Too good to be free – legal issues arising from AI image generators

Ian Liu

Some experts have described the rise of AI as the next industrial revolution and AI image creation services such as Midjourney and Stable Diffusion have become the latest subject of debate. With ChatGPT catching the attention of the world, OpenAI, the team who created ChatGPT has also developed a powerful AI text-to-image generator called DALL·E 2. Not unexpectedly, the use of such image creation services give rise to a myriad of novel legal issues. In this article, we will discuss how text guided image generators, such as Stable Diffusion, work and the intellectual property concerns that arise when training and using the image generator.

How do image generators work?

The new image generation AI technique uses neural networks (a collection of computer-simulated neurons, modelled on the neurons of the human brain, that are designed to recognise patterns) to generate, refine and improve images based on text input by a user. The model is "trained" from a dataset of images and the systems will continue to refine a generated image, guided by the text input, for a stipulated number of iterations. As a result, a high-quality and contextually relevant image is created after a number of refinement steps of the image.

Text guided image generators, such as DALL·E 2, are based on a combination of two machine learning techniques: reconstructive generative modelling and latent space manipulation by way of natural language supervision. The generators use a technique called "diffusion model" to train an AI by using a noise vector (random visual clutter) to obscure an image, e.g. an image of a cat. The trained AI learns to recover the data, e.g. the image depicting a cat that is obscured by the noise injected into the image. The AI can then be used to subtract noise from an image to recover the data and adjust the features of the image to produce a clear result. It is through iterations of this successive diffusion process that the model can produce high quality images from a wide range of subjects and styles.

The second machine learning technique is a method for linking image-to-text by training the neural networks to assess the similarity between an image and input text. The similarity score is used as a signal to steer the latent diffusion image generation model to refine images and hence increase the relevancy of the content generated. This helps the generator to re-create relationships between objects in an image based on the words used to describe it.

Using the combined techniques, text guided image generators have produced impressive results. The images produced by latent diffusion image models, such as Midjourney, Stable Diffusion and DALL·E 2, are surprisingly lifelike, with high levels of detail and realism. The "text-to-image generation" capability is extremely powerful and useful as it means that users can use the model to generate images specific to the textual instructions given to the image generator. For example, see the images that have been generated using Stable Diffusion v1.5 with the text inputs: "A picture of a flying cat with wings, facing rainbows, background blue sky, cartoon style, realistic, high detail, 4k" and "A picture of a dog sitting on a pink gym ball, realistic photograph".

Legal concerns in training an image generator

Copyright infringement

High quality image generators can only be trained by an enormous amount of images. Text guided image generators require a vast training set of image-text pairs, and such a huge quantity of good quality text labelled images are not easily obtainable. Naturally, the internet has been used as a resource for such data. Many images available on the internet are protected by copyright laws and subject to use conditions under specific licensing terms. Use without permission, or in breach of licensing terms, may constitute infringement.

Getty Images, the famous stock photo supplier, has recently commenced legal proceedings in the UK and US against Stability AI, a company that provides AI tools including an image generator based on Stable Diffusion. Getty Images alleges that Stability AI has copied millions of photographs from its collection, choosing "to ignore the viable licensing options and long-standing legal protections in pursuit of their stand-alone commercial interests." Getty Images believes that its image archive is especially suitable for the training of image generators, due to the high quality of the pictures, as well as the accompanying content-specific captions and metadata. The case is one of the lawsuits claiming that AI generators are being trained on copyright works without consent, and without giving credit or compensation to the original artists or creators.

There have been reports that image generator AIs may "memorize" some images used for training[1]. Although neural networks generally do not store a full copy of the training image, it is possible for a neural network to memorize some images in substantial detail. In such circumstances, the neural network has inadvertently made a copy of the training image, and there may be risk of further infringement if the neural network is used to generate images.

In the US, the issue of using images for AI training may depend on the doctrine of fair use, which is yet to be decided by the courts in the context of AI. In Hong Kong, the question will also be interesting as there is no general defence of "fair use"; the concept of "fair dealing" under Hong Kong law applies to a strictly defined set of permitted acts which are subject to specific conditions and limitations.

Possible trade mark infringement

Some images used by the image generators may contain trade marks which may give rise to trade mark infringement issues depending on the circumstances. In its US lawsuit, Getty Images is also alleging infringement of its famous trade marks, claiming that the reproduction of Getty Images' trade mark in some of Stability AI's images creates "confusion as to the source of the images and falsely implying an association with Getty Images." In addition, the incorporation of Getty Images' trade marks in images dilutes those marks in addition to violation of federal and state trade mark laws.

Who owns the generated images?

Bearing in mind the potential infringement problems where generated images incorporate existing works, it is still unsettled whether the resulting images generated by machine learning algorithms are considered to be original works themselves and, therefore, protectable by copyright. A number of legal issues in relation to copyright in AI generated content still need to be considered, as well as the exploration of complex factual questions, including the source of training data, exactly how the training data set is used to train, use of fine tuning and base model training techniques, and how the image generators are used to generate contents. Even if the images are regarded as original works, should the owner of the AI that trains the model, or the person who inputs the text prompts, or nobody, own the images? The US Copyright Office has recently rejected a request to grant copyright to an AI-created painting on the grounds that it did not include an element of "human authorship". During the consultation exercise leading up to the recent amendments to the Copyright Ordinance, the Hong Kong Government specifically mentioned AI and copyright as an emerging issue that needs to be addressed in future legislation.

Other legal concerns

Scraping material from all publicly available data may give rise to other legal issues such as privacy and consent since online data may include private photos, images of celebrities, or sensitive information, such as medical images.

Given the amount of information scraped off the internet, it is not surprising that the AI generators may be trained on images that may contain pornographic, harmful or illegal content. It is feared that AI image generators may be used to create obscene content, contribute to discrimination or hate by reproducing damaging stereotypes, or to generate misinformation. Although steps may be taken to remove undesirable and not safe for work content from the training dataset, and to filter problematic images, this is not an easy task and it is unlikely that all harmful content will be caught.

Moreover, a number of trained AI models are easily available to the public for download and are granted under permissive open source licences, such as CreativeML Open RAIL-M. Although the licence terms restrict uses that violate any laws or regulations and harmful purposes, the permissiveness of the licensing terms provide users with a sense of that they are free to use the AI models and works created by the AI. However, there may be hidden mine fields.

The way forward

As with any disruptive technology, the advent of AI generators raises many concerns. There are likely to be growing pains as we consider how to manage the novel legal and ethical implications of generative AIs. All stakeholders need to watch closely the development of case law and legislation which should address the ongoing legal uncertainties surrounding infringement, ownership and the status of generated images.

[1] https://80.lv/articles/stable-diffusion-may-memorize-some-images-used-for-training/?comment=10859

Want to know more?

Ian Liu
Partner
ian.liu@deacons.com
+852 2826 5360

www.deacons.com