

AI Legal Briefing: Input and Output Intellectual Property

Submitted by: Codified Legal

Monday, 24 April 2023

AI Briefing

Part 2: Input and Output Intellectual Property

Introduction

In recent years there has been the rapid rise of artificial intelligence (AI). Developments such as the emergence and growth of ChatGPT since November 2022 and Bard by Google announced in February 2023 have thrust AI well and truly into the spotlight. These are examples of the most commonly used type of AI tool and known as a large language models (LLM). Large language models are essentially a way of guessing what comes next in a line of text, but on a hugely sophisticated scale. They have been around for about five years but only recently has their disruptive potential come to public attention.

There are many wonderful things that AI can help to accomplish such as assisting vaccine creation or inventing new technology. However, whilst there is good to be gained from using AI, its use does create legal issues. We have produced a series of briefing notes on these, covering regulation, IP and data protection. In this part we'll look at the intellectual property issues posed and protection of AI creations in the current legal framework.

How LLM AI works

AI tools that are used to generate new material such as literary content, images, or even new software code replace the human element of creation relying on "neural networks" to manipulate existing materials available to the tool. The speed and depth at which they can use the source material to create their output is obviously far in excess of human capability.

An AI tool requires "training" before it can produce meaningful outputs, the process of which involves scanning a data set or the internet and 'scraping' potentially millions of lines of text or images.

Most AI tools use a mix of publicly available content (e.g. web-scraped material) and licensed data sets.

The AI tool then uses a node system to process the source material, able to analyse multiple pieces of data at the same time. It looks for patterns in the data and "learns" how to construct the output taking into account the context of the source material.

Once the training is complete the AI tool is fine-tuned, perhaps to make it more focussed on its primary purpose, and feedback is provided on the accuracy of the test output until it is ready for use. The tool can continue to improve and "learn" from feedback or information provided to it once it is operational.

Input IPR: Does training AI infringe existing intellectual property rights?

The way LLM or Generative AI models work can cause conflict with intellectual property law. Under UK copyright law, as soon as someone creates an original piece of art or music for example, they have

protection of that piece of work. If the whole of, or a substantial part of, that work is copied then there is likely to be an infringement of the right of the original owner. Websites may also qualify for protection under similar protection known as the database right.

So does mining the internet to train the AI tool create an IP infringement? These 'input' materials may be subject to intellectual property rights and the use of them to train an AI tool may be an infringement of those rights. We have been here before in the early days of the Internet, through to the Napster music streaming case and we have a number of cases concerning web spidering and news aggregation to look at.

Assuming that some or all of the materials in the training data set are copyright materials the AI tool owners will either have to rely on an exemption to copyright infringement or need a licence to use the materials for the purposes of training the AI tool. There is a specific copyright exemption in the UK which permits the making of copies for the purposes of data mining in non-commercial research. The UK government has previously proposed expanding this exemption to data mining for all purposes but these plans were shelved after a backlash from the creative industry lobby groups. Therefore, as it stands, if the AI tool is going to be used on a commercial basis the data mining exemption is not viable.

There are other 'fair dealing' exemptions under UK copyright law, certain activities that can be carried out without permission of the copyright owner as long as the use is 'fair'. This includes; quotation, criticism and review, news reporting, research and private study for non-commercial purposes. It seems unlikely that a commercial focussed AI tool will be able to take the benefit of these exemptions although some specific tools may be able to.

It is important to note that UK and US copyright laws diverge on the point of fair dealing, with the US concept of 'fair use' being a more general defence than the more limited UK exemptions.

This leaves those responsible for the AI tool with the task of ensuring they are adequately licensed to use the source material to train the tool. There are a few options in this regard:

- Materials that are 'open source' can be used. For example, Wikipedia materials are subject to open source licensing and various government data is also available on an open source basis. Different open source licenses take different approaches but assuming they are of the more "permissive" nature then the AI Tool will be free to use this information provided that the open source conditions are met. This usually means, there is attribution of the source materials, and in many cases the source materials must be onward licensed on the same basis (but not necessarily products created through the use of the materials).

- Materials subject to specific licences. The AI tool developers may have decided that there is a certain data set will give their tool a specific advantage and have taken a licence from the owner of that data to use it for the purposes of the AI tool. In this case it becomes an issue whether the licence is wide enough to allow for the intended use, especially if that evolves over time (AI has a habit of moving into areas the developers had not intended). There is also a risk that the licence is not wide enough to cover the end customer use of the tool output if not properly reviewed.

- Implied licences. It is possible that materials can be used on an implied licence basis, i.e. because the IP owner published the material on the Internet, he or she intended that others could use the material without restriction. This is the least attractive area for AI tool developers as there is much less certainty as to whether the materials are subject to a licence or not. Many websites contain terms of use that state the extent of the permitted use and an implied licence could not realistically be inferred if the use was outside of these restrictions.

Major complaints are already surfacing with Getty Images leading the charge in its cases filed against Stability AI in both the UK and USA for infringement on the copyright of millions of images it owns. Getty argues that it offers AI data mining licences and Stability AI has circumvented that by copying the source materials without permission. We have yet to see the detail of Stability AI's defence.

It seems that in the absence of a widening of the data mining exemption in the UK, AI tools are likely to be infringing the IP of rights holders unless they have been carefully trained to use only open source or licensed materials. However, the way generative AI creates text or images means that the outcome can often look quite different to the text or images the AI was trained on. In such a scenario it could be practically difficult for rights holders to enforce their rights because it is difficult to prove that a copyrighted image or text of that individual has actually been used in the training phase of the AI tool.

In addition to the IP issues, it is possible that the training phase of the AI tool constitutes a breach of data protection laws. We discuss this in Part 3 of our AI briefing.

Output IP: The current UK legal framework for AI creations

This section looks at what intellectual property rights are contained in the content created by AI tools.

Patents

The UK Government has clarified that AI inventions can be protected under the Patents Act 1977 so long as they fulfil the normal criteria of being novel, having an inventive step, and being capable of industrial application just like any other invention.

There are several issues that could affect the current patent system once AI inventions become more commonplace: an overcrowding of the patent system, AI becoming too knowledgeable compared to the reasonably skilled person in the art so as to challenge the inventive step test, and the question as to who should be assigned the IP rights of an invention created largely or wholly by AI. The UK Supreme Court is currently considering if a patent can be granted without a named human inventor.

Overcrowding

The speed at which AI is able to produce new works could put pressure on the Intellectual Property Office and other patent registries around the world. The greatly increased output could be a step change for

innovation and development but if there are so many going through these systems then, processing them will become delayed.

The Inventive Step

A successful patent application requires an inventive step to be taken, i.e. something not obvious to someone reasonably skilled in the art. The European Patent Office has stated that simply using AI to solve a problem is considered obvious and therefore, in itself is not enough to meet the inventive step requirements.

This means that patent applications involving AI must show that an inventive step has been taken within the AI algorithm or the AI training set. Several applications have failed because the applicants have simply referred to AI being used to solve a technical or medical issue. In order for the application to be successful there must be disclosure about the AI training, so that a skilled person could in theory carry out the invention.

The Patent Owner

To date the inventor of a patentable technology has always been a human "legal person". Patent law assumes this to be the case and considers the inventor to be the first owner of the creation. Machines are not a legal person and therefore the assumptions on which patent law is established are called into question. Patent applications can be refused if the owner of the patent is not identifiable or if someone who isn't an inventor is listed as an inventor.

This leaves a patent applicant with the question as to who the listed inventor should be when AI is the inventor. UK patent law doesn't allow the machine to be the owner but arguably also can't give the patent to the inventor of the machine because they aren't the actual inventor. One of the major incentives of investing in research is the potential reward of a patent but if a patent is at risk of not being granted (due to a lack of a legally recognised inventor) then investors may decide the risk isn't worth taking.

At the time of writing we are awaiting publication of a UK Supreme Court decision in an appeal by Dr Stephen Thaler. Dr Thaler has attempted to apply for a patent on behalf of the machine called DABUS, the application was refused by the UK Patent Office and this decision was upheld through High Court and Court of Appeal rulings. The Supreme Court will now decide if a machine can be listed as an inventor in a patent application and whether an application can be made by a human on behalf of that machine. The case has been litigated in the European Patent Office, US and Australia, with the Supreme Court being the highest level court to decide. Its decision will therefore be a significant step in deciding how AI inventions are treated and may lead to legislative amendments.

It seems unlikely that the courts will want to call for machines to have legal personality, as this seems to raise more questions than answers. A more sensible approach seems to be that the owner of the patent would be the legal person that commissioned the AI to create the invention (as is permitted under UK copyright law) thereby acting as an inventor by proxy.

Either way, it would be beneficial to have increased clarity in patent law in the form of primary legislation so that the language reflects an understanding of modern AI capabilities rather than trying to include AI inventions under the current wording in the Patents Act 1977.

Copyright

Under copyright law in the UK there is some protection offered for 'computer-generated works'. This is somewhat unusual in comparison with other jurisdictions, including the US. Whilst this isn't the exact same protection offered for the work created by a human, with a shorter protection period of 50 years from the date the work is made (rather than 70+ years for other types of work) this is still a substantial period of protection.

This protection makes reference to a human author where the Copyright, Designs and Patents Act 1988 s9(3) states that 'In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken'. Although the copyright legislation couldn't have envisaged the full potential of AI it does go some way to dealing with the ownership of machine generated copyright.

The application of section 9(3) has not been tested very much in the UK courts. An AI tool used by an organisation to create output on its behalf would not seem to have any issues with applying the section 9(3) to argue that it is the first owner of the copyright created by the tool. The application of the section becomes less clear when the AI tool is licensed to another party for use. In this case, is it the licensed party that the arrangements for the creation of the work are undertaken, or is it still the AI tool provider? These issues can be regulated in a commercial context by the use of an AI tool licence to determine which party will end up owning the output (through an assignment arrangement if necessary). In the consumer context, AI tool providers will need to decide if they are happy for customers to own the output of the tool made available to them via terms and conditions.

This UK copyright protection probably only accounts for the current types of AI tool available, i.e. generative AI or large language models where there is a human behind the machine, rather than the machine being sentient in its own right.

The development of sentient AI in the future could cause issues for the copyright authorship idea. There could be a point reached where the AI tool would be able to create things of its own accord without further training or request from humans. This AI could become so far removed from humans that the reference to the 'the person by whom the arrangements necessary for the creation of the work are undertaken' would no longer apply. This is some way off yet but something copyright law will need to adapt to, as generative AI becomes more powerful.

What IP issues to consider when using AI

AI is proving an increasing challenge to existing intellectual property regimes. Governments and regulators may need to adjust or replace existing rules rather than rely on interpretation of existing legislation. In the meantime there are some things businesses can do when developing or commissioning AI.

If developing an AI tool it is important to understand the risks around infringing the intellectual property rights of rights holders when using training data. In the absence of an increased exemption for data mining, this means ensuring that training data is either open source or is adequately licensed.

If commissioning an AI tool from a third party then due diligence on the source of training data is advisable, as well as warranties and indemnities aimed at protecting you from claims that the tool infringes the intellectual property of rights holders.

If the goal of the AI tool is to create or assist with a patentable invention then consideration should be given to the inventive step that is required to be shown, this may require the step to be made within the AI algorithm or the training data rather than simply using AI to solve a problem.

It is also important to remember that in UK copyright law the default position of ownership of computer generated works will be the legal person 'by whom the arrangements necessary for the creation of the work are undertaken', potentially exposing companies if the contract is not worded with sufficient clarity. Contracts could therefore seek to designate who is the undertaker of the necessary arrangements for the purposes of Copyright, Designs and Patents Act 1988 s9 (3) or draft IP assignments appropriately with this in mind.

For more information on the issues raised in this note or for any of your IT or data legal issues please get in touch with us:

Codified Legal 24 April 2023
7 Stratford Place
London
W1C 1AY

The information contained in this briefing note is intended to be for information purposes only and is not legal advice. You must take professional legal advice before acting on any issues raised in this briefing.