Haseltine Lake Kempner

Drafting patent applications for AI

October 2019

Artificial Intelligence has been described as the last great invention. Despite that, any patent attorney can tell you there is no shortage of inventive activity in the field. This presents a pressing question: What can we do to ensure that applications being filed today have the best chance of standing up to the tests of tomorrow?

In this article, Caroline Day and Gemma Robin explore current AI techniques and, with the help of our colleagues Joanne Addison and Catherine Williamson, consider how lessons learnt from drafting patent applications in the field of chemistry can be applied to writing patent applications for AI.

Imagine better™

AI – Technical background

Categorising AI

There are more ways of categorising AI than Alexa has weak jokes. For the sake of this article, we will be discussing just three of these categories on the basis that – for now at least –they appear to underlie the majority of AI inventions (where the term AI is used herein to be inclusive of machine learning). These categories are supervised learning, unsupervised learning and reinforcement learning. If you recognise all those terms, we recommend you skip ahead to section 3. For readers less familiar with the topic, a high level introduction follows below.

Consider a set of handwritten ones and zeros, thousands of each, written by people of different nationalities, ages and propensity for neatness, collected together in a single muddled list.

You, honoured reader, could look at that list and immediately recognise that it was made up of two different types of shape- a stick-like shape and loopy shape. In other words, despite the fact that French number ones look rather like English number sevens and some zeros are circles while others are ovals, some open at the top, or closed with an extra loop, it is not a hard task for you to separate the ones from the zeros. You wouldn't even stop there. You would ascribe meaning to their shapes- these are numbers. Numbers could mean a host of things but you may know that ones and zeros together can mean something special. Ones and zeros together with nothing else- that could be binary code.

So from the handwritten scrawls of strangers, you have categorised, understood and contextualised information at first sight, likely all in less than a second.

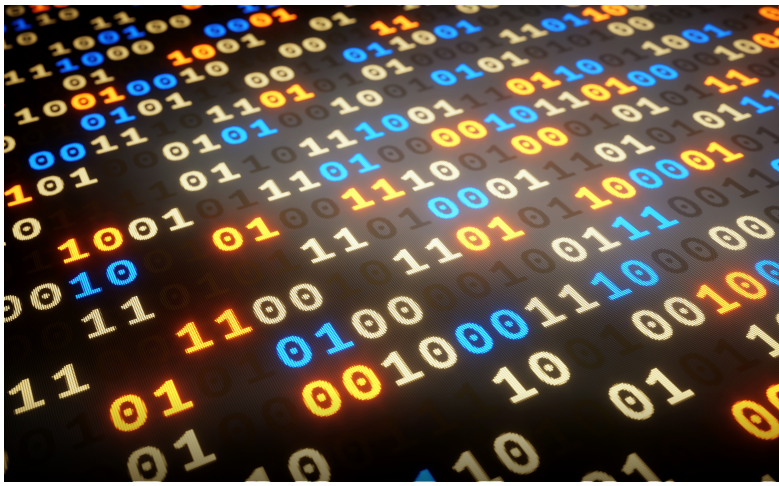
This is intelligence and it's not as easy as you make it look.

Let's consider the task from a point of view of a machine. To this machine, each one and each zero may be a grid of pixels, and each of those pixels may be associated with a value: for example, 0 for black pixel and 255 for a white pixel, with shades of grey somewhere in between. And that's all it knows. It doesn't know that shape is important here, or that there is any attempt to convey meaning. It doesn't understand what a number is. It certainly doesn't know that we are, quite literally, talking its language. It just knows the values and a pixel index.

So how to get from that to a machine which might be anywhere as good as we are at telling the ones from the zeros?

Supervised Learning

One thing we could do is help the machine along. We could take a hundred of those ones and a hundred of those zeros and tell the machine what they are using labels. This is supervised learning and the labelled ones and zeros are referred to as a training dataset. The machine's task is to generalise the characteristics of the ones and zeros



from the training dataset and to make guesses about future ones and zeros it may encounter.

To do this, the machine may take the values associated with each pixel and look for patterns: Is it more likely that, if pixels 1038, 53, 132 and 506 are white, the label associated with that example is a zero? If so, that combination of pixels may have a positive relationship with the label "zero" whereas, if the opposite is true, it may have a positive relationship with the label "one". Hundreds, thousands or millions of such groupings may be assessed, and their relative importance – the strength of the relationship – deduced and encoded in a model.

An interesting observation may be made here. If a human was asked to teach a machine how to categorise the ones and zeros, they might consider how to recognise the shapes. For example, you could consider a one as being made up of lots of little lines which are generally in the same direction whereas a 0 is made up of lots of little lines which change direction. Therefore, you might be tempted to write a program which divides each shape into little lines and considers how they are orientated with respect one another- perhaps a zero has around about an equal number of lines in each direction, whereas the ones favour one direction over another substantially. However, research has shown that, in a

"The machine is concerned only with getting the right answer, not with developing an analytical understanding of the solution."

model derived by a machine, pixels can be associated seemingly at random. The machine is concerned only with getting the right answer, not with developing an analytical understanding of the solution.

Unsupervised learning

Kind as it was of you to label up a training set, what if you can't you bothered? Another option would be to present the dataset un-labelled and ask the machine to work out what was there. To do this, the machine could be tasked with dividing the images in to sets, and hopefully, for example by associating pixels as described above, the machine could identify set A: loopy things and set B: sticky things.

Techniques of this type may use techniques such as clustering or self-organising maps which can loop forwards and backwards through the dataset trying to find features in common and draw inferences from these. It might not deduce the groups we're expecting though: it could for example decide to divide the dataset into those with more black in the top right and those with black on the left. It may divide things into three groups, or four. The usefulness of the solution it derives could well depend on factors such as a well-constructed training dataset and any initial assumptions (which may be referred to as bias) provided to the machine.

Reinforcement learning

In reinforcement learning, there is an underlying understanding of what the 'right' answer looks like. To continue our example, the 'right answer' may be correctly assigning each digit into one

of two sets. Thus, in (a slightly strained) example of reinforcement learning, the machine may initially divide the digits into two sets randomly and receive a score of how well it had done compared to the optimal result. The machine may then try this several more times and work out if there are any attributes associated with a better score rather than worse score. The machine can now develop a policy for dividing the ones and zeros based on attributes of its good guesses and excluding attributes of its bad guesses.

A better analogy here might be a self-driving car simulation, with the car finding its way past obstacles, trialling different actions available to it according to its current state and environment in order to improve a distance travelled. A greater distance results in a higher score for the computer, and so better and better policies for deciding how to move may be identified over a number of trials, in some cases iteratively improving in a manner comparable to evolution.

AI in practice

Elements of these different learning techniques can be combined to generate AI as we know it.

For example, an Alexa device may ask you if a response has answered your question. This can be used either to evaluate a supervised learning result or in reinforcement learning, which could improve a solution which may have initially been developed using supervised or unsupervised learning, and so on.

Designing AI machines

When considering inventorship, we often think in terms of how an inventor has solved a problem. For AI inventions, we would suggest a small shift in perspective: how did the inventor create the right conditions to allow the machine to solve the problem? In other words, for all that the solution may be derived by a machine, there will be at some point a directing hand of human which contributed to that solution being a good one. There are many decisions made in designing AI machines, and it is those decisions that make the difference between a machine of significant technical and commercial value, and an expensive waste of processing power and resources.

Let's start with the training data set.

There is a story, likely apocryphal, in relation to early image recognition techniques which serves to illustrate a point. The task at hand was to identify army tanks which had been camouflaged amongst trees. 50 images of camouflaged tanks and 50 images of woodland were provided, duly labelled and the derived model worked brilliantly at classifying further images. The 'proven' model was then deployed in practice and found to be utterly useless. Analysis revealed that the images of tanks had been acquired on sunny days (when tanks are at their prettiest) and the images of woodland had been taken on cloudy days. Rather than identifying covert enemy activity, the algorithm was instead detecting picnic weather.

In other words, design of a training dataset can be critical to the success of the algorithm. It is exceptionally easy to introduce bias at this point: if the training dataset contains 60 images of red wine and 40 images of white wine, then it appears to the computer that red wine is more common than white wine and

"The human hand may be found throughout AI machine design."

thus a bias towards red wine (shared by these authors) will be inherent in the model. Whilst this example is trivial, there are decidedly more serious examples: concerns such as racial profiling or stereotyping being built into AI are very real and can have massive consequences.

Therefore, the composition of the training dataset- its size, how data outliers are handled, how it is acquired- may differentiate a good machine from a bad one. Deciding how the dataset is represented to the machine is also a key choice. For example, should your ones and zeros be presented as arrays of pixels, or as vector models or in some other way? Both the efficiency with which the machine is developed and the usefulness of its output may be influenced by such decisions.

Designing the model used to derive the AI is another important consideration. This can start at a pretty high level: Is this a neural network, a genetic algorithm, a decision tree or something else? How is it structured? Is there any assumption about what is likely to be relevant in the data? Are there any assumptions about the number or type of output expected? Inevitably, we could go on.

Once these decisions have been made, it is common practice to evaluate the performance of an AI machine, for example using a labelled dataset which was held back during training. The evaluation may be comparative: does this model perform better than another (or better than a human), or just as well but more efficiently? In other examples, the standard may be absolute: does this model result in a threshold accuracy of categorisation, or decision making? In still other examples, the evaluation may be more qualitative: is this useful, and if so how?

The human hand may be found throughout AI machine design. A human may start with a hypothesis, for example that data with certain attributes is likely to form a good dataset, or a given type of model or reward function is likely promote desirable behaviour. This may be followed by a test stage, in which the machine may be built based on the hypothesis. Finally, there may be an evaluation of the result, and in some cases it may take a human mind to understand what a good result looks like.

AI and the Chemistry Analogy

Patent Attorneys working in the field of chemistry have long had to grapple with a problem: for all the research, and the identification of likely candidates and the painstaking manipulation of experimental conditions, it may well be the case that the inventor doesn't know ***why*** composition A works better than composition B, or ***why*** the yield of X increases when you do Y. The inventor just knows that they worked long and hard to get to a point where they are sure that it does. Although it may be relatively easy for a chemist to make a random new composition, it is much more difficult to make new compositions that have a particular new or improved result.

This is true of AI too: an immaculately thought-out training dataset, an admirable reward function and a flawless set of assumptions may lead you to a machine of remarkable utility without its inventor ever understanding ***why*** the numbers underlying that particular machine are so successful. Returning to our example of handwritten ones and zeros, the machine's focus on getting to the right answer, as opposed to understanding the problem itself, means it may not be apparent why there is such a strong weighting associated with a particular combination of pixels in a particular stage of a neural network. But by testing the machine, and checking the results independently, the inventor can be sure that their AI machine is a good one, in just the same way that a drug which results in increased survival is a good drug, even if the precise mechanism underlying its efficacy is not known.

In chemical patent preparation, this has led to the practice of ensuring experimental data is included in the patent application. This may include experimental data to demonstrate how the invention is carried out and the effect of the invention, and may also include comparative experimental data to demonstrate an improved effect of this invention in the context of what has gone before.

This standard practice in the world of chemical drafting is, we suspect, likely to be useful to the AI field. In the following discussion, we take a look at some of the uses of experimental data in chemical drafting practice, and explore how the insights gained from the use of such data may be applied to the field of AI.

Sufficiency of Disclosure

In order for a patent application directed to a new chemical composition to meet the requirement of sufficiency of disclosure, the patent application must provide enough information for the skilled person to be able to produce the composition. A description of the composition per se may not be enough, and it is often necessary to include full experimental details of at least one way of producing a particular product, for example including starting materials and reaction conditions, in order to satisfy the sufficiency of disclosure requirement.

Returning for a moment to AI, this may be compared to describing the principles underlying the data included in the dataset, and any assumptions or parameters built into your model.

Now let's consider the case of a chemical application where a particular result is limiting on the scope of the claims. For example, in the case of claims directed to a method that provides a particular result, or to use of a composition for a particular purpose, it is generally necessary to provide data to prove that this result is actually obtained by the invention in order to satisfy the sufficiency of disclosure requirement.

This in turn can be compared to an AI scenario: if the machine is designed to improve distinguishing between handwritten numbers, is this result achieved, and can you prove this?

By thinking of the AI design as a chemistry experiment, it may be possible to identify components analogous to initial compositions (e.g. a training dataset), the experimental conditions (model assumptions and design) and to evaluation of the effectiveness of the result. Including these components in a draft application in the same way as a chemistry attorney would could add substantial weight to your disclosure by showing how the skilled person created the conditions for success.

Demonstrating a Technical Effect Across the Scope of the Claims

Without clearly knowing why a new chemical composition works better than already known compositions, it may be difficult to justify a broad definition of the new composition. However, the applicant may want to obtain patent protection covering similar compositions that are expected to show similar improvements.

In practice, when chemical claims are drafted broadly, Examiners can object that it is not plausible that a technical effect that has been demonstrated for a particular embodiment, e.g. one particular composition, would be obtained for all embodiments within the scope of the claims. In some circumstances, plausibility of different embodiments can be argued based on known scientific principles. However, additional experimental data relating to a number of embodiments across the scope of the claims can be helpful to dismiss any concerns the Examiner may have.

Let's say you want to claim a composition comprising a metal in general. Good practice in chemical drafting would be to include examples demonstrating that a particular technical effect is obtained with more than one type of metal in the application. It would be particularly useful to include examples demonstrating that a particular technical effect is obtained with different metals having different chemical natures in the application (e.g. metals from different groups of the periodic table rather than metals from a single group).

Now let's apply this to AI. Let's say you've designed a classifier which

used a training dataset including pictures of cats and dogs, and a particular model embodying a set of assumptions. Your model is not just effective, but fast and lightweight in terms of processing resources. What you want to claim is the model as a classifier per se, for categorising any two objects into two camps. But have you demonstrated that this is plausible? Or is there something about cat and dog images which isn't true of sets of images of bolts and screws, or oil slicks and rainbows? Selecting diverse examples of datasets may help to show plausibility in such an example.

In the AI field, as in the chemical and life sciences field, including experimental data relating to a range of embodiments covered by the claims could be useful to show that such embodiments provide a similar technical effect, and consequently that the technical effect is demonstrated across the claim scope.

Inventive Step – Setting the Technical Problem

Data demonstrating a particular technical effect associated with the invention is also useful to support inventive step arguments, even when the technical effect is not limiting on the claim. This may arise during drafting, or much further down line once objections have been raised in examination.

In the EPO's problem-solution analysis of inventive step, once a technical effect has been demonstrated, it is used to set an "objective technical problem", i.e. the problem addressed by the invention in view of the prior art. For example, the objective technical problem may be set as the provision of a composition that provides a new or improved technical effect, rather than simply the provision of an alternative composition to provide the same technical effect as previous compositions. Demonstrating a new or improved technical effect associated with the composition therefore strengthens the arguments for inventive step of the invention when the prior art does not provide any teaching in relation to that new or improved technical effect.

Where the technical effect is an improvement of a technical effect already disclosed in the prior art, **comparative** experimental data demonstrating the improvement provided by the invention compared to embodiments falling outside the scope of the claims, for example embodiments described in a known piece of prior art, is useful. And since new prior art is often discovered after the application has been filed, additional data can be filed during examination, provided that it supports a technical effect that is described in the application as originally filed.

For example, the invention may relate to a composition comprising a Group I metal. The closest prior art that you are aware of may generally disclose a product comprising transition metals. The EPO may consider selecting a Group I metal to be obvious when the technical problem is set as merely providing an alternative product to provide the same technical effect as the prior art. However, if data is included that demonstrates that Group I metals provide a particularly good technical effect compared to transition metals, the objective technical problem can be set as providing an improved product, and it can be argued that it would not have been obvious that the claimed Group I metals would provide this improved effect.

The data may additionally demonstrate that, within Group I, lithium and sodium provide a particularly good technical effect compared to potassium and caesium. If prior art was later found disclosing a product comprising potassium, you could fall-back to the product comprising lithium or sodium and argue that it would not have been obvious from the prior art that lithium and sodium would provide this particularly enhanced technical effect.

It may well be the case in AI that nudging a knob in a model may result in significant changes in efficacy. Comparative examples demonstrating the effectiveness of the adjustment may be powerful tool to combat an examiner's assertion that such 'tweaks' are a mere workshop variation.

Comparative data- supplied in an application on filing or possibly in response to prior art raised by an examiner- which demonstrates the effectiveness of including a novel feature may provide a means to enhance an inventive step argument.

Conclusions

While we cannot claim to foresee all the challenges which may lie ahead in the field of protecting AI inventions, we believe that established practices from the field of chemical drafting may be useful in tackling at least some of these challenges.

Therefore, it is our view that deepening understanding of chemical drafting practises, and in particular the usefulness of including experimental data and discussion of experimental conditions, could serve attorneys working in the field of AI well in the coming years.

Contact us

Caroline Day
Partner
cday@hlk-ip.com
+44 (0) 117 910 3200

Joanne Addison
Senior Associate
jaddison@hlk-ip.com
+44 (0) 117 910 3200

Gemma Robin
Senior Associate
grobin@hlk-ip.com
+44 (0) 117 910 3200

Catherine Williamson
Senior Associate
cwilliamson@hlk-ip.com
+44 (0) 117 910 3200