/arXiv.2009.01575).
- Bresnahan, Timothy F (2023). “What innovation paths for AI to become a GPT?” In: *Journal of Economics & Management Strategy*. DOI: [10.1111/jems.12524](https://doi.org/10.1111/jems.12524).
- 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](https://doi.org/10.1016/0304-4076(94)01598-T).
- 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](https://doi.org/10.3386/w24001).
- (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](https://doi.org/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](https://doi.org/10.3386/w24449).
- Cockburn, Iain M, Rebecca Henderson, and Scott Stern (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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1093/oxrep/grab012).
- Dafoe, Allan (2018). *AI governance: a research agenda*. Tech. rep. Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK. URL: <http://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf>.
- Ducuing, Charlotte (2020). “Data as infrastructure? A study of data sharing legal regimes”. In: *Competition and Regulation in Network Industries* 21.2, pp. 124–142. DOI: [10.1177/1783591719895390](https://doi.org/10.1177/1783591719895390).
- 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>.
- 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.
- Fujii, Hidemichi and Shunsuke Managi (2018). “Trends and priority shifts in artificial intelligence technology invention: A global patent analysis”. In: *Economic Analysis and Policy* 58, pp. 60–69. DOI: [10.1016/j.eap.2017.12.006](https://doi.org/10.1016/j.eap.2017.12.006).
- Geels, Frank W (2005). *Technological transitions and system innovations: a co-evolutionary and socio-technical analysis*. Edward Elgar Publishing.
- 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](https://doi.org/10.2139/ssrn.3866793).
- Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis (2023). “Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings”. In: *Research Policy* 52.1, p. 104653. DOI: [10.1016/j.respol.2022.104653](https://doi.org/10.1016/j.respol.2022.104653).
- Granstrand, Ove (2009). “Innovation and Intellectual Property Rights”. In: *The Oxford Handbook of Innovation*. Ed. by Jan Fagerber and David Mowery. DOI: [10.1093/oxfordhb/9780199286805.003.0010](https://doi.org/10.1093/oxfordhb/9780199286805.003.0010).
- 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](https://doi.org/10.4337/9781845427924.00022).
- Hennemann, Moritz (2020). “Artificial Intelligence and Competition Law”. In: *Regulating Artificial Intelligence*, pp. 361–388.

- Hötte, Kerstin and Su Jung Jee (2022). “Knowledge for a warmer world: A patent analysis of climate change adaptation technologies”. In: *Technological Forecasting and Social Change* 183, p. 121879. DOI: [10.1016/j.techfore.2022.121879](https://doi.org/10.1016/j.techfore.2022.121879).
- 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](https://doi.org/10.1016/j.rser.2020.110654).
- 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 Gaétan De Rassenfosse (2017). “Patent citation data in social science research: Overview and best practices”. In: *Journal of the Association for Information Science and Technology* 68.6, pp. 1360–1374. DOI: [10.1002/asi.23731](https://doi.org/10.1002/asi.23731).
- Jee, Su Jung and So Young Sohn (2023). “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](https://doi.org/10.1093/icc/dtad002).
- Jelinek, Thorsten, Wendell Wallach, and Danil Kerimi (2021). “Policy brief: the creation of a G20 coordinating committee for the governance of artificial intelligence”. In: *AI and Ethics* 1.2, pp. 141–150. DOI: [10.1007/s43681-020-00019-y](https://doi.org/10.1007/s43681-020-00019-y).
- Jovanovic, Boyan and Peter L Rousseau (2005). *General purpose technologies*. Working Paper. National Bureau of Economic Research. DOI: [10.3386/w11093](https://doi.org/10.3386/w11093).
- Jung, Su, Kerstin Hötte, Caoimhe Ring, and Robert Burrell (2023). *Intellectual Property Rights, Climate Technology Transfer and Innovation in Developing Countries*. INET Oxford Working Paper No. 2023-14. Institute for New Economic Thinking, University of Oxford. URL: <https://www.inet.ox.ac.uk/files/intellectual-property-rights-2023-14.pdf>.
- 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](https://doi.org/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](https://arxiv.org/abs/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](https://doi.org/10.1093/qje/qjw040).
- Kovács, Balázs, Gianluca Carnabuci, and Filippo Carlo Wezel (2021). “Categories, attention, and the impact of inventions”. In: *Strategic Management Journal* 42.5, pp. 992–1023. DOI: [10.1002/smj.3271](https://doi.org/10.1002/smj.3271).

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1002/smj.3145).
 - (2021). *Reliance on Science in Patenting (Version v30) [Data set]*. Zenodo. DOI: [10.5281/zenodo.4850763](https://doi.org/10.5281/zenodo.4850763).
- Mazzucato, Mariana, Marietje Schaake, Seb Krier, and Josh Entsminger (2022). *Governing artificial intelligence in the public interest*. Working Paper Series, IIPP WP 2020-12. UCL Institute for Innovation and Public Purpose. URL: <https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12>.
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.respol.2020.104013).
- Roche, Cathy, PJ Wall, and Dave Lewis (2023). “Ethics and diversity in artificial intelligence policies, strategies and initiatives”. In: *AI and Ethics* 3.4, pp. 1095–1115. DOI: [10.1007/s43681-022-00218-9](https://doi.org/10.1007/s43681-022-00218-9).
- Schmitt, Lewin (2022). “Mapping global AI governance: a nascent regime in a fragmented landscape”. In: *AI and Ethics* 2.2, pp. 303–314. DOI: [10.1007/s43681-021-00083-y](https://doi.org/10.1007/s43681-021-00083-y).
- 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](https://doi.org/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](https://doi.org/10.1145/2740908.2742839).
- Stuart, Russell and Peter Norvig (2003). *Artificial intelligence: a modern approach*. Upper Saddle River, NJ, USA: Prentice Hall.
- Tambe, Prasanna, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson (2020). *Digital capital and superstar firms*. NBER Working Paper 28285. National Bureau of Economic Research. DOI: [10.3386/w28285](https://doi.org/10.3386/w28285).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1080/10438599700000006).
- Turing, Alan (1950). “Computing machinery and intelligence”. In: *Mind* 59.236, p. 433.
- USPTO (2020). *Inventing AI: Tracing the diffusion of artificial intelligence with U.S. patents*. U.S. Patent and Trademark Office. URL: <https://www.uspto.gov/sites/default/files/documents/OCE-DH-AI.pdf>.
- 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](https://doi.org/10.2139/ssrn.3482150).
- Whittaker, Meredith (2021). “The steep cost of capture”. In: *Interactions* 28.6, pp. 50–55. DOI: [10.1145/3488666](https://doi.org/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 Appendix, Figure C.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 $i \in \{\text{keyword, science, WIPO, USPTO}\}$ at time t , indexed by year. We compute the growth rate as

$$\frac{N_{i,t} - N_{i,t-1}}{N_{i,t-1}}. \quad (\text{A.0})$$

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, 2017). 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

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

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_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.

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

where $\mathbf{1}(\#ncites_{p,j,t} \geq 0) = 1$ 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 $d \in \{1, 3, 4\}$ in the code. N_i is the number of patents in approach i . 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

$$\frac{\#CPCs_{i,d,t}}{N_d} \quad (\text{A.0})$$

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

$$\frac{1}{N_{i,t}} \sum_p \#CPCs_{p,i,d,t} \quad (\text{A.0})$$

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.⁸ 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., 2018.⁹ 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.

⁸<https://bulkdata.uspto.gov/data/patent/classification/cpc/>

⁹While 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.¹⁰

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

¹⁰Note 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 IndexTM. 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. Keywords in detail

C.1. Words used in the 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		

C.2. 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 count the unique number of matching documents and what percentages of the AI patents match according to this keyword. Figures C.1, C.2, and C.3 below illustrate both counts and shares.

Figure C.1: Symbolic keywords: in full texts

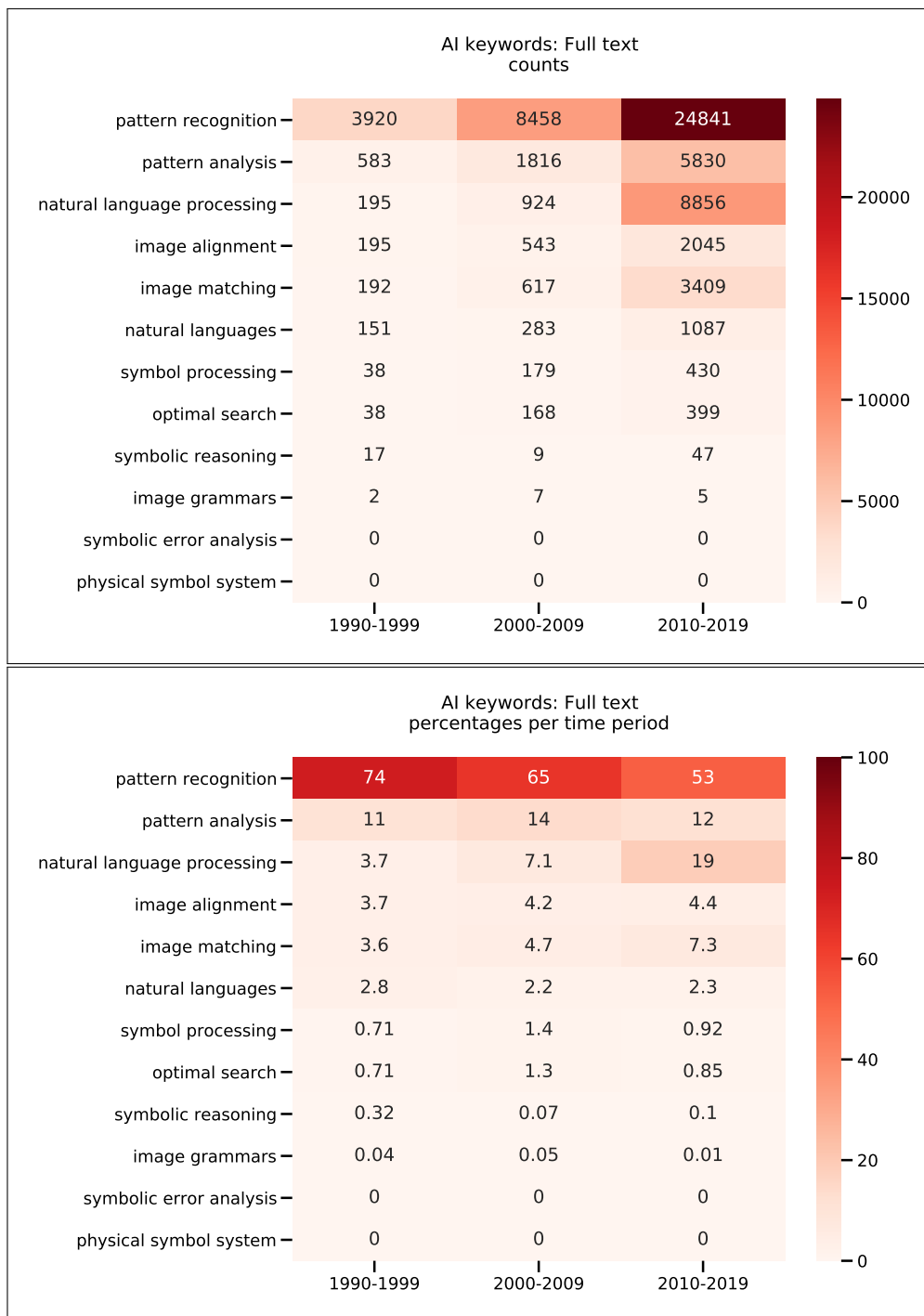


Figure C.2: Learning keywords: in full Texts

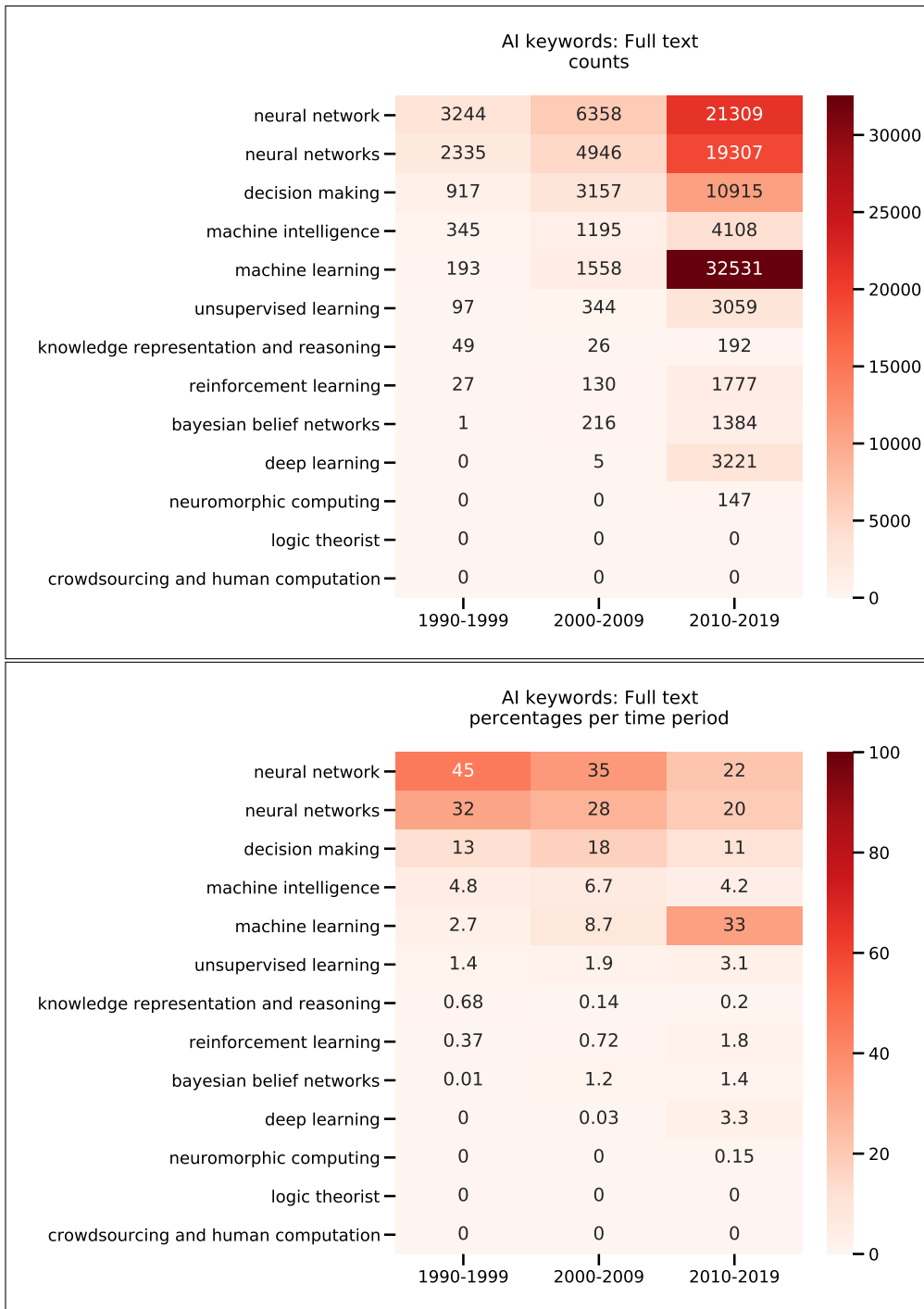
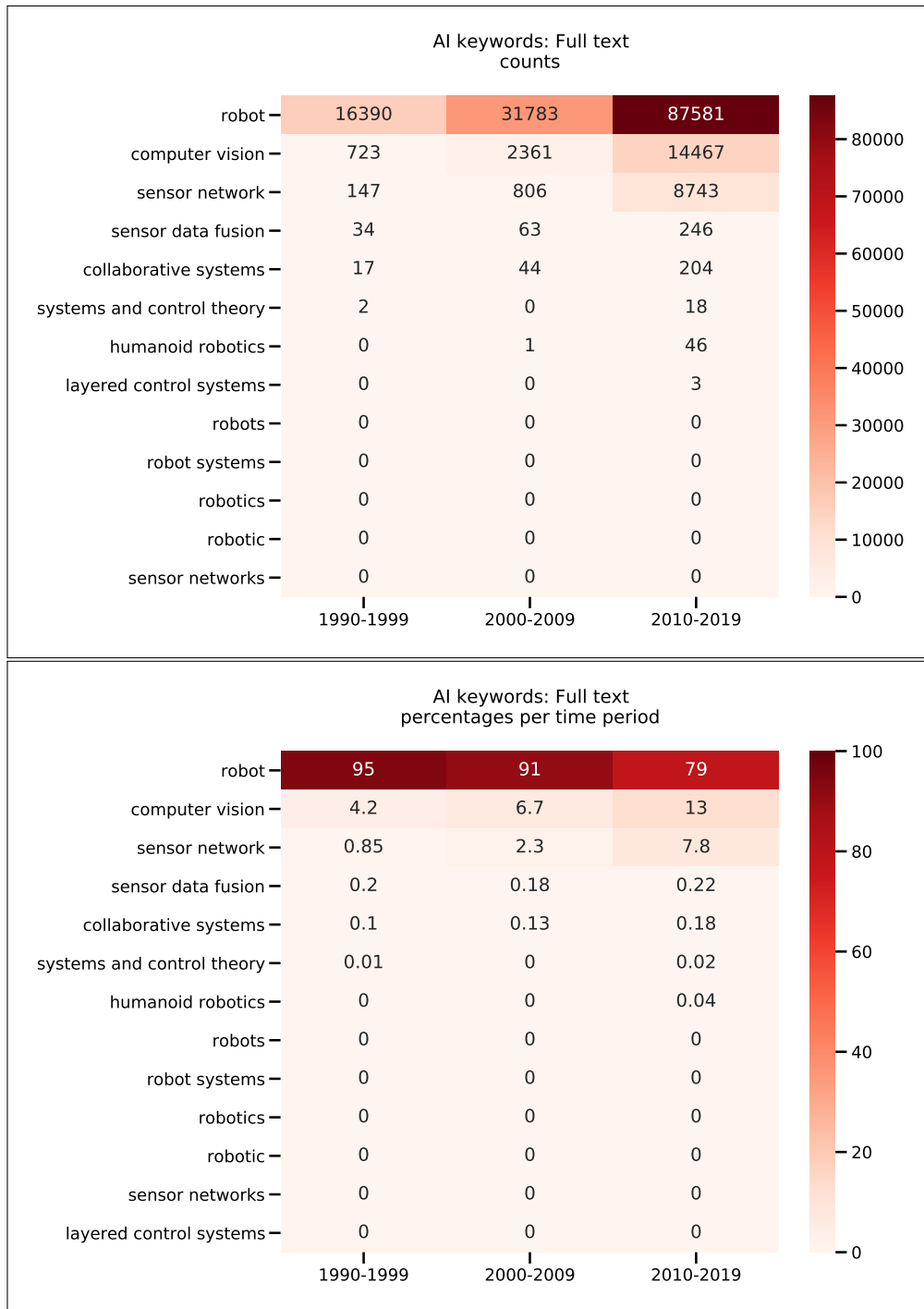


Figure C.3: Robotics keywords: in full Texts



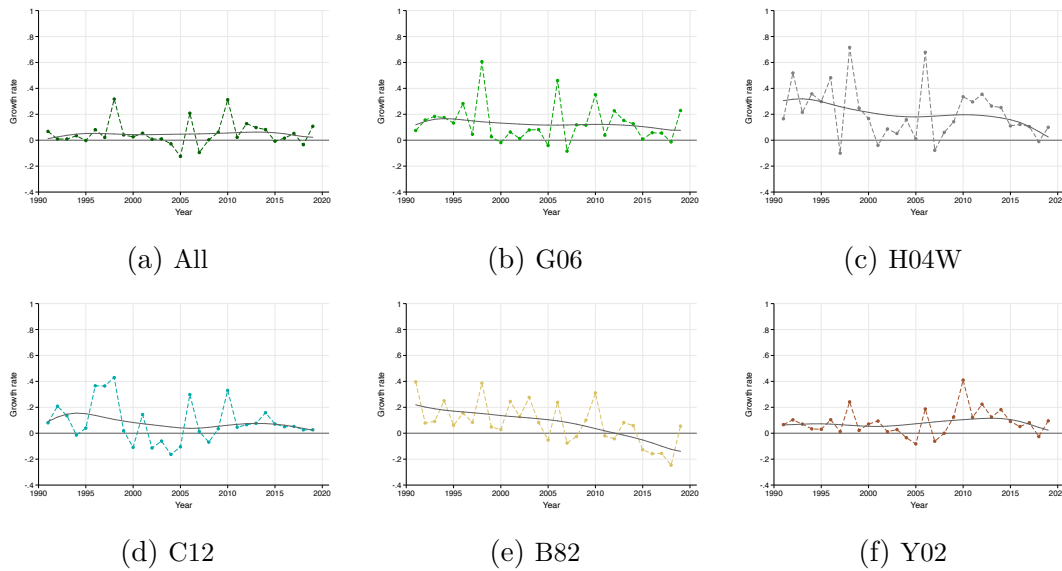
D. Additional results

D.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.

D.1.1. Growth

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



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.

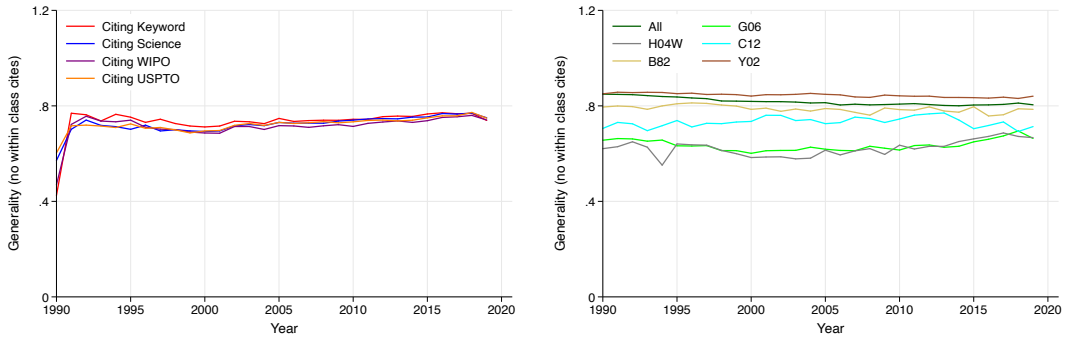
D.1.2. Generality

Table D.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 D.5: Generality Index at the 1-digit CPC-section Level



(a) AI citing patents

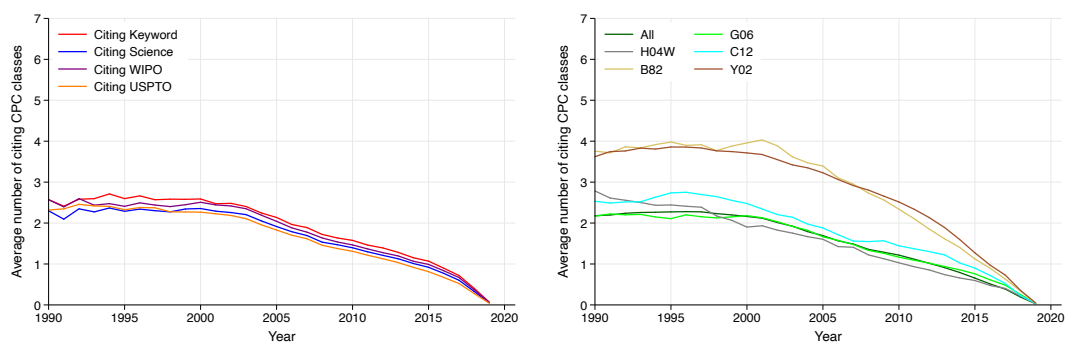
(b) Benchmark

Table D.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 D.6: Average Number of CPC Classes Citing AI



(a) AI citing patents

(b) Benchmark

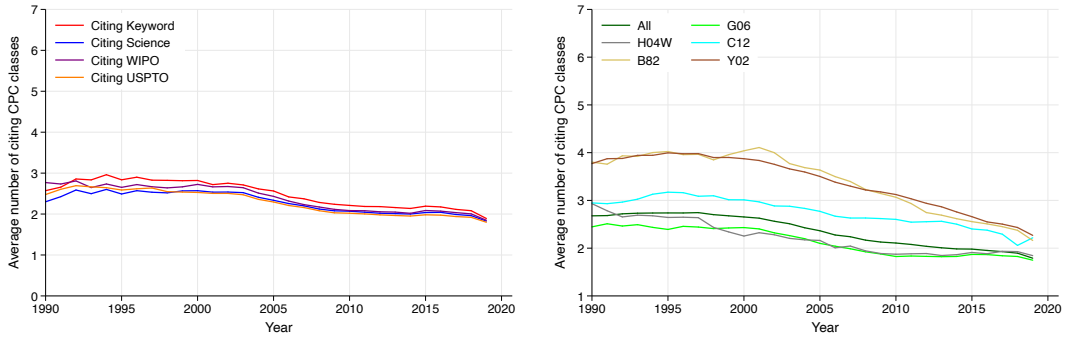
Note: ‘All’ refers to all patents, G06, H04W, B82, C12, and Y02 refers to computing, wireless communications, biochemistry/genetic engineering, nanotechnology, and climate invention-related patents, respectively.

Table D.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 D.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.

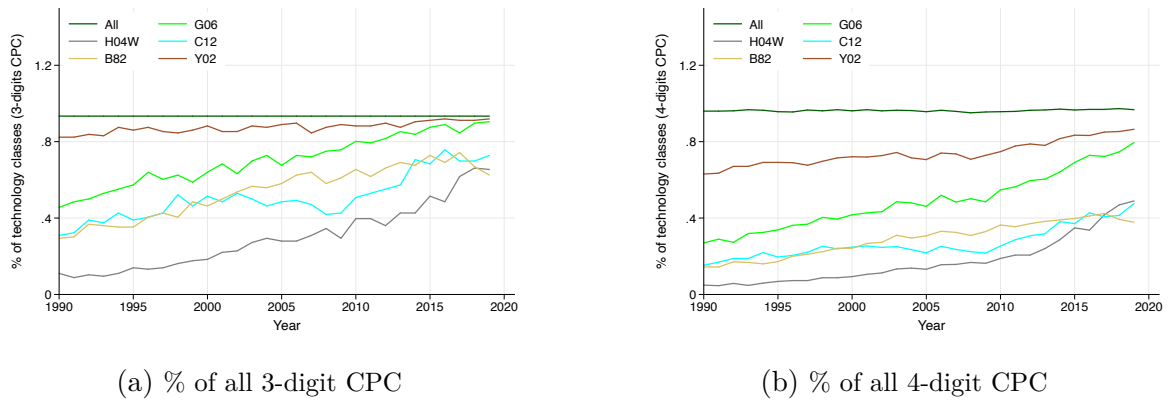
Table D.5: Average Citation Lags by Patents in Benchmark Categories

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.

D.1.3. Complementarity

Figure D.8: Share of technology classes: diversity of benchmark categories



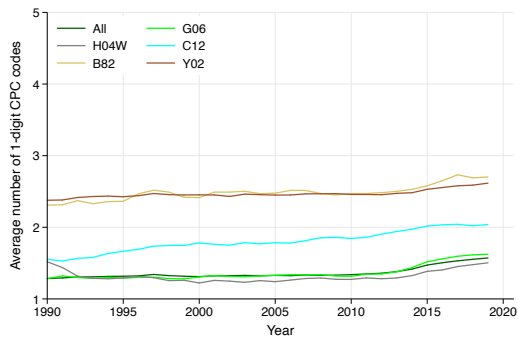
Note: Panel (a) shows the percentage of 3-digit CPC codes and panel (b) shows the percentage of 4-digit CPC as a share of all codes in the respective category. Note that the total number of 3-digit and 4-digit CPC codes are 136 and 674, respectively. ‘All’ refers to all patents, G06, H04W, B82, C12,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

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

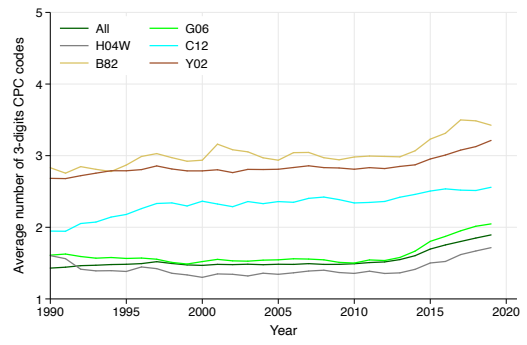
	All	G06	H04W	C12	B82	Y02
1 digit	1.36	1.36	1.32	1.80	2.48	2.47
3 digit	1.54	1.62	1.43	2.32	3.03	2.85
4 digit	1.80	1.81	2.26	2.93	3.50	3.39

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.

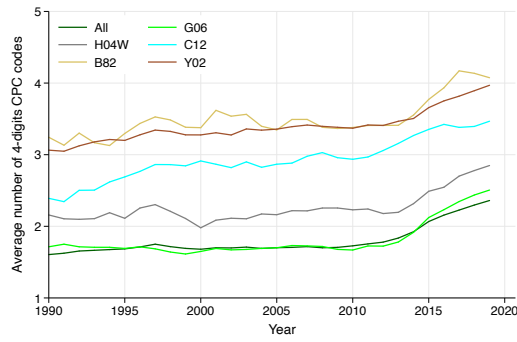
Figure D.9: Average Technology Classes: Patent-Level Diversity of Benchmark Categories



(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.

D.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, D.40, D.41, D.39, D.42 are significant.

D.2.1. Growth

Table D.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	0.02	0.00	0.00	0.00					
1990-2019	G06	1.00	1.00	1.00	1.00	0.00				
1990-2019	H04W	0.27	0.62	0.74	0.17	0.00	0.04			
1990-2019	C12	1.00	0.08	0.40	0.76	1.00	0.75	0.01		
1990-2019	B82	1.00	0.29	1.00	0.69	1.00	0.86	0.04	1.00	
1990-2019	Y02	1.00	0.60	0.32	0.67	0.03	0.86	0.00	1.00	1.00
1990-1999	Science	0.18								
1990-1999	WIPO	1.00	0.85							
1990-1999	USPTO	0.62	1.00	1.00						
1990-1999	All	1.00	0.45	1.00	0.18					
1990-1999	G06	0.32	1.00	1.00	1.00	0.45				
1990-1999	H04W	0.45	1.00	1.00	1.00	0.45	0.85			
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	0.32	1.00	0.32	1.00	0.18	0.45	1.00	0.45
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	0.71							
2010-2019	USPTO	1.00	1.00	0.85						
2010-2019	All	0.09	0.71	0.09	1.00					
2010-2019	G06	1.00	1.00	0.30	1.00	0.09				
2010-2019	H04W	1.00	1.00	1.00	1.00	0.13	1.00			
2010-2019	C12	0.55	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	0.85	0.71	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 D.8: Summary Statistics for Growth Rates

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

Table D.9: Growth Rates (Citing AI)

period	pair	Keyword	Science	WIPO	USPTO	G06	H04W	C12	B82
1990-2019	Science	1.00							
1990-2019	WIPO	1.00	1.00						
1990-2019	USPTO	0.02	0.00	0.01					
1990-2019	G06	0.14	0.00	0.00	1.00				
1990-2019	H04W	0.91	0.89	1.00	0.01	0.00			
1990-2019	C12	0.89	0.62	0.96	1.00	1.00	0.10		
1990-2019	B82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
1990-2019	Y02	0.07	0.12	0.96	1.00	1.00	0.05	1.00	1.00
1990-1999	Science	1.00							
1990-1999	WIPO	1.00	1.00						
1990-1999	USPTO	1.00	0.14	0.26					
1990-1999	G06	1.00	1.00	0.14	1.00				
1990-1999	H04W	1.00	1.00	1.00	1.00	0.14			
1990-1999	C12	1.00	1.00	1.00	1.00	1.00	0.88		
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	0.68	0.68	1.00					
2000-2009	G06	1.00	1.00	1.00	1.00				
2000-2009	H04W	1.00	1.00	1.00	1.00	0.90			
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	0.07	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	0.33	0.62	1.00					
2010-2019	G06	0.14	0.21	0.62	1.00				
2010-2019	H04W	1.00	1.00	1.00	1.00	0.45			
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 D.10: Summary Statistics for Growth Rates (Citing AI)

	Keyword	Science	WIPO	USPTO	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.95	0.75	1.74	0.53	0.96	1.29	0.70	1.67	0.70
Median 1990-2019	0.14	0.13	0.12	0.09	0.11	0.19	0.08	0.12	0.10
St.dev. 1990-2019	3.80	2.56	8.01	1.71	4.08	5.34	2.58	7.77	2.65
Mean 1990-1999	2.81	2.15	5.38	1.52	2.91	3.89	2.07	5.14	2.07
Median 1990-1999	0.57	0.58	0.68	0.54	0.44	0.73	0.43	0.39	0.40
St.dev. 1990-1999	6.69	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	0.12	0.10	0.12	0.09	0.09	0.12	0.04	0.13	0.09
St.dev. 2000-2009	0.17	0.16	0.16	0.13	0.13	0.19	0.22	0.17	0.15
Mean 2010-2019	0.10	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	0.13	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.

D.2.2. Generality

Table D.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	0.00							
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	0.00	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	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	1.00							
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	1.00	1.00	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	0.11	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	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.92	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.32
2010-2019	Science	0.09								
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	0.09				
2010-2019	H04W	0.09	0.09	0.09	1.00	0.09	0.12			
2010-2019	C12	0.67	0.09	0.09	0.09	0.67	0.09	0.09		
2010-2019	B82	1.00	0.09	0.09	0.09	1.00	0.09	0.09	0.09	
2010-2019	Y02	1.00	0.09	0.09	0.09	1.00	0.09	0.09	0.09	1.00

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

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.81	0.77	0.76	0.73	0.83	0.7	0.68	0.78	0.82	0.81
Median 1990-2019	0.81	0.77	0.76	0.72	0.83	0.7	0.68	0.79	0.82	0.81
St.dev. 1990-2019	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.01
Mean 1990-1999	0.82	0.79	0.79	0.76	0.84	0.73	0.68	0.78	0.83	0.81
Median 1990-1999	0.82	0.79	0.79	0.76	0.84	0.73	0.68	0.78	0.83	0.81
St.dev. 1990-1999	0.01	0.02	0.01	0.02	0	0.01	0.01	0	0	0
Mean 2000-2009	0.81	0.77	0.76	0.72	0.83	0.7	0.66	0.79	0.82	0.81
Median 2000-2009	0.81	0.77	0.76	0.72	0.83	0.7	0.66	0.79	0.82	0.81
St.dev. 2000-2009	0	0.01	0.01	0.01	0	0.01	0.01	0.01	0	0.01
Mean 2010-2019	0.8	0.76	0.74	0.7	0.81	0.68	0.7	0.78	0.81	0.81
Median 2010-2019	0.81	0.76	0.74	0.7	0.82	0.68	0.7	0.78	0.81	0.81
St.dev. 2010-2019	0.03	0.01	0.01	0.01	0.03	0.01	0.02	0.02	0.02	0.01

Table D.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	0.74							
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	0.22	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.05	0.00	0.00	0.01		
1990-2019	B82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.74	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	0.39							
1990-1999	USPTO	0.09	0.29	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	0.29			
1990-1999	C12	0.09	0.09	0.09	0.26	0.09	0.29	0.39		
1990-1999	B82	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.39	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	1.00							
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	1.00	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	0.09	0.09	0.09					
2010-2019	G06	0.09	0.09	0.09	1.00	0.09				
2010-2019	H04W	0.09	0.21	0.12	1.00	0.09	0.37			
2010-2019	C12	0.09	0.09	0.09	0.09	0.09	0.09	1.00		
2010-2019	B82	0.74	0.67	0.37	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	1.00	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 D.14: Summary Statistics for Generality Index at 3-Digit Level

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.94	0.9	0.9	0.87	0.95	0.86	0.87	0.88	0.93	0.94
Median 1990-2019	0.94	0.9	0.9	0.87	0.95	0.86	0.86	0.88	0.93	0.94
St.dev. 1990-2019	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.02	0
Mean 1990-1999	0.94	0.91	0.91	0.89	0.96	0.87	0.88	0.88	0.93	0.94
Median 1990-1999	0.94	0.91	0.92	0.9	0.96	0.88	0.88	0.88	0.93	0.94
St.dev. 1990-1999	0	0.01	0.01	0.02	0	0.01	0.01	0.01	0	0
Mean 2000-2009	0.94	0.9	0.89	0.86	0.95	0.85	0.85	0.88	0.93	0.94
Median 2000-2009	0.94	0.9	0.89	0.86	0.95	0.85	0.85	0.88	0.93	0.94
St.dev. 2000-2009	0	0.01	0.01	0.01	0	0	0.01	0	0	0
Mean 2010-2019	0.93	0.89	0.9	0.86	0.94	0.86	0.87	0.88	0.92	0.94
Median 2010-2019	0.94	0.89	0.9	0.86	0.95	0.86	0.88	0.88	0.93	0.94
St.dev. 2010-2019	0.02	0	0.01	0.01	0.02	0	0.02	0.01	0.03	0

Table D.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	0.05							
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	0.00			
1990-2019	C12	0.00	0.00	0.00	0.69	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	0.09								
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	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	1.00	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.17							
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	0.17	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	0.63								
2010-2019	WIPO	0.33	0.27							
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	0.65	0.09				
2010-2019	H04W	0.09	0.09	0.09	0.21	0.09	0.16			
2010-2019	C12	0.09	0.09	0.09	1.00	0.09	1.00	0.65		
2010-2019	B82	0.65	0.59	0.52	0.09	0.09	0.09	0.13	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 D.16: Summary Statistics for Generality Index at 4-Digit Level.

	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82	Y02
Mean 1990-2019	0.97	0.96	0.95	0.95	0.98	0.94	0.91	0.95	0.97	0.98
Median 1990-2019	0.97	0.96	0.95	0.94	0.98	0.94	0.9	0.95	0.97	0.99
St.dev. 1990-2019	0.01	0	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0
Mean 1990-1999	0.97	0.96	0.96	0.96	0.99	0.94	0.91	0.95	0.97	0.99
Median 1990-1999	0.98	0.96	0.96	0.96	0.99	0.94	0.91	0.95	0.97	0.99
St.dev. 1990-1999	0	0.01	0.01	0.01	0	0.01	0.01	0	0	0
Mean 2000-2009	0.97	0.95	0.95	0.94	0.98	0.93	0.9	0.95	0.97	0.98
Median 2000-2009	0.97	0.95	0.95	0.94	0.98	0.93	0.9	0.95	0.97	0.98
St.dev. 2000-2009	0	0	0	0	0	0	0	0	0	0
Mean 2010-2019	0.97	0.96	0.95	0.94	0.98	0.94	0.92	0.94	0.96	0.98
Median 2010-2019	0.97	0.96	0.96	0.94	0.98	0.94	0.92	0.94	0.97	0.98
St.dev. 2010-2019	0.02	0	0.01	0	0.01	0	0.02	0.01	0.02	0

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

Period	I-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	0.00	0.74				
1990-2019	H04W	0.00	0.00	0.00	0.00	1.00	1.00			
1990-2019	C12	0.00	0.00	0.00	0.00	1.00	1.00	1.00		
1990-2019	B82	0.74	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	0.09								
1990-1999	WIPO	0.09	0.56							
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	0.39	0.15			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.39		
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	0.10							
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	0.13		
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	0.19								
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	0.48				
2010-2019	H04W	0.09	0.09	0.09	0.09	1.00	0.09			
2010-2019	C12	0.09	0.09	0.09	0.19	0.84	1.00	1.00		
2010-2019	B82	0.09	1.00	1.00	0.58	0.19	0.48	0.41	0.12	
2010-2019	Y02	0.09	1.00	1.00	0.48	0.09	0.41	0.19	0.09	0.48

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 D.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	2.98	2.53	2.57	2.38	2.14	2.17	2.14	2.15	3.08	2.77
Median 1990-2019	3.38	2.85	3.03	2.72	2.45	2.51	2.43	2.26	3.79	3.3
St.dev. 1990-2019	1.24	1.1	1.15	1.17	1.06	1.03	1.06	1.08	1.5	1.24
Mean 1990-1999	4.04	3.58	3.59	3.5	3.16	3.11	3.22	3.28	4.31	3.81
Median 1990-1999	4.03	3.57	3.61	3.53	3.17	3.11	3.26	3.29	4.33	3.81
St.dev. 1990-1999	0.12	0.09	0.08	0.08	0.04	0.04	0.15	0.1	0.09	0.07
Mean 2000-2009	3.41	2.82	2.94	2.68	2.43	2.49	2.36	2.28	3.73	3.24
Median 2000-2009	3.38	2.85	3.03	2.72	2.45	2.51	2.43	2.26	3.79	3.3
St.dev. 2000-2009	0.52	0.42	0.55	0.54	0.44	0.46	0.36	0.43	0.6	0.38
Mean 2010-2019	1.48	1.2	1.17	0.94	0.84	0.9	0.85	0.88	1.2	1.25
Median 2010-2019	1.61	1.31	1.27	0.99	0.85	0.97	0.88	0.95	1.18	1.27
St.dev. 2010-2019	0.86	0.65	0.64	0.57	0.58	0.55	0.54	0.58	0.89	0.89

Table D.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	0.01	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	0.65	0.91			
1990-2019	C12	0.00	0.00	0.00	0.00	1.00	0.00	0.00		
1990-2019	B82	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.02	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1990-1999	Science	0.09								
1990-1999	WIPO	0.09	0.96							
1990-1999	USPTO	0.09	1.00	0.25						
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	0.59			
1990-1999	C12	0.09	0.09	0.09	0.09	0.25	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	1.00	1.00	1.00	0.09	0.09	0.10	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.25	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	0.64	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.09	0.09	0.09		
2000-2009	B82	0.15	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.64	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.17								
2010-2019	WIPO	0.10	1.00							
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	0.10	0.17				
2010-2019	H04W	0.10	0.09	0.09	0.10	1.00	0.10			
2010-2019	C12	0.09	0.09	0.09	0.10	1.00	0.10	1.00		
2010-2019	B82	0.09	1.00	1.00	0.58	0.11	0.49	0.38	0.10	
2010-2019	Y02	0.09	1.00	1.00	0.38	0.09	0.38	0.17	0.09	0.41

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 D.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	6	4.82	5.07	4.57	3.59	4	3.92	3.56	5.88	5.12
Median 1990-2019	6.73	5.31	5.92	4.95	4	4.41	4.11	3.64	7.29	6.06
St.dev. 1990-2019	2.9	2.47	2.62	2.67	1.95	2.22	2.4	2	3.23	2.59
Mean 1990-1999	8.65	7.37	7.56	7.36	5.54	6.24	6.65	5.75	8.54	7.49
Median 1990-1999	8.58	7.27	7.54	7.44	5.57	6.27	6.78	5.74	8.64	7.57
St.dev. 1990-1999	0.4	0.44	0.26	0.24	0.1	0.14	0.68	0.29	0.39	0.23
Mean 2000-2009	6.88	5.22	5.78	4.97	4.01	4.44	3.96	3.69	7.29	5.94
Median 2000-2009	6.73	5.31	5.92	4.95	4	4.41	4.11	3.64	7.29	6.06
St.dev. 2000-2009	1.4	1.05	1.47	1.42	0.91	1.19	0.9	0.86	1.74	0.99
Mean 2010-2019	2.46	1.86	1.88	1.38	1.21	1.32	1.16	1.24	1.82	1.91
Median 2010-2019	2.6	1.96	2.01	1.41	1.19	1.39	1.18	1.3	1.69	1.83
St.dev. 2010-2019	1.54	1.07	1.1	0.88	0.87	0.83	0.76	0.85	1.44	1.44

Table D.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	0.10	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	0.88	0.07	0.88	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.00	0.02	0.00	0.00		
1990-2019	B82	0.21	0.05	0.42	0.00	0.00	0.00	0.03	0.00	
1990-2019	Y02	0.00	0.88	0.01	0.44	0.00	0.00	0.88	0.00	0.02
1990-1999	Science	0.17								
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	1.00	0.17	1.00	0.45	0.09	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	1.00	0.09		
1990-1999	B82	1.00	0.71	1.00	0.38	0.09	0.09	1.00	0.09	
1990-1999	Y02	0.09	1.00	0.09	0.38	0.09	0.09	0.29	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.64	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	0.53	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.09	0.64	0.09	0.09		
2000-2009	B82	0.92	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.64	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.16								
2010-2019	WIPO	0.09	0.19							
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	0.09				
2010-2019	H04W	0.09	0.09	0.09	1.00	0.09	0.30			
2010-2019	C12	0.09	0.09	0.09	0.09	0.11	0.74	0.09		
2010-2019	B82	0.09	0.19	0.11	1.00	0.16	1.00	1.00	0.19	
2010-2019	Y02	0.09	1.00	1.00	0.33	0.09	0.33	0.30	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 D.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	9.26	7.82	8.42	7.5	5.26	6.15	7.62	5.72	8.81	7.87
Median 1990-2019	10.62	8.63	9.9	8.04	5.84	6.74	7.93	5.66	10.77	9.37
St.dev. 1990-2019	4.55	4.13	4.47	4.55	2.93	3.51	4.85	3.38	5.16	4.12
Mean 1990-1999	13.32	12.11	12.56	12.27	8.2	9.67	13.16	9.52	13.2	11.65
Median 1990-1999	13.44	11.9	12.6	12.44	8.24	9.67	13.47	9.58	13.29	11.8
St.dev. 1990-1999	0.57	0.68	0.41	0.31	0.2	0.16	1.33	0.59	0.89	0.5
Mean 2000-2009	10.78	8.48	9.72	8.15	5.87	6.85	7.64	5.78	10.84	9.21
Median 2000-2009	10.62	8.63	9.9	8.04	5.84	6.74	7.93	5.66	10.77	9.37
St.dev. 2000-2009	2.28	1.85	2.69	2.56	1.43	1.99	1.91	1.48	3.07	1.71
Mean 2010-2019	3.68	2.88	2.98	2.09	1.7	1.93	2.06	1.86	2.39	2.76
Median 2010-2019	3.81	3	3.17	2.12	1.67	2.04	2.07	1.92	2.18	2.56
St.dev. 2010-2019	2.4	1.7	1.78	1.37	1.24	1.23	1.39	1.3	1.93	2.14

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

Period	I-digit	Keyword	Science	WIPO	USPTO	All	G06	H04W	C12	B82
1990-2019	Science	0.00								
1990-2019	WIPO	0.00	0.26							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.42					
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	0.42	0.32	0.01	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	0.42	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	1.00							
1990-1999	USPTO	0.09	0.34	0.29						
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	0.26	0.32	0.97	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	1.00							
2000-2009	USPTO	0.09	0.09	0.09						
2000-2009	All	0.09	0.09	0.09	1.00					
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	1.00	1.00	0.65	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	0.39	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	0.09							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.09	0.09	0.55	0.09					
2010-2019	G06	0.09	0.09	0.09	0.53	0.09				
2010-2019	H04W	0.09	0.09	0.22	0.53	0.26	0.53			
2010-2019	C12	0.56	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	0.26

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 D.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	3.44	2.99	2.96	2.83	2.81	2.67	2.53	3.05	3.74	3.4
Median 1990-2019	3.67	3.17	3.19	2.96	2.92	2.79	2.54	3.07	4.05	3.57
St.dev. 1990-2019	0.78	0.67	0.71	0.74	0.6	0.58	0.6	0.5	0.76	0.56
Mean 1990-1999	4.14	3.66	3.66	3.62	3.42	3.25	3.24	3.59	4.38	3.95
Median 1990-1999	4.12	3.66	3.65	3.63	3.44	3.24	3.28	3.6	4.4	3.95
St.dev. 1990-1999	0.1	0.07	0.06	0.07	0.03	0.04	0.14	0.08	0.09	0.07
Mean 2000-2009	3.68	3.14	3.12	2.94	2.91	2.78	2.49	3.09	4.04	3.55
Median 2000-2009	3.67	3.17	3.19	2.96	2.92	2.79	2.54	3.07	4.05	3.57
St.dev. 2000-2009	0.41	0.35	0.45	0.42	0.3	0.35	0.27	0.2	0.39	0.26
Mean 2010-2019	2.5	2.16	2.11	1.95	2.08	1.97	1.86	2.48	2.78	2.71
Median 2010-2019	2.61	2.16	2.12	1.94	2.06	1.97	1.83	2.52	2.73	2.68
St.dev. 2010-2019	0.43	0.22	0.19	0.18	0.22	0.15	0.12	0.25	0.33	0.25

Table D.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	0.08							
1990-2019	USPTO	0.00	0.00	0.00						
1990-2019	All	0.00	0.00	0.00	0.01					
1990-2019	G06	0.00	0.00	0.00	0.00	0.08				
1990-2019	H04W	0.00	0.00	0.00	0.00	0.54	0.08			
1990-2019	C12	0.00	0.00	0.00	0.08	0.00	0.31	0.05		
1990-2019	B82	0.54	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	0.09								
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	0.09	0.09	0.09	0.15	1.00			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	0.84	0.95		
1990-1999	B82	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
1990-1999	Y02	0.09	1.00	1.00	1.00	0.09	0.09	0.09	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.32							
2000-2009	USPTO	0.09	0.10	0.09						
2000-2009	All	0.09	0.09	0.09	0.09					
2000-2009	G06	0.09	0.09	0.09	0.09	0.34				
2000-2009	H04W	0.09	0.09	0.09	0.09	0.09	0.09			
2000-2009	C12	0.09	0.09	0.09	0.34	0.09	0.77	0.09		
2000-2009	B82	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2000-2009	Y02	0.09	0.09	0.34	0.09	0.09	0.09	0.09	0.09	0.09
2010-2019	Science	0.15								
2010-2019	WIPO	0.11	1.00							
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	0.15				
2010-2019	H04W	0.15	0.11	0.15	0.79	0.52	0.74			
2010-2019	C12	0.15	0.18	0.15	0.09	0.09	0.09	0.11		
2010-2019	B82	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	
2010-2019	Y02	1.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	1.00

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 D.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	6.76	5.54	5.71	5.27	4.57	4.76	4.47	4.9	6.85	6.1
Median 1990-2019	7.3	5.91	6.23	5.37	4.78	4.89	4.29	4.96	7.79	6.56
St.dev. 1990-2019	2.25	1.89	2	2.14	1.37	1.65	1.83	1.28	2.27	1.67
Mean 1990-1999	8.86	7.55	7.69	7.6	6.01	6.51	6.69	6.29	8.67	7.76
Median 1990-1999	8.82	7.43	7.67	7.65	6.04	6.52	6.81	6.24	8.76	7.82
St.dev. 1990-1999	0.36	0.39	0.25	0.23	0.09	0.14	0.67	0.26	0.4	0.23
Mean 2000-2009	7.4	5.78	6.12	5.42	4.79	4.94	4.18	4.98	7.88	6.49
Median 2000-2009	7.3	5.91	6.23	5.37	4.78	4.89	4.29	4.96	7.79	6.56
St.dev. 2000-2009	1.22	0.96	1.3	1.25	0.73	1.03	0.79	0.55	1.4	0.8
Mean 2010-2019	4.03	3.28	3.32	2.79	2.93	2.82	2.55	3.42	4.02	4.06
Median 2010-2019	4.22	3.24	3.35	2.76	2.87	2.84	2.49	3.44	3.89	3.98
St.dev. 2010-2019	1.04	0.49	0.49	0.43	0.43	0.34	0.26	0.53	0.85	0.62

Table D.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	0.04	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	0.37	0.06	0.90	0.00	0.00			
1990-2019	C12	0.00	0.00	0.00	0.07	0.00	0.01	0.19		
1990-2019	B82	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1990-2019	Y02	0.00	0.04	0.83	0.04	0.00	0.00	0.16	0.00	0.03
1990-1999	Science	0.11								
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	1.00	0.59	1.00	1.00	0.09	0.09			
1990-1999	C12	0.09	0.09	0.09	0.09	0.09	1.00	0.09		
1990-1999	B82	1.00	1.00	1.00	0.90	0.09	0.09	1.00	0.09	
1990-1999	Y02	0.09	1.00	0.11	0.59	0.09	0.09	0.90	0.09	0.09
2000-2009	Science	0.09								
2000-2009	WIPO	0.09	0.09							
2000-2009	USPTO	0.09	0.22	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	1.00		
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	0.30								
2010-2019	WIPO	0.46	0.30							
2010-2019	USPTO	0.09	0.09	0.09						
2010-2019	All	0.10	0.09	0.09	1.00					
2010-2019	G06	0.10	0.09	0.09	1.00	1.00				
2010-2019	H04W	0.39	0.54	0.46	1.00	1.00	0.46			
2010-2019	C12	0.22	1.00	0.46	0.09	0.09	0.09	1.00		
2010-2019	B82	0.39	0.39	1.00	0.09	0.09	0.09	0.46	1.00	
2010-2019	Y02	1.00	0.09	0.46	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 D.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	10.4	8.94	9.43	8.56	6.66	7.26	8.56	7.75	10.12	9.29
Median 1990-2019	11.53	9.61	10.42	8.72	6.96	7.49	8.29	7.7	11.51	10.14
St.dev. 1990-2019	3.62	3.28	3.53	3.79	2.15	2.71	3.89	2.38	3.93	2.87
Mean 1990-1999	13.64	12.4	12.79	12.66	8.89	10.1	13.25	10.42	13.4	12.07
Median 1990-1999	13.76	12.2	12.82	12.78	8.94	10.09	13.52	10.41	13.53	12.26
St.dev. 1990-1999	0.54	0.61	0.42	0.28	0.21	0.17	1.31	0.55	0.91	0.51
Mean 2000-2009	11.6	9.4	10.29	8.86	7	7.6	8.05	7.77	11.68	10.05
Median 2000-2009	11.53	9.61	10.42	8.72	6.96	7.49	8.29	7.7	11.51	10.14
St.dev. 2000-2009	2	1.71	2.43	2.31	1.19	1.77	1.7	1.04	2.61	1.43
Mean 2010-2019	5.97	5.03	5.21	4.17	4.1	4.09	4.4	5.05	5.28	5.75
Median 2010-2019	6.19	4.97	5.29	4.17	4.03	4.16	4.29	5.08	5.02	5.52
St.dev. 2010-2019	1.69	0.88	0.9	0.76	0.65	0.56	0.45	0.91	1.13	1.06

D.2.3. Complementarity

Table D.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 D.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	0.58	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 D.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 D.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 D.33: Average Diversity at 1-Digit Level

Period	I-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 D.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 D.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 D.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 D.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.

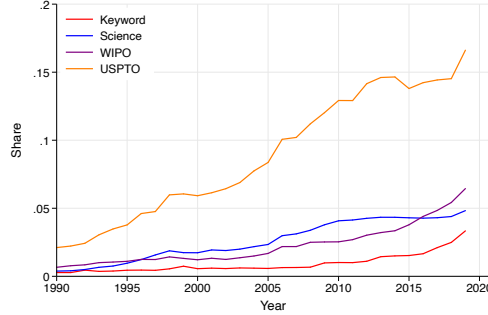
Table D.38: Summary Statistics for Average Diversity at 4-Digit Level

	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

D.3. Additional results

D.3.1. Volume and time trends

Figure D.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.

D.3.2. Generality

Table D.39: Average number of citing CPCs (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 D.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

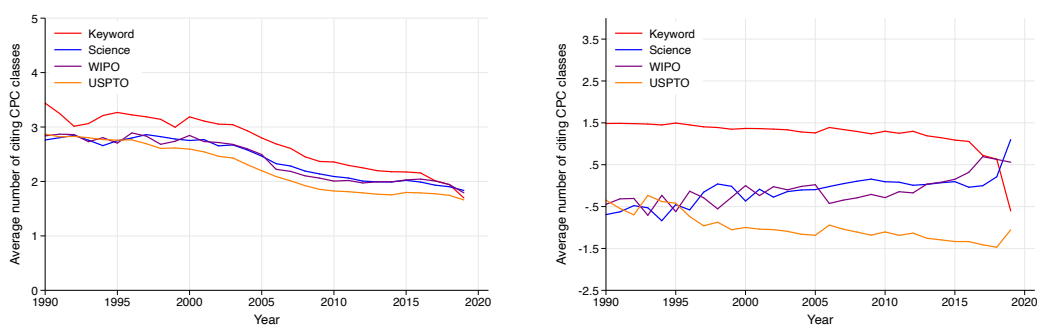
Notes: The generality index is defined as share of citations to patents in different CPC classes at different aggregation levels (see A.2). Citations within the same class are excluded.

Table D.41: Average Number of Citing CPCs (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 D.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.

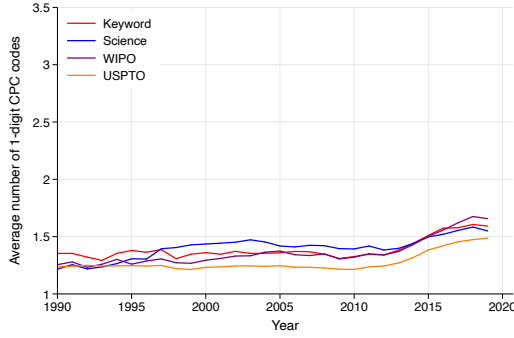
D.3.3. Complementarity

Table D.42: Average Number of 1-, 3- and 4-digit CPCs 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 D.12: Patent-Level Diversity - Average Technology Classes



(a) Average number of 1-digit CPC

D.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 widespread 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 D.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 D.44: Average Number of Citing CPCs: 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 D.45: Average Number of Citing CPCs (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 D.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.