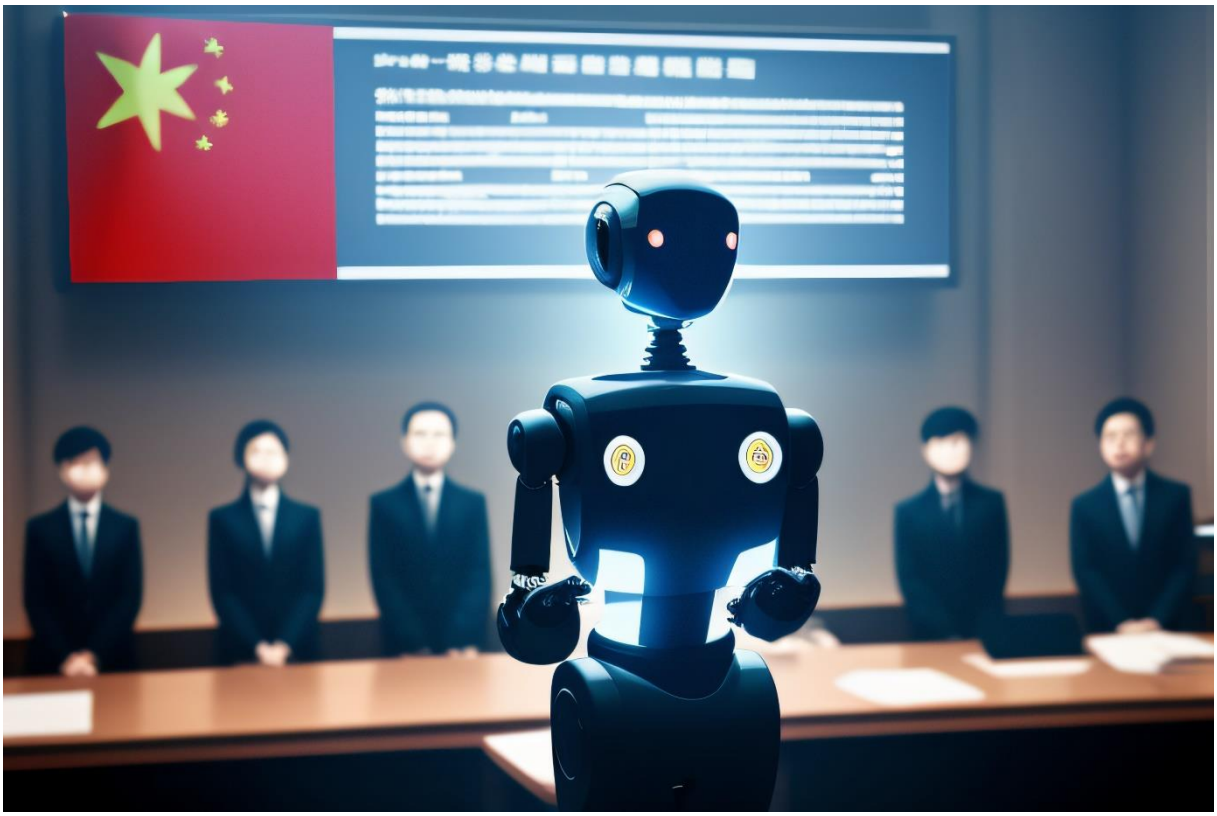

Threshold of originality for AI output in copyright law

Posted on: https://www.linkedin.com/pulse/threshold-orignality-ai-output-copyright-law-jakub-wyczik-hh9sf/

The last few days have been overwhelmed by the news that a Chinese court has declared an AI output protected by copyright. According to the machine translation of the decision, the plaintiff used one of the open source implementations for Stable Diffusion models. As first evidence, the plaintiff provided access to an account on a certain platform and a generated image posted from that account. The defendant used this image as an illustration for her original poem, but stated that she had simply found it somewhere on the Internet. The defendant had no problem apologizing if the image belonged to the plaintiff.

To prove his creation, the plaintiff presented a video in which he carefully reproduced the entire process of generating the image. He chose a specific implementation of Stable Diffusion (e.g., one with some special predefined settings, including different VAE), models (base and LoRa), and a positive prompt. He added a negative prompt, and from the translation it appears that only some of the negative prompts were written by him - the rest were copied from prompts shared by other users on the forum, which is a common practice. He selected a number of 33 iterations to generate the image (steps), changed the aspect ratio to the standard 2:3 proportions, changed the CFG scale to 9 (the parameter describing how much the model follows the prompt), and selected the seed 2692150200 (usually the seed is chosen randomly,

alternatively users choose a seed that other users have found "good" and posted on the forum; choosing such a complicated seed suggests that it is more likely that the plaintiff actually left the value as a random -1). The result was the following image.

Plaintiff then stated that he changed the LoRa parameter to 0.75 and got this result.

He then subtracted one from the seed value, showed the intermediate result, and then added a few words to the positive prompt. This is how the controversial image was created.

According to the plaintiff, he is the author of the work because, first of all:

1) by choosing models, he was able to express his sense of aesthetic taste;

2) by choosing positive prompts based on the general pattern: art type + subject + environment + composition + style - he selected specific elements according to his preference and then shaped the negative prompt accordingly, adding especially the relevant phrases that are the opposite of imagegraphy, in order to avoid a result in the style of painting, cartoon, animation, anime, etc. [but VAE was "animevae"];

3) by choosing other parameters, in particular the aspect ratio "obtained by the plaintiff through many experiments" [the 2:3 ratio is the most common ratio used to create realistic images of people];

and generally arguing the above, that such an original creation was created especially as a result of the plaintiff's intellectual efforts and many trials and errors.

Also noteworthy is the evidence of the defendant's use of the same image, which, according to the translation, involved checking the invisible watermark stored in the image's data. According to the plaintiff, the image on the platform also had a normal visible watermark that had been removed - but it is not clear by whom, as the defendant did not remember whether she had done so. The hidden watermark embedded in the file indicated the origin of the platform on which the plaintiff posted the image and contained a specific identifying number that pointed to the plaintiff.

Aside from the issue of proving use of the same image, it appears that, contrary to popular belief, Chinese jurisprudence is still rooted in the "sweat of the brow" doctrine. The court found that such decisions by the plaintiff were sufficient to create a work. This suggests that the case does not fit into an EU, US or even international framework.

The basis of international copyright law is the separation of unprotected ideas from protected forms of expression (see TRIPS, Article 9; WCT, Article 2). What constitutes the subject matter of protection is only the result of the choice of specific forms of expression which,

when put together, form a work. Thus, it is the combination of words into sentences, sentences into paragraphs, paragraphs into chapters, and so on. Specific shapes and colors into patterns, patterns into objects, objects into compilations of objects, and so on (see the analysis of the Berne Convention Guide posted on IPKat). Looking at the approach of the Chinese court, the case did not recognize the problem of separating idea from expression at all, so it can be said that the case unfortunately does not add much more to the international discussion than the Chinese Dreamwriter case. Nevertheless, the decision is interesting and worth reading. It inspired me to write this article and to point out some very practical problems.

The first is relevant to this article. I did not copy the above images from the court decision. In fact, I found the model in the image and was reminded that I had also arranged a image shoot with her some time ago. To prove it, I even have more images from that session than the plaintiff.

Of course, all the images were actually re-created after reading the judgment based on the information contained therein. The court's description of the defendant's actions basically provided everything that was needed. The three images above, on the other hand, are alternative versions that differ only in the order of several prompts (1), the choice of a different sampler (2), or were created using the disputed image as a base for the reference mode in ControlNet (3).

The question then becomes, to what extent could the above images be considered derivative works of the first image? Most copyright cases so far have been based on the argument of high probability of access to the original work and substantial similarity between the original work and the alleged derivative work. But how do you apply these arguments in the world of AI, when in fact it is very easy to have some repetitiveness? I won't even mention that substantial similarity arises specifically from the use of a particular LoRa.

Current models are more stable, so if you have access to the right tools and information, you can reproduce almost the same images as others. I wrote "almost", why? Because a trained eye will notice that I used a little trick. The images at the beginning of the article are not identical to those presented in the decision. Several elements in the generation process are somewhat "random" and depend, for example, on the equipment you have, which was not provided in the decision because even the court was probably not aware of it (and even if it was, the chance of reproducing exactly the same circumstances is small). However, the

images above are quite similar. Without a doubt, one could say that they meet the "substantial similarity" standard.

For reference, below left is an image from the decision. On the right is a "reinterpreted" version.

This raises another interesting question.

This time the settings are the same as in the example, but with a different (random) seed. The seed is the value that creates the original noise that the AI denoises and creates the result from. As you can see, changing the seed makes the result significantly different, even with the same prompts and settings.

The seed is most often chosen at random, so how can one prove that the creator of the image chose this particular seed with this, and not some other, expression in mind?

How do you separate the idea (which is the same in all of the images) from the very different forms of expression in these images?

Is the selection of one of the results presented to us (as among the patterns drawn by the wind on the sand) sufficient to consider the selected creation a work?

Or is the role of the court in cases involving AI creations to assess the likelihood of a similar outcome?

Should we even bother to examine the creative process to decide whether something deserves legal protection?

Perhaps the threshold of originality never made sense, and the creators of the "sweat of the brow" doctrine were right?

Obviously, the selection of appropriate parameters, prompts, models, etc. gives the output a certain quality, and therefore the output deserves some protection. The question, however, is whether we can still rely on the threshold of originality and the idea-expression dichotomy in such circumstances. In my opinion, this theory has long been abandoned, and even in EU or US jurisprudence, the boundaries of this dichotomy are not clearly visible. It seems that this test is already used selectively in practice to motivate the sense of justice of the person deciding a case. So should we rely on the cherry-picking of a particular person, or should we rely on more verifiable standards? Should we rely on something that exists only on paper, especially when everyone treats virtually everything that constitutes text, images, video, etc. as if it were copyrighted?

This article was written as part of the results of a research project funded by the National Science Centre, Poland, grant no. 2021/41/N/HS5/02726.