THE UNIVERSITY OF CHICAGO

1111 East 60th Street | Chicago, Illinois 60637
phone 773-702-5188 | fax 773-702-0730
e-mail jmasur@uchicago.edu
www.law.uchicago.edu/faculty/masur

Jonathan Masur
John P. Wilson Professor of Law

5730 South Ellis Street | Chicago, Illinois 60637
phone 773-702-2322
e-mail feamster@uchicago.edu
people.cs.uchicago.edu/~feamster/

Nick Feamster
Neubauer Professor of Computer Science

December 16, 2019

Director Andrei Iancu
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Director Iancu:

We write to offer comments in response to the Patent and Trademark Office's Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation, at 84 Fed. Reg. 58,141. We appreciate this opportunity. We are, respectively a Professor of Computer Science and a Professor of Law at the University of Chicago. One of us is an expert in the applications of artificial intelligence; the other of us is an expert in intellectual property. In particular, we write to comment on the following questions:

1. Should a work produced by an AI algorithm or process, without the involvement of a natural person contributing expression to the resulting work, qualify as a work of authorship protectable under U.S. copyright law? Why or why not?

2. Assuming involvement by a natural person is or should be required, what kind of involvement would or should be sufficient so that the work qualifies for copyright protection? For example, should it be sufficient if a person (i) designed the AI algorithm or process that created the work; (ii) contributed to the design of the algorithm or process; (iii) chose data used by the algorithm for training or otherwise; (iv) caused the AI algorithm or process to be used to yield the work; or (v) engaged in some specific combination of the foregoing? Are there other contributions a person could make in a potentially copyrightable AI-generated work in order to be considered an "author"?

An "AI algorithm" should not be thought of as an isolated process that makes predictions on data, but rather as part of a systematic pipeline that includes gathering, labeling, and manipulating data as input to that model, both when the model is developed (i.e., trained) and deployed in practice.

An AI algorithm is itself necessarily the product of substantial human intervention. One or more humans design the algorithm and establish how that algorithm will operate. As part of this design process, humans must decide how to train the algorithm. They must decide which data or inputs to use in training the algorithm, as well as how that data is represented to the algorithm. They must further decide how to format, arrange, and clean that data in order to make it useable by the algorithm. Ultimately, the "AI algorithm" that takes observations as input and produces predictions or classifications is a small part of the overall process. In many cases, the steps that involve processing the data that are inputs to the algorithm will ultimately determine whether the model can work in practice. An algorithm that relies on inaccurate data (or data that is not normalized or represented in the appropriate way) will ultimately produce bad results, regardless of the algorithm itself.

Accordingly, when an algorithm produces a work (i.e., a trained model), that algorithm will have been the result of substantial authorial choices by the humans involved in designing and training the algorithm. Consequently, the work produced by the algorithm will be the result of all of the human authorial choices that went into the design and training of the algorithm in the first instance. These authorial choices, which were used to produce the algorithm that then produced the work, render the work a work of authorship that should is protectable under U.S. copyright law.

An appropriate (though not perfect) analogy is to a player piano. When a player piano operates, it produces a copyrightable work (if the sound it makes is fixed in a tangible medium). Player pianos can operate autonomously; one only need press a button, and the player piano will run without further human intervention. However, the player piano itself is the result of substantial human creativity and authorial design choices. Someone has designed the player piano and decided on the basic framework for how it will operate.

The design of a player piano is analogous to the design of the AI algorithm. Someone has created the roll that will tell the player piano which sounds or notes to produce. This is analogous to choosing which data to use in training the algorithm. The imperfection in the analogy is that the human's actions in designing and "training" a player piano are perfectly determinative, and the human will be able to predict in advance the work that the player piano will create. This will not necessarily be true for an algorithm. Nonetheless, the design and training of the algorithm will involve authorial choices by humans, just as would the design and training of a player piano. Accordingly, the work produced by the AI algorithm as a result of training on the input data of the designer's choosing should be copyrightable in the same fashion as the work produced by the player piano.

All of this is to say that an AI algorithm is a machine built through human effort. Any works that result from that machine should be considered works of authorship attributable to the humans who constructed that machine, just as they would be for other types of machines.

This then raises the issue posed in question #2 as to which individuals should be credited with authorship of the work. As we have described, the design of the algorithm

involves authorial choices that will affect the work that is eventually produced. Accordingly, anyone who was substantially involved in the design of the algorithm should be considered an author of the resulting work. These are the people in categories (i) and (ii) described in the question. In addition, the decisions regarding which data to use in training the algorithm—and not only which data to use, but how to clean and format that data, and which pieces of information to include or exclude—will have a significant effect on the eventual output of the algorithm. These choices as to data similarly involve significant and critical authorial discretion. Accordingly, anyone who has made substantial discretionary choices regarding the type and format of data used to train the algorithm should be considered an author of the resulting work. These are the people described in category (iii) above. However, the individual who "caused the AI algorithm or process to be used to yield the work" (category (iv))—that is, the individual who "turned the crank" and ran the algorithm once it had been designed and trained—should not be considered an author. This individual has done nothing more than press the start button on a machine built by others. He or she has made no authorial choices that would affect the eventual work produced.

3. To the extent an AI algorithm or process learns its function(s) by ingesting large volumes of copyrighted material, does the existing statutory language (e.g., the fair use doctrine) and related case law adequately address the legality of making such use? Should authors be recognized for this type of use of their works? If so, how?

In the course of training an algorithm on data, the algorithm will frequently make a copy of the data in computer memory, so that it might read the data (sometimes multiple times). This copying process is, in some sense, a "mental" process by the algorithm. The copy of the data is never disseminated or publicized. The process is akin to a human reading a set of data and memorizing the various elements of that set, in order to be able to think about organizing those elements or detecting patterns within them. Moreover, in most cases the copy of the data will reside in memory only so long as the algorithm is being trained on the data. Once the algorithm has been trained, there is no residual copy that must remain, either in volatile memory or in persistent storage. That is, in many cases, once the algorithm is trained, it suffices to delete or destroy the copy of the data that was used to train the algorithm. The algorithm has incorporated whatever information was contained within the data—again, like a human learning from a dataset—and the original data can be discarded.

We believe that the act of copying that occurs when an algorithm is trained on data should be considered fair use, so long as any copy of the data that is made during training is destroyed when the data is no longer needed to train the algorithm. This type of copying does not implicate any of the principles underlying copyright law. The copy of the data will never be disseminated in any fashion, and whether or not the data will actually be copied is a mere artifact of how the algorithm is written. This copying will not in any way interfere with the market for the data, because the data will not be shared with any other potential user or customer. Any copy that is made will be entirely internal to the algorithm making the copy. To hold otherwise would be to force algorithm designers either to negotiate licenses in situations in which no license should be needed, or to design their algorithms in such a way that the algorithm does not create a copy when it operates. That type of algorithm design may be more difficult or inefficient than other

methods, and there is no principled reason to require algorithm designers to adopt that type of approach.

10. How, if at all, does AI impact trade secret law? Is the Defend Trade Secrets Act (DTSA), 18 U.S.C. 1836 et seq., adequate to address the use of AI in the marketplace?

11. Do any laws, policies, or practices need to change in order to ensure an appropriate balance between maintaining trade secrets on the one hand and obtaining patents, copyrights, or other forms of intellectual property protection related to AI on the other?

In some cases, it will be possible to protect algorithms as trade secrets. Once trained, an AI algorithm is simply a piece of code that can be stored in a secure location. Even if the algorithm is operating in public—that is, even if the algorithm is being used in a public setting to produce results that are observable by the public—in some cases the algorithm will nonetheless be impossible to reverse-engineer. For instance, imagine a model that is used to produce bail or parole recommendations in a particular criminal jurisdiction. In some cases, even if the results of the model are made public, as they often must be in the course of a criminal proceeding, the operation of the model, including the aspects or features of the observed data that factored most significantly in making those recommendations, will be impossible to decipher. Many algorithms are sufficiently complex that the manner in which they operate cannot be determined merely from observing the results.

In other cases, however, it will be possible to reverse-engineer an algorithm based upon the results it provides in response to various inputs. If the results of the model are made public, as in the bail/parole example from the preceding paragraph, it might be possible for an expert observer to reverse-engineer and then copy the model. In these instances, trade secret law will not be sufficient to protect the intellectual property contained in the algorithm. Instead, patents will be necessary, and thus the PTO should continue to allow patents on AI algorithms, particularly in instances where the possibility of reverse-engineering the algorithm is substantial.

However, the patents that have been granted on AI to date have often been very weak, based on vague claims that do not adequately describe the details of the algorithm, the process of training the algorithm, or the processing of the data to produce the inputs to that algorithm. In many cases, they do not enable the inventions. Even someone with far greater than "ordinary" skill in the art would not be able to recreate the claims based on the limited and frequently abstract information provided in the specifications. In particular, without information on which data were used to train the algorithm, it is often impossible to create the invention described in the claims. Some examples of particularly weak patents and applications are as described below. While these examples pertain specifically to applications of AI in intrusion detection for IoT, these examples are illustrative and the general observations apply more broadly.:

- "Intrusion Detection Model for an Internet-of-Things Operations Environment", US2019/0239483 A1 (August 1, 2019). This patent application describes "generating" and "executing" a "machine learning model", without details pertaining to the model that is used or the data (or

representations of that data) that should be used to train the unspecified model.

- "Method for Protecting IoT Devices from Intrusions by Performing Statistical Analysis", US2019/0182278 A1 (June 13, 2019). This patent describes in general terms the process of extracting and aggregating featres from network traffic to detect intrusions on Internet of Things (IoT) devices. The patent application discusses possible exemplary training windows for gathering statistics but does not precisely describe which time windows are likely to be effective, nor how such a model might be reduced to practice if the time windows prove to be large. The claims also exhaustively list general properties of network traffic (including traffic volumes) but do not specify how each of these features might be encoded, aggregated, windowed, or otherwise manipulated to train the algorithm. In fact, Claim 19 ultimately describes the process by which *any* intrusion detection system or algorithm might be designed, and the *process* of arriving at design decisions, without describing the embodiment of any specific system..

The PTO should require prospective patentees to describe how to enable the claims in greater detail, at a lower level of abstraction and a greater level of detail, such that someone reading the patent could produce a working algorithm in practice. In particular, the PTO should require that patentees describe which data were used to train the model and show how those data were represented and organized. Without this information, the patent is of very little informational value to anyone reading it.

Thank you once again for the opportunity to comment on these questions.

Sincerely,

Nick Feamster
Neubauer Professor of Computer Science

Jonathan Masur
John P. Wilson Professor of Law