Patent
Seekers

Patent Searching at the AI-Life Science Intersection

Table of Contents

Introduction

The AI-Life Science intersection has seen a rapid development in recent years. The utilisation of Artificial Intelligence (AI) in the field of life sciences can have immeasurable benefits, certainly where it relates to speeding up processes such as clinical trials and drug development. However, the term AI is very broad and encompasses an abstract array of technology and applications, whilst the field of life sciences is equally as complex, for a number of different reasons (which we will discuss later on). Combining these two fields opens up a veritable minefield of pitfalls and unforeseen troubles that could cost you and/or your client a significant amount of time and money. Here we hope to delve deeper into the AI-Life Science intersection, to try and shed some light on its history, the aforementioned pitfalls and discuss some strategies that will hopefully set you on the right path towards a beneficial outcome when dealing with these two complex technologies.

What is the AI-Life Science Intersection?

The subject matter we will be discussing is, by its very nature, abstract and complex. Therefore, in order to better understand what we are discussing it serves to briefly clarify what is meant by AI, Machine Learning and Deep Learning and how exactly these concepts interface with the life sciences.

Artificial Intelligence in its simplest form is the science of getting computers to exhibit intelligent behaviour indistinguishable from a human.

Machine Learning is an application of Artificial Intelligence that comprises systems with the ability to access data and learn from said data and experience, so as to improve without being programmed. The systems do require human input in the form of structured/labelled data.

Deep Learning is a subset of Machine Learning, wherein the systems do not require structured/labelled data. The systems have networks capable of learning independently from unlabelled/unstructured data. Deep Learning requires a significantly larger data set in order to operate effectively.

The intersection is essentially the application of AI and machine/deep learning to existing biotechnology applications, medical apparatus or pharmaceuticals, as well as applying it to the development of new advancements in those fields. For example, using machine/deep learning to analyse medical images to assess for abnormalities such as cancerous growths or to analyse a database of pharmaceuticals that may be applied to treat a new disease or disorder.

AI: A Patent History

A hoard of patent information is out there relating to AI. The quantity of potentially relevant prior art is a staggering prospect, with one of the earliest patents that discloses the term "artificial intelligence" having a priority date from 1962 - US3388381A, entitled "Data processing means", is owned by the US Navy. Interestingly, the term "Machine Learning" does nothing to narrow the searching (in fact the field appears larger when looking at Figures 1-3), with one of the first disclosures of the term hailing from 1961 (US3254324A), which also makes further reference to machine learning by stating:

"It is quite conceivable that the machine could be programmed to improve its own formulas on the basis of experience."

This comment alone embodies the abstract nature posed by such technologies, and the subsequent difficulties that follow when patent searching within those fields.

"Deep Learning" has a much more recent footprint in the patent landscape with the first significant event occurring in 2015-2016. The earliest priority appears to hail from 1998 and is attributed to US2004018476A1.

As clearly shown by figures 1-3, the fields of AI, machine learning and deep learning have exploded in recent years, with a clear spike first occurring in 2017 and then continuing onwards in this fashion.

It's important to note that the apparent drop in 2019 is most likely the result of applications yet to be published that would be claiming these years as priority.

To try and get some indication of the first instance of when the intersection between life sciences and AI occurs, we can search for patents that have been filed in at least one of the more common AI patent classifications, such as:

International Patent Classifications (IPCs)
G06N* COMPUTER SYSTEMS BASED ON SPECIFIC COMPUTATIONAL MODELS

United States Patent Classifications (USPCs)
706 Data processing: artificial intelligence

AND in at least one of the more common life sciences patent classifications, such as:

International Patent Classifications (IPCs)
A61* MEDICAL OR VETERINARY SCIENCE; HYGIENE
C07K* PEPTIDES

United States Patent Classifications (USPCs)
424 Drug, bio-affecting and body treating compositions
435 Chemistry: molecular biology and microbiology
436 Chemistry: analytical and immunological testing

The earliest patent from this dataset appears to be US3546601A – NEURONAL EVENT RECOGNIZER. Figure 4 shows a similar trend to that of figures 1-3. For each of figures 5-8, the US is the clear top contender when it comes to filing patents in relation to AI, deep learning, machine learning and the AI-Life Science intersection. China appears to be the second most proactive territory, however, they are still significantly further behind than the US, with the exception of the deep learning field.

It is also important to note that figures 5-8 do not show EPO or WO patents.

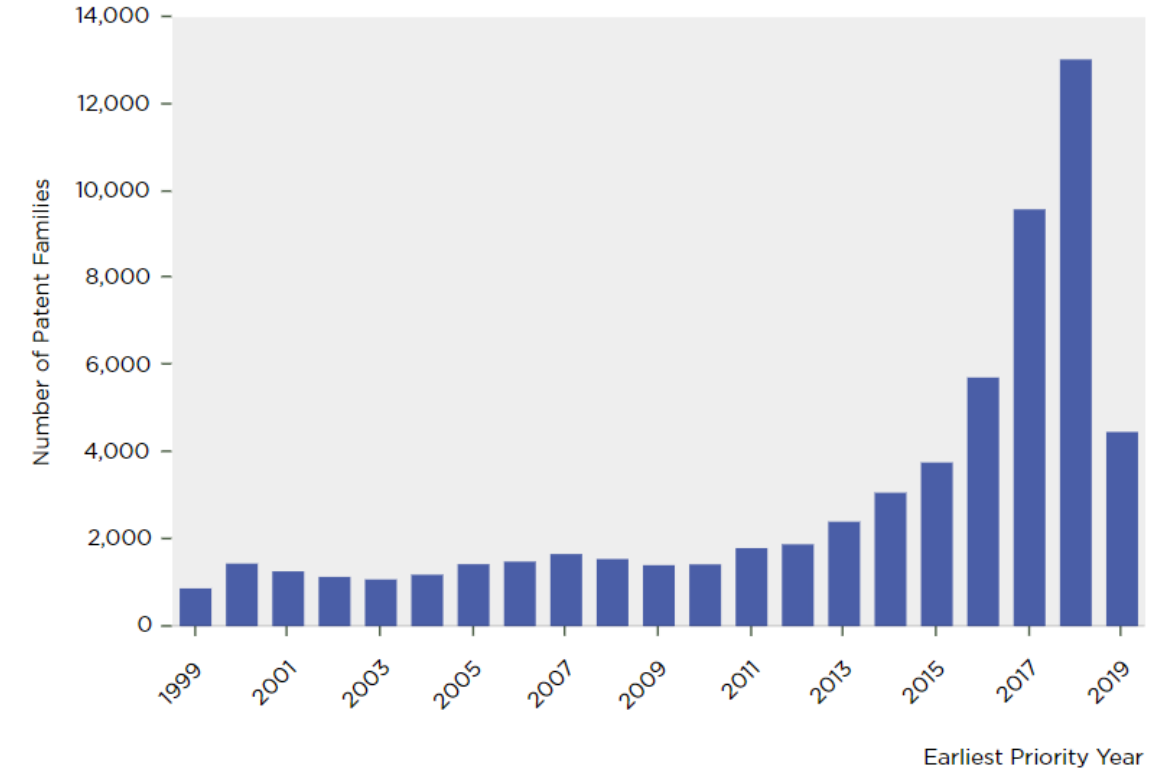

*Figure 1: Earliest Priority Year vs Number of Patent Families – Artificial Intelligence**

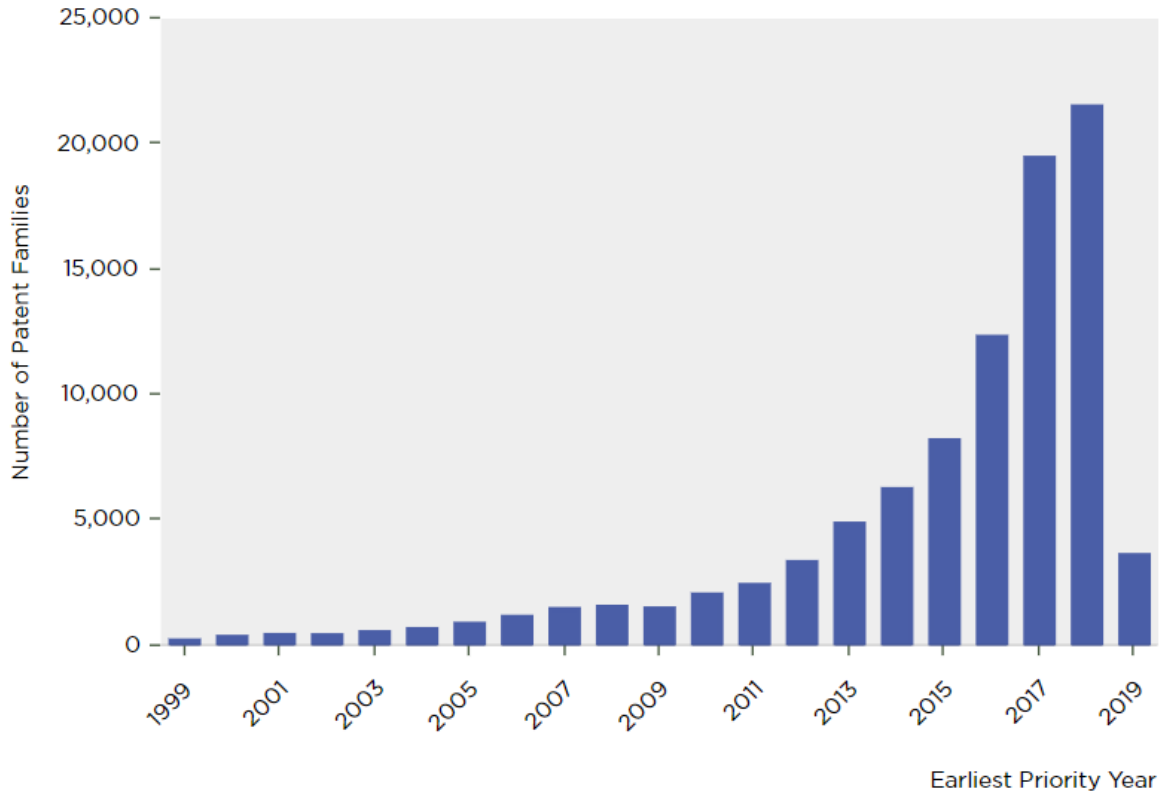

*Figure 2: Earliest Priority Year vs Number of Patent Families – Machine Learning**

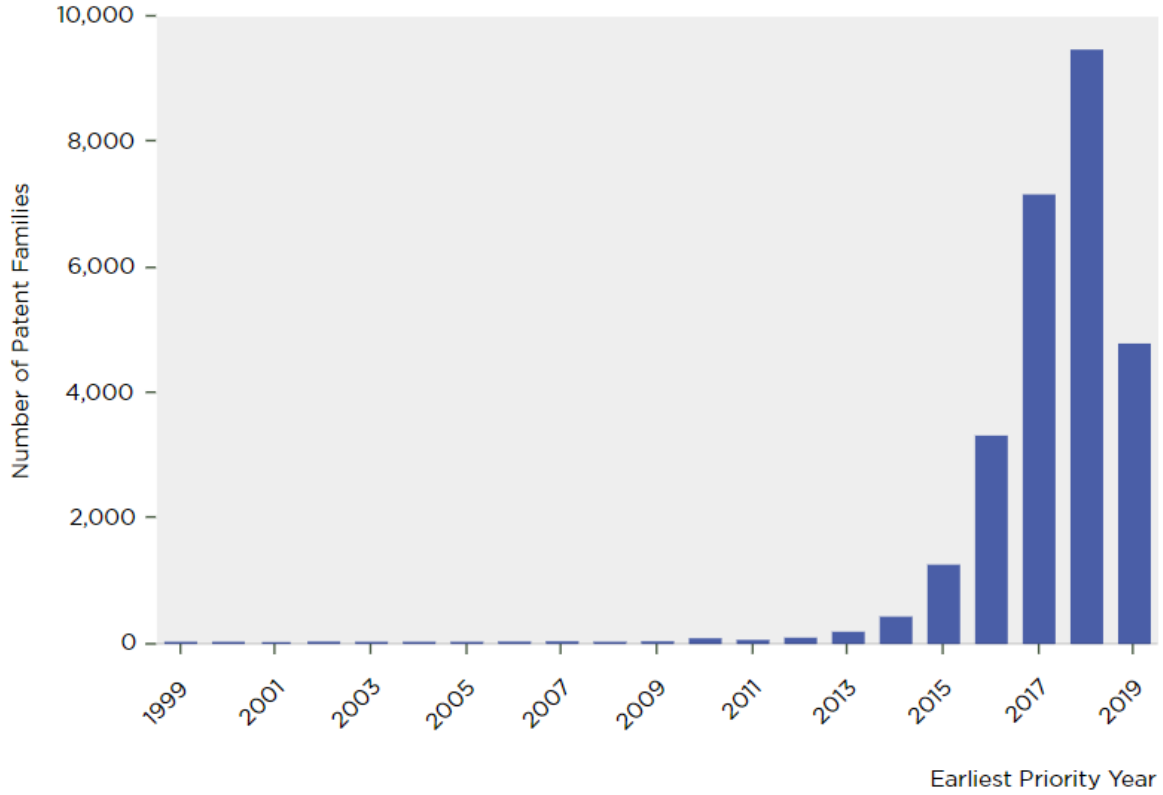

*Figure 3: Earliest Priority Year vs Number of Patent Families – Deep Learning**

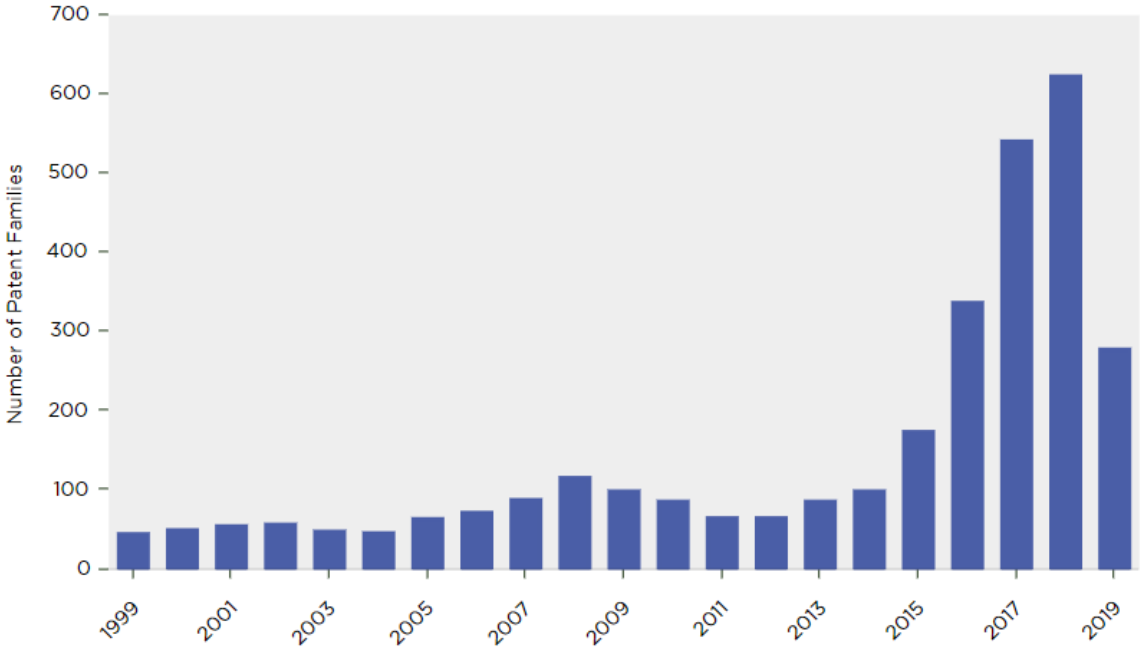

*Figure 4. Earliest Priority Year vs Number of Patent Families – AI-Life Sciences Intersection**

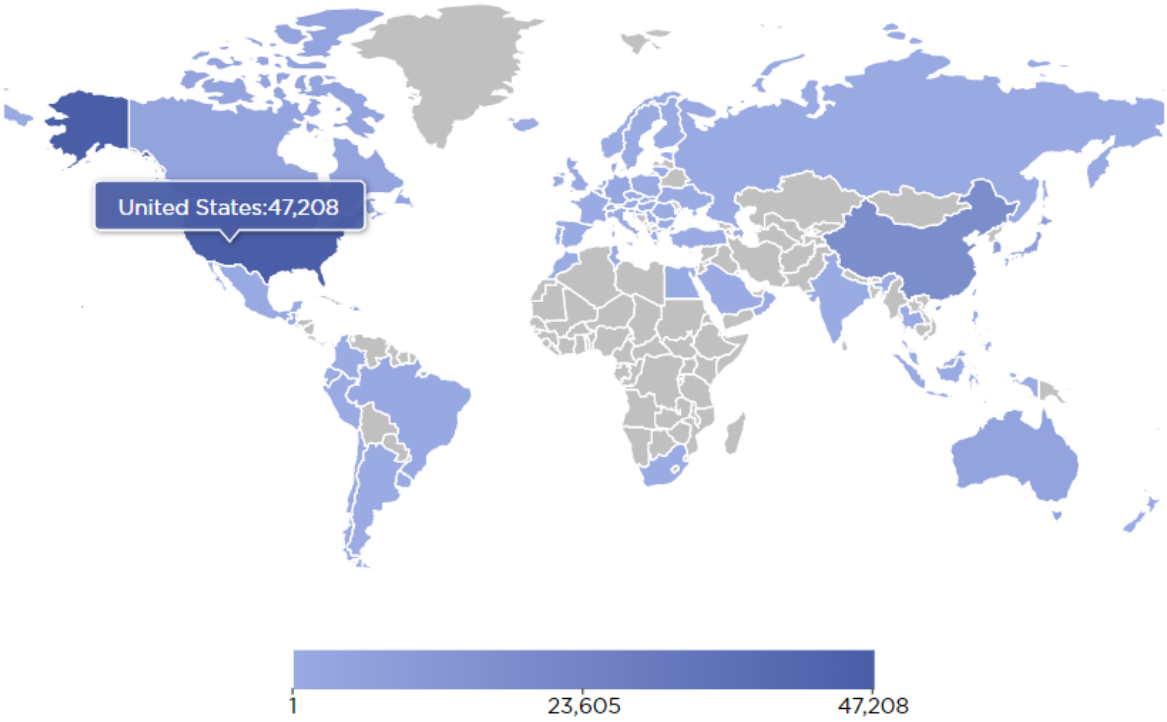

*Figure 5. Global distribution of Artificial Intelligence related patent filings.**

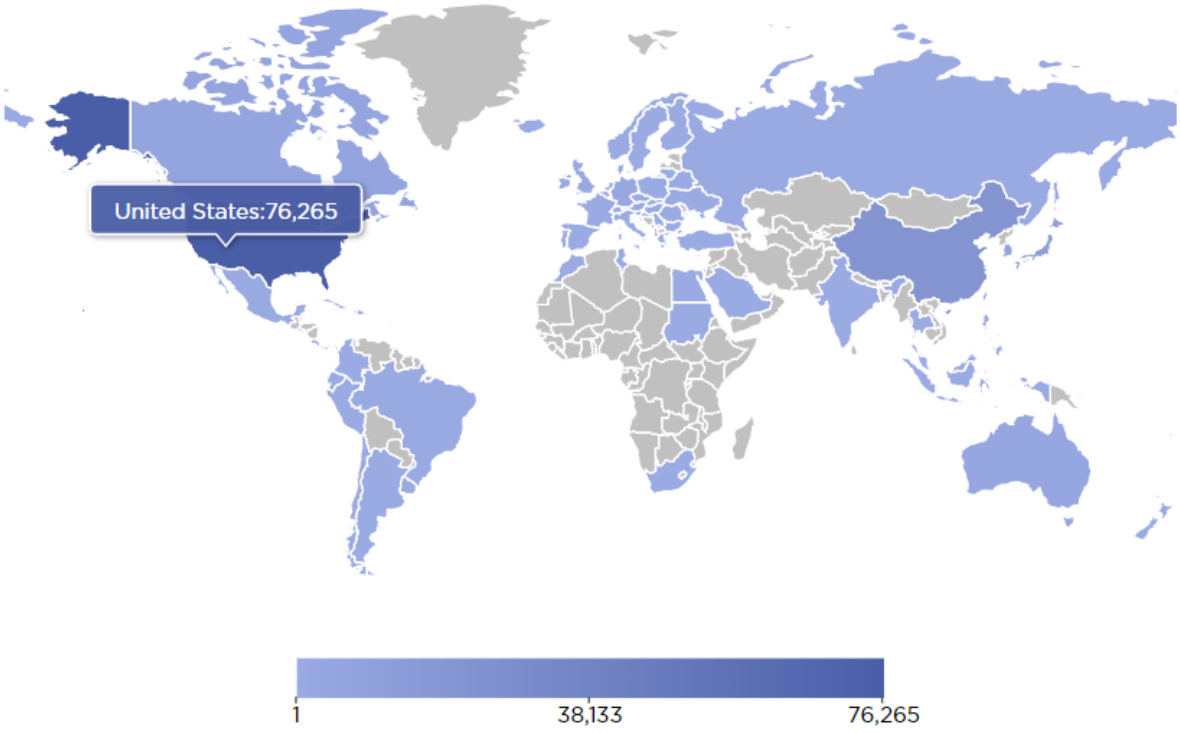

*Figure 6. Global distribution of Machine Learning related patent filings.**

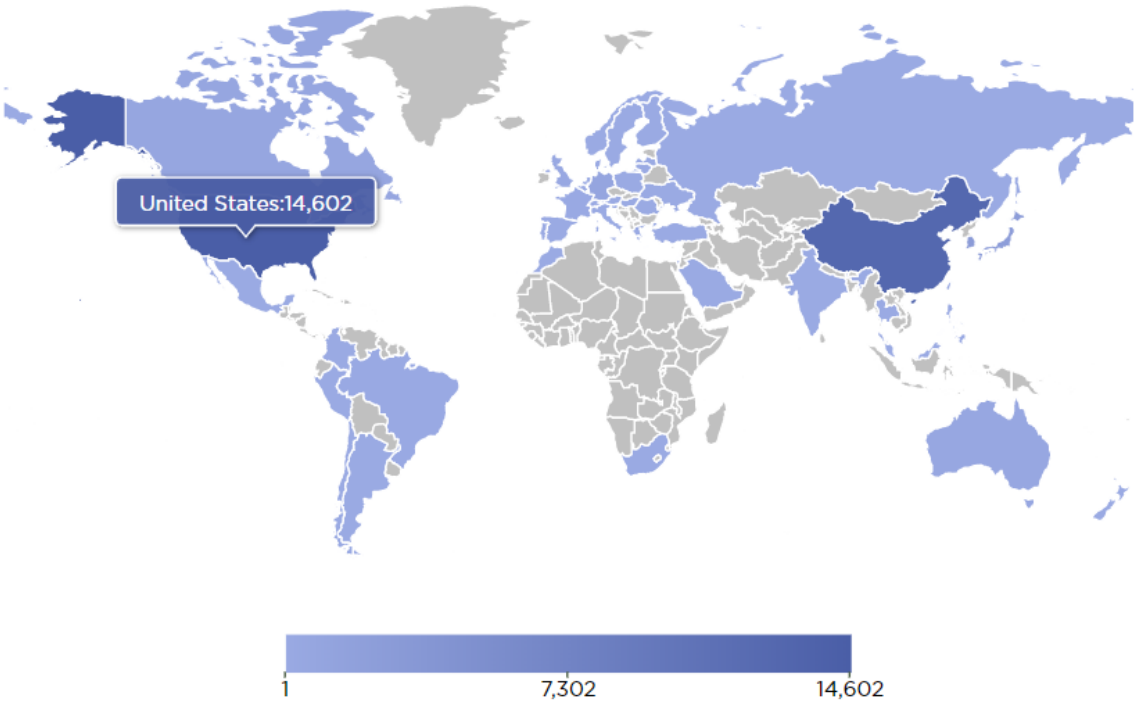

*Figure 7. Global distribution of Deep Learning related patent filings.**

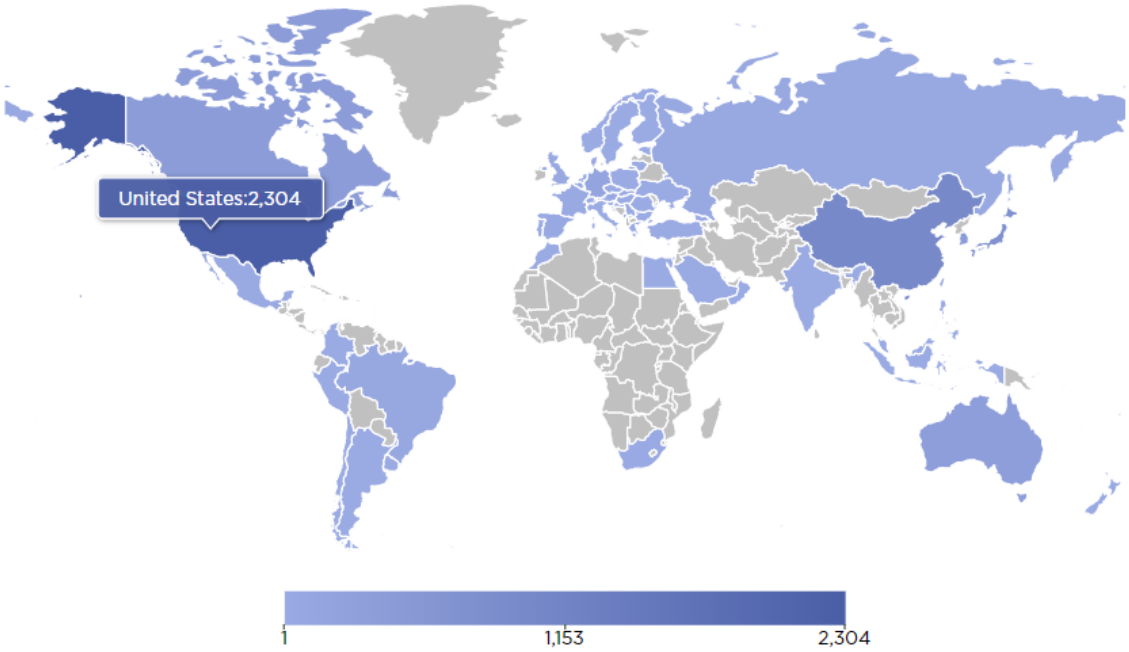

*Figure 8. Global distribution of AI-Life Sciences related patent filings.**

All figures within this report were produced using PatWorld.

Search Problems

Searching for patents relating to machine learning poses its own set of unique problems, as does searching for particular subject matter within the life sciences field (particularly the type of subject matter that tends to be associated with the AI-Life Science intersection). Searching for patents that pertain to the intersection between the two areas only serves to exacerbate the issues.

Here we'll discuss some of the common problems and pitfalls when searching within these two fields, such that you may (hopefully!) avoid them should you ever find yourself in this space.

Sequence Searching

The biggest issue that presents itself when conducting a search based around a biological sequence is the presentation of biological sequences within patents themselves. Sequences can be presented in a number of different ways, including: as images; in separate, non-searchable files; as "SEQ ID:#"; and incorrectly (due to uploading and searching database errors). These numerous formats each require a different technique to overcome, these techniques can involve using a specialised biological sequence searching database, clever use of truncation or using broad keywords so as to find potentially relevant results in the area and manually examine the sequences.

Synonyms

Using a wide variety of keywords when searching is always important, it is even more important when it comes to the machine learning and life sciences fields. With respect to machine/deep learning, the sheer variety in which a machine/deep learning algorithm can be described means that you need to cast a wide net in order to be assured that you have conducted an effective search. Life science patents have different issues: molecules may have a plethora of different names; some organisms and molecules may have had their names changed over the years, potentially causing older patents to pass through your radar undetected; patents may try to use broad terminology to describe molecules, such as describing antibodies as proteins, or describing interleukins as cytokines.

Symbols

Symbols are a prevalent issue within the life sciences field, symbols such as α and β can be written and displayed in multiple formats within databases. Common displays include: the letters a and B, for alpha and beta, however, certain other symbols often can be interpreted as the most similar character the database can find; occasionally the symbols may actually be displayed appropriately, though as is often the case when this happens, the database doesn't support actually searching for the symbols; lastly, the database may simply replace it with another random symbol or text because it doesn't recognise the symbol. Much like with sequence searching, the most efficient way around these issues is with clever truncation and broad keyword searching.

Abstract concepts

This point is very similar to that made in the synonyms section, and it is that machine learning and deep learning are by their very nature, abstract concepts. From a searching perspective this means that a wide range of keywords and arrangement of these keywords can be used to describe the algorithm and/or the task that it performs. It is therefore always of the utmost importance that you bear this in mind when conducting a search in this field and always use a broad selection of keywords. It is also important to utilise patent classifications effectively during your search. These classifications are assigned by the patent examiners and their respective offices and therefore should provide a somewhat solid foundation from which to develop your search around. The flip side of this information is that assigning appropriate classifications can be a subjective task.

Rapidly expanding field

Both AI and the life sciences are rapidly expanding fields, there is even more truth to that statement when you take a look at the AI-Life Science intersection. Unfortunately, due to the 18-month publication delay ("black hole") there will be potentially relevant patent literature that is inaccessible to

anyone, that can't be found by any means. That said, this pitfall can be somewhat mitigated, by "topping-up" your search after 18 months.

AI-Life Science Search Strategy

To a degree, a search is only as strong as the words you use during said search. Research is the first place to start with any patent search and it is vital that you research all the terminology that may be used to describe the novel parts of your invention. In most cases, it is likely that an individual will be an expert in either AI/machine learning OR life sciences, if this is the case, then it could be a helpful activity to reach out to someone within your organisation with the appropriate background (if possible of course!) such that they can offer their expertise to compliment your own. That said, professional patent search companies should have multiple teams with varying backgrounds and disciplines, so you should always be safe in these instances (it never hurts to ask though – and you'll feel all the more reassured knowing that their teams have experience collaborating effectively on matters as complicated as these!).

Now, it is important to bear in mind how abstract the definitions and/or descriptions of AI/machine learning can be. To overcome this, you need to open up the breadth of your "key words" to encompass the broader terminology used in the patent literature. Whilst terms such as "machine learning", "deep learning" and "neural network" may be essential to the search, they are by no means all-encompassing. Key words such as: train, teach, taught, learn and experience, should also be considered. These words enable you to cast a significantly wider net when searching and help mitigate the difficulties of navigating such an abstract field. The other side of the coin here is that the broader you go the more likely you are to find a larger volume of irrelevant results, thus wasting more of your time. So, it really is important that you get the balance right.

An equally important aspect of patent searching is the use of patent classifications. Patent classifications are going to be a useful resource to help overcome some of the problems mentioned previously, especially with machine learning. If you have a patent search carried out and classifications are not used, then you can be assured that it was not of the quality you need. To help you, here are a few of the more common classifications associated with machine learning and AI:

<u>International Patent Classifications (IPCs)/ Cooperative Patent Classifications (CPCs)</u>
G06N3/00 Computer systems based on biological models
G06N5/00 Computer systems using knowledge-based models
G06N20/00 Machine learning

<u>United States Patent Classifications (USPCs)</u>
706 Data processing: artificial intelligence

The term "life sciences" is far more encompassing than machine learning so these classification suggestions will just be a useful starting place:

<u>International Patent Classifications (IPCs)</u>
A61 MEDICAL OR VETERINARY SCIENCE; HYGIENE
C07K PEPTIDES

<u>United States Patent Classifications (USPCs)</u>
424 Drug, bio-affecting and body treating compositions
435 Chemistry: molecular biology and microbiology
436 Chemistry: analytical and immunological testing

It is important to remember however, that classifications are assigned as the patent office sees fit, so whilst these are a useful starting point, they are not exhaustive.

A useful tip in this instance is to familiarise yourself with the patent language as well, try and identify some initial patents that are in the same "area" as yours, these will be a treasure trove of relevant terminology and classifications.

The databases you use for searching will also heavily influence the quality of your result set. Make sure you are aware of the limitations of each database within your repertoire. Free databases in particular will often have significant limitations to them, thus hindering your search. The complexity of

AI and life sciences related patents often means that searching within such limitations could leave you with an incomplete (or even non-existent!) result set, placing you in an ill-informed and dangerous situation. Commercial databases, as you'd expect, are significantly superior to their free counterparts, in terms of coverage and searching capability. A truly useful perk of some databases is the ability to search for biological sequences, affording you an extra avenue of coverage and search completeness. Biological sequence searching can also aid you in avoiding some of the pitfalls mentioned in the previous section.

Conclusion

It has been shown, without a doubt, that the intersection of AI and life sciences is both a rapidly expanding and complicated field. These two factors exemplify the dire need for a comprehensive, professional patent search. With ever-increasing technological advancements in each field, this trend is sure to continue.

There are so many nuances to AI and the life sciences, as well as patent searching, that to combine all those fields together in one big endeavour is no easy feat. With that in mind the benefits offered by a professional search firm are invaluable, their experience, access to numerous commercial databases and knowledge of the subject matter and importantly, their ability to collaborate interdepartmentally make for a superior outcome every time. Always make sure to 'put the screws' to a search firm, ask for search samples – both showing different types of cases and showcasing different subject matter, and make sure to ask for background on the search teams. There is often a large amount of money involved in the fields of AI and life science so don't be afraid to ask these things of a search firm. In the intersection of AI and life sciences a prior art search is invaluable, critical and can offer a wide array of useful information, however, when done ineffectually, the repercussions could prove costly.

When the attorney and searcher have open discussions about the invention, the strategy and the desired outcome, exhibiting true collaboration, that's when you begin to have genuinely beneficial and prosperous experiences.

If you found the topic discussed in this paper to be beneficial and/or interesting then please check out the free webinar Patent Seekers, Torrey Pines Law Group & IPWatchdog collaborated on here: https://www.ipwatchdog.com/webinars/torrey-pines-patent-seekers-patent-strategy-life-sciences-ai/

Contacts

Author
Alec Griffiths
IP Manager
Biotechnology Department
Patent Seekers (UK and North America)
alecgriffiths@patentseekers.com

With thanks to:
Christian Hartmann
Head of Business Development
Patent Seekers (North America)
chartmann@patentseekers.com

Kevin W. Buckley
Founder
Torrey Pines Law Group, P.C.
kevin@torreypineslaw.com

Neil Thompson
Patent Agent
Torrey Pines Law Group, P.C.
neil@torreypineslaw.com

Further Contact Details:

Patent Seekers North American Office:
120 Adelaide Street West,
Suite 2500,
Toronto, Ontario,
M5H 1T1,
Canada

NA@patentseekers.com
+1 416 847 7309

Patent Seekers UK Office:
Innovation House, 6 Cedar Court,
Celtic Springs, Newport,
South Wales,
United Kingdom,
NP10 8FY

mail@patentseekers.com
+44 (0) 1633 816601

PatWorld North American Office:
120 Adelaide Street West,
Suite 2500,
Toronto, Ontario,
M5H 1T1,

Canada

salesNA@patworld.com
+1 416 847 730

PatWorld UK Office:
Innovation House, 6 Cedar Court,
Celtic Springs, Newport,
South Wales,
United Kingdom,
NP10 8FY

sales@patworld.com
+44 (0) 1633 816601