IGE | IPI ip-search
Eidgenössisches Institut
für Geistiges Eigentum

AI and Patent Searches

Whitepaper

Swiss Federal Institute of Intellectual Property
Stauffacherstrasse 65/59g, 3003 Bern, Switzerland
+41 31 377 77 77, info@ip-search.swiss, www.ip-search.swiss

AI and Patent Searches

Artificial intelligence (AI) is changing the way patent searches are conducted. If all users of the patent system have access to specialised AI-powered search platforms, what value remains in professional patent searching? More broadly, what opportunities does AI offer and where are its limitations? How can qualified experts create additional value? To what extent will human input be replaced by AI? Will the rapid development of AI-based searching continue, or will it slow down? And what about data protection?

These are questions that concern many people when it comes to the use of AI in patent searches.

It should be noted that developments in the AI field are currently occurring so rapidly that it seems almost impossible to adequately reflect all the latest advances in an article with review character such as the present one. A few figures may illustrate this challenge: according to Hugging Face (https://huggingface.co/models) on 10 March 2026, the number of available AI models was 2,698,318. It is simply impossible to test them all in a timely manner.

Historical Background

But let us start at the beginning. AI is not actually new. As early as the 1940s (McCulloch and Pitts, 1943) and the 1970s, algorithms and theoretical concepts underlying modern AI already were developed. What was lacking, however, was computing power, and it took some time for the first practically applicable AI models to produce meaningful results. With the availability of powerful computing clusters and high-capacity storage systems, data processing has reached a new level of performance over the last few decades. The ability to store and process vast amounts of data gave rise to the new field of data science as a branch of computer science. Efficient access to large datasets became essential, leading to the development of new database systems and highly effective search engines. Lucene appeared in 2000, Solr in 2006, and Elasticsearch in 2010. All three are search engines based on indexed entries and segmented ranking functions. At the same time, the field of machine learning (ML) – often marketed as AI – experienced significant growth, fuelled by the aforementioned newly available computing power. While most scientific applications focus on the machine processing of numerical data, developments in natural language processing (NLP) have opened up broad application possibilities. In particular, large language models (LLMs), implemented through neural networks, have significantly expanded capabilities since the emergence of the transformer architecture in 2017–2018. In contrast to index searches, the new models can incorporate semantic context. As a result, synonyms can be recognised automatically, and machine learning has significantly improved. However, this ‘learning’ is based on statistical analysis and therefore requires large quantities of training data, i.e. texts.

Approaches based on recently developed LLMs have proven particularly promising in the field of patent searching. The first experiments with AI-assisted patent searching at the Swiss Federal Institute of Intellectual Property (IPI) took place about 15 years ago, comparing various machine-learning approaches. At that time, the results were disappointing, and only isolated cases demonstrated real value. Only in recent years dramatic advances have been achieved through LLMs. Nevertheless, significant shortcomings remain, besides the high resource requirements. Despite the name, there is one thing that AI is not: AI is not intelligent. Put simply, AI is essentially complex, cleverly designed, adaptable statistics – but statistics nonetheless. Whilst humans can learn from a single experience, AI generally becomes more reliable the more training data that is available. Modern AI agents attempt to overcome this limitation by combining various non-statistical techniques, such as Boolean searching and reference hunting (search for citing and cited references).

Data Protection and Confidentiality

Whether AI is used in a patent search depends on various factors. The most important issues are, data protection and the safeguarding of professional confidentiality and maintaining office secrecy. When searches involve publicly available data and information, their use is generally unproblematic. However, when unpublished technical information is involved, such as invention disclosures or product descriptions in freedom-to-operate (FTO) searches – it becomes crucial to prevent sensitive information from reaching third parties. Protecting confidential information then becomes a top priority.

The legal issues are far from trivial. Different providers implement different safegards. The Swiss goverment maintains a list of countries that are considered to offer an adequate level of data protection. The location of the servers - in Europe or on another continent, and provider's headquarters, are also important factors. Wherever possible, the Swiss Federal Institute of Intellectual Property (IPI),– and thus its commercial services provided under the ip-search label – conclude specific agreements with providers to minimise risks. When web-based AI tools are used, the search input generally leaves both the IPI and Switzerland. Even where contractual arrangements regarding data protection, security and confidentiality exist, there is still a risk that input data could be intercepted and misused. For this reason, and because providers operate under different legal settings, ip-search provides its clients with a consent form regarding AI tools, allowing them to determine the extent to which specific tools may be used or whether they should be excluded entirely.

AI tools are often web-based and cloud-hosted because of their substantial computing requirements. In principle, some tools can alternatively be operated on premise. Such options are available to ip-search only to very limited extent. Whether a solution is deployed on premises, or via cloud-based browser access, is often not entirely a matter of choice. Local solutions generally offer advantages in terms of data protection but are more difficult to maintain and may therefore lag behind in the latest developments. Given the rapid improvements in model performance and data access, this can affect quality. In such situations, data protection risks must be weighed against quality risks. Furthermore, local solutions are usually considerably more expensive and often less powerful.

Rather than purchasing ready-made search platforms, organisations can acquire datasets and LLMs to develop their own AI systems. Whilst this may be preferable from the perspective of data protection, it also entails even greater risks in terms of quality. Experience shows that despite the rapid evolution of AI, the same providers consistently deliver the best results in terms of patent-search quality. Developing an equivalent AI system requires extensive experience and expertise. It appears unlikely that organisations can quickly develop AI systems themselves that offer the same quality as those of the market leaders. Today, for the reasons just statetd, most AI-tools used are browser based cloud solutions. They are continously updated, and, once a procurement decision has been made, they can be rapidly deployed with minimal additional technical effort .

AI can often provide additional quality – but not always. The search quality of AI tools and that of humans differ and complement each other. At the same time, the more AI is used, the more sensitive data-protection issues become important.

At ip-search, clients can help determine how quality and data protection should be optimised. They can also receive detailed explanations regarding the role of AI tools and human-added value in meeting their specific requirements.

The Capabilities and Limitations of AI

Sakari Arvela, a co-founder of IPRally, argued in a presentation at the IP Service World Conference in Munich in November 2025 that, AI is only as good as its training data. He illustrated this with an example: patent search engines, for instance, are trained on official search reports. These searches are conducted within predefined frameworks and cost-benefit constraints. Consequently, they are not exhaustive and often involve uncertaintiesin the evaluation of cited documents. Arvela suggested that future AI systems could be trained using documents generated during opposition proceedings instead. Because more effort has already been invested at that stage, both the quality of the cited documents and their evaluations is likely to improve. This would greatly reduce uncertaintyand enhance the quality of AI searches.

The outcome of purely AI-based searches therefore depends heavily on the quality and scope of the training material. In validity searches for granted patents, such as European patents, AI operating solely through an LLM is unlikely to identify a uniquedocument employing unusual terminology. If AI incorporates additional search methods, such as citation mining, it may discover and correctly rank such a document. However, this presupposes that a human searcher has previously classifiedthe document correctly. Without prior human input that can be leveraged by the AI, the document is unlikely to be found.

When properly applied, AI currently identifies approximately two thirds to three quarters of relevant documents in a typical patent search, with performance continuing to improve (e.g. Ali et al., 2024). Some providers claim their AI systems already find 95–100% of all available relevant documents. However, it should be noted that these documents are delivered within large result sets containing substantial amounts of irrelevant material. Human review often reduces this by a factor of ten, meaning that only one out of ten provided documents is actually relevant. Even with leading AI patent search tools, several hours of expert review are required to produce a truly useful search result.

Unlike traditional patent search engines, which generate results through transparent and reproducible search steps involving keywords, classifications and reference hunting and others, AI systems remain black boxes. Consequently, it is difficult to assess the completeness of a search or identify areas requiring further investigation. Models and training datasets may also contain errors that can only be detected using non-AI verification methods. Therefore, tests, validation queries and practical experience are necessary to avoid AI-induced quality risks and supplement AI searches efficiently with human expertise. Ultimately, the expertise of the the person skilled in the art is crucial, as only a human expert can determine whether AI-generated results are correct and whether further investigation is required.

Testing AI and Measuring Quality

In the past, the quality of search engines was assessed using concepts such as ‘precision’ and ‘recall’. In principle, this remains possible today. However, due to the above-mentioned rapid evolution of AI tools – for example, through the deployment of new AI modules – the search performance of these systems can change substantially within a very short period. In our internal tests, we repeatedly observed that rankings of the performance of different tools can change so significantly over a matter of weeks that their relative qualitative order is occasionally reversed. Consequently, extensive testing campaigns provide only limited value, and quickly become invalid. Therefore, there is no truly reliable method of determining the absolute or relative performance of AI tools with certainty. Depending on the individual client's approval, ip-search may use several AI-tools, a limited numberof tools, or none at all during a search. Whenever an AI tool is used, the delivered search results are examined to determine whether the documents delivered were already included in the AI search. These data are then evaluated quantitatively on a tool-specific basis. This approach has several weaknesses. The higher the quality of an AI tool, the more attention the searcher will pay to its suggested documents. Furthermore, searchers use different AI tools in different ways depending on their personal experience and the tools' technical suitability. Despite these limitations, some tools largely maintain their relative positions over extended periods. Notably, certain market leaders manage to retain their leading positions despite substantial technical changes. Comparisons between AI-generated results and the delivered documents are not evaluated solely on a tool-specific basis. The position of documents within the tool's internal ranking is also recorded and compared with the relevance assessment contained in the search report.

AI tools can be used either exclusively at the beginning of a search or at various stages throughout the process. Using them at the beginning of the search process is particularly efficient and important. While subsequent use of AI tools can still improve the search to a limited extent, typically yielding an additional 10–15% relevant documents compared with the initial AI-assisted search, according to surveys among search professionals, the greatest benefits are generally realised when AI is employed at the outset. Discussions with other search service providers revealed that searchers tend to use the tools with which they are most familiar. This creates a potentially significant bias based on personal preference. To avoid this bias and ensure properobjective use of AI tools, ip-search has developed standardised search queries that are centrally executed by patent administration staff before the expert begins their work. The results of these preliminary AI searches are then provided to the individual search professionals in a standardised format for further evaluation. This approach has proven effective so far: as when ip-search informed interested clients about its experience with AI tool testing, they have generally provided positive feedback, thereby offering modest confirmation of the methodology.

It has become evident across all these tests, particularly when defining internal AI query standards such as whether to provide the AI with an entire document or only part of it, that such standards are gradually losing importance, as is prompting itself. Consequently, detailed discussions of prompting are becoming unnecessary.

Human Added Value – "with AI and beyond"

In practice, ip-search frequently receives search assignments relating to subjects that have already been investigated by clients themselves, by other search providers, by patent offices, or by opposing parties. AI is used extensively in most prior art searches. Patent offices, large corporations and patent law firms now have access to a wide range of AI-based search tools. However, there are still law firms and SMEs for whom the acquisition of costly professional AI tools is not viable. In such cases, even a quick AI search can add considerable value. When AI-generated results are available, the challenge lies in creating value beyond what AI has already produced. This is likely to be the direction in which professional patent searching will continue to evolve. This is what ip-search refers to as "with AI and beyond".

If an AI search does not yield the desired results, simply repeating the same search is insufficient. So, how can additional relevant material be found?

AI searches themselves can be expanded using AI. Multiple AI tools can be used in a complementary manner, with the results generated by one tool being re-examined using the ranking mechanisms of another. To extract maximum value from AI tools, ip-search specialists develop tool-specific methodologies. AI tools can be used in the subsequent searches for one-off queries, iterative workflows or different prompts, as required. However, experience has shown that while iterative AI use and alternative prompting can provide some benefits after the initial AI search, the additional value is generally modest. As AI tools continue to improve, the incremental benefit of iterative AI use and prompting continues to decline. Document sets generated by AI can also be expanded through simple reference hunting. In practice, such approaches continue to identify additional relevant documents despite increasingly capable AI systems.

As already noted, AI tools are increasingly effective black boxes that can accomplish many things – but not everything. Experience shows that search subjects can often be investigated without using the defining technical features. If the underlying technical problem and/or application can be described without explicitly referring to the distinguishing features or specific keywords of the invention, an alternative search strategy emerges. Such alternative approaches are unlikely to be included in standard AI searches, and they often yield additional relevant documents.

Another effective strategy involves conducting extremely broad searches using only a small number of terms after reviewing the AI-generated results. Such searches typically produce enormous numbers of documents, making comprehensive review impossible. However, advanced database ranking algorithms such as Lucene and Elasticsearch can be used to organise the review process efficiently. Surprisingly, even limited sampling of such broad searches regularly uncovers relevant documents that had not been identified previously. Depending on the nature of these results, the search can then be expanded and the focus adjusted in a targeted manner. This is particularly useful when search subjects with similar technical functions are employed across different industries. For example, braking systems may be used in both automobiles and aircraft, even though the actual implementation forms and dimensions differ significantly. An AI search that initially only produced hits relating to automotive technology may therefore be supplemented with manual searches in the aerospace field, in particular with respect to novelty.

In this context it should be noted that success of AI use also depends on the technical domain. Different technical fields such as mechanical engineering, chemistry or biology use different modes of disclosure in the literature, and, thus, need specific search competences. Here, AI trained on a general corpus of documents reaches its limits. Further difficulties are related to how essential inventive features are described in the text – LLMs mostly rely on text based evaluations only. Even in the same technical field the differences can be important e.g. geometric special arrangements in mechanics may be difficult to assess by text, whereas figures often are better suited for such a purposes. Although image analysis shows substantial progress, at least as patents are concerned, for the time being humans remain superior to the machine.

Depending on the technical field – such as telecommunications, semiconductor technology or optics – searches in the scientific literature remain indispensable, particularly for novelty and validity searches. In these areas, the limitations of current AI tools are especially apparent. To obtain quality results professionals continue to rely on complementary traditional Boolean search strategies within specialised scientific databases or analyses of technical standards, where the decisive text passages are often found. Perhaps such sources will eventually be fully integrated into AI systems, but due to copyright issues adequate access to these sources remains challenging.

The search approaches described above are only examples, and they are by no means exhaustive. Many comparable supplementary search methods exist. Selecting and applying such methods requires considerable experience. Where appropriate, ip-search also consults with clients regarding specific issues, as clients often possess deeper technical knowledge of their own field than external search providers. For this reason, particularly when conducting expanded searches, it is worthwhile aligning the assignment not only with the client's technical objectives but also with the client's own search experience.

Because AI tools evolve rapidly, ip-search adapts its quality assurance measures and training to changing circumstances. Internal guidelines specify which tools should be used, in what order, and under what conditions. The experience of the individual search professionals is pooled and made available to all searchers. Nevertheless, quality remains a resource issue. Although the probability of success often increases only asymptotically as more resources are invested, experience shows e.g., for validity searches that searching can be extended meaningfully for up to two weeks using technically sound approaches. For reasons of cost and efficiency, such extensive efforts are generally not advisable. However, this observation demonstrates that AI is far from being exhaustive with regard to all available search possibilities and that greater effort often continue to yield better results.

What about AI Developments to come?

It should be remembered that the quality of AI searches depends on various factors, including database coverage and the quality of training data. Ultimately, AI search tools are based on language models, that operate largely on statistical principles. Whenever the volume of data is statistically reduced, unusual or highly specialised documents with unusual terminology tend to be lost. Despite improvements in AI models and greater accessibility of larger datasets, AI-based patent searching is approaching a plateauin practice. Simple searches are likely to be requested less frequently, and AI can already perform a substantial proportion of the work in such cases. However, aside from small performance gains, there is currently no indication that AI tools will overcome the fundamental statistical limitations and uncertainties inherent in their training data. Therefore, for professional patent searching, a combination of AI and human expertise appears unlikely to be replaced by *purely* AI-based searching in the foreseeable future.

This is the outlook for AI in the field of patent search. For generative AI the situation is different. Development is progressing rapidly in tasks involving the drafting of new documents or the comparison of existing documents according to predefined rules. Numerous examples already demonstrate that generative AI can deliver significant efficiency gains, and these gains are expected to increase in the future.

If you have any specific questions about patent searches using AI, without AI, or 'beyond AI', please do not hesitate to contact our team.

Dr Peter Bruns

Peter Bruns is Head of Commercial Services. Prior to this role, he led the Engineering Team. He was previously responsible for EPOQUENET, the European Patent Office’s search tool, and for the implementation of its successor platform, ANSERA-based SEARCH. Today, his technical focus lies on the integration of AI tools and the optimisation of patent searching both with AI and ‘beyond AI’, particularly with respect to the added value provided by human expertise.

Dr David Rees

David Rees is Head of Materials and Medical Technology Team, and he holds a PhD in Chemistry as well as a Master's degree in Natural Sciences. Prior to his current role, he was Head of Customer Interaction at ip-search. Before joining the Swiss Federal Institute of Intellectual Property, he held positions in the patent departments of Novartis Pharma AG in Basel and Henkel KGaA in Düsseldorf and at the European Patent Office.

Reference works

Warren McCulloch and Walter Pitts, A logical calculus of the ideas immanent in nervous activity. In: Bulletin of Mathematical Biophysics, Vol. 5 (1943), pp. 115–133.

Ali, Amna et al. (2024), Innovating Patent Retrieval: A Comprehensive Review of Techniques, Trends, and Challenges in Prior Art Searches. Applied System Innovation 7(5), 91.

Reference works not cited:

Rossitza Setchi, Artificial intelligence for patent prior art searching, World Pat. Inf., 64 (2021), Article 102021, 10.1016/j.wpi.2021.102021, View PDFView articleView in ScopusGoogle Scholar

B.L. Genin, Similarity search in patents databases. The evaluations of the search quality, World Pat. Inf., 64 (2021), Article 102022, 10.1016/j.wpi.2021.102022, View PDFView articleView in ScopusGoogle Scholar

Renukswamy Chikkamath, Is your search query well-formed? A natural query understanding for patent prior art search, World Pat. Inf., 76 (2024), Article 102254, 10.1016/j.wpi.2023.102254, View PDFView articleView in ScopusGoogle Scholar

Eleni Kamateri, Will AI solve the patent classification problem? World Pat. Inf., 78 (2024), Article 102294, 10.1016/j.wpi.2024.102294, View PDFView articleView in ScopusGoogle Scholar

Anna Maria Villa, Manuel Wirz,

A sequential patent search approach combining semantics and artificial intelligence to identify initial State-of-the-Art documents, World Pat. Inf., 68 (2022), Article 102096, 10.1016/J.WPI.2022.102096, View PDFView articleView in ScopusGoogle Scholar

Lothar Walter, A review on digitalization trends in patent information databases and interrogation tools, World Pat. Inf., 69 (2022), Article 102107, 10.1016/j.wpi.2022.102107, View PDFView articleView in ScopusGoogle Scholar

Homaira Huda Shomee, A Comprehensive Survey on AI-Based Methods for Patents, Arxiv.org (2024), http://arxiv.org/abs/2404.08668

Anon, AI-Assisted Patent Prior Art Searching - Feasibility Study Intellectual Property Office (2020), ISBN: 9781910790809, Google Scholar