8/31/23

To Whom It May Concern:

I am writing primarily as a citizen who is concerned with the way copyright law can be used to manipulate and/or foster creative work. I am a writer of short fiction, although not professionally, and a lawyer, although intellectual property is not my field so my knowledge of the area of law is limited. I have an educated layperson's understanding of AI and have not worked in the field.

My comments relate particularly to the questions raised by the request for guidance on copyrightability. Most of my comments can be regarded as pertaining specifically to the "AI" systems known as "large language models" or LLM, as my knowledge of AI models which create and manipulate visual or audio material is far more limited. I am also very unfamiliar with the specifics of how a human would alter or manipulate the visual or audio output of an AI and am therefore ill-equipped to judge how much work by a human being would be required to confer authorship. On the other hand, I feel that I understand the operation of LLMs and the type of manipulation that would be provided by an end user sufficiently to have an opinion in this area.

In reference to **question 18**, I do not believe that a human operating an AI system via either programming or by entering prompts can be deemed to have a copyright in any resulting textual output from the AI. The person who programs the AI model or system would obviously have a copyright in the specific code written; however, the output eventually produced by the AI would be so attenuated from the original actions of the human that the human could not be fairly said to have any agency over the end result. Selecting the material for training cannot be fairly considered a significant contribution to the eventual output, in large part because LLMs require such a large amount of input to be useful: billions and billions of words at minimum.

Even in cases where a secondary trainer is selecting training datasets to adapt a general AI model to a more specific purpose, developing or selecting training data merely amounts to the conglomeration of massive amounts of material. Usually this material was not produced or owned in any way by the person training the AI model, but a compilation of the copyrighted work of thousands or even millions of other human beings. A person who trained an AI model exclusively on their own intellectual property might have a claim that the output of the model should be subject to copyright by the author of the IP used to train it, but frankly it seems unlikely that an LLM would ever be trained to such a limited extent because at least under the current state of the technology this small a dataset would not result in a useable AI model. When the eventual output of the model is based on the vast amounts of other people's intellectual property, the output is if anything derivative of the training datasets, not the person who selected or compiled the datasets.

I do not believe that a human entering prompts into an AI system can hold a copyright in the output, either. They can certainly hold a copyright in things that the model has produced that they have themselves manipulated and altered after the fact. But merely entering a prompt cannot be sufficient exercise of creative control to change the machine-produced text that results into a human-authored work. AI systems have been known to produce vast amounts of cogent text based on a prompt of a few words or a sentence- why would asking a machine to write you a story about a dragon rescuing a princess give you a copyright in the resulting story, when asking another person to write such a story would not? Unlike a human being generating a story

based on a prompt, the machine exercises no creative judgment; there is no thought because machines can't think. The term "artificial intelligence" is a misnomer that implies these systems have agency or the ability to think when in fact all they do is make probabilistic conclusions about which word would be most likely to come next based on their previous examination of millions of examples of humans assembling sentences. That is why the model cannot hold a copyright in its output. But the human being entering the prompt ALSO has not exercised any creative judgment or thinking in the assembly of the result the model produces. On this basis, it is my opinion that the raw output of an LLM-based system or model is not copyrightable at all.

In reference to **question 19**, any language used in the law is by necessity going to be subject to interpretation based on unique situations that arise. But the standard for allowing the output of an AI model to be copyrighted ought to be that a human being has *substantially* altered the output, meaning that they have exercised some creative judgment or control over that output, beyond mere proofreading or copyediting.

In reference to **questions 20 and 21**, there would be no public interest served by allowing the copyrighting of the output of AI models/systems. Inputting a prompt is not a sufficiently creative act that treating the AI output as copyrightable would encourage the arts and sciences. Indeed, because AI systems are not capable of independent thought, the output that they produce is often low-quality or even gibberish. It does not further the arts and sciences, and a copyright scheme that encourages ever-greater production of valueless drivel of this type does not further the arts and sciences either. In addition, it is easy to foresee ways in which AIs could foster a whole new genre of copyright trollery. For example, AIs could be used to easily and quickly create stories or ideas that cover a broad range, which could then be the basis for the owner of that output to sue infringers who come upon similar ideas. If the AI generates similar output based on a similar prompt, would the person who used the prompt first be able to sue the second for causing the AI to reproduce his copyrighted material?

Giving the creator or publisher of the AI model/system a copyright interest in the output could be even more disastrous, leading to lawsuits against anyone who uses the system and then attempts to use the output to their own benefit. Companies are eagerly packaging AI models/systems into all kinds of products; some, such as Microsoft and Google, are even making this technology available free of charge to users, with the apparent intent of encouraging adoption of and reliance on the technology. Granting the owner of an AI system an IP interest in the output creates a risk that ever-larger swathes of human creative endeavor would come under the control of the model/system owner. If a word processor with an AI helps you write a letter or a story, does the publisher of the word processor have a legal ownership right in the work produced using that processor? Is the company that produced the AI built into your search engine entitled to authorship credit on articles where the AI model compiled research in the field based on a request by the article's writer? If a movie studio uses an AI model to generate a script for the next blockbuster, is the AI model's creator entitled to a cut of the box office receipts?

It should be clear from the present crowded state of the AI field that no copyright protection is necessary to encourage the development of the technology. The primary way in which the creators/owners of AI models and systems generate profit is by licensing the use of their systems and models to businesses and individuals who wish to use them. As with any software product, the future development will likely be geared towards enhancing the utility of AI to the target users, such that they will continue to use and pay for AI systems. At present, the ability of the creators/owners of AI models and systems to profit off their work is based on their copyright in the computer code that operates these systems, which is what allows them to

maintain the exclusive ability to license the use of their creations. This is more than sufficient to promote development of AI research and development in future.

I appreciate that the Copyright Office is exploring this area with an awareness of the broad spectrum of implications generative AI has for copyright law. I hope that the Office will consider the points above in tuning their position and guidance on the copyrightability of AI generated works.

Yours,

Lindsay Keipper