Organisation for Economic Co-operation and Development

Restricted Use **English - Or. English**

AI and subsequent innovation

Preliminary and incomplete draft – Please do not circulate

Flavio Calvino (Flavio.CALVINO@oecd.org)
Jelmer Reijerink (Jelmer.REIJERINK@oecd.org)
Lea Samek (Lea.SAMEK@oecd.org)

The views expressed here are those of the authors and cannot be attributed to the OECD or its member countries

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Restricted Use - A usage restreint

Acknowledgements

This work contributes to the OECD programme on AI-WIPS, supported by the German Federal Ministry of Labour and Social Affairs (BMAS).

Restricted Use - À usage restreint

Table of contents

Acknowledgements 2

1 Introduction 5

2 Existing evidence 6
- Patents as a proxy of innovation 6
- AI as a driver of innovation 8

3 Data and methodology 10
- Patent selection 10
- Identifying (generative) AI patents 11
- Characterising the distinct patent groups 17

4 Innovation dynamics 19
- Impact on subsequent innovation: magnitude and timing 19
- The characteristics subsequent innovation 25
- Breadth of subsequent innovation 27

5 Accelerated innovation 30
- Identifying feedback chains 30
- How exceptional are feedback chains? 31
- Characteristics of knowledge flows 33

6 Concluding remarks 34

References 35

Annex A. Labelling patent datasets 39

Annex B. Robustness checks 41

Annex C. Comparison other technologies 42

FIGURES

Figure 3.1. Total number of USPTO patents per year, 2014-2024 11
Figure 3.2. Framework for identifying both AI and generative AI patents 13

Restricted Use - À usage restreint

Figure 3.3. Shares of generative AI, other AI, and non-AI patents by filing year, 2014-2024 17
Figure 3.4. Distribution of WIPO technology fields per patent group 18
Figure 4.1. Distribution of normalised forward citations 21
Figure 4.2. Evolution of 3-year forward citation differences across patent groups 24
Figure 4.3. AI relatedness of follow-on innovation 26
Figure 4.4. Within versus outside field 26
Figure 4.6. Distribution of generality by patent group 27
Figure 5.1. Example of a patent citation network with feedback chains 31
Figure 5.2. Visualisation of the comparison process for feedback chains 32

TABLES

Table 3.1. Algorithm for labelling patents 14
Table 3.2. Performance of classification framework 16
Table 3.3. Citation activity and age structure across patent groups 18
Table 4.1. Regression results citation impact (1) 22
Table 4.2. Regression results citation impact (2) 23
Table 4.3. Regression results generality index 28
Table 5.1. Number of feedback chains 32

Restricted Use - À usage restreint

1 Introduction

1. AI is increasingly regarded as a transformative technology across multiple sectors. Its rapidly growing impact across different fields, such as education, healthcare, law and creative industries suggests that AI may be more than just a regular technological advance. Many now view AI as a potential General-Purpose Technology (GPT) – technologies that are characterised by widespread diffusion, continuous improvement, and capacity to spawn complementary innovations (Bresnahan and Trajtenberg, 1995[1]). Recent studies underscore the alignment of AI with these features, highlighting its potential to significantly impact economies and societies (Calvino, Haerle and Liu, 2025[2]; Hötte et al., 2022[3]).

2. While previous research has largely focused on AI as a whole, the rise of generative AI has introduced new patterns that distinguish it from the broader pool of AI technologies. Additionally, while some studies have looked at the general diffusion of generative AI, few have investigated the specific nature of innovations emerging from the technology. Despite industry reports and anecdotal evidence pointing to the transformative potential of generative AI, there is still a lack of extensive empirical analysis. Some studies suggest that AI patents drive innovation in other sectors, which may, in turn, influence the development of future generative AI technologies. However, these claims remain largely untested, and little attention has been given to the characteristics and impacts of innovation feedback chains - beyond the potential existence of such a mechanism.

3. This study seeks to fill these gaps. It aims to examine how AI influences innovation across different fields through a deeper exploration of patent data and citation networks. By taking a quantitative perspective, this research aims to provide stronger insights into the impacts of AI inventions, moving beyond earlier descriptive findings. Furthermore, the analysis provides a more granular view of how generative AI affects various technological domains. By comparing patenting trends across different sectors, the study offers a comparative perspective. Through this, the study aims to shed light on the broader implications of AI as a technology that spawns innovation across diverse fields and industries, aiming at helping policymakers, businesses, and researchers better understand how to harness the full potential of AI and generative AI for innovation and technological progress.

Restricted Use - À usage restreint

2 Existing evidence

Patents as a proxy of innovation

4. Patents are widely used as a proxy of innovation, because they represent legally recognised inventions that often involve novel technical contributions. By analysing patent data, researchers can track the rise of new technologies, monitor how these technologies spread, and assess their impact across different sectors. Patent data have long been used as a critical tool to evaluate technological progress and understand innovation dynamics (Tseng and Ting, 2013[4]; Iori, Martinelli and Mina, 2021[5]; Hall, Jaffe and Trajtenberg, 2001[6]). Central to this analysis is the citation structure. According to the OECD Patent Statistics Manual (OECD, 2009[7]); if patent B cites patent A, it indicates that patent A has existing knowledge upon which B builds, but over which B can have no legal claim. Consequently, patent citation data offer both quantitative and qualitative insights into the evolution of technology. By examining these backward and forward-looking indicators, researchers can distinguish core inventions from incremental developments and study the global diffusion of knowledge across sectors and countries (Jaffe and Rassenfosse, 2016[8]; Calvino et al., 2023[9]).

5. Various metrics derived from patent data can be used to evaluate a technology's impact on innovation. For example, the number of patent applications gives a basic measure of inventive output. In contrast, quality measures, often based on citation analysis, offer a deeper understanding of impact. Forward citations, which reflect how often a patent is cited by subsequent patents, are frequently used as an indicator of technological impact and relevance for future developments. For instance, Calvino et al. (2023[9]) define core AI patents based on their high number of AI-related forward citations, arguing that these patents are more likely to be at the centre of AI advancements.[1] Additionally, Tseng and Ting (2013[4]) use certain citation ratios (such as the "Citation Impact Index") to compare the patent quality of AI technologies across different countries.

6. The process of applying for a patent typically involves applying to the appropriate patent office providing details about the invention, and providing information on how it is new and useful. The application is reviewed by patent examiners to check if it meets the necessary criteria, such as novelty and applicability. There are various patent offices around the world, such as the United States Patent and Trademark Office (USPTO) in the United States, the European Patent Office (EPO) in Europe, and the World Intellectual Property Organization (WIPO). Inventors must decide whether to file in just one or multiple countries, depending on where they want protection. Applying for patents for software inventions, and AI inventions in particular, is a complex process. Many legal and technical challenges can arise, such as determining the patentability of AI-related inventions, which often involve complex algorithms and abstract concepts. The rapidly evolving nature of AI adds further complexity to this procedure. This paper focuses on patents filed at the USPTO because it allows software to be patented, where other jurisdictions have stricter requirements for demonstrating a technical contribution or inventive step for this (see Box A.1 for a further discussion about the patentability of AI inventions).

[1] The analysis focuses on the normalised forward citations count to account for the fact (among others) that older patents are more likely to have a higher forward citation count.

Restricted Use - À usage restreint

Box A.1. Patentability of AI inventions

Patenting software inventions comes with significant challenges, one of which is the difficulty in defining the boundaries of what makes a patentable invention. AI inventions, a subset of software inventions, introduce additional complexities. Unlike traditional patents for physical inventions, software patents are often more abstract. The patentability of a software invention depends on legal criteria, which vary across jurisdictions.

In the United States, the USPTO allows patents for new and useful inventions. However, a U.S. Supreme Court's decision (2014) established that abstract ideas implemented on a computer are not patentable unless they involve an "inventive concept" that transforms the abstract idea into a patent-eligible invention.[2] This ruling has made software patentability more challenging, but it remains possible to protect innovative software as intellectual property. In July 2024, the USPTO released updated guidance on AI inventions, highlighting the importance of clear claims that frame AI inventions as technological improvements. The guidance urges applicants to show that their AI inventions solve technical problems, which aligns with existing case law and helps address eligibility challenges. This is key to strengthening patent protection for AI innovations.[3]

In the European Union, the EPO states that software can only be patented within the context of "Computer implemented inventions" (CII) and may be patentable if they solve a technical problem using technical means.[4] This criterion extends to AI inventions too: AI applications must solve technical problems before they are eligible for patenting. The EPO applies a "technical character" test, which means that AI inventions must demonstrate a technical effect that contributes to the advancement of technology in order to meet patent eligibility requirements. In this regard, AI algorithms on their own, without providing a technical solution to a real-world problem, are generally not patentable under EPO rules.

Hence, while both offices reject purely abstract inventions, the USPTO allows for a broader range of AI inventions to be considered patentable, as long as they are more than abstract ideas, while the EPO applies a more restrictive test, requiring a clear technical contribution to the field.

7. Inventors face many considerations when deciding whether to apply for a patent. Patents offer protection by preventing others from using or copying innovations, providing them a competitive advantage. However, patenting involves costs like filing fees, legal expenses, and the risk of disclosing sensitive information to the public. Inventors must also decide whether to patent internationally or focus on key markets. As a result, not all inventions are patented, either due to strategic decisions or unmet patentability criteria. Research by Hughes and Mina (2010[10]) confirms this, and shows that small firms are less likely to use patents as a way to protect their investments compared to alternatives like confidentiality, secrecy, or relying on time-to-market strategies. Furthermore, the choice of legal protection depends on the technology; for instance, Dernis et al. (2019[11]) show that AI-related inventions are significantly more likely to be protected by patents than by trademarks. In software markets, Cockburn and MacGarvie (2011[12]) suggest that patents influence competition and market entry, by showing that

[2] Alice Corp. v. CLS Bank International, 573 U.S. 208 (2014).

[3] https://www.uspto.gov/about-us/news-updates/uspto-issues-ai-subject-matter-eligibility-guidance

[4] https://www.epo.org/en/legal/guidelines-epc/2025/g_ii_3_6.html

Restricted Use - À usage restreint

incumbents' patents can increase entry costs, particularly when there are many overlapping patents, known as "patent thickets". Jaffe and Rassenfosse (2016[8]) highlight several pitfalls associated with patent citation data, including office-specific effects, technology dependencies, and strategic citation behaviour. Therefore, while patent data is widely used, it is important to recognise its limitations as a sole measure of innovation and to consider it alongside other indicators of innovation.

AI as a driver of innovation

8. The role of AI in driving innovation has gained significant attention, particularly regarding how AI patents contribute to technological advancements across various sectors. Many studies look at general AI patents, where there is no specific focus on generative AI. Evidence on this broader field of AI suggests that AI patents appear across a diverse range of fields. For instance, Van Roy, Vertesy, and Damioli (2019[13]) analyse global AI patenting trends and find a significant increase in patent filings since 2013, particularly between 2015 and 2016. They highlight that AI patenting is concentrated in software and electronics sectors but has increasingly expanded into non-tech industries. This trend underscores AI's growing role as an enabler of innovation across a wide range of sectors. In line with this, Liu et al. (2021[14]) further explore the impact of AI patents, revealing that although AI innovation is expanding rapidly, it is largely driven by a small number of organisations. Moreover, AI technologies are not limited to computing and information sectors but extend into other areas like healthcare, finance, and education. This underscores the broad applicability of AI and its potential to transform a wide range of industries. In this context, recent evidence from Calvino, Reijerink and Samek (2025[15]) suggests that generative AI is further transforming the economic landscape by significantly impacting productivity, innovation, and entrepreneurship. In line with this, Santarelli, Staccioli, and Vivarelli (2023[16]) analyse the distribution of AI patents using USPTO patent data, showing that AI patents are considerably widespread across different fields. The paper also notes that AI innovation is building on earlier ICT technologies, suggesting that AI is deepening existing technological trends rather than a completely radical shift. Fujii and Managi (2018[17]) highlight the shift in the nature of AI patents, from early innovations inspired by biological systems and general knowledge to more recent patents focused on specific mathematical techniques. This shift indicates a growing complexity in AI technologies and their applications, suggesting that AI patents are becoming increasingly specialised and sophisticated.

9. Patent citation data also provide valuable insights into the global distribution of AI innovation. Dernis et al. (2021[18]) explore the profiles of these innovators, identifying the specific types of firms – from startups to multinationals – that are at the forefront of developing AI-related goods and services. Complementing this firm-level perspective, Tseng and Ting (2013[4]) use citation data to compare the development of AI technologies across different countries, finding that the United States leads in both the volume and the influence of AI patents. Other countries such as Japan, Germany, Korea, and the United Kingdom are significant followers in terms of technological output. Forward citations, which measure how influential patents are in subsequent innovations, demonstrate the global nature of AI innovation and highlight the United States' central role in shaping AI technologies worldwide.

10. While previous studies have focused on identifying the sectors where AI innovations are most prominent, other research suggests that AI serves as a foundational technology, driving innovation across multiple domains. For instance, using normalised citation counts, Calvino et al. (2023[9]) find that core AI patents tend to be more general, meaning they are cited by patents belonging to a wider range of technology fields. However, while the analysis identifies application specific patent fields - such as autonomous driving and robotics - it does not track the adoption of these technologies into broader non-technological sectors. Their network analysis confirms that while these application fields primarily lead to innovation within their own field, many others (such as those related to algorithms as well as recognition and detection technologies) act as inputs cited across the broader AI landscape. Along similar lines, Biggi et al. (2025[19]) analyse patent applications to study whether AI acts as a GPT, and show that AI is

Restricted Use - À usage restreint

becoming a key technology to tackle sustainability challenges and is helping to drive green innovation. These findings align with the perspective of Cockburn, Henderson and Stern (2018[20]), who argue that AI shares characteristics with GPTs, influencing a wide range of industries and potentially transforming innovation processes in ways not yet fully realised. In line with this, Damioli et al. (2025[21]) observe that AI patents have spread into non-ICT sectors, particularly in service industries. Although their study does not isolate the exact mechanisms by which AI patenting influences innovation in non-AI fields, the growing presence of AI patents across diverse industries suggests that AI fosters a cross-sectoral flow of knowledge and technology. This is particularly true in areas like healthcare, where AI's ability to analyse large datasets has led to advances in diagnostics and drug development. Recent evidence focusing on generative AI patents substantiates this idea. Calvino, Haerle and Liu (2025[2]) show in fact that generative AI applications span a wide range of areas beyond just a few sectors.

11. The nature of innovation spawned by AI patents is another key area of research. Several studies suggest that AI accelerates innovation through a combination of mechanisms, including enhancing knowledge creation, improving learning and absorptive capacities, and boosting R&D investment. Liu et al. (2020[22]) argue that AI promotes technological innovation by improving knowledge transfer and facilitating new discoveries, particularly in sectors like manufacturing, where AI adoption has been shown to significantly increase innovation in low-tech industries. In addition, Agrawal, McHale and Oetll (2023[23]) provide evidence that in areas like drug discovery and materials science AI potentially improves the efficiency of discovery pipelines by augmenting human creativity and extracting insights from large datasets. Furthermore, AI-driven innovation goes beyond small improvements, bringing significant changes in the innovation process itself. For instance, Cockburn, Henderson and Stern (2018[20]) suggest that AI, particularly deep learning techniques, represents a new "method of invention", where the focus is on leveraging large datasets and predictive algorithms to generate new technologies (see also Bianchini et al. (2022[24])). This shift indicates a fundamental change in how research and development are conducted, with AI enabling faster and more accurate predictions, which in turn increase the potential for new inventions.

12. Another key element of AI and innovation is the potential feedback mechanism between AI patents and other technological fields. Several studies highlight how innovations in AI stimulate new discoveries in other sectors, creating a dynamic cycle of cross-sectoral innovation. Bresnahan (2024[25]) emphasizes the importance of feedback mechanisms in the GPT discussion, yet their quantifiable effects remain under-examined in the context of generative AI. Calvino, Haerle, and Liu (2025[2]) provide initial evidence on a potential existence of a feedback chain, by showing that generative AI patents are not only cited across various fields but also drive further advancements in AI technologies themselves. This feedback mechanism reinforces the general-purpose nature of AI, where progress in one area accelerates development in others. However, the feedback chains between generative AI patents and innovations in other fields have not yet been fully explored and quantified.

Restricted Use - À usage restreint

3 Data and methodology

Patent selection

13. To analyse technological process and innovation dynamics, this study uses patent data from the USPTO. As explained in Box A.1, the USPTO allows for a broader range of patents to be considered patentable, more specifically for software-related inventions. Therefore, compared to patents filed at other patent offices, these patents provide a more comprehensive source for studying how AI – notably including generative AI – inventions shape technological progress. Data is derived from the *Spring 2025 Edition* of the Worldwide Patent Statistical Database of the EPO, also known as the PATSTAT database (European Patent Office, 2025[26]). Within the OECD STI Micro-data Lab, this database has been further adapted and curated to ensure consistency, harmonised applicant and inventor information, and reliable linkage with additional datasets. These refinements make it easier to perform analytical work.

14. The study focuses on patent applications rather than solely granted patents in the dataset. Patent applications are a critical component for studying emerging technologies before they are granted. Applications provide early insights into technological trends and allow for an exploration of ideas that might not have been granted a patent (yet), offering a more comprehensive analysis of technological developments. This is particularly relevant for generative AI inventions, which are likely to be relatively recent and may not yet be fully reflected in the stock of granted patents. Furthermore, applications are filed for various reasons, such as filing early to secure rights, attracting investment, or signalling innovation activity to competitors (Cohen, Nelson and Walsh, 2000[27]). Even if these applications ultimately do not result in a granted patent, including them ensures that the analysis captures more potential innovations.

15. To capture the evolving impact of AI patents, this study includes patent applications filed from 2014 through 2024, the most recent available year. This starting point follows evidence that AI patenting begins to increase noticeably around that time (Aranda et al., 2025[28]). In this selection, the filing date of the application is used, as it best reflects when the invention formally entered the USPTO system and became visible as a potential source of follow-on innovation. Using 2014 as a starting point ensures coverage of the early stages of generative AI while also incorporating the most recent developments that may be particularly important for understanding subsequent innovation and feedback effects. Rather than excluding earlier years from the dataset, all available recent data are retained, because trimming the dataset would risk omitting valuable insights from more recent inventions. At the same time, the use of recent data requires careful consideration. Patents from the most recent years are less complete in the database and have had less time to accumulate citations. These limitations are addressed in the quantitative analysis, for example by normalising citation counts or by controlling for application year.

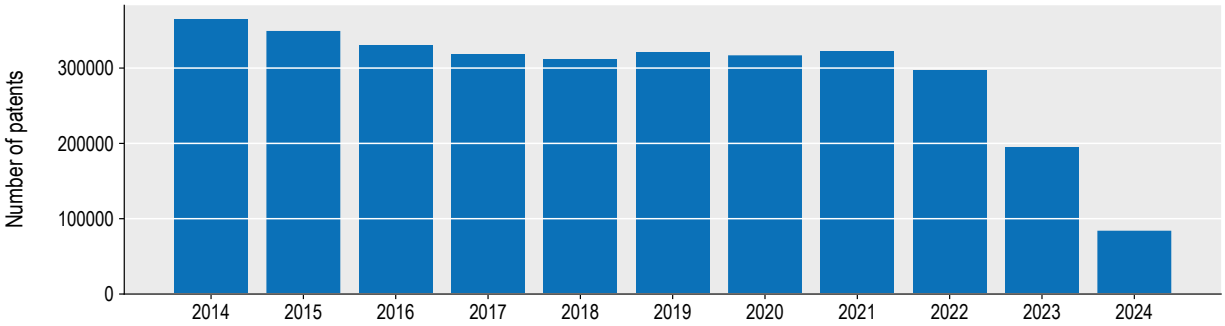
16. Beyond basic patent information, such as filing date and patent application number, additional variables are extracted from the database to enable a more detailed analysis. For example, the Cooperative Patent Classification (CPC), International Patent Classification (IPC) and WIPO technology

Restricted Use - À usage restreint

fields provide structured information on the technological domain of each patent.[5] Finally, several patent quality indicators are included, namely generality, originality, scope and radicalness. This follows the methodology introduced in Squicciarini, Dernis and Criscuolo (2013[29]), which outlines their construction and measurement. Taken together, these additional variables provide the basis for a deeper analysis of technological developments and allow the analysis to go beyond simple counts of applications.

17. Last, to avoid double counting and ensure that each invention is represented only once, the analysis is conducted at the level of DOCDB patent families. A DOCDB patent family groups together all patent applications that protect the same invention, using the family definitions maintained by the EPO for global patent data. For each family, the earliest (first) application is taken as the reference record. If information is missing for this application, data are supplemented using other applications from the same family, so that the final dataset is as complete as possible.

Figure 3.1. Total number of USPTO patents per year, 2014-2024



Note: Data refer to unique DOCDB patent families filed at the USPTO.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025.

18. Having established the data foundation, Figure 3.1 provides an overview of the final dataset used in this study, showing the total number of new USPTO patents per year between 2014 and 2024, measured at the level of unique inventions. Although the overall trend is relatively stable, with annual new filings ranging between 300 thousand and 350 thousand, the number of patent applications per year drops after 2022. This drop reflects the publication lag: since patent applications are generally published 18 months after the filing date, many recent filings are not yet visible in the dataset. It is important to consider this incompleteness in the quantitative analyses that will follow, to ensure that recent trends are interpreted correctly.

Identifying (generative) AI patents

Existing methodologies for patent classification

19. To analyse how (generative) AI influences subsequent innovation, it is first necessary to identify which patents concern AI, and within that subset, which specifically involve generative AI. The identification of AI patents has been the subject of a growing body of literature. An early effort in this area is done by WIPO (2019[30]), which developed a large-scale taxonomy of AI patents by combining AI keywords and CPC/IPC codes. Furthermore, Baruffaldi et al. (2020[31]) identify AI-related keywords through text mining

[5] The WIPO technology categories aggregate detailed IPC patent classes into 35 broader and comparable technology groups (Schmoch, 2008[46]).

Restricted Use - À usage restreint

of recent AI scientific publications, and then use these keywords to determine relevant patent classes and search for AI-related terms in titles, abstracts and claims. More recently, Aranda et al. (2025[28]) updated and expanded this approach. In their paper, they introduce a semi-automated process where they use recent AI papers on arXiv to extract emerging AI terminology and link these terms to the CPC classes in which they most often appear in patent titles and abstracts.[6] This new approach incorporates a broader and more adaptable framework, also allowing for the inclusion of more recent applications, such as generative AI.

20. Although these studies provide a robust foundation for identifying patents related to AI, they are rule-based and rely on keyword and patent classes. This means they may miss patents that use different wording to describe AI or that have not been placed in AI-related patent classes. This limitation is particularly important for studies that go beyond descriptive statistics and aim to assess the impact of (generative) AI on subsequent innovation, since misclassified or missing AI patents can affect the underlying measures and therefore bias the results. Some recent work has in fact started using other methods to identify AI patents. For example, Giczy, Pairolero and Toole (2022[32]) depart from this approach by using machine learning (ML) to identify AI patents directly from patent texts. They trained supervised models to classify patents as AI or non-AI, based on linguistic and semantic patterns in the full text, allowing them to capture patents that describe AI inventions in less conventional wording, or that have not been assigned to specific IPC or CPC classes. While this method points in a promising direction, it may not fully capture the most recent AI developments, given the speed at which both the technology and the AI-based analytical tools are evolving.

21. For generative AI patents, more specifically, fewer studies exist. This largely reflects the recent uptake of the technology. Furthermore, whereas statistical AI can be more easily identified through existing taxonomies and keywords, the boundaries of what defines generative AI are more difficult to describe as the technology continues to evolve. This also makes it more difficult to find patents related to this technology. Given these challenges, and reflecting the broader shift from rule-based toward model-driven approaches, a notable effort in this direction is WIPO (2019[30]), which fine-tuned a BERT-based language model on a manually labeled set of patents to distinguish between generative and non-generative AI patents.[7] Such model was finetuned and tested on a preselected set of patents identified through keyword-based searches and expert selection within specific generative AI-related domains. Its performance on a more diverse, real-world sample beyond that initial set of patents remains to be explored.

Implemented patent classification framework

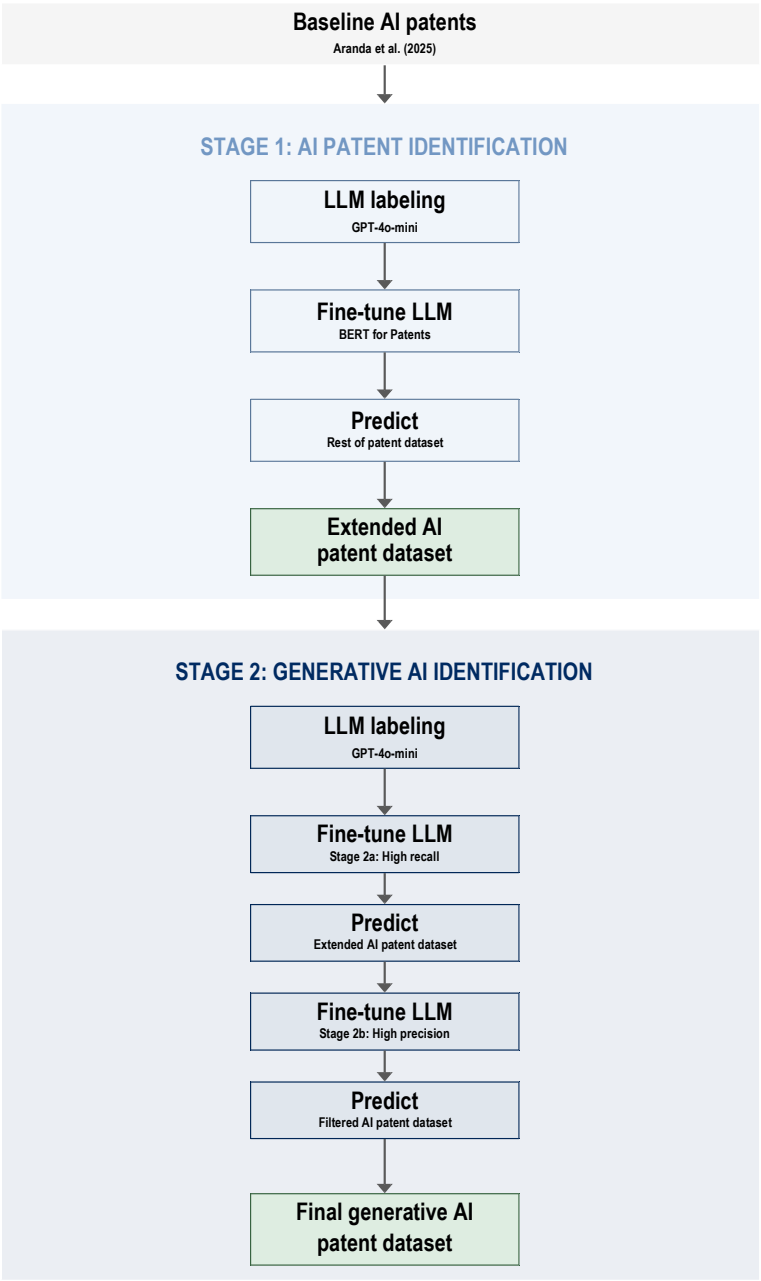
22. To address the limitations and challenges outlined above and to ensure a comprehensive identification of both AI and generative AI patents, this study develops a new multi-stage classification approach, which integrates large language models (LLMs) and ML techniques. It proceeds in two main stages: first, the construction of a comprehensive set of AI patents, and second, the identification of the subset of those patents that relate specifically to generative AI. Please see Figure 3.2 for a schematic representation of the classification framework.

[6] arXiv is an online repository where researchers share recent research papers that have not yet been peer reviewed.

[7] BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based language model developed by Google (Devlin et al., 2019[43]). It learns contextual relationships between words in a text, which makes it effective for understanding language patterns for tasks such as classification, question answering, and information retrieval.

Restricted Use - À usage restreint

Figure 3.2. Framework for identifying both AI and generative AI patents



23. The starting point for the classification is the set of AI patents defined by Aranda et al. (2025[28]). This dataset provides a high-quality baseline, which is then complemented and expanded through the development of a new labelling and prediction pipeline. First, all patents that were not already classified as being AI-related are selected to find any additional AI patents. Next, a large language model is prompted to label a random sample of those patents as either AI-related or not. The labelling process uses the GPT-4o-mini model with a structured prompt design to ensure consistent and reproducible classifications across

Restricted Use - À usage restreint

all patents.[8] The algorithm in Table 3.1 outlines the full sequence of steps used to label the patents. Each patent is presented to the model with the patent title, main CPC code, and the abstract, allowing the LLM to consider both textual content and technical context. While running the prompts, the model temperature can be adjusted, which determines how much variation the model allows in its responses. In this study, the temperature was set to zero to avoid randomness and to ensure deterministic responses. Furthermore, patents with low confidence are excluded to maintain a high-quality labelled dataset.

Table 3.1. Algorithm for labelling patents

Algorithm
Require: A: Dataset of patent abstracts
Ensure: Structured JSON output for each patent, containing *label* (1 or 0) and *certainty* (0.0 to 1.0)
1: Initialise connection with GPT-4o-mini model using OpenAI API, set temperature to 0
2: **for each** abstract a in A
3: **Message: System**
4: *"You are an expert in AI and patents. Classify whether a patent abstract describes an invention that uses or is based on …"*
5: **Message: User**
6: Content: the prompt containing the title, the main CPC code and the abstract of a patent
7: Send messages to GPT-4o-mini
8: Retrieve response from GPT-4o-mini
9: Parse returned text to extract JSON containing *label* and *certainty*
10: Append original patent data together with parsed *label* and *certainty* to csv file
11: Exclude rows where *certainty* < 0.9
12: **end for**

Note: This table summarises the procedure used to label patents with the GPT-4o-mini model. Complete prompt texts can be found Box A.1 for AI and Box A.3 for generative AI in Annex A.

24. The approach used in this study to label AI patents builds directly on definitions, established by the OECD, of an AI system. They describe AI systems as machine-based systems that infer from inputs how to produce outputs, such as predictions, content or decisions (OECD, 2019[33]; 2024[34]) The prompt builds on this foundation by translating the broad definition into operational criteria, while extending it with concrete examples and guardrails designed to improve precision in a patent context, including explicit inclusion and exclusion rules. The full prompt template used in the labelling step is provided in Box A.1 in Annex A. The structured approach of the prompt improves the labelling process by making the interpretation of patent abstracts more consistent and transparent, while also reducing ambiguity linked to the wide range of terminology found in patent abstracts.[9]

25. The resulting labelled sample dataset then forms the training data for fine-tuning the BERT for patents model (Srebrovic and Yonamine, 2020[35]), a model based on BERT that has been trained

[8] GPT-4o-mini is a compact version of the GPT-4o model developed by OpenAI, optimised for efficient inference while maintaining strong performance at lower computational cost. Technical documentation is available at https://platform.openai.com/docs/models/gpt-4o-mini

[9] A robustness check using an alternative prompt based strictly on the OECD definition was also performed. This prompt, provided in Box A.2 (Annex A), relies only on the OECD wording and requires clear evidence that the invention involves a machine-based system performing inference. When applied to the same set of 2 500 randomly sampled abstracts, more than 96 percent of the patents classified as AI by the main prompt were also classified as AI under this OECD-only version. This shows that the positive labels are reasonably stable even under a narrower interpretation of the definition. The remaining differences reflect intentional adjustments in how the definition is applied and help to improve the overall quality of the labelling.

Restricted Use - À usage restreint

specifically on patent documents. It allows to understand the technical language and structure that is typically present in patent texts. Finetuning the model on the labelled data enables it to learn how to label patents on its own and to apply these labels consistently across the full dataset. Once finetuned, the model is applied to all patents that were not already identified as AI-related by Aranda et al. (2025[28]) to predict which additional patents are related to AI technologies with high confidence. Only less than 1 in 25 patents were classified as AI-related in this additional step. While this share is relatively small, given the large overall number of patents in the dataset, it still represents a substantial number of additional AI patents. This prediction therefore augments the baseline dataset, resulting in a broader set of AI patents that aims to combine the high precision of the established rule-based approach with the wider coverage of the ML method.

26. By definition, all generative AI patents are part of this subset of AI patents. Therefore, the classification of generative AI patents can also be restricted to this specific subset, which makes the task more focused and efficient since it limits the analysis to patents that are already likely to involve AI technologies. This not only reduces unnecessary computation but also helps avoid noise from patents that are clearly unrelated to generative AI. In addition, the share of generative AI patents is higher within the AI subset than in the full patent dataset, improving model performance and reliability of the results.

27. The identification of generative AI patents follows the same overall approach as the AI stage. Similarly, an LLM is prompted to label a random sample of AI patents. Only this time, it determines whether they describe generative AI technologies or not. This prompt uses a definition that is in line with commonly used descriptions of generative AI, including the OECD definition, which describes generative AI models as systems that create new outputs such as text, code, audio, images or video in response to prompts (OECD, 2023[36]). The prompt, reported in Box A.3 in Annex A, applies this definition directly and makes the classification criteria more explicit, similar to the labelling step in the AI phase.[10]

28. The resulting labelled dataset is then used in a two-stage finetuning and prediction process (stage 2a and 2b in Figure 3.2, respectively). The implementation of a two-stage process is important for two main reasons. First, the relatively small share of generative AI patents within the broader AI subset creates a class imbalance that is often difficult for single-stage models to deal with, potentially leading to higher false positive rates. Second, the model now needs to distinguish between patents that are generative AI and those that are non-generative AI, which are often quite similar since they fall within the same overall AI domain. The two-stage approach addresses both challenges. In the first step, a BERT for Patents model is finetuned on the labelled sample and acts as a broad filter that identifies all patents with potential generative AI features. This step is designed to ensure that as few relevant patents as possible are missed. In the second stage, a separate BERT for Patents model is finetuned on the filtered subset from the first step but then applies a stricter threshold to distinguish actual generative AI patents from related but non-generative AI inventions. This second stage focuses on reducing false positives while ensuring that not too many relevant patents are excluded.[11] After completing these finetuning steps, the two-stage pipeline

[10] Following the approach used for the AI labelling step, a robustness check was also carried out for the generative AI labels. The alternative prompt, shown in Box A.4 in Annex A, relies strictly on the OECD definition. When applied to the same set of 2 500 randomly sampled abstracts, more than 82 percent of the patents classified as generative AI by the main prompt were also classified as such under this stricter version. Therefore, that the two definitions capture a largely overlapping set of patents while also showing some differences. Taken together, the results suggest that the main findings are not highly sensitive to the choice between these two definitions.

[11] In fact, there is always a trade-off between reducing false positives and reducing false negatives. For example, labelling every patent as generative AI would eliminate false negatives entirely but produce an extremely high false positive rate. Conversely, labelling only the patents with highest confidence as generative AI would minimise false positives but miss many true generative AI patents. The goal is therefore to achieve a balance that limits both false positives and false negatives.

Restricted Use - À usage restreint

is applied as a waterfall across the entire AI patent subset, sequentially filtering patents to identify those most likely associated with generative AI.

29. To test the performance of both models, two separate evaluation datasets were assembled, one for the identification of AI patents, and one for generative AI patents. These evaluation datasets were randomly sampled from the full dataset used in the current study, ensuring they reflect the true prevalence of AI and generative AI patents. Patents that were used during the training of the models were excluded from the evaluation sets, to avoid inflating evaluation results. Table 3.2 presents the performance of both models on the evaluation datasets. For the identification of AI patents, the model has a precision of 0.80 and a recall of 0.86. In other words, among all patents predicted to be AI, 80% are correctly identified, while the model successfully detects 86% of all actual AI patents. For the identification of generative AI patents, the model has a precision of 0.73 and a recall of 0.77. These performance metrics fall within the range typically achieved by LLMs in patent classification tasks, showing respectable performance particularly given the added complexity of identifying more specialised and recently emerged categories such as AI and generative AI.[12] Given the aims of this study, the approach deliberately prioritises high recall while maintaining a precision level of approximately three quarters. This ensures that the majority of patents related to generative AI are captured, while still limiting the inclusion of false positives to an acceptable degree. This balance is crucial because subsequent analyses build directly on these subsets to examine their impact.

Table 3.2. Performance of classification framework

Model	Precision	Recall	F1	Prevalence (%)
AI identification (Stage 1)	0.80	0.86	0.83	6.04
Generative AI identification (Stage 2)	0.73	0.77	0.75	0.94

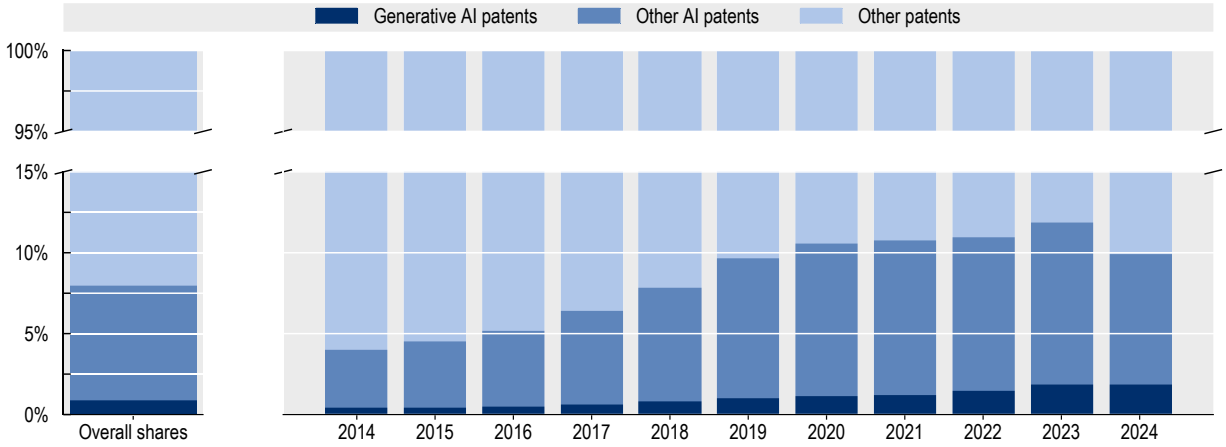
Note: This table reports the precision, recall, and F1 score for both classification models. Evaluation metrics are calculated on a random sample of 10 thousand USPTO patents drawn from the full patent dataset, excluding any patents used in model training. Prevalence refers to the share of patents in the evaluation sample that are labelled as positive by the model.

30. Based on the classification framework described above, the final dataset of USPTO patent applications is divided into three mutually exclusive groups: generative AI patents, other AI patents, and other (non-AI) patents. Generative AI patents are defined as those that specifically relate to content-generating inventions, while other AI patents cover a broader range of AI applications that do not involve this generative functionality. All residual patents are grouped in the remaining category. As Figure 3.3 illustrates, across all USPTO filings between 2014 and 2024, around 7 percent of unique patent families are identified as AI-related. In addition, around 1 percent of all inventions are labelled as generative AI.

[12] See for example Lee and Hsiang (2020[47]) and Haghighian Roudsari et al. (2021[48]) who also apply LLM-based methods for patent classification, and show that performance varies with (among others) task design and dataset characteristics. Yet, they generally fall within a similar overall range to the performance reported in the current study.

Restricted Use - À usage restreint

Figure 3.3. Shares of generative AI, other AI, and non-AI patents by filing year, 2014-2024



Note: Shares are based on unique DOCDB patent families filed at the USPTO. Patent families are categorised into three mutually exclusive groups. An axis break between 15 and 95 percent is applied to emphasise variation within the AI segments.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025.

31. Looking at the filing trends over time, the share of AI-related patent families in the dataset increased steadily, from around 4 percent in 2014 to 11 percent by 2023, after which it drops slightly in the most recent year.[13] This development is consistent with broader global trends. For example, WIPO statistics report a strong increase of AI patenting activity after 2013, following the diffusion of deep learning methods (WIPO, 2019[30]). In contrast, generative AI inventions emerge only a few years later. Their presence in the dataset remains negligible until 2017, after which it gradually increases. Measured as a share of all patent families, they account for almost 2.5 percent in 2024. This corresponds to roughly 12 percent of AI patent families in the dataset. Although the overall share remains modest relative to the broader AI category, the upward trajectory is clear and consistent with international evidence (WIPO, 2024[37]).

Characterising the distinct patent groups

32. Building on the three mutually exclusive groups defined above, Table 3.3 provides an overview of their citation patterns and age profiles. Generative AI patents stand out as the youngest group, with an average age of just over six years. They also show a relatively high share of patents without any forward citations, at nearly 47 percent. This uncited share is almost 6 percentage points higher than what is observed for other AI patents. This difference is consistent with their more recent emergence, giving them less time to accumulate citations, and with the smaller presence of generative AI patents in earlier years. Patents outside the AI domain are older on average and have the highest proportion of uncited families overall. These contrasts highlight the differing stages of technological maturity across the groups and the early lifecycle position of generative AI inventions.

[13] The 2024 dip in AI patent families should be interpreted with caution. Even though evidence by IFI CLAIMS (2024[44]) shows a decline in applications in the CPC G06N class, which is often considered the most relevant AI patent class according to e.g. the Japan Patent Office (2024[45]), the 2024 data is still incomplete at the time of writing due to publication lags. It is therefore too early to determine whether this reflects in fact a structural change.

Restricted Use - À usage restreint

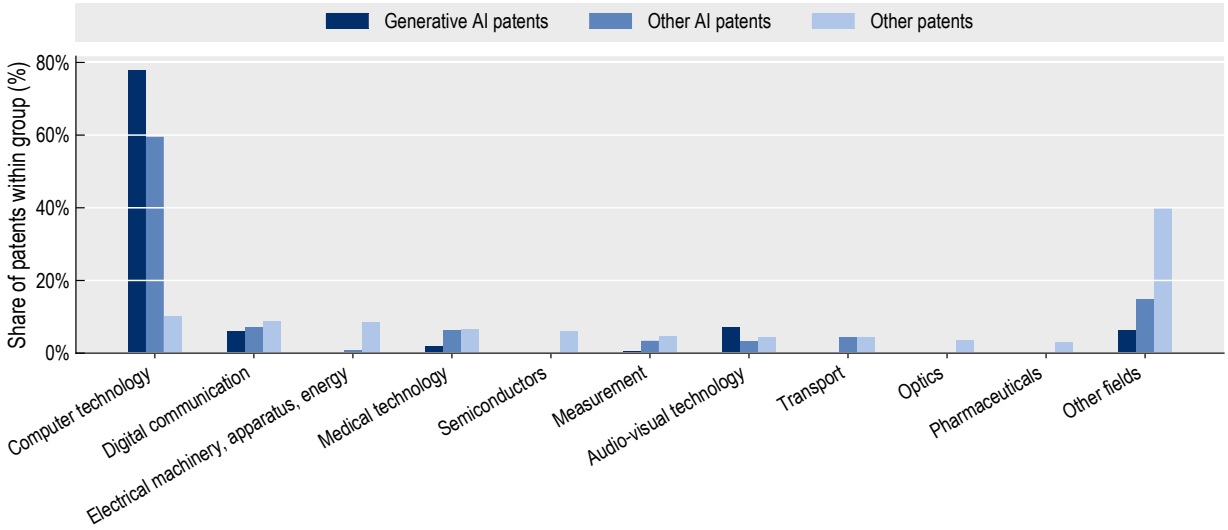
Table 3.3. Citation activity and age structure across patent groups

Group	Share of patents without forward citations	Mean age (years)
Generative AI patents	46.9%	6.2
Other AI patents	43.0%	6.7
Other patents	48.0%	7.7

Note: Shares are based on unique DOCDB patent families filed at the USPTO.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025.

33. Differences across the three groups also emerge in the technological domains in which they are active. Figure 3.4 shows that generative AI patents are overwhelmingly concentrated in computer technology, with nearly eighty percent of filings assigned to that field, while other AI patents display a similar but less strong concentration. Non-AI patents, in contrast, are distributed much more broadly across technology areas. This concentrated technological distribution highlights the specialised nature of (generative) AI patents and provides important context for analysing how innovation develops from these groups.

Figure 3.4. Distribution of WIPO technology fields per patent group



Note: Shares are based on unique DOCDB patent families filed at the USPTO. Only the ten most frequent fields are displayed for interpretability.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025. WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Restricted Use - À usage restreint

4 Innovation dynamics

34. Building on established literature, much of which has been discussed in Section 2, this study uses forward patent citation counts as a proxy for the impact of the invention on subsequent innovation. This section analyses these forward citations to compare follow-up innovation patterns across patents in the dataset.[14]

35. To construct this measure, all forward citations are retrieved from the patent database from the OECD STI Micro-data Lab. To capture the full scope of follow-up innovation, citations are collected at the level of all patent applications within each DOCDB patent family rather than focusing on only one application. Patent equivalents are also included.[15] Finally, two filters are applied during the construction of the citation dataset. First, self-citations are excluded, which are instances where a patent cites another patent by the same inventor.[16] Second, citation links are dropped where the application date of the cited patent is earlier than the application date of the citing patent, since these are likely inconsistencies that do not represent a real knowledge flow.

36. A central distinction throughout this section is between AI technology and sector or field. AI and generative AI are no patent domains that are tied to a single industry; they can in principle be applied across a wide range of technological and economical domains. By contrast, sectoral diffusion is often assessed using broad technological domains, such as WIPO technology fields or IPC codes. This distinction is important when interpreting citation patterns, as a technology might generate substantial innovation impact without immediately diffusing widely across sectors.

37. This section proceeds in three steps. First, it examines the magnitude and timing of citation impact. Second, it analyses the extent to which follow-on innovation is self-reinforcing within AI. Third, it studies the scope of diffusion by looking at the breadth of the follow-on innovation. Together these analyses shed light on how AI-driven innovation is currently characterised.

Impact on subsequent innovation: magnitude and timing

Descriptive evidence

38. This analysis begins with a descriptive investigation of the innovation dynamics of the three patent groups: generative AI, other AI, and other (non-AI) patents. Existing research shows that absolute citation counts are strongly influenced by factors such as filing year and the general citation intensity of a technology field (Bornmann, Marx and Barth, 2013[38]; Squicciarini, Dernis and Criscuolo, 2013[29]). To

[14] Henceforth, we will often refer to innovation dynamics as proxied by the observed invention patterns, building upon the discussion carried out in the previous section, keeping relevant limitations about patentability in mind.

[15] Patent equivalents refer to the same invention filed in different jurisdictions. Including those ensures that citations from outside the USPTO are also counted, which gives a more comprehensive measure of the influence of a given patent (Webb et al., 2005[49]).

[16] According to Hall, Jaffe and Trajtenberg (2001[6]), self-citations tend to reflect internal knowledge flows rather than external diffusion. As such, they might introduce bias, so they are excluded from the citation dataset in this study.

Restricted Use - À usage restreint

account for these factors, the descriptive analysis focuses on normalised forward citation counts rather than raw citation numbers. More specifically, the total number of forward citations received over the full period is divided by the average citation count of a reference group of patents filed in the same year and within the same technology field, as defined by the WIPO technology classification.[17]

39. The distribution of forward citations is highly skewed: a small fraction of patents accounts for a disproportionate share of all citations. In the dataset used in this study, the top 1 percent of patents generate more than 60 percent of total forward citations, indicating that there is a small but highly influential set of patents. To capture this, and in line with Calvino et al. (2023[9]), patents are ranked by their normalised forward citation counts, and the top 1 percent patents are flagged as core ones. This flagging facilitates citation patterns to be examined in the full dataset as well as in a version that excludes the most highly cited patents, allowing the analysis to focus on citation behaviour that is more representative of the broader population of patents.

40. In addition, grouping only by filing year and technology field may not be sufficient to ensure a fair comparison. In fact, the three technologies under scrutiny likely differ in maturity: generative AI technologies tend to be younger than other AI applications and less mature than other technologies. This is also true for the relevant patents, as shown in Table 3.3. Older technologies may naturally be better positioned to attract citations, even if patents are filed in the same year of those related to a less mature technology. Furthermore, the most recent part of the dataset is still incomplete in terms of recorded citations. Without correcting for these characteristics, citation counts may confound data availability and technological maturity/patent age. To address this, 3- and 5-year fixed citation windows are computed so that each patent is observed over the same period after its publication.[18] Patents that have not yet had enough time to complete the full 3- or 5-year citation window are excluded from the analysis, so that citation counts are not mechanically lower simply because a patent is too recent. This helps ensure comparability over time and across technologies.

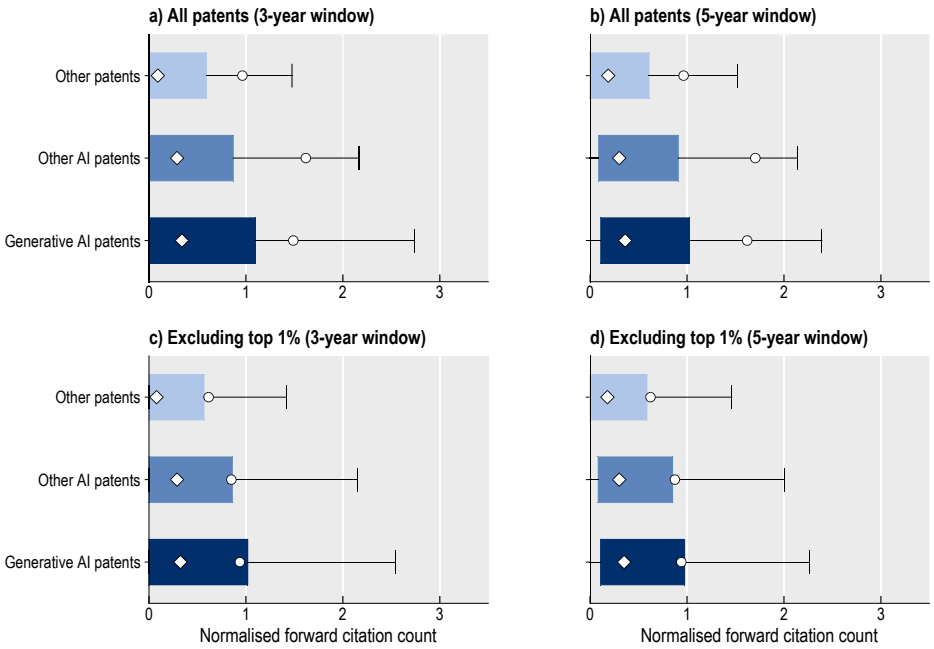
41. The distribution of normalised forward citation counts for the three patent groups is shown in Figure 4.1. The figure presents the citation dynamics across a 3- and 5-year citation window, both for the full dataset and for a subset with the top 1 percent cited patents excluded. The boxplots indicate that generative AI patents exhibit consistently higher median normalised forward citation counts compared to both other AI patents and non-AI patents. This pattern holds across both citation windows and both patent sets. For the bottom 99 percent patents more specifically, also the means are higher for the generative AI group compared to the other groups. Across all panels, the distributions are right skewed, with means above medians, which indicates that a relatively small group of highly cited patents drive the upper tail of the distribution, even after excluding the top 1 percent citations. Taken together, the descriptive evidence suggests AI is to be a high-impact technology in terms of subsequent innovation. Generative AI is at the upper end of this distribution, suggesting an additional citation premium relative to other AI inventions.

[17] WIPO published an IPC-WIPO Technology concordance table on www.wipo.int/ipstats/en/statistics/patents, allowing for a classification of all patents in each of the 35 WIPO Technology Fields according to their most frequently assigned IPC class.

[18] While the initial patent selection is based on the filing dates of the patent application, as explained in Section 3, the 3- and 5-year citation windows are calculated from the publication dates of the patents, since citations only become observable after publication and this ensures a consistent observation window.

Restricted Use - À usage restreint

Figure 4.1. Distribution of normalised forward citations



Note: This figure shows the distribution of normalised forward citations. Citation counts are normalised relative to the mean number of forward citations for patents in the same WIPO technology field and filing year. The figure reports normalised citation distributions over 3-year and 5-year citation windows. Patents that are too recent to have accumulated citations over the full 3- or 5-year window are excluded. Diamonds represent medians, and circles indicate mean values.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025. WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Model based evidence

42. The consistency of the higher generative AI citation counts across both time windows, and after removing top performing patents, suggests this pattern may reflect systematic differences in technological impact. These differences in citation impact across patent groups are further evaluated using regression models. As indicated, forward citation counts are highly skewed, with means above medians and a relatively small group of patents accounting for a large share of all citations. For this reason, an Ordinary Least Squares (OLS) regression does not appear to be the most suitable model in this setting. Standard Poisson count models are also problematic because they require the mean and variance of the count variable to be equal, whereas the citation data in this study consistently shows that variance is larger (overdispersion). Therefore, a negative binomial regression is used.[19] This model explicitly accounts for this overdispersion and is also well suited to the non-normal and highly skewed nature of the citation data.

43. The regression uses the regular 3- and 5-year forward citation counts as the dependent variable, and patents that could not have completed these citation windows are excluded. Fixed effects for filing year and WIPO technology fields ensure that observed differences in citation impact are driven by field specific or cohort specific effects. While the fixed citation windows standardise citation exposure time, filing year fixed effects additionally control for changes in baseline citation intensity across cohorts. As the regressions already account for differences across years and technology fields, raw citation counts are

[19] This approach is standard practice in existing literature (e.g. Hausman et al. (1984[50])).

used, and no separate normalisation step is required. More specifically, the following negative binomial model is estimated:

$$E[C_i|X_i] = \exp\left(\beta_0 + \beta_1 GenAI_i + \beta_2 Others_i + \sum_j \gamma_j YEAR_{ij} + \sum_k \delta_k WIPO_{ik}\right) \quad \text{Equation 1}$$

44. Where C_i is the number of forward citations received by patent i within a 3- and 5-year citation window, X_i is the vector of explanatory variables for patent i. $GenAI_i$ equals 1 if patent i is classified as a generative AI patent and 0 otherwise. $Others_i$ equals 1 if patent i is in the group of other (non-AI) patents, and 0 otherwise. Finally, $YEAR_{ij}$ is a set of indicator variables for the application filing year j, and $WIPO_{ik}$ is a set of indicator variables for each WIPO technology field k.

45. Table 4.1 reports the regression results for the 3- and 5-year citation windows. The coefficients for generative AI patents are positive and statistically significant across both windows, and the corresponding exponentiated coefficients (IRR) are above one. This indicates that, relative to other AI patents, generative AI patents receive more forward citations within both time frames considered. Other (non-AI) patents show significantly lower citation counts and IRR values below one, confirming that they have fewer citations compared to the baseline group of other AI patents. These results are consistent with the descriptive evidence and imply that AI patents tend to outperform non-AI patents in terms of forward citations, with generative AI patents showing the highest citation impact within the AI category.

Table 4.1. Regression results citation impact (1)

Citation impact of patent groups over 3- and 5-year windows

Variables	Citation count 3y	IRR 3y	Citation count 5y	IRR 5y
Group: generative AI patents	0.143***	1.154***	0.106***	1.112***
	(0.0147)	(0.0169)	(0.0183)	(0.0203)
Group: other patents	-0.463***	0.629***	-0.479***	0.619***
	(0.00604)	(0.00380)	(0.00693)	(0.00429)
Observations	2,546,532	2,546,532	1,915,584	1,915,584
Filing Year FE	Yes		Yes	
WIPO Field FE	Yes		Yes	
AIC	1.010e+07		1.040e+07	
log-likelihood	-5.071e+06		-5.189e+06	
Pseudo R2	0.0518		0.0364	

Note: This table presents results from negative binomial regressions with a log link, using other (non-generative) AI patents as the baseline group. The dependent variable is the forward citation count. Patents that are too recent to have accumulated citations over the full 3- or 5-year window and the top 1 percent most-cited patents are excluded. All models include filing year and WIPO field fixed effects. Exponentiated coefficients (IRR) are reported and indicate the expected proportional change in citation counts relative to the baseline group. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Restricted Use - À usage restreint

46. As an additional robustness check, the specification from Equation 1 is re-estimated by allowing filing year fixed effects to vary by WIPO technology field through interacted filing year and WIPO field fixed effects. The main results from the baseline regressions remain largely unchanged, indicating that the results are not driven by different citation time trends across technologies. The full regression results are reported in Table B.1.

47. Table 4.2 extends the regression from Equation 1 by adding patent- and inventor-level control variables. First of all, geographic controls are added to the model, because citation behaviour and patent impact vary systematically across location. Besides the United States and the People's Republic of China (hereafter "China"), this is captured using grouped country shares rather than individual country indicators, both to limit model complexity and to reflect systematic geographic differences in inventive activity across major regions rather than specific countries.[20] In addition, a measure for the number of inventor countries is also included, which captures whether inventive activity took place within a single country or whether it is spread across multiple countries.

Table 4.2. Regression results citation impact (2)

Citation impact of patent groups over 3- and 5-year windows, including control variables

Variables	Citation count 3y	IRR 3y	Citation count 5y	IRR 5y
Group: generative AI patents	0.126***	1.135***	0.0845***	1.088***
	(0.0145)	(0.0164)	(0.0181)	(0.0197)
Group: other patents	-0.465***	0.628***	-0.475***	0.622***
	(0.00578)	(0.00363)	(0.00662)	(0.00412)
Backward citations (log)	0.323***	1.382***	0.300***	1.351***
	(0.00136)	(0.00187)	(0.00140)	(0.00190)
Inventor team size (log)	0.187***	1.206***	0.169***	1.184***
	(0.00288)	(0.00347)	(0.00304)	(0.00359)
Inventor share: United States	0.520***	1.683***	0.550***	1.732***
	(0.00360)	(0.00605)	(0.00377)	(0.00653)
Inventor share: China	-0.280***	0.755***	-0.365***	0.695***
	(0.00640)	(0.00484)	(0.00738)	(0.00512)
Inventor share: Europe	-0.200***	0.819***	-0.169***	0.844***
	(0.00538)	(0.00441)	(0.00556)	(0.00470)
Number of inventor countries	0.190***	1.209***	0.203***	1.225***
	(0.00442)	(0.00534)	(0.00476)	(0.00583)
Observations	2,546,435	2,546,435	1,915,549	1,915,549
Filing Year FE	Yes		Yes	
WIPO Field FE	Yes		Yes	
AIC	9.863e+06		1.010e+07	
log-likelihood	-4.932e+06		-5.065e+06	

[20] In the *Europe* and the *other countries* groups, citation levels differ does differ across economies, but these levels are spread across several large countries rather than concentrated in countries with systematically high or low citation rates. This suggests that within-group country heterogeneity does not bias the regression results.

Restricted Use - À usage restreint

Pseudo R2	0.0777		0.0594	

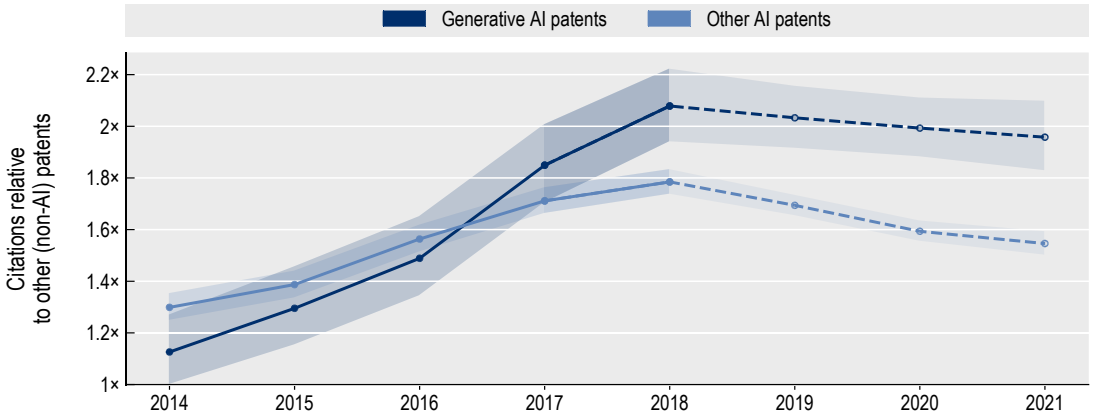
Note: This table presents results from negative binomial regressions with a log link, using other AI patents (non-generative) as the baseline group. The dependent variable is the forward citation count. The regressions exclude patents that are too recent to have accumulated citations over the full 3- or 5-year window and the top 1% most-cited patents. All models include filing year and WIPO field fixed effects. Exponentiated coefficients (IRR) are reported and indicate the expected proportional change in citation counts relative to the baseline group. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

48. After adding the full set of control variables to the model, the results remain largely unchanged. Non-AI patents continue to receive significantly fewer forward citations than AI patents, and standard controls such as backward citations, team size, and inventor location present expected signs. Once these controls are included, the coefficients for generative AI also remain significant, although they become slightly smaller.

Temporal dynamics

To examine the timing of citation differences more directly, this subsection examines how relative forward citation counts evolve over the filing years. Rather than focusing on averages over the full sample period, the analysis allows citation differences between patent groups to vary by filing year. The negative binomial regressions from Equation 1 are therefore re-estimated with interactions between patent group indicators and filing year fixed effects. Using the resulting estimated coefficients, year-specific relative citation counts for each patent group are calculated with respect to the baseline group of other (non-AI) patents, together with their 95 percent confidence intervals. The results of this analysis are presented in Figure 4.2.

Figure 4.2. Evolution of 3-year forward citation differences across patent groups



Note: This figure shows differences in expected 3-year forward citation counts for generative AI and other AI patents relative to a baseline group of all other patents. Estimates are derived from negative binomial regressions with a log link that interact patent group indicators with filing year. Values above 1 indicate that patents in the corresponding group receive more citations, on average, than the baseline group in the same year. Shaded areas represent 95 percent confidence intervals, and dashed segments indicate filing post-2018 filing years for which citations may be incomplete due to publication and citation lags. The sample excludes patents that are too recent to have accumulated citations over the full 3-year window and the top 1 percent most cited patents. All specifications include control variables and filing year and WIPO technology field fixed effects.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Restricted Use - À usage restreint

49. For the early part of the sample, Figure 4.2 shows that generative AI patents do not receive significantly more forward citations than other AI patents. A statistically significant citation premium is visible starting between 2016 and 2017, which increases until 2018, and then remains stable in subsequent filing years. In the later years of the period, generative AI patents receive more than twice as many forward citations as non-AI patents on average. Other AI patents also exhibit a positive citation premium compared to non-AI throughout the period, but this premium is smaller and stabilises at a slightly lower level after 2018. However, forward citations measured over a 3-year window for the most recent filing years should be interpreted with some caution. Although the dataset includes patents up to the end of 2024, some citations that fall within the 3-year window may not yet be observed due to publication lags and delays in citation recording. This might lead to undercounting forward citations for these most recent years. Overall, these results indicate that the higher citation impact of generative AI patents is a relatively recent development. This timing helps explain why the generative AI coefficient becomes smaller in longer citation windows, as the years for which the citation premium are strongest are only partially observed when using longer citation windows.

50. Figure C.1 in Annex C presents a further extension of this analysis: including several WIPO technology fields in the patent group indicator allows placing the findings in perspective with respect to other technologies. In this analysis, the patent groups are constructed hierarchically and mutually exclusive: patents are first classified as generative AI or other AI patents, and patents that are not assigned to either AI category are subsequently classified into selected WIPO technology fields. All remaining patents form the baseline group. This hierarchical assignment ensures that the WIPO technology groups capture patents in those fields that do not involve AI technologies. Furthermore, this approach reflects the fact that AI technologies may exist across many technological fields, whereas WIPO technology fields are intended to capture sector-specific innovation. Looking at other technology fields, Figure C.1 indicates that citation premia are not unique to AI, but their size and timing differ. Digital communications and semiconductors show substantially higher citation levels than the baseline throughout the period, with a peak around 2018 followed by a mild decline. The comparatively lower citation premium observed for semiconductors may partly reflect the fact that this is a more mature GPT, whose major diffusion phase predates the 2014-2024 observation window. Mechanical engineering, by contrast, remains consistently below the baseline level, with no citation premium. Overall, while several technology fields attract more citations than average, the sharp and relatively sudden rise in generative AI citations stands out as both larger and more recent than in these other fields.

The characteristics subsequent innovation

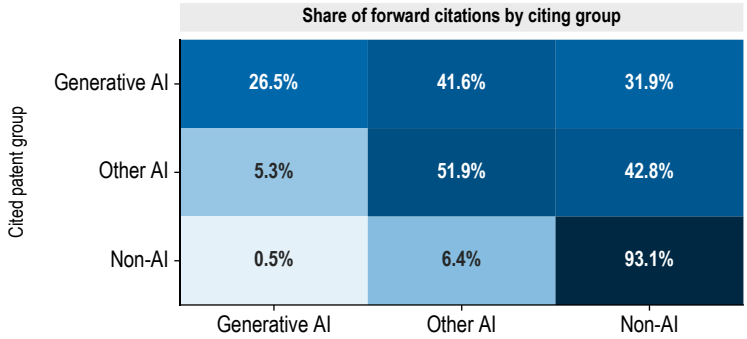
51. The analysis so far has established that AI-related inventions generate substantially more follow-on innovation than non-AI patents, and that generative AI shows a recent acceleration within these broader innovation dynamics. This section further studies the characteristics of subsequent innovation.

Is subsequent innovation AI-related?

52. While AI and generative AI are no patent domains but rather techniques that can be applied across many technological domains, it is still informative to examine the type of follow-on innovation they generate. Figure 4.3 addresses this by showing that, while follow-on innovation spawned by AI-related patents is predominantly AI-related, a considerable share of citations (more than 40% for other AI patents) comes from non-AI patents.. This indicates on the one hand that AI technologies build upon themselves, consistently with their considerable improvement over time. On the other hand, it also suggests that – even in this early stage of technological development – AI innovations are relevant for a considerable share of non-AI ones. In contrast, non-AI patents are cited almost exclusively by other non-AI patents, but this is not surprising given the limited relevance of several non-AI technology for AI inventions.

Restricted Use - À usage restreint

Figure 4.3. AI relatedness of follow-on innovation



Note: The figure reports the percentage distribution of forward citations received within three years of filing, by the AI classification of the citing patent. Rows indicate the technology group of the cited patent, while columns indicate the technology group of the citing patent. Shares sum to 100 per cited patent group

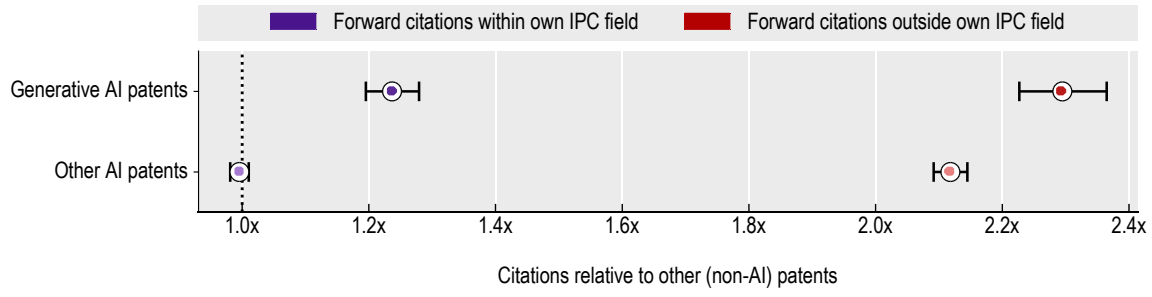
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Direction of subsequent innovation

53. Beyond the technological classification of citing patents, it is informative to examine where follow-on innovation takes place relative to the cited patent's own technological field. To do so, the analysis distinguishes between forward citations originating from within the same IPC4 subclass as the cited patent and those coming from outside that subclass. This distinction is useful because within-field citations are more likely to reflect cumulative development within a narrow technological area, while outside-field citations capture the spread of ideas across different technical domains.

54. Figure 4.4 presents the estimates from negative binomial regressions comparing within-field and outside-field forward citation counts for generative AI and other AI patents relative to non-AI patents. Generative AI patents receive more citations than non-AI patents both within and outside their own IPC4 field, but the relative advantage is substantially larger for citations originating outside the field. Within-field citations for generative AI patents are almost 1.3 times higher than for non-AI patents, whereas outside-field citations are more than twice as high. For other AI patents, the within-field citation premium is small and close to zero, while the outside-field citation premium is positive but smaller than for generative AI patents. These results indicate that the higher citation impact of generative AI and other AI patents are driven primarily by citations originating outside their core technological field, pointing to a broader pattern of technological influence and innovation spawning compared with non-AI patents.

Figure 4.4. Within versus outside field



Note: This figure shows differences in expected 3-year forward citation counts received within and outside a patent's own IPC4 field for generative AI and other AI patents, relative to non-AI patents. Estimates are derived from separate negative binomial regressions with a log link for within field and outside field citations. All models include filing year fixed effects, WIPO technology field fixed effects, and the full set of control variables. Values above 1 indicate that patents in the corresponding group receive more citations, on average, than non-AI patents. Horizontal bars represent 95 percent confidence intervals. The sample excludes that are too recent to have accumulated citations over the full 3-year window and the top 1 percent most cited patents.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Breadth of subsequent innovation

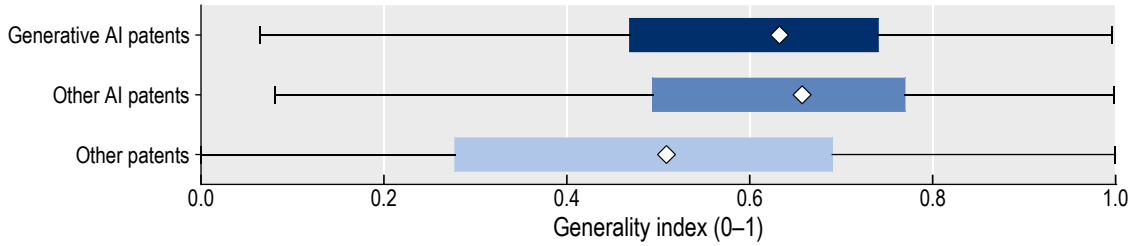
Generality of subsequent innovation

55. Forward citation counts alone provide limited information about the nature of the influence of patents on subsequent innovation. While the previous analyses characterised the volume, direction, and distribution of cross-field citations, it does not indicate whether subsequent innovation is concentrated or spread within those sections at a more granular level of technological detail. To address this, the analysis is extended by focusing on the generality index.

56. Following Trajtenberg et al. (1997[39]) and Squicciarini, Dernis and Criscuolo (2013[29]), the generality index is a way to measure the technological and economic value of a patent. This index builds on a modification of the Hirschman-Herfindahl Index and uses information on both the number and the technological distribution of forward citations received by a patent. The measure therefore captures whether follow-up inventions build on the focal patent within a small set of closely related technologies, or across a wider range of technological fields. The generality index is constructed at the 4-digit IPC level and uses forward citations within a five-year window. For each citing patent, all assigned IPC symbols are observed on a 4-digit level. To ensure that citing patents with multiple IPC classifications do not inflate the measured breadth, the calculation of the generality index is based on a fractional approach using their relative frequency.

57. Higher values of the generality index indicate that a patent is cited by subsequent inventions that span a wider range of IPC4 subclasses, while lower values reflect more concentrated follow-up innovation. For the dataset analysed in the current study, Figure 4.5 shows that both generative AI and other AI patents have higher median generality scores than non-AI patents. This suggests that follow-up innovation building on AI-related inventions spans a wider range of fields.

Figure 4.5. Distribution of generality by patent group



Note: This figure shows the distribution of the generality index by patent group. Diamonds represent median values.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025.

58. To assess whether the observed differences in technological breadth across patent groups reflect systematic variation rather than differences in filing years, technology fields or citation intensity, a regression analysis is performed. As indicated, the generality index is bounded between zero and one by construction. Although bounded outcomes may raise concerns for linear regression models, the distribution of generality scores in this dataset shows that there is substantial variation across the full range, with no clustering at the lower- or upper bounds. Therefore, an OLS estimation provides a reasonable baseline and allows the coefficients' interpretation to be straightforward. In addition, this approach is consistent with the patent literature (e.g. Hall, Jaffe and Trajtenberg (2001[6])).

59. The regression uses the generality index as the dependent variable, and patents that have not completed a full 5-year forward citation window are excluded. This restriction reflects the construction of the generality index, which is based on forward citations counts within 5-years, too. Fixed effects for filing year and WIPO technology fields ensure that the differences in generality are not driven by variation across years or technology fields.[21] More specifically, the following model is estimated:

$$Generality_i = 1 - \sum_{h=1}^{M_i} \left(\frac{C_{ih}}{\sum_h C_{ih}} \right)^2 \qquad \text{Equation 2}$$

$$Generality_i = \beta_0 + \beta_1 GenAI_i + \beta_2 Others_i + \sum_j \gamma_j YEAR_j + \sum_k \delta_k WIPO_k \qquad \text{Equation 3}$$

60. In Equation 2, the dependent variable is the generality score of patent i. In this expression, h denotes the 4-digit IPC subclasses, and M_i is the number of different technology fields from which patent i receives at least one citation. C_{ih} represents the number of forward citations to patent i coming from IPC field h, and the term $\sum_h C_{ih}$ represents the total number of citations received by patent i across all fields. Equation 3 shows the regression model, where the dependent variable is the generality score. $GenAI_i$ takes the value 1 if patent i is classified as a generative AI patent and 0 otherwise, while $Others_i$ equals 1 for the group of other patents. In addition, $YEAR_j$ is a set of indicator variables for the application filing year, and $WIPO_k$ is a set of indicator variables for the WIPO technology field.

Table 4.3. Regression results generality index

Generality index and patent groups

Variables	Generality (1)	Generality (2)
Group: generative AI patents	-0.00871***	-0.0109***
	(0.00220)	(0.00219)
Group: other patents	-0.166***	-0.165***
	(0.000812)	(0.000810)
Backward citations (log)		0.0110***
		(0.000181)

[21] No interactions between year and WIPO technology field fixed effects are included. Year fixed effects account for changes in citations over time, while WIPO field fixed effects capture differences across technologies. Since the generality index reflects the distribution of citations across fields rather than citation counts, interacting year and field fixed effects would not provide additional variation.

Restricted Use - À usage restreint

Inventor team size (log)		0.00176***
		(0.000402)
Inventor share: United States		0.0439***
		(0.000500)
Inventor share: China		-0.0199***
		(0.000938)
Inventor share: Europe		0.00422***
		(0.000708)
Number of inventor countries		0.00251***
		(0.000613)
Constant	0.654***	0.605***
	(0.00125)	(0.00153)
Observations	1,548,834	1,548,810
R-squared	0.059	0.068
Filing Year FE	Yes	Yes
WIPO Field FE	Yes	Yes

Note: This table presents results from an OLS regression, using AI patents (non-generative) as the baseline group. The dependent variable is the patent generality index. The regressions exclude patents that are too recent to have accumulated citations over the full 5-year window and the top 1% most-cited patents. Both models include filing year and WIPO field fixed effects. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

61. Table 4.3 reports the regression results for the generality index over a 5-year citation window. Both generative AI and other AI patents have significantly higher generality scores than non-AI patents, indicating that follow-up innovation building on AI-related inventions is more technologically diverse. The difference between generative AI and other AI patents is significant but very small, suggesting that the main distinction in technological breadth lies between AI and non-AI patents rather than within the AI group.

Comparison with other technologies

62. To place these results in perspective, Figure C.2 in Annex C compares estimated differences in generality across groups relative to a baseline of all other patents. The patent groups are constructed to be hierarchical and mutually exclusive, in line with the earlier comparison between several technologies shown in Figure C.1 in Annex C. Generative AI and other AI patents are excluded from the remaining technology groups by construction: as a result, the digital communications, semiconductors, and mechanical engineering categories shown in the figure represent non-AI patents within those technological domains. All remaining patents form the baseline. Against this baseline, both generative AI and other AI patents stand out with substantially higher generality scores. Non-AI mechanical engineering patents have a positive but much smaller difference relative to the baseline, non-AI digital communications patents have no difference, and non-AI semiconductors show negative differences, indicating that follow-on innovation in these non-AI fields is more concentrated at the IPC4 level. This comparison helps to place AI-related patents within a broader technological landscape and shows that their higher generality is not a mechanical feature of all technologically advanced fields. At the same time, the interpretation of these differences requires caution. The generality index is constructed at the IPC4 level, which captures relatively fine-grained technological distinctions. As a result, higher generality scores often reflect recombination across closely related subclasses rather than across fundamentally distinct technological domains.

Restricted Use - À usage restreint

5 Accelerated innovation

63. Studying direct citation impact aims at capturing the relevance of certain patents for subsequent innovation. However, it does not reveal how innovation dynamics evolve after that. To study these deeper technological dynamics, this study further focuses on feedback chains in the citation network. A feedback chain is defined as a citation sequence where an initial patent influences an intermediary patent in another domain, which is subsequently cited by a newer patent in the same domain as the initial patent. In this case, knowledge does not just flow outward to another field but becomes part of its further development. Detecting these chains changes the focus of the analysis from measuring direct impact to a broader perspective. This network perspective makes it possible to track how knowledge flows and re-enters the domain where it started, reflecting the broader shift in the literature from counting citations to examining how they structure technological changes over time (Martinelli and Nomaler, 2014[40]). If AI patents repeatedly appear in these recursive chains, this suggests they follow a self-reinforcing trajectory even beyond the one depicted by the direct AI-to-AI citations. As discussed in Section 2, this is one of the defining features of a GPT (Calvino, Haerle and Liu, 2025[2]; Bresnahan, 2024[25]).

Identifying feedback chains

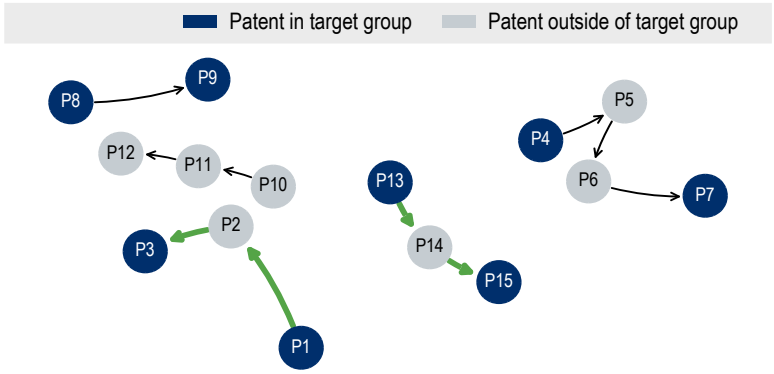
64. In the previous subsection, single-step citation links were used to quantify direct subsequent innovation. In contrast, this section focuses on recursive feedback chains, where knowledge originating in one domain returns to that domain after passing through an intermediate patent in another domain. To ensure the interpretability of these flows, the path length is restricted to three patents. This restriction on path length is in line with existing work on patent citation networks. For example, Jaffe, Trajtenberg and Fogarty (2000[41]) show that, while a citation link significantly increases the likelihood that inventors report a knowledge dependence, up to half of all citation links do not correspond to any perceived knowledge flow. This indicates that interpretability decreases quickly once multiple citation steps are accumulated. In addition, Atallah and Rodríquez (2013[42]) calculate the cumulative influence index, which aggregates all indirect citations along a citation chain, and show that meaningful influence is concentrated in the very first few citation steps before it decreases to zero.

65. To detect feedback chains, the dataset of citation links is used to track knowledge flows between individual patents. [22] A chain is identified as a sequence that originates from a target patent and concludes when a subsequent target patent reconnects to the chain. The definition requires that only the first and last patents in this sequence belong to the target group, while all intermediate patents must be non-target patents. To ensure the analysis captures only forward knowledge flows, every link within the path must follow a consistent chronological order based on filing dates. Figure 5.1 shows a stylised example of several citation chains, with the feedback chains highlighted by green arrows between the patents.

[22] As with the construction of measures used in the previous section, the citation data are cleaned by excluding self-citations and by removing citation links with inconsistent date sequences. Furthermore, the top 1 percent cited patents are excluded, as well as citation links that fall outside of the 3-year citation window.

Restricted Use - À usage restreint

Figure 5.1. Example of a patent citation network with feedback chains



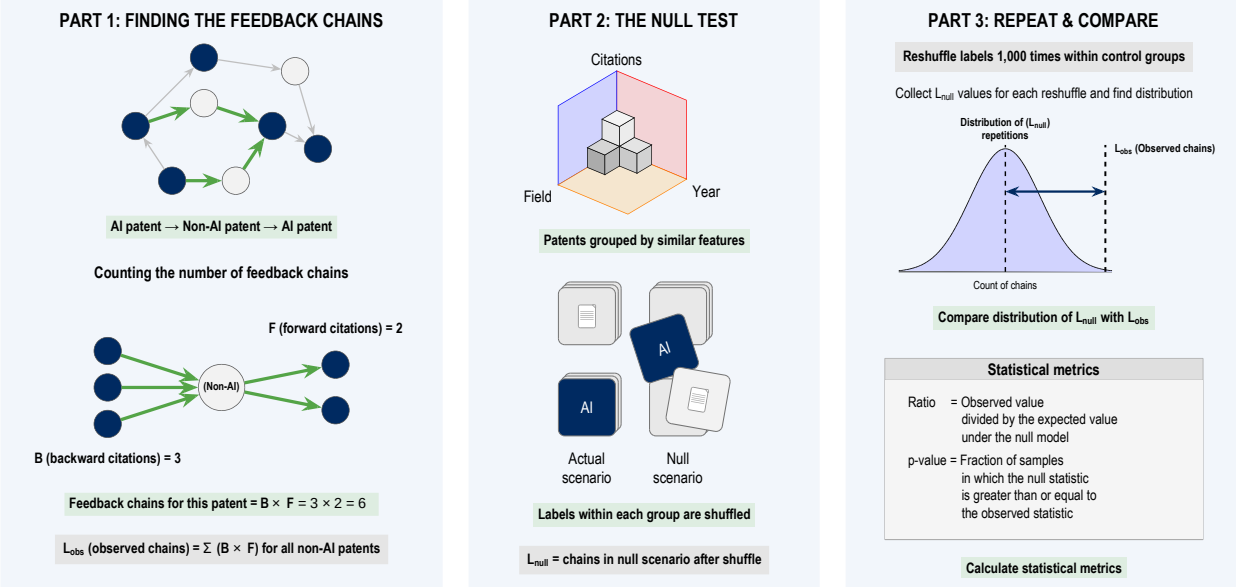
Note: This figure illustrates a simplified patent citation network highlighting the interaction between target and non-target patents, using synthetic data for illustrative purposes. Feedback chains are depicted with thicker green edges to show the specific citation chains where a target innovation influences subsequent developments through intermediate non-target nodes.

How exceptional are feedback chains?

66. To draw meaningful conclusions from observed citation chains, it is necessary to establish whether they are exceptional relative to a baseline. Simply counting the number of chains and comparing them across technologies would be straightforward, but that would be of limited information given the structural properties of citation networks. Patents observed in citation networks are often driven by the network's own structure. In particular, more citation chains are naturally present for larger technologies, because more patents create more opportunities for citations. In addition, older fields have had more time to accumulate links. Furthermore, variation in citations across fields and in the density of local citation networks introduce additional difficulties. A meaningful assessment of the exceptionality of feedback chains therefore requires benchmarking observed chain counts against an explicit null model that accounts for these factors, so that any excess feedback can be attributed to distinctive technological complementarities rather than to age, size, or popularity.

67. To evaluate whether a specific technological group shows an unusually high level of knowledge reuse, this methodology leverages a framework based on feedback chain counts. A visualization of this framework is presented in Figure 5.2. First, citation chains are identified in which a patent from the target group is cited by a non-target bridging patent and then cited again by another target patent. The total number of observed chains (L_{obs}) is calculated by summing, across all central patents, the products of backward and forward citations involving the target group. Next, a benchmark is constructed (L_{null}) by randomly reshuffling technological labels, while holding key patent characteristics fixed. Labels are reassigned only within groups of patents that are similar in application year, technological field, and citation structure. In this way, size, age, and citation profile of each technology are preserved, while any systematic association between patents and their original labels is removed. This procedure is repeated 1 thousand times to generate a null distribution. This allows the observed number of chains to be compared to what would be expected randomly, making it possible to compute an effect-size ratio (L_{obs}/L_{null}) and a p-value, and to assess whether the observed level of feedback chains likely reflects a technological feature or not.

Figure 5.2. Visualisation of the comparison process for feedback chains



Note: This figure illustrates the methodology used to identify the feedback chains and test how exceptional they are compared to a null scenario. The process illustrated in this figure shows the feedback chains for AI patents, but the same framework can be applied to other technological categories in similar fashion.

Analysing the indirect self-reinforcing nature of AI patents

68. First, this analysis focuses on the full spectrum of AI patents, which includes both generative AI as well as other AI patents. This group consists of feedback chains in which an initial AI patent is cited by a non-AI patent, which is then cited again by a subsequent AI patent (AI → non-AI → AI). As indicated above, the aim is to assess whether these chains occur more often than would be expected by chance. In the null model that we compare it with, the AI and non-AI labels are randomly reassigned across patents that share the same application year, WIPO technology field, and citation profile.

Table 5.1. Number of feedback chains

	L_{obs}	Mean(L_{null})	Ratio	p-value
All AI feedback chains	137,857	76,869	1.793	p < 0.001

Note: This table reports the observed number of AI → Non-AI → AI feedback chains in the patent citation network and compares it to the expected number obtained from the null benchmark.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

69. The results of this exercise are reported in Table 5.1. In the observed citation network, a total of just under 138 thousand feedback chains have been found. This is substantially higher than the average of almost 77 thousand chains obtained from all repetitions of the null baseline. The resulting ratio of 1.793 indicates that these recursive knowledge flows occur almost 1.8 times more frequently than would be expected under the null model. The associated p-score ($p < 0.001$) confirms that this difference is highly statistically significant. Taken together, these findings suggest that AI patents are disproportionately involved in feedback chains.

Restricted Use - À usage restreint

Decomposition generative AI and Other AI

70. Ongoing work: this section aims to study the feedback chains for generative AI and other AI technologies specifically.

Comparison with other technologies

71. Ongoing work: this section aims to compare the presence of feedback chains for AI technologies with other technology fields.

Characteristics of knowledge flows

72. Ongoing work: this section aims to study the characteristics of the patents that are present in the feedback chains.

Restricted Use - À usage restreint

6 Concluding remarks

73. Ongoing work.

Restricted Use - À usage restreint

References

Agrawal, A., J. McHale and A. Oettl (2023), "Superhuman science: How artificial intelligence may impact innovation", *Journal of Evolutionary Economics*, Vol. 33/5, pp. 1473-1517, https://doi.org/10.1007/S00191-023-00845-3. [23]

Aranda, L. et al. (2025), "Identifying emerging AI technologies using patent data", *OECD Science, Technology and Industry Working Papers, forthcoming*. [28]

Atallah, G. and G. Rodríguez (2013), "Indirect patent citations", *Scientometrics*, Vol. 67/3, pp. 437-465, https://doi.org/10.1556/Scient.67.2006.3.7. [42]

Baruffaldi, S. et al. (2020), "Identifying and measuring developments in artificial intelligence: Making the impossible possible", *OECD Science, Technology and Industry Working Papers*, No. 2020/05, OECD Publishing, Paris, https://doi.org/10.1787/5f65ff7e-en. [31]

Bianchini, S., M. Müller and P. Pelletier (2022), "Artificial intelligence in science: An emerging general method of invention", *Research Policy*, Vol. 51/10, p. 104604, https://doi.org/10.1016/j.respol.2022.104604. [24]

Biggi, G. et al. (2025), "Green intelligence: the AI content of green technologies", *Eurasian Business Review*, pp. 1-38, https://doi.org/10.1007/S40821-024-00288-1/TABLES/12. [19]

Bornmann, L., W. Marx and A. Barth (2013), "The normalization of citation counts based on classification systems", *Publications*, Vol. 1/2, pp. 78-86, https://doi.org/10.3390/publications1020078. [38]

Bresnahan, T. (2024), "What innovation paths for AI to become a GPT?", *Journal of Economics & Management Strategy*, Vol. 33/2, pp. 305-316, https://doi.org/10.1111/JEMS.12524. [25]

Bresnahan, T. and M. Trajtenberg (1995), "General purpose technologies 'Engines of growth'?", *Journal of Econometrics*, Vol. 65/1, pp. 83-108, https://doi.org/10.1016/0304-4076(94)01598-t. [1]

Calvino, F. et al. (2023), "What technologies are at the core of AI?: An exploration based on patent data", *OECD Artificial Intelligence Papers*, No. 6, OECD Publishing, Paris, https://doi.org/10.1787/32406765-en. [9]

Calvino, F., D. Haerle and S. Liu (2025), "Is generative artificial intelligence a general-purpose technology? Implications for productivity and policy", *OECD Artificial Intelligence Papers*, No. 40. [2]

Restricted Use - À usage restreint

Calvino, F., J. Reijerink and L. Samek (2025), "The effects of generative AI on productivity, innovation and entrepreneurship", *OECD Artificial Intelligence Papers*, No. 39, OECD Publishing, Paris, https://doi.org/10.1787/b21df222-en. [15]

Cockburn, I., R. Henderson and S. Stern (2018), "The Impact of Artificial Intelligence on Innovation", https://doi.org/10.3386/W24449. [20]

Cockburn, I. and M. MacGarvie (2011), "Entry and Patenting in the Software Industry", *Management Science*, Vol. 57/5, pp. 915-933, https://doi.org/10.1287/MNSC.1110.1321. [12]

Cohen, W., R. Nelson and J. Walsh (2000), "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)", https://doi.org/10.3386/W7552. [27]

Damioli, G. et al. (2025), "Is artificial intelligence leading to a new technological paradigm?", *Structural Change and Economic Dynamics*, Vol. 72, pp. 347-359, https://doi.org/10.1016/J.STRUECO.2024.12.006. [21]

Dernis, H. et al. (2019), "World Corporate Top R&D investors: Shaping the Future of Technologies and of AI", https://doi.org/10.2760/472704. [11]

Dernis, H. et al. (2021), "Who develops AI-related innovations, goods and services? : A firm-level analysis", *OECD Science, Technology and Industry Policy Papers*, No. 121, OECD Publishing, Paris, https://doi.org/10.1787/3e4aedd4-en. [18]

Devlin, J. et al. (2019), "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1/Long and Short Papers, pp. 4171-4186, https://doi.org/10.18653/V1/N19-1423. [43]

European Patent Office (2025), *PATSTAT: The EPO Worldwide Patent Statistical Database, Spring 2025 edition*. [26]

Fujii, H. and S. Managi (2018), "Trends and priority shifts in artificial intelligence technology invention: A global patent analysis", *Economic Analysis and Policy*, Vol. 58, pp. 60-69, https://doi.org/10.1016/J.EAP.2017.12.006. [17]

Giczy, A., N. Pairolero and A. Toole (2022), "Identifying artificial intelligence (AI) invention: a novel AI patent dataset", *Journal of Technology Transfer*, Vol. 47/2, pp. 476-505, https://doi.org/10.1007/S10961-021-09900-2/TABLES/6. [32]

Haghighian Roudsari, A. et al. (2021), "PatentNet: multi-label classification of patent documents using deep learning based language understanding", *Scientometrics 2021 127:1*, Vol. 127/1, pp. 207-231, https://doi.org/10.1007/S11192-021-04179-4. [48]

Hall, B., A. Jaffe and M. Trajtenberg (2001), "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools", https://doi.org/10.3386/W8498. [6]

Hausman, J. et al. (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", https://EconPapers.repec.org/RePEc:nbr:nberte:0017 (accessed on 26 January 2026). [50]

Hötte, K. et al. (2022), "Measuring artificial intelligence: a systematic assessment and implications for governance", https://arxiv.org/pdf/2204.10304 (accessed on [3]

Restricted Use - À usage restreint

27 January 2026).

Hughes, A. and A. Mina (2010), "Impact of the Patent System on SMEs", *Working Paper*, No. 411, Centre for Business Research, University of Cambridge. [10]

IFI CLAIMS Patent Services (2024), *2024 Trends & Insights*, https://www.ificlaims.com/rankings/trends-2024 (accessed on 13 October 2025). [44]

Iori, M., A. Martinelli and A. Mina (2021), *The direction of technical change in AI and the trajectory effects of government funding*, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy. [5]

Jaffe, A. and G. Rassenfosse (2016), "Patent Citation Data in Social Science Research: Overview and Best Practices", https://doi.org/10.3386/W21868. [8]

Jaffe, A., M. Trajtenberg and M. Fogarty (2000), "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors", *American Economic Review*, Vol. 90/2, pp. 215-218, https://doi.org/10.1257/AER.90.2.215. [41]

Japan Patent Office (2024), "Recent Trends in AI-related Inventions". [45]

Lee, J. and J. Hsiang (2020), "Patent classification by fine-tuning BERT language model", *World Patent Information*, Vol. 61, p. 101965, https://doi.org/10.1016/J.WPI.2020.101965. [47]

Liu, J. et al. (2020), "Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors", *Technological Forecasting and Social Change*, Vol. 158, p. 120142, https://doi.org/10.1016/J.TECHFORE.2020.120142. [22]

Liu, N. et al. (2021), "Mapping technological innovation dynamics in artificial intelligence domains: Evidence from a global patent analysis", *PLOS ONE*, Vol. 16/12, p. e0262050, https://doi.org/10.1371/JOURNAL.PONE.0262050. [14]

Martinelli, A. and Ö. Nomaler (2014), "Measuring knowledge persistence: A genetic approach to patent citation networks", *Journal of Evolutionary Economics*, Vol. 24/3, pp. 623-652, https://doi.org/10.1007/s00191-014-0349-5. [40]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en. [34]

OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *OECD Digital Economy Papers*, No. 352, OECD Publishing, Paris, https://doi.org/10.1787/13d38f92-en. [36]

OECD (2019), *Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [33]

OECD (2009), *OECD Patent Statistics Manual*, OECD Publishing, Paris, https://doi.org/10.1787/9789264056442-en. [7]

Santarelli, E., J. Staccioli and M. Vivarelli (2023), "Automation and related technologies: a mapping of the new knowledge base", *Journal of Technology Transfer*, Vol. 48/2, pp. 779-813, https://doi.org/10.1007/s10961-021-09914-w. [16]

Schmoch, U. (2008), "Concept of a Technology Classification for Coun-try Comparisons 1 [46]

Restricted Use - À usage restreint

Final Report to the World Intellectual Property Organisation (WIPO)".

Squicciarini, M., H. Dernis and C. Criscuolo (2013), "Measuring Patent Quality: Indicators of Technological and Economic Value", *OECD Science, Technology and Industry Working Papers*, No. 2013/3, OECD Publishing, Paris, https://doi.org/10.1787/5k4522wkw1r8-en. [29]

Srebrovic, R. and J. Yonamine (2020), *Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery*, https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html (accessed on 7 October 2025). [35]

Trajtenberg, M., A. Jaffe and R. Henderson (1997), "University versus Corporate Patents: A Window on", *Economics of Innovation and New Technology*, Vol. 5/1, pp. 19-50. [39]

Tseng, C. and P. Ting (2013), "Patent analysis for technology development of artificial intelligence: A country-level comparative study", *Innovation*, Vol. 15/4, pp. 463-475, https://doi.org/10.5172/IMPP.2013.15.4.463. [4]

Van Roy, V., D. Vertesy and G. Damioli (2019), "AI and Robotics Innovation: a Sectoral and Geographical Mapping using Patent Data". [13]

Webb, C. et al. (2005), "Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks", *OECD Science, Technology and Industry Working Papers*, No. 2005/9, OECD Publishing, Paris, https://doi.org/10.1787/883002633010. [49]

WIPO (2024), *Patent Landscape Report - Generative Artificial Intelligence (GenAI)*, World Intellectual Property Organization, https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/index.html. [37]

WIPO (2019), *WIPO Technology Trends 2019 - Artificial Intelligence*, World Intellectual Property Organization, https://doi.org/10.34667/TIND.29084. [30]

Restricted Use - À usage restreint

Annex A. Labelling patent datasets

Box A.1. Prompt for labelling AI patents

The OECD uses the following definition for artificial intelligence (AI):

"*An AI system is a machine based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*."

You are an expert in AI and patents. Based on this definition, decide whether a patent abstract describes an AI invention.

- Classify as AI (1) if the abstract describes or clearly implies AI methods, AI models, or AI systems for learning, reasoning, perception, prediction, decision making, or content generation. This includes classical machine learning methods such as clustering, decision tree learning or support vector machines when they involve training or learning a model from data. It further includes AI methods such as deep learning, neural networks, natural language processing, computer vision, reinforcement learning, and AI based methods for recognition, detection, classification, pattern analysis, or optimisation. Generative AI (genAI) systems, such as AI models for text generation, image synthesis, or audio and video creation, are a subset of AI and should also be included.
- Classify as not AI (0) if the abstract only describes fixed algorithms, simple statistical formulas, data transmission, or mechanical processes without evidence of model based learning, adaptive behaviour, or content generation.

Return a single JSON object exactly in this format:

{ "is_ai": 1 or 0, "certainty": 0.0 to 1.0 }

Box A.2. Alternative prompt for labelling AI patents

The OECD uses the following definition for artificial intelligence (AI):

"*An AI system is a machine based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*."

You are an expert in AI and patents. Based on this definition, decide whether a patent abstract describes an AI invention.

- Classify as AI (1) if the abstract describes or clearly implies an AI invention.
- Classify as not AI (0) if the abstract does not describe or clearly imply an AI invention.

Return a single JSON object exactly in this format:

{ "is_ai": 1 or 0, "certainty": 0.0 to 1.0 }

Restricted Use - À usage restreint

Box A.3. Prompt for labelling generative AI patents

The OECD uses the following definition for artificial intelligence (AI):

"GenAI models create new outputs (e.g. text, code, audio, images, video), often in response to prompts, based on their training data."

You are an expert in genAI and patents. Based on this definition, decide whether a patent abstract describes a genAI invention.

- Classify as genAI (1) if the abstract describes or clearly implies a genAI invention. This includes generating or adapting text, speech, images, video, audio, designs, code, 3D data (excluding embeddings and representations), molecular structures, synthetic data, or other digital content. Examples include large language models, text to image or text to audio models, multimodal generation, dialogue or chat systems, and AI models for content adaptation or synthesis.
- Classify as not genAI (0) if the abstract only analyses, detects, recognises, classifies, forecasts, or predicts without creating new content, or if it describes content creation that is not based on a trained generative model or learning from data, or if it describes rule based or statistical systems.

Return a single JSON object exactly in this format:

{ "is_gen_ai": 1 or 0, "certainty": 0.0 to 1.0 }

Box A.4. Alternative prompt for labelling generative AI patents

The OECD uses the following definition for artificial intelligence (AI):

"GenAI models create new outputs (e.g. text, code, audio, images, video), often in response to prompts, based on their training data."

You are an expert in genAI and patents. Based on this definition, decide whether a patent abstract describes a genAI invention.

- Classify as genAI (1) if the abstract describes or clearly implies a genAI invention.
- Classify as not genAI (0) if the abstract does not describe or clearly imply a genAI invention.

Return a single JSON object exactly in this format:

{ "is_gen_ai": 1 or 0, "certainty": 0.0 to 1.0 }

Restricted Use - À usage restreint

Annex B. Robustness checks

Table B.1. Regression results citation impact with interacted fixed effects

Citation impact of patent groups over 3- and 5-year windows with interacted fixed effects

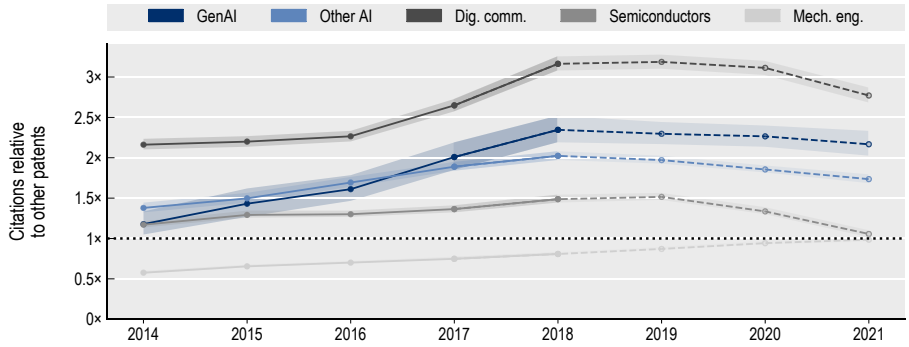
Variables	Citation count 3y	IRR 3y	Citation count 5y	IRR 5y
Group: generative AI patents	0.121***	1.128***	0.0901***	1.094***
	(0.0147)	(0.0166)	(0.0184)	(0.0201)
Group: other patents	-0.439***	0.644***	-0.461***	0.630***
	(0.00608)	(0.00392)	(0.00698)	(0.00440)
Observations	2,546,532	2,546,532	1,915,584	1,915,584
Filing Year * WIPO Field FE	Yes		Yes	
AIC	1.010e+07		1.040e+07	
log-likelihood	-5.052e+06		-5.180e+06	
Pseudo R2	0.0552		0.0380	

Note: This table presents results from negative binomial regressions with a log link, using other AI patents (non-generative) as the baseline group. The dependent variable is the forward citation count. The regressions exclude patents without a full 3- or 5-year citation window and the top 1% most-cited patents. All models include interacted filing year and WIPO field fixed effects. Exponentiated coefficients (IRR) are reported and indicate the expected proportional change in citation counts relative to the baseline group. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Restricted Use - À usage restreint

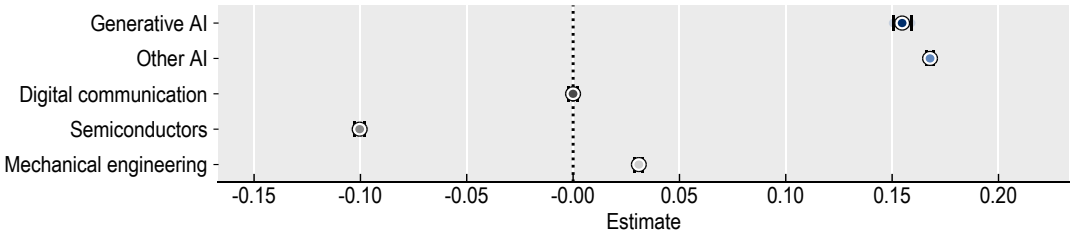
Annex C. Comparison other technologies

Figure C.1. Evolution of 3-year forward citation differences for other technologies



Note: This figure shows differences in expected 3-year forward citation counts for generative AI, other AI, and selected non-AI technological groups relative to a baseline group of all other patents. Estimates are derived from negative binomial regressions with a log link that interact patent group indicators with filing year. Patent groups follow a hierarchical and mutually exclusive classification into generative AI, other AI, selected non-AI WIPO technology fields. All remaining patents form the baseline. Values above 1 indicate that patents in the corresponding group receive more citations, on average, than the baseline group in the same year. Shaded areas represent 95 percent confidence intervals, and dashed segments indicate filing post-2018 filing years for which citations may be incomplete due to publication and citation lags. The sample excludes patents that are too recent to have accumulated citations over the full 3-year window and the top 1 percent most cited patents. All specifications include control variables and filing year and WIPO technology field fixed effects; for categories used in the group construction, these effects are uniquely identified by the patent group indicators to avoid multicollinearity.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Figure C.2. Regression analysis of patent generality across other technologies



Note: This figure shows coefficient estimates for the patent group derived from an OLS regression, where the dependent variable is patent generality. Patent groups follow a hierarchical and mutually exclusive classification into generative AI, other AI, and selected non-AI WIPO technology fields. All remaining patents form the baseline. The regressions exclude patents that are too recent to have accumulated citations over the full 5-year window and the top 1% most-cited patents. All specifications include control variables and filing year and WIPO technology field fixed effects; for categories used in the group construction, these effects are uniquely identified by the patent group indicators to avoid multicollinearity. Robust standard errors are used. Coefficients greater than zero indicate that patents in the corresponding group have higher generality scores than the baseline group. Horizontal bars represent 95 percent confidence intervals.
Source: OECD, STI Micro-data Lab: Intellectual Property Database, http://oe.cd/ipstats, September 2025, WIPO IPC-Technology Concordance Table, http://www.wipo.int/ipstats/en/statistics/patents, September 2025.

Restricted Use - À usage restreint