Artificial Intelligence – can it deliver improved patent search?

a personal view.

Magister Ltd, Roche, United Kingdom

Stephen Adams

Managing Director of Magister Ltd.

stephen.adams@magister.co.uk

1 Personal background

In considering whether artificial intelligence (AI) could replace the human being in the conduct of intellectual processes such as a patentability search, I have to start by giving some personal background, since this article is an essentially personal view of the possible impact upon my chosen profession.

I completed my Masters degree in Information Science in 1982. Almost all of our research work, and our understanding of information retrieval and the use of scientific and technical documents, was paper-based. Nonetheless, I did complete my first literature search in that year, on the Dialog ® online system. It presented two clear advantages; it was fast, and it was interactive. The speed made it possible to complete a search across many year's-worth of the Chemical Abstracts database in one step, instead of repeating my manual search in a succession of 5-year or 10-year segments, using the cumulative indexes. The degree of interactivity, such as reading the titles of some search results, trying different search terms, or recombining search sets, enabled me to modify my ideas of "a relevant answer" in mid-search. In a paper-based search, this modification of my objectives would result in me having to abandon my initial search strategy and my half-finished list of answers, and start again – a huge waste of effort.

With these first-generation search systems, the final quality of the search was still entirely dependent upon the decisions of the human searcher. All the important tasks, such as selecting the appropriate database, looking up controlled vocabulary or indexing terms, and judging relevance of answers, formed part of the "art" of literature search, not the "science". Information professionals learnt to rely upon these search systems in their everyday work, to deliver results in response to the human searcher's skill, confident that the machine was not in any way replacing the human. The technology behind an electronic search was a simple algorithm which carried out essentially the same work processes as a paper-based search. A searcher who understood their paper-based search tools could quickly understand what was going on "behind the keyboard", and to trust the computer as a reliable "co-worker". Electronic information systems were therefore seen as little real threat to human jobs, because

they were entirely reliant upon human skill to operate effectively, just like a dumb robot.

Over the last three decades, there have been many developments in electronic literature searching. From the point of view of this article, I would highlight two in particular; one relates to the data, the other to the process of search.

The first change is the development and use of databases which have never existed in a paper-based form before. My first search results from online Chemical Abstracts improved in direct proportion to my skill and understanding of how to use the corresponding paper indexes. The more I used the paper form of Chemical Abstracts, the better I could understand the electronic search fields, how they had been created and their significance for efficient searching. By contrast, if an electronic database can be searched in a way which has no human analogue, it may be harder for the searcher to conceive of the possibilities – or pitfalls – of that searching technique. For example, any full text database – newspapers, journals or even patents – offers a search capability which is simply not available to the human searcher, because it is physically impossible to read every word of every record and extract a meaningful result from the corresponding paper sources. The ability to search *inside* thousands or even millions of documents and to locate a single exact word has given us unimagined possibilities for retrieval. However, since the human searcher has never tried to carry out this process manually, they may struggle to develop an accurate mental model of what the machine is actually doing – and more importantly, what can go wrong. Frequently, a searcher learns more about a full text database when their search results are unexpected, or irrelevant, than when they appear to be correct. Smaller errors may be entirely invisible to the inexperienced searcher.

The second change has been the expansion of different forms of search process. The operations which are carried out on our databases today are substantially different to the simple duplication of paper-based search steps, which were used by the first generation of online systems. These new processes include the use of built-in taxonomies or thesauri, word-stemming, proximity searching, automatic synonym or cross-lingual term selection, relevance ranking and more advanced developments such as latent semantic analysis or the replacement of Boolean matching with a Support Vector Machine. The challenge for the information professional in this area is not “How well do I understand the database?”, but “How well do I understand what my search engine is doing?” (or even “How much do I *need* to understand?”).

The outcome of these two changes means that the online keyboard can no longer be treated as if it was still the same dumb robot of 30 years ago. Some of the intellectual skill of achieving a good search result has already been taken away from the human. In the patent search environment, searchers have been used to taking a high degree of responsibility for the quality of their output, and we rightly take a professional pride in the significance of what we do. With these changes in search – and now the appearance of AI – it can appear that the human searcher is being asked to hand over even more responsibility to a ‘black box’ of poorly-understood technology, which is operating on an inconceivably complex database structure.

It is no wonder, then, that some are questioning whether AI systems are worthy of that degree of trust, either now or in the foreseeable future. The older generation of searchers may be inherently more sceptical than the younger generation, but they are also more able to probe beyond the hype.

2 Technology background

In the last few years, it seems as if many different industries have begun to be aware of the possibility of using AI to improve their business processes. Some authors have proposed that AI is capable of penetrating to all areas of intellectual property work, including the drafting, proof-reading and prosecution of applications, portfolio management, renewal decisions, current awareness etc. As a professional searcher, my concern in this article is whether AI will be able to perform in the area of patent search, including patentability, state-of-the-art and freedom to operate or clearance searches.

At first sight, all AI developments may appear to be simply the next logical step in the use of any sort of machine to complete a task. It is already well known to replace a human being with a machine in some aspects of life, and these developments have largely been accepted as beneficial. A machine may be able to perform functions which are dangerous or impossible for a human being, such as activities which take place under water, in a confined space or exposed to high temperatures or toxic fumes. For example, it is routine to use a remotely-controlled vehicle to inspect gas pipelines, or a mini-submarine to take underwater pictures. It is even known to install miniature cameras into a pill which can be swallowed by a patient, to record the condition of their digestive system[(1)]. In these applications, the machine does not in any sense 'replace' a human worker, since it is carrying out operations which were not possible for a human at all.

A slightly more controversial use of machines – at least in the early periods of the Industrial Revolution of the 19th and early 20th century – is to replace humans in repetitive manual tasks, such as building products on a factory assembly line. In these cases, humans are capable of performing the same tasks, but machines offer the chance to increase productivity and often to drive down manufacturing costs. As a consequence, these sort of developments have often resulted in the fear of "Will this machine take my job away?". Robots can also produce a more consistent product, as well as more quantity in less time. A welding robot, for example, can be programmed to repeat the same sequence of welds on a car body time after time, without needing to take a break or losing concentration. Because of this, robotics has been at the heart of improvements in manufacturing quality in many industries. Indeed, some quality control initiatives, such as Six Sigma, have at their heart the principle of *"improving the quality of the output of a process by .... removing the causes of defects and minimizing variability..."* [(2)].

However, AI is rather different to both of these uses of machines. In the AI field, attempts are being made to develop "learning systems". These are defined in a recent WIPO report as *'machines that can become better at a task typically performed by humans, and to achieve that level of performance with limited or no human intervention'* [(3)]. The popular press has not always understood this distinction.

Newspaper reports have sometimes described a particular computer-based system as exhibiting "artificial intelligence", even when its functionality has been achieved merely by virtue of advances in computational power and speed, allied to a huge base of reference data. True AI systems are more than "robots on a chip", and the tasks which they may be able to perform extend far beyond mere high-speed, tireless repetition. They hold out the potential of being able to penetrate into areas of intellectual and creative work, which up to now have been seen as the sole preserve of the human mind. True AI – so-called "super-intelligence", where AI systems become able to take on any intellectual task without guidance – has not been achieved yet, and may be a long time coming. As the writer of the foreword to the same WIPO report, Andrew Ng, concedes[4], *"Effective 'unsupervised learning' – learning without labelled data – remains a holy grail of AI."*

3 Irrational fear, or reasonable doubt?

In the course of preparing this article, I have assumed that it is unlikely that my generation of information professionals will ever have to make a binary decision; *"Do I use AI in patent search, or not?"*. Like most technologies or work practices, it is likely to penetrate the profession gradually, initially working in parallel with existing systems before – possibly – replacing them altogether. Therefore, I think it is reasonable for today's patent search professional to examine the promises of AI in the light of what the human searcher is already capable of, when performing at their highest level of competence. If AI systems are being designed to replace the human searcher, we must ensure that their performance is *at least as good* as our current tools in the hands of a skilled human searcher. Put simply, if an AI system is to gain widespread acceptance, it must be able to beat the best of what a human can do, and not just once, but continually.

I will consider these factors in the form of three questions. These are:

A. Will AI-based patent search systems be AGNOSTIC but AGILE?

B. Will AI-based patent search systems be BIASED?

C. Will AI-based patent search systems be CONSISTENT?

4 Agnostic

I am using the term 'agnostic' in its more recent sense of 'not holding a strong opinion one way or the other'. Why is this an important factor for search? It arises out of the recognition that a search in a literature database is never a standard product. Searchers are not sitting on a production line, turning out batches of results which have been 'manufactured' using the same tools and look much the same each time, with only slight variation or customisation. Each patentability search is unique, designed to fulfil a single objective – to establish the absolute novelty and/or inventive step underlying each new patent application. The search stands alone, and the results must not be influenced by what was done before, even if the search is carried out on behalf of the same applicant company working in the same narrow sector of technology for many years. This is in complete contradiction to the development trends of new search algorithms and search engines, which actively retain what they 'learn' about the searcher's preferences. The phenomenon of the 'filter bubble'[5] is now well recognised

– the fact that many common search engines retain information about user behaviour, and over a period of time will start to give greater prominence to results which are similar to what the user has accepted in the past. Electronic news feeds begin to produce answers which are in line with the user's known political or social viewpoint, and to suppress answers which present a contradictory view. The technology behind 'cookies' is now being extended between online platforms, such that you begin to see advertisements on one website relating to goods that you have searched for on another website. It is arguable that there may be helpful aspects to this for day-to-day search and retrieval, but it is definitely unhelpful in the scientific field.

If AI is to be truly helpful for scientific or patent search, then the technology behind the search process must remain *entirely neutral* with respect to how the search is conducted and how the results are sorted and presented to the user. For example, a chemical structure-based search may sometimes yield results for molecules which are useful in more than one industry, such as drugs or agrochemicals. If the machine 'learns' that the user is mostly interested in drug applications, there is a risk that eventually, any references relating to agrochemicals may get filtered out. Then, on the isolated occasion when the same user wants to focus on the agrochemical uses, the algorithms behind the search machine may be blocking them from view, or worse still, not retrieving them at all.

A second aspect of this lack of 'search neutrality' is seen today in a typical Google search. As soon as a user begins to type in a search term, it is now auto-completed with the closest term(s) used in other recent searches; in other words, certain trending topics are suggested as being the most likely target of your search. This encourages the user to become lazy, and complete their search using the same strategy as previous users. If professional searchers begin to rely on search systems like this, there is a risk that they will lose the skills and confidence necessary to over-ride these suggestions, and simply adopt the path of least resistance by choosing the suggested search terms. In this way, a searcher is steered away from an *ab initio* approach to the data, and towards a search which 'looks like' other common searches at the same time, in the same information source, or by the same user. To my way of thinking, this is a completely unacceptable feature for a patent search process; neither the search input nor the retrieval mechanism must be influenced by any 'previous experience' or 'retained knowledge' about what is, or is not, relevant or current (trendy).

A third aspect of the influence of machine learning on effective search is perhaps best described by another 'A' – instead of asking systems to be 'AGNOSTIC', they must also be 'AGILE'. A skilled human searcher knows that they will often have to adapt their search strategy as they move between different information databases, or even different forms of literature. For example, an efficient search in a patent database may use a pre-defined classification scheme such as the International Patent Classification (IPC) or the Japanese FI system. In order to complete the same subject search in the non-patent literature (NPL), it is usually necessary to replace the IPC or FI classes with some other equivalent search terms (e.g. text or a parallel

controlled vocabulary scheme). That is, the search strategy has to be adapted, to utilise the strengths of each individual database or publication type. Given that AI systems need a base set of training material in order to learn how to search, will AI systems have to be completely re-trained in order to discover corresponding NPL during the same search session? If this is an inefficient process, or is not allowed for in the building of the search engine, there is a risk that future searching will tend to ignore all the valuable meta-data which has been built into some of the major patent and NPL databases, and concentrate solely on the minimum common search parameter, which is simple text. No matter how 'intelligent' search algorithms are, any system which is reduced to searching solely on this basis will only perform to the same level of efficiency as a naïve beginner at the beginning of their training as a patent searcher. Such an AI system will be a very unsatisfactory replacement for the skilled patent searcher.

A further aspect of an AGILE search system is one which is capable of adapting its search strategy in response to the user's perception and feedback of how well the initial results fulfil the search goals. One way of achieving this is by moving up and down the ontology of a given scientific or technical discipline during the search. It is somewhat concerning to me to note that the vast majority of research – as measured by the patent applications – in the area of deep learning is based on the analysis of original text structures. This approach will result in systems which have only a limited understanding of how scientific terminology develops over time, or how sub-disciplines within a technical field mature and adapt. Less than 1% of the AI patent families reported in the 2019 WIPO study related to ontology engineering or probabilistic reasoning. Any AI search system which hopes to replace the human searcher will have to learn about not only today's scientific publications, but also the wider context of the technical field. They will also probably have to be re-trained and brought up to date (as indeed humans are) as the field changes and develops. Will an AI system which has learnt about past technologies be able to recognise a new invention which breaks the paradigm? This goes to the heart of patentability search, and the establishment of inventive step (non-obviousness) arguments during patent prosecution. There are anecdotal stories that when the first patent applications for the hovercraft were submitted to the UK Patent Office, it took some time before it was agreed which examiner should consider the case; was it a boat or an aeroplane, since it shared characteristics of both? In the end, the first of Sir Christopher Cockerell's patents[6] was classified under the IPC in sub-class B64D (for aircraft) but today would be placed in B60V (for air-cushion vehicles). The ability to place a given patent application in the wider context of science is a crucial step in successful adaptation of retrieval strategies. A good quality search strategy must be scaleable and adaptable, both to user feedback and to the age of literature which it is searching, to take account of different terminology over time and the state of common general knowledge.

5 Biased

The question of bias in AI systems is already well known. Current algorithms depend upon sets of learning data in order to build a background of experience, and to deliver future results. If the learning data contains certain

biases, this can generate a negative feedback loop, such that the AI inherits a biased viewpoint on how future data should look. The system may then suggest actions which tend to confirm its predictions. This problem has already been encountered in predictive policing systems such as PredPol[7]. Similar problems have been found in recruiting systems, which used AI to pre-select candidates for a job interview[8].

Experts in AI may challenge this viewpoint as 'scaremongering', which may be partly justified – in the popular imagination, there may be only a short step between reading about shortcomings of the PredPol system and remembering the plotline of the science-fiction film "Minority Report". Once such problems are detected, they can usually be remedied, and past mistakes are not a strong argument for stopping the development of newer and better search engines. However, I believe that these are real issues which need to be addressed in the specific context of developing AI for patent search. What are the possible implications for a patent search if there is bias (conscious or unconscious) built into either the selection of an appropriate database, or the building of a search strategy which is optimised for a specific database?

Human patent searchers already encounter such challenges to their skills every day, and grow in experience of how to address them. For example, when the first databases of full text PCT documents were launched, only certain languages were available. Only documents published in English, French or German were truly available in full text. Later, texts published in Spanish, Russian or Asian languages like Japanese or Chinese began to be added, but initially the character-coded texts for these other languages were not available, and were effectively invisible to the searcher. Inexperienced searchers who were unaware of this might attempt a search using English search terms only. Any records corresponding to English-language specifications would be certainly located, but also some non-English documents by virtue of the existence of an English abstract; this gives the impression that the entire database has been searched to the same depth. The searcher would then close their search session, unaware that their results were significantly incomplete and biased. During the analysis phase, a search system based on relevance ranking of results by word frequency would bring the English full-text specifications to the top of the list, simply because of the larger volume of text for those records and the higher hit-term frequency within them. As a result, any highly relevant answers which had been located purely on the basis of the English abstract would be relegated to a position much lower down in the relevance-ranked list, and could easily be overlooked.

Learning sets may have built-in biases, which will get reproduced, perpetuated and possibly even enhanced by machines. However, when so-called 'expert systems' were being built in the 1980s, their underlying structure was a rules base which attempted to encapsulate a human skill as a series of logic decisions. One criterion for a successful expert system was that it could not only arrive at a decision, but also be able to justify *how* it arrived at that point. By contrast, the design of modern AI systems appears to have as its goal the ability of a system to function autonomously; that is, it will not only *fail* to deliver such accountability, but will be *unable* to deliver it, thus denying the human user any insight into the search process

or any opportunity to correct it.

A further aspect of how bias may creep into patent searching relates to the form of information which is found in a complex document such as a patent application. It appears to me that current AI system developers are working on the assumption that the entire teaching, or representation, of a new invention in the form of a patent application is accessible through analysis of the text. However, anyone who has done patent searching for any length of time quickly discovers that many patent applications use text very loosely; they are totally unlike any other word-based document, such as a newspaper, book or even a scientific journal article. Not only does a patent use a highly specialised form of scientific language, but it is also found that much more information about the invention is not expressed in words at all, but embedded in entirely different 'languages', including tabular data, figures and diagrams, chemical and biochemical structures, equations, software code etc. An AI system which can only search text, and which will give the most prominence to those records which contain the "most relevant" text (i.e. containing the greatest volume of useful words or word pairs), is highly likely to miss any answers which describe the same invention in non-text language, and to mis-represent what it does find. Teaching a machine to "understand text" is only part of the solution towards judging content and thereby understanding relevance. Although there have been some advances in direct searching of non-text data, they are a long way from being able to contribute significantly to the search process[9]. The sole exception to this is in the chemical field, where there is now a variety of tools for searching by chemical structure.

6 Consistent

There are two principal aspects to consistency of search which I think are of concern in the development of AI-based patent search systems. The first is the question of whether the same search engine will deliver the same results to every searcher, irrespective of their location or circumstances. The second is whether it will be possible to archive the processes of search at a particular point in time, such that it is always possible to reconstruct what was retrieved on a certain date. Both of these questions are important whenever there are challenges concerning the completeness or competence of a search (for example, in litigation when one party has to show what they knew on or before the filing date of a contested patent application).

The first question is a special aspect of search localisation, and is also relevant in my discussions above about search neutrality and bias. We can only argue by analogy with the popular search systems of today, but if AI search systems follow the same approach, we could find that the "same" search engine is delivering different results sets to different users of the same database on the same day. For example, would the machine 'learn' that a searcher located in Japan is 'always more interested' in Japanese-language documents, and act in a way which directs their search towards a Japan-centric view of the prior art? Google has already implemented such search localisation, so that a searcher located in the United Kingdom cannot effectively search the German version of Google (www.google.de) even when they wish to concentrate on German-language results, without making changes to their strategy or browser settings[10]. This is

clearly not satisfactory when the user wishes to make a global search, as in the case of patentability.

The second issue relates more to the storage and re-use of search strategies. In many companies, it is standard procedure to 'top-up' an initial patentability search during the period that a patent application is on file but unpublished (typically between 12-18 months). With current search systems, this can be achieved by archiving the exact sequence of search commands used, together with information about the date of search and the update status of the database(s) used. Repeating the search to locate any new material is then simply a matter of running the same strategy and limiting the results to only those which have been added since the previous date – a procedure which is quite straightforward in Boolean systems. However, this becomes much more complex in the case of the present generation of semantic searches, and is likely to be completely impossible with more advanced AI systems. This is because the search algorithms are constantly evolving, and cannot be rolled back to the state they were in at some point in the past. Therefore, re-running the same strategy (even if it could be retrieved and archived from the 'black box' on the original date) has the potential to retrieve different results from the original time period of the search, as well as any new material added to the database since then. In this environment, it becomes effectively impossible to answer the question, "What was searched, and how, on this date?". This means that the original searcher has lost access to much of the evidence which they need in order to demonstrate that their past search was conducted in a competent manner, in line with best practice. This factor may be highly relevant during patent litigation, in order to establish what was in the public domain or part of the common general knowledge. It remains to be seen whether a judge would consider that the use of an AI-based search system absolves the user of any degree of responsibility for becoming aware of prior disclosures.

7 Is AI the only way forward?

I have tried to consider in this article whether there is a reasonable chance of the development of AI support for the work of patent searching. My arguments above are based on some 30 years of experience in patent searching, and are intended to highlight some of my expectations for what AI can deliver, but also some factors which – as far as I know – are not well recognised by system developers thus far. However, these technical arguments may eventually be over-ridden by a more pressing economic one; namely, "Can we afford NOT to invest in AI ? – is it the only hope for the future of patent searching?".

One of the stated motivations for developing AI to handle search work is that "the human being cannot cope". This argument is frequently backed up by statistics on the number of patent applications being published, or patents granted, each year around the world, and the likely rise in the numbers of candidate answers which a typical search will be expected to retrieve. This is indeed a major problem facing anyone doing a patentability search today, but the solutions which are being proposed commonly focus on automatic screening, relevance ranking or classification (clustering or so-called 'faceting') of a very large answer set, rather than the creation of the answer set in

the first place – that is, to improve post-search analysis rather than the search process itself. It may turn out that the way forward will be continued improvement in this aspect, combined with genuine changes in the process of search, using well-designed AI systems. I prefer at this point to think that the approach adopted by IBM, called Augmented Intelligence instead of Artificial Intelligence, will remain the preferred route for some time. The human searcher will still be needed for the critical first stages of a patent search, which are the reference interview with the client in order to draw out the real information need through informed discussion, subject analysis and iterative trial searches. One of my principal concerns about the entire automation model is that it is based on an attempt to achieve an optimum answer in a single step. In my experience, simply throwing a 'relevant patent' into a full-text based search engine and expecting to get closely related disclosures may yield some relevant results, but generally falls far short of the quality which major corporate applicants expect from their (human) colleagues. Provided the "money vs. time" arguments do not over-rule the need to develop usable systems, the intelligent human patent searcher will have a job for the foreseeable future.

References

(1) Reduced size imaging device. US 7662093-B2. 16 February 2010.

(2) Six Sigma. Wikipedia entry, available at https://en.wikipedia.org/wiki/Six_Sigma [Accessed 2019.08.13]

(3) Artificial Intelligence. WIPO Technology Trends 2019. Publication No. 1055E/19. Geneva: WIPO, 2019., p.39.

(4) Ibid, p.8.

(5) Filter Bubble. Wikipedia entry, available at https://en.wikipedia.org/wiki/Filter_bubble [Accessed 2019.08.27]

(6) Improvements in or relating to vehicles for travelling over land and/or water. GB 854211-A. 16 November 1960.

(7) D. Ensign et al. Runaway feedback loops in predictive policing. 2017. https://arxiv.org/abs/1706.09847v3 [Accessed 2019.08.27]

(8) J. Dastin. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters Business News, 10 October 2018. https://www.reuters.com/article/ us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [Accessed 2019.08.27].

(9) S. Vrochidis et al. Concept-based patent image retrieval. World Patent Information, 34(4), 2012, 292-303.

(10) A. Barysevich. How you can see Google search results for different locations. 26 February 2019. SearchEngineJournal.com. Available at www.searchenginejournal.com/see-google-search-result-different-location/294829/#close [Accessed 2019.08.28].

人工知能―改善された特許検索を提供できるか？　個人的な見解（日本語抄録）

Stephen Adams

【個人的な背景】

筆者は 1982 年に情報科学の修士号を取得した。私達の研究活動と理解は紙ベースだった。だから、データの仕組みと、それらがどのように作られるかが理解できた。

最近の情報検索の 2 つの利点は、高速とインタラクティブである。検索はワンステップで完了するに変わり、検索中に「関連する回答」の検索方針を変更することができるように変わった。

最初のデータの変更は、レコードのすべての単語を読むだけである。だから論文から意味のある結果を抽出することは不可能である。検索結果が予期しない、または無関係である場合には、経験の浅い検索者には、小さなエラーを全く見えなくしている。

2 番目の検索プロセスの変更は、さまざまな形式の検索プロセスの拡張である。新しいプロセスは、シソーラスの使用、単語ステミング、近接検索、自動同義語または言語間の用語選択、関連性ランキング、潜在セマンティック分析への置き換えなどを含み、ベクター（行列計算）マシンに変わった。

データの変更とプロセスの拡張性という変更の結果は、知的スキルの一部を人間から奪い、技術の「ブラックボックス」の責任を情報プロフェッショナルに負わせている。私達は、AI システムが信頼に値するかどうか疑問視し、よく調査すべきである。

【技術の背景】

一見、AI 開発は、あらゆる種類のマシンを使用してタスクを完了する2つの論理的なステップにすぎないように見える。①人間にとってまったく不可能な操作を実行している場合（人間に置換できない場合）と、②人間と同じタスクを実行できる場合である。後者は、「このマシンが私の仕事を奪ってしまうのではないか？」という恐怖をもたらした。しかし、現状の AI は、これらのマシンの使用の両方とはかなり異なる。

WIPO レポートでは、AI は「人間が通常実行するタスクでより良くなり、人間の介入が制限されているか全くない状態でそのレベルのパフォーマンスを達成できるマシン」と定義している。知的および創造的な仕事の分野に侵入することができる可能性を保持している「真の AI」、いわゆる「スーパーインテリジェンス」は、まだ達成されておらず、実現には長い時間がかかる。

「効果的な「教師なし学習」- ラベル付きデータなしで学習することは、AI の聖杯である」と AI 研究の大御所（Andrew Ng 氏）は述べている。

【不合理な恐怖、または合理的な疑い？】

特許検索の専門家は最高の能力を発揮して検索調査すると並行して、AI の可能性を検討するのが合理的と筆者は思う。

AI を検討する際には、３つの質問を投げるのが良い。

A. AI ベースの特許検索システムは偏見なき中立者で、かつ、機智者か？

B. AI ベースの特許検索システムにはバイアスが掛かるか？

C. AI ベースの特許検索システムには一貫性があるか？

【偏見なき中立者】

筆者は、現状の AI は「偏見なき中立者」であるべきと思う。決して AI は標準的な製品ではないと認識する。

特許検索は、絶対的な新規性および / または進歩性を確立するためにユニークな業務であり、毎回同じように見えるが、決してそうではない。本来、結果は以前に行われた特許検索に影響されてはいけない。

しかしながら、AI は検索者の好みについて「学習」した内容を積極的に保持するから良くない。科学分野（特許分野）ではそのやり方は間違いなく役に立たない。

この「検索中立性」の欠如は、他の最近の検索で使用された最も近い用語で自動補完され、特定のトレンドトピックが検索の最も可能性の高いターゲットとして提案されてしまう。他の一般的な検索と「似ている」検索に向かって誘導されてしまう。これは特許検索プロセスにとって全く受け入れられない機能である。

熟練した人間の検索者は、異なる情報データベース間、

または異なる形式の文献間を移動するときに、検索戦略を適応させる。特許情報を検索する際には、AI システムを完全に再トレーニングする必要がある。そうでなければ、初心者の初心者と同じレベルの効率でのみ検索実行がされてしまう。

現状の AI システムは、熟練した特許調査員にとって非常に不十分な代替品である。AGILE 検索システムにおいては、質の高い検索戦略が履行できるように設計すべきである。①ユーザーの認識と②フィードバックに応じて、検索戦略を適応させることができるものを開発すべきである。

【バイアス】

現在のアルゴリズムは、一般経験の背景を構築し、将来の結果を提供するために、学習データのセットに依存し過ぎている。学習データに特定のバイアスが含まれて、偏った視点を継承してしまう。不完全な予測システムを実装しているから問題である。

特許検索用の AI を開発するという特定のコンテキストで対処する必要がある。

現状の AI は、テキストは読むが、非テキスト言語は読めなく無視してしまう。単語の頻度による結果で、ランキングに基づいてリストの先頭に表示されたりする。バイアスは、マシンによって再現、永続化、さらには強化される。

最新の AI システムの設計は、システムが自律的に機能する能力を目標としている。その結果、人間のユーザーが検索プロセスに対する洞察やそれを修正する機会を否定する。

発明に関するより多くの情報が非テキストで表現されて、表形式のデータ、図、図表を含む全く異なる「言語」に埋め込まれていることもわかっているが、AI はそれを考慮していない。唯一の例外は化学構造による検索である。

【一貫性】

AI ベースの特許検索システムの開発では、検索の一貫性で 2 つ問題がある。

1 つ目は、同じ検索エンジンが、場所や状況に関係なく、すべての検索者に同じ結果を提供できないという検索のローカリゼーション問題である。

2 つ目は、特定の時点で検索されたものを常に再構築できるように、特定の時点での検索プロセスをアーカイブで再現できないという問題です。

例えば Google は検索ローカリゼーションを実装していて、AI 検索システムが同じ検索エンジンが同じデータベースの異なるユーザーに同じ日に異なる結果セットを配信している。例えば、日本の検索者は、日本語の文献は検索できるが英語の文献を AI は検索に行かない。

2 番目の問題は、検索戦略の保存と再利用に関連する。検索アルゴリズムが常に進化しており、過去のある時点での状態にロールバックできない。この環境下では、元の検索者が過去の検索がベストプラクティスに沿って適切な方法で行われたことを示すために必要な多くの証拠にアクセスできなくなった。特許訴訟に関連して重要な課題である。

【AI は前進する唯一の方法ですか？】

AI を開発する動機の 1 つは、特許出願の数、特許の数に関する統計、候補回答の数の増加に対して「人間は対処できない」という理由からです。この時点で、人工知能の代わりに拡張知能と呼ばれるアプローチを優先ルートとして筆者は好む。

特許調査の重要な第 1 段階では、人間のサーチャーが必要で、インフォームドディスカッション、主題分析、反復試行調査を通じて実際の情報ニーズを引き出すべきだ。

筆者の主な懸念の 1 つは、「現在開発中の AI の品質は、人間・同僚に期待する品質にはほど遠い」ことである。知的な人間の特許検索者は予見可能な未来に向けて仕事を行い、AI に代わることはない。