
From: jbash aka John Bashinski <jbash@jbash.com>
Sent: Monday, November 18, 2019 1:10 PM
To: aipartnership
Subject: Comments on docket number PTO-C-2019-2038, intellectual property and AI
Attachments: signature.asc

Below please find the comments of John R. Bashinski in response to the USPTO's inquiry in the Federal Register of October 30, 2019, docket number PTO-C-2019-0038, under the title "Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation"

John R. Bashinski

Copyright

=====

Eligibility and Authorship

As to questions (1) and (2), works produced by computer programs, in general, are not and should not be protected by copyright. This arises from the fundamental principal that copyright protects only authors.

The author, owner, or operator of a computer program does not generally act as an author of the program's output, and therefore can receive no authorial protection with respect to that output. The program itself does act as an author, but has no legal personality, cannot have legal rights, and therefore cannot be granted copyright protection, even through the proxy of its owner. This state of affairs is appropriate and desirable, as will be argued below.

The only exception is, and should be, the case in which all of the following apply:

- (a) A natural person uses a computer program as a tool to produce a specific work,
- (b) That person designs the program itself, and/or provides its input parameters, in a way intended to produce specific characteristics in the resulting work, and
- (c) The final work in fact possesses the intended characteristics.

This exception could equally be described as the case in which a human "manually" exercises control over some part of the produced work, and therefore makes authorial decisions with respect to that work.

Question (2) could perhaps be better framed in terms of which specific elements of an algorithmically produced work are or should be protected by copyright, rather than assuming that there is some threshold at which the entire work becomes subject to some form of copyright protection. [1]

A subsequent work should be considered derivative of an algorithmically produced work, and therefore potentially infringing on any copyright in that work, only insofar as the new work reproduces significant elements or characteristics originally intended and specified by the algorithmic work's human author, as described above, assuming any such elements exist.

Reproduction of any element or elements introduced, without specific authorial direction, by a computer program, should not be considered to create a derivative or otherwise infringing work, because such elements are not part of any original work of authorship. [2]

A person who simply runs a computer program does not engage in any creative process, does not select creative elements or consider options, and therefore does not create a work of authorship.

More significantly, even a person who writes a computer program or selects its inputs, without planning or directing specific elements of the program's output, does not thereby engage in any authorial activity with respect to the output. Such a person is clearly the author of the computer program itself, but that is irrelevant to authorship of the output.

Not only is algorithmic output not copyrightable under existing law, but any legal change to permit it to be copyrightable would be a deeply unwise and dangerous public policy, tending to entrench oligarchic control of intellectual property. This is further addressed later in this letter under "Goals of Copyright".

In the United States, creation of copyrights without a requirement of true human authorship would also fall beyond the scope of the Constitution's copyright clause.

The same arguments provide a still clearer answer to question (5). An organization or institution, qua institution, is incapable of engaging in the process of authorship. This fact is recognized by the existing legal requirement that every work have a natural person as its author. Ownership or operation of a computer program does not cause a company to perform the mental processes of authorship. A company therefore cannot, and should not be permitted to, originate a copyright via a computer program or in any other way.

Should computer programs, in some distant and hypothetical future, be granted legal personality of their own, then a program, which does in fact engage in authorial action, might be permitted to originate and own a copyright. Even in that case, authorship and ownership should lie with the program, not with any corporation or similar fictive legal entity, and not even with the program's own author.

Sub-questions of question (2)

With respect to the specific sub-questions of question (2):

- (i) Mere design of an algorithm or process confers, and should confer, no authorial rights at all over the output, except for parts of the output deliberately "hardwired" into the specification of the algorithm.
- (ii) Contributions to the design of an algorithm may influence the division of any of the authorial rights mentioned in (i), but does not and should not influence the question of which authorial rights exist.
- (iii) Choice of training data confers, and should confer, authorial rights only insofar as those training data were chosen with the specific intent of producing particular, identifiable characteristics of the output.

There is no important distinction between the training data and the algorithm itself, except perhaps that it is more difficult to intentionally introduce intended elements into output of a machine learning algorithm through training data

than through code, and therefore more difficult to achieve even limited authorial rights over the output through the training data than through the code.

- (iv) Running a program, or causing it to be run, does not, and should not, confer any authorial rights of any kind.
- (v) How these elements are combined should have no direct bearing on the authorial rights conferred by any one of them.

Copyrighted Training Data, and Infringement by Algorithmic Works

With respect to question (3), the fair use doctrine would seem to enter the question only if an algorithm's output actually reproduced protected elements of its training data. This seems unlikely to be the norm.

The sub-question about recognition of the contributions of authors of training data is inapt. "Recognition" is not a significant part of the subject matter of copyright per se, even though recognition, like any other matter, may be a condition of a license granted under copyright law.

With respect to question (4), there seems to be no reason to expect that existing laws are in any way inadequate for identifying infringement or for assigning liability for infringement.

Like machine learning systems, human authors are "trained" by "ingesting" large volumes of information, much of which is protected by copyright. This does not imply that every work of a human author is derivative of that person's "input" for any copyright-related purpose.

There is no obvious, or even likely, reason to use different criteria for determining whether the output of a computer program infringes a copyright in some part of its input, than for determining whether the identical work would be infringing if it had been created by a human author. The means of production are simply irrelevant to the question of infringement. It would be unwise and dangerous to treat an algorithmic work as infringing when a "manual" work would not be so treated, or vice versa.

As for assigning liability, liability must lie with whatever legal entity causes infringement. A machine learning system lacks both volition and legal personality, and cannot infringe by itself. Machine learning systems may produce, and more importantly publish, infringing material, but they will do so only at the direction of their operators, who bear any liability for the results.

There are of course secondary questions. For example, we might expect "manual" infringement to be more often willful, whereas algorithmic infringement might be comparatively more likely to be negligent. Existing law has well-honed mechanisms for dealing with this sort of distinction and assigning liability accordingly.

There is no clear reason to believe that there would be anything wrong with those mechanisms in a world with more algorithmically produced works. Greater liability is applied for willful infringement for fundamentally punitive reasons, not because the resulting money awards or other sanctions have intrinsic social or economic value. If a greater proportion of infringement moves into the negligent category, that is not in itself a reason for concern. No party has a "right" to collect compensation in excess of actual proven damage to that party.

Goals of Copyright

The only valid goal of copyright is to create incentives for the production of valuable works. In the present context, the "valuable works" in question include both the algorithmically-produced output, and the programs and learned models which produce that output.

Output works

With respect to the output works, the most important consideration is to limit chilling effects on that might be caused by fears of infringing others' copyrights or other intellectual property rights.

Because the intrinsic marginal cost of producing each output work will generally be very low, costs associated with managing such liability risks, or related regulatory costs, could easily dominate all other costs for producers. This is very different from the case of manually created works, whose dominant cost is that of the authorial effort involved in creating them, and for which liability and regulatory costs are insignificant in comparison.

Therefore, any approach that increases the risk of infringing rights in another work, even accidentally, will surely have a dramatic negative effect on the types and number of algorithmic works that can practically be produced or used.

Any attempt to extend copyright to elements of algorithmic works not arising from human authorship would create precisely such a dangerous chilling effect. Because of the very low cost of producing algorithmic works without human guidance, the volume of such works in existence will probably be large. If all elements of all of those works were protected by copyright, an operator whose algorithm produced a new work would be at relatively high risk of accidental infringement, compared to the case in which only human authorship were protected.

Indeed, the cost of production is likely to be so low that it is possible that some entity or entities might be able to erect a "copyright thicket" by attempting to encumber a large fraction of all possible valuable works. It would obviously be impossible for such an entity to produce all possible valuable works, but it might be possible to produce works containing elements that were essential, or at least natural, to a large fraction of valuable works.

Thus, extending copyright to algorithmically created elements would actively damage the goals of copyright by discouraging the creation of new algorithmic works.

Authorial programs as works

With respect to the production of programs that create works ("authorial" programs), the case is somewhat different. The prospect of exclusive control over all of a program's output would surely be a positive incentive to create that program.

Nonetheless, a program capable of producing large volumes of valuable output works is a very significant asset even if its owner has no exclusive rights over its output. There is every reason to expect authorial programs to be produced in more than adequate numbers regardless what copyright protection is extended to their output.

Although no output protection exists at present, and although many if not most developers are aware of that fact, a large and increasing number of such programs are already being developed. The evidence of experience shows that the incentives for producing such programs are already sufficient. Claims to the contrary in comments should be treated with extreme skepticism.

It is true that no large businesses have yet been created around such programs, but the goal is the creation of the works, not the creation of large businesses. Indeed, if an entity is able to create a large business around such a program, that may be evidence that that entity is being permitted to capture too much of the value created by the program, rather than too little. Such excessive capture frequently reduces the total value created, rather than increasing it.

Trade Secrets

=====

Machine learning and other "AI" technologies would seem to create no obvious need to adjust any major part of trade secret policy.

Confidentiality can be applied to a computer program, or to a trained model, or to output produced by these, as easily as to any other information. Damages can be collected for dissemination or misuse of trade secrets in these forms as well as in any other. Protections are thus in place should trade secrets find their way into these data by illegitimate means.

The mens rea requirement of 18 USC 1832 should, if properly interpreted, protect operators of machine learning systems from draconian criminal penalties for any inadvertent incorporation of trade secret information into a trained model. Although those penalties are arguably excessive and unwise in general, they do not appear to disproportionately burden automated systems.

There is, however, a risk that certain parties might wish to suggest inappropriate changes to policy in the case where a machine learning system or other "AI" program infers or deduces a "trade secret" from public or properly obtained information. It is possible that such a system could make an inference that an unaided human would be unable to reach. Some might wish to extend trade secret law to treat such inferences as misappropriation, or to create some new category of prohibition for them. This idea is a pernicious one which should be resisted.

Improvements in technology have always made it possible to seek out former "trade secrets" in previously impractical ways. Modern analytical chemistry can make short work of longstanding "secret" formulae, if there is a real economic incentive to do so. Modern imaging technology makes it far easier than in the past to determine how a product is designed or manufactured. An entire reverse engineering industry is in robust operation. Even business relationships are easier to discover than they were in the past, in no small part due to the interwoven nature of the modern economy. The impact of "AI" is unlikely to be noticeable compared to that of these things.

In contrast to the arena of personal privacy, where there are obvious public welfare reasons to restrict the use of mechanical inferences and other formerly private information made discoverable by new technologies, there is no good reason to further the enclosure of ideas by creating new "trade secret" burdens on the use of properly obtained information. In no case should "AI" be used as a pretext to expand the already excessive reach of trade secret protection.

This is especially true if it is a goal to encourage the use of patents over that of trade secrets.

About the commenter

=====

I am a foreign citizen and resident who work with technology, and is affected by the great influence of the United States on international intellectual property policy. I was in the past, but am no longer, a US citizen.

I lived and worked for many years in the US, and developed firsthand personal experience with the practical application of US intellectual property policy to product design, development, and deployment, as well as with the impact of such policy on Internet security and on issues around business ethics and accountability.

Unlike many, and perhaps most, other potential similarly situated individual commenters, I am, at present, neither fully absorbed by starting a new business, nor employed by an existing large business whose interests I might be bound to respect in providing comments on questions like this one. I believe that I can therefore offer an under-represented perspective on these matters.

I am deeply concerned that certain interests seem to be seeking to muddy the clear concepts of authorship and inventorship with respect to the products of "AI". I regret not noticing the August 27 question on patents, in response to which I would have submitted comments exactly analogous to my comments concerning copyright.

Endnotes

=====

[1] It may be worth noting that the elements most often protected in algorithmic works are likely to differ from those most often protected in manually produced works.

In a manually written book, for example, the specific order and arrangement of particular words is usually very clearly protected, since each word is individually chosen by a human author. In algorithmically produced works, on the other hand, such details are likely to be mechanical and therefore unprotected.

In general, we should expect relatively limited, if any, copyright protection to exist for any particular algorithmically generated work. When protection exists, it is more likely to apply to "higher level" elements, such as the specific parts of the subject matter to be included or excluded, or the specific traits of a character, since these are more likely to represent actual authorial choice.

[2] Treating authorially specified elements as subject to copyright without applying copyright to mechanically produced material is consistent with the existing treatment of the compiled forms of computer programs themselves as derivative works of their source code. Compiler output contains many stereotyped, mechanically produced elements that are not specified by source code authors, but that does not mean that every object program produced by a given compiler is derivative of every other object program produced by the same compiler. Only the programmer-specified elements are protected.