iam

AI inventions and sufficiency of disclosure –
when enough is enough
NLO
Harm van der Heijden

Yearbook
2020

Building IP value in the 21st century

HOW TO FORTIFY YOUR INNOVATIONS, IDEAS AND TRADEMARKS

Innovations, designs and trademarks are a valuable asset in any business. They enhance a company's distinguishing capacity, boost its impact, strengthen its competitive position and make it more attractive to investors. Therefore, there is every reason to give Intellectual Property a prominent place in your business strategy. We can support you by protecting and enriching your innovations, designs and trademarks, thanks to our extensive range of services and the expertise of our trademark, design and patent specialists who find solutions where others don't. We are the largest consultancy in Intellectual Property in the Netherlands and Belgium and one of the most pioneering firms in Europe. If you would like to know more, visit **www.nlo.eu**

NLO

EUROPEAN PATENT AND
TRADEMARK ATTORNEYS

www.nlo.eu

AI inventions and sufficiency of disclosure – when enough is enough

By Harm van der Heijden, NLO

Recently, there has been increased interest in artificial intelligence (AI), both in the mainstream news and the IP world. From self-driving cars to the European Patent Office (EPO) organising a seminar on patenting AI – the technology, or at least the promise thereof, seems to be everywhere.

Alongside increasing AI sophistication comes increasing complexity. The increasing complexity provides challenges to the patent practitioner. How can we adequately patent AI-related inventions? This chapter will explore the disclosing of AI inventions, particularly in view of the requirement that a patent application discloses an invention in sufficient detail for a skilled person to work that invention.

Where this chapter talks about 'sufficiency of disclosure', the meaning of the term found in Article 83 of the European Patent Convention (EPC) and related case law is intended. For the United States, similar requirements apply.

Overview of AI technologies

It is important to distinguish between different forms of AI, since sufficiency of disclosure is not equally relevant to every form. This chapter will follow the overview of AI in Goodfellow, Bengio and Courville's Deep Learning (MIT Press, 2016), as illustrated in Figure 1.

On the outside, a generic AI example is formed by knowledge bases (also known as 'expert systems'), which is essentially a storage of data and a set of rules to draw logical conclusions from this data. Both the data and the rules must be supplied by the operators of the AI.

FIGURE 1. Venn diagram of AI techniques
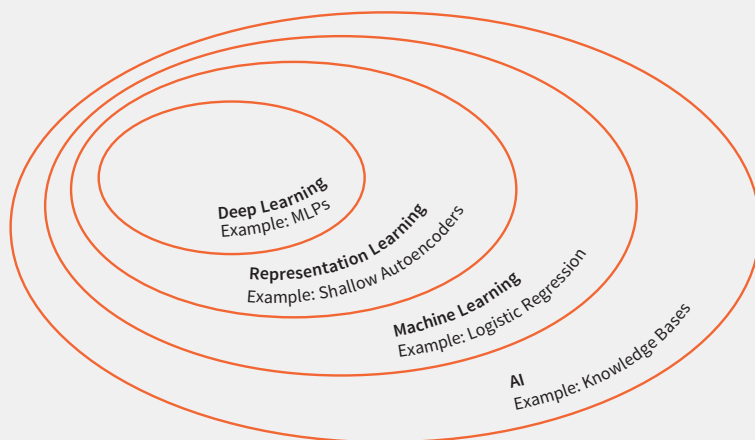
The next level is machine learning, in which the AI will use input and output data presented by an operator and will try to find a rule (eg, using logistic regression) which maps the input data to the output data, so that it can start making predictions for input data for which no output data is available. The procedure of finding a rule is typically called 'training' and the used data is known as 'training data'.

Next in sophistication is 'representation learning', which is a specific form of machine learning. Compared to the logistic regression example above, the AI now also learns to transform the input data into a form better suited for the specific problem at hand.

This is vital when the data becomes more unstructured. For example, suppose an AI model should be trained to recognise if a digital image shows a cat or a dog. Without representation learning, a team of specialists would be needed for this task: a biologist to list key physiological differences between cats and dogs; a graphical artist to draw these physiological differences in various aspects (eg, seen from front, back and side); a mathematician to design a means to calculate the degree of matching the drawings to the pixels in the digital image; and a programmer to program the matching algorithm. The contribution of the machine learning, devising a rule to resolve the detected match values into a binary cat-or-dog output, would be a relatively minor feat.

Representation learning AI technologies can provide superior results compared to basic machine learning. However, this is also where sufficiency of disclosure may become a significant factor. Representation learning comes up with a representation which may not be readily understood, or may at least be difficult to describe.

Finally, deep learning is a subset of representation learning using a model with a number of layers (known as the 'depth'). A general term is 'multi-layer perceptron', which essentially means that a number of relatively simple mathematical operations are applied, each operation adding a layer.

A slightly more whimsical (but perhaps illuminating) definition is that older AI technologies, such as knowledge bases and logistic regression, are typically useful for "things that are hard for humans, but easy for computers" (eg, applying pre-determined rules to large data sets and least squares optimisations), while newer AI technologies, such as representation learning and in particular deep learning, are useful for "things that are easy for humans, but hard for computers" (eg,

Harm van der Heijden
Partner
vanderheijden@nlo.eu

Harm van der Heijden is primarily active in the fields of electronics, communications technology and software. As a partner of NLO, he leads a team of junior and senior patent attorneys, helping clients to maintain their competitive edge by protecting and asserting their intellectual property. Dr van der Heijden has a PhD in plasma physics and a previous career as an industry researcher and developer in signal processing software, medical devices and integrated circuit design. Besides drafting and prosecuting patent applications, he is also involved in patent litigation procedures and licence negotiations.

pattern recognition, image processing or natural language processing).

The deep learning AI technologies have seen dramatic improvements in this century and are the main reason for the current excitement around AI.

Claiming AI

Much has been written already on AI by various commentators. This chapter provides a brief summary of the situation for European patent applications, with some comments on the US position.

Fundamental AI technology

In a nutshell, the EPO has indicated that the approach it has developed for computer implemented inventions also applies to AI. In effect, this means that an AI-enabled invention can be patentable provided that the claimed technical features are inventive (ie, any claimed non-technical features are not considered for inventive step).

FIGURE 2. Diagram indicating layers of a CNN
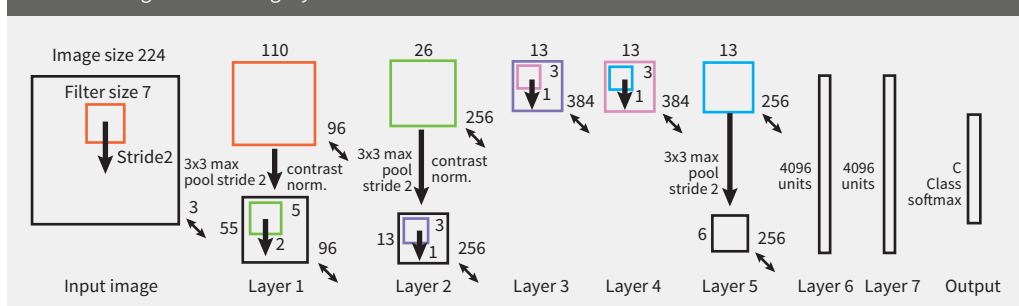
Any claimed AI-related features as such are not considered technical (being mathematical in nature) and are considered only to contribute to an inventive step if they support a technical effect or purpose.

This approach immediately closes the door on the patentability of fundamental AI algorithms (ie, an AI algorithm that is not directly coupled to a specific application). While this is certainly understandable in the case of AI technologies involving relatively simple and well-known mathematics (eg, logistic regression), it is at least questionable whether this treatment is also suitable for the far more complicated, multi-layered models of deep learning, even though every layer by itself is still mathematical in nature.

While patentability of fundamental AI algorithms is effectively ruled out in Europe, in the United States the door is slightly ajar. The two-prong approach of the Mayo framework will work on the assumption (prong one) that a fundamental AI algorithm as a mathematical concept is an abstract idea and thus not eligible for patenting, but (prong two) provides a way out, in that a claim is eligible if "the claim, as a whole, integrates the recited judicial exception into a practical application of that exception" (source: 2019 Revised Patent Subject Matter Eligibility Guidance, www.govinfo.gov/content/pkg/FR-2019-01-07/pdf/2018-28282.pdf).

Applied AI technologies
When AI technology is used for a technical goal, it can in principle be claimed in Europe. However, the question of inventive step will depend largely on the definition of 'technical', which can sometimes be surprising for those not fluent in EPO case law. For example:
- a computer implemented method using AI algorithms for classifying digital images, videos,

audio or speech signals is patentable (technical purpose); or
- a computer implemented method using AI algorithm for classifying text documents is not patentable (linguistic purpose).

Both examples come from the EPO Guidelines for Examination, Part G, Section II 3.3.1, "Artificial intelligence and machine learning" (www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_3_3_1.htm). With respect to classifying text documents, the guidelines cite EPO Board of Appeal Case T 1358/09 and remark that "classifying text documents solely in respect of their textual content is, however, not regarded to be *per se* a technical purpose but a linguistic one".

There is a list of purposes that is considered 'technical' by the EPO, which can be found in the Guidelines for Examination, G II 3.3 (www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_3_3.htm). As a general rule, any purpose that is related to one of the exclusions of patentability under Article 52(2) of the EPC will be considered non-technical. Most notable exclusions are mathematical methods (the reason for excluding fundamental AI technologies), methods for performing mental acts or doing business (this rules out most applications of AI in finance) and presentations of information.

In the United States, the situation seems slightly different in the Mayo framework. The purpose of the invention does not play such a prominent role as in Europe, but rather questions whether the claim relates to a practical application of a judicial exception.

Disclosing AI
Basic structure of AI model
A typical example of a deep learning model is a convolutional neural network (CNN).

FIGURE 3. Examples of filter visualisations in a fully trained CNN, showing image features which elicit a large response from the various filters. In Layer 1 (left), only basic edge detectors are found (such as might be designed by hand); Layer 3 (middle) shows mostly pattern detectors; while in Layer 5 (right) highly specialised filters that trigger on assorted recurring features of the training set (eg, dogs' faces, bicycle wheels and birds' eyes) have (surprisingly) emerged (Zeiler and Fergus).

Layer 1   Layer 3   Layer 5
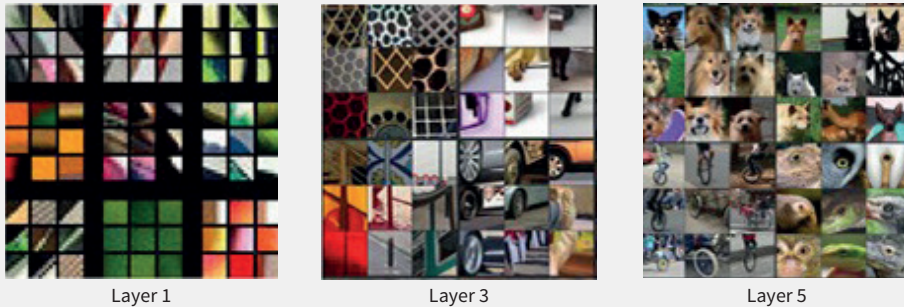
An example CNN is shown in Figure 2 (from: Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", 2013). These types of model are widely used for image analysis, but can also be applied to video analysis, natural language processing and drug discovery, among others.

Without going into too much detail, a diagram such as Figure 2 will tell the skilled person exactly how many layers the model has and how each layer should be configured. Programming libraries for AI models have also developed a handy shorthand for defining models, which could be used to disclose the model's structure in a patent application (eg, Keras' Model API, https://keras.io/models/model/).

In case a claimed AI model uses a component that is not standard (eg, a custom activation layer), this novel component would have to be described exactly, either in mathematical form, pseudo code or actual computer code. This, too, should present no major challenge. The same holds for novel optimisation schemes or non-standard feedback loops.

Training and trained coefficients
Once the basic model is adequately described, the skilled person still does not have enough information to implement the model. In other words, the invention is not sufficiently disclosed. What is still needed is at least one of the following:
- a description of the way that the model is trained, including a reference to the training data; or
- every learned coefficient or weight of the model.

The importance of this cannot be overstated. The only parts of the structure in a deep learning model (such as the one shown in Figure 2) that are pre-determined are the input image at the beginning and the output values at the end. All layers between input and output are called 'hidden layers', which already convey the notion that it is not well known what exactly occurs in these layers after they have been formed during training. In fact, the earlier cited Zeiler and Fergus paper is considered important in the field for the reason that it was one of the first to actually investigate and visualise what these layers do in a fully trained model. Figure 3 shows a few examples of the (surprising) specialisations that emerged in the trained filters of the layers of Figure 2.

Regarding option 1, describing the method of training is generally not difficult. The training data can be a mixture of publicly available data (eg, ImageNet data, http://www.image-net.org/) combined with domain specific data (which will typically not be publicly available). It can also consist exclusively of domain-specific data.

For domain-specific data, the question arises whether a description of the said data suffices (1,000 pictures of cats and 1,000 pictures of dogs), or whether the training data itself must be made available to the public to ensure sufficiency of disclosure. If the latter, how should it be made available? A large library like ImageNet contains millions of images, so it is not a workable proposition to include such a dataset with a patent application. Even smaller data sets used for domain

FIGURE 4. Various options for disclosing an AI model
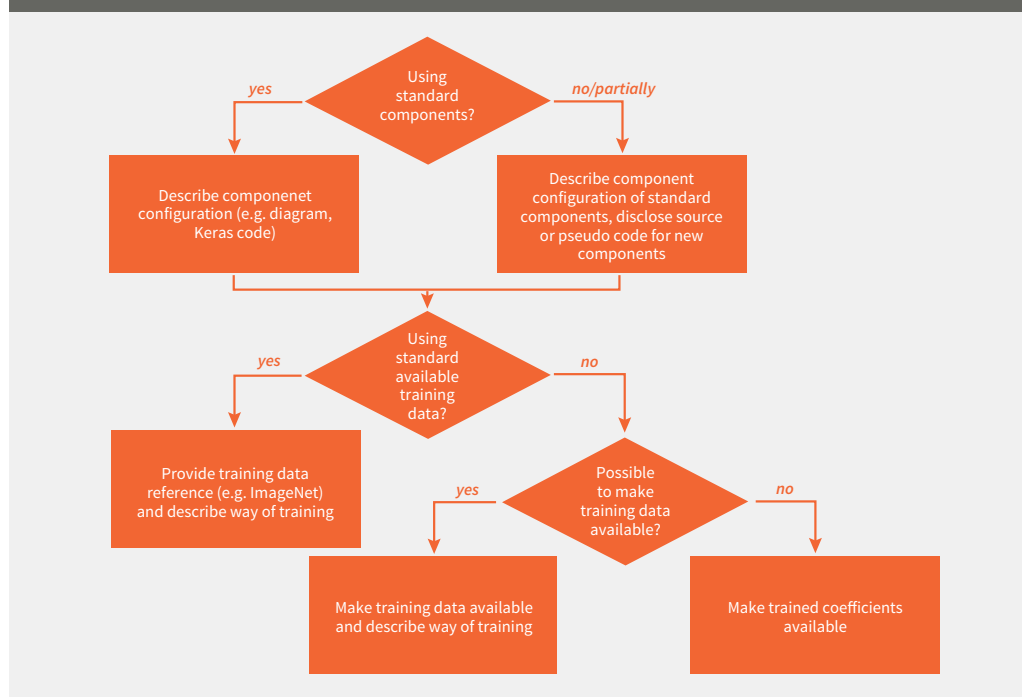
specific training typically include thousands of images.

Regarding option 2, a drawback is that the amount of data to be disclosed in the description can be quite high. For example, in the CNN of Figure 2, Layer 1 will consist of 96 x 7 x 7 = 4,704 trained values; Layer 2 of 256 x 5 x 5 = 6,400; and so on. However, it is quite possible to include this data in tables to fully disclose at least one embodiment of a fully trained model.

However, a bigger drawback is the fact that the trained coefficient data is not particularly useful as a basis for continued research. It is thus imaginable that future case law could decide that such tables do not comply with the spirit of the obligation to disclose the invention. After all, the goal of patent publications is to further global knowledge (in exchange for a temporary monopoly for the applicant), and merely disclosing the trained coefficients might be just enough to enable the skilled person to reproduce a particular embodiment of the invention, but it hardly allows them to improve it (eg, by tweaking the model structure), since that would require access to the actual training

data and training methodology in order to train a modified model.

Based on the above, it is recommended to disclose the method of training and the training data, rather than only the trained model coefficients. The available options for disclosing the structure and the content of the AI model are shown in Figure 4.

A drawback of disclosing the training data is that the applicant may have spent a significant effort in meticulously gathering, and in the case of supervised learning, labelling training data. The applicant may not be inclined to make this data set available to the public, reasoning that a competitor could use it to quickly train a different AI model (carefully selected to avoid infringing the applicant's claims) and thus gain an unfair competitive advantage. This is possibly one reason why it is often only a description of the training method that is included in the description, while the training data is omitted.

We are not currently aware of any present case law that has held an AI-related patent (application) invalid for lack of disclosure of the training data. However, such case law may well develop in

the future and may then adversely affect patent applications being drafted today. Therefore, it is recommended to attempt to future-proof AI-related patent applications in this respect.

Microorganisms to the rescue?
A similar disclosure problem has already been addressed a long time ago in an entirely different field in patenting, namely for inventions involving microorganisms.

A biotechnology invention might use certain microorganisms to produce a useful substance from basic materials, just like an AI invention might use a trained AI model to make useful predictions based on input data. In the case of microorganisms, there is the similar problem of how to allow the public access to these microorganisms in order to work the invention. This has resulted in the system of biological material deposits (under the Budapest Treaty 1977). Access to biological material is given under strict conditions, such as for research purposes.

While not a perfect solution, a similar system for AI training data deposits would give the public access to proprietary data for research, while attempting to safeguard the interests of the applicants who have collected said data.

Comment
With the increasing attention that AI inventions are receiving, it is to be expected that new case law will develop in the coming years. With increasing complexity of AI models, sufficiency of disclosure may no longer be a given.

Certain precautions can already be taken today. The structure of an exemplary AI model should, at least, be clearly described. In addition, the skilled person should have all the required information needed to either train the model or set the model's coefficients with properly trained values. It remains to be seen whether the required information for training also includes the used training data.

As the neural networks are actually modelled on their biological counterparts, it is perhaps unsurprising that patenting neural networks may ultimately come to inherit certain traits from patenting biotechnology, such as depositing training data similar to depositing biological material. *iam*

NLO
PO Box 29720
The Hague 2502 LS
The Netherlands
Tel   +31 70 331 2500
Web   www.nlo.eu