January 10, 2020

The Honorable Andrei Iancu
Under Secretary of Commerce for Intellectual Property
Director of the United States Patent and Trademark Office
600 Dulany Street
Alexandria, VA 22314
Via email: AIPartnership@USPTO.gov

Via email: AIPartnership@USPTO.gov

AUTM's Comments on Intellectual Property Protections for Artificial Intelligence Innovation (Docket No. PTO-C-2019-0038)

Dear Under Secretary Iancu:

Thank you for the opportunity to provide comments regarding intellectual property protections for artificial intelligence innovation posted in the Federal Register on October 30, 2019.

This letter is sent on behalf of AUTM in response to the USPTO's request for comments regarding whether laws that serve to protect artificial intelligence innovations and the assets related thereto, e.g., algorithms and the training datasets, should be modified to address certain unique attributes of AI innovations. While AUTM does not provide direct responses to all of the USPTO's questions, it would like to use this opportunity to highlight a few of the challenges university technology transfer offices have been wrestling with given the lack of clear, effective mechanisms for protecting AI innovations.

I. Given the lack of certainty as to whether owners of AI innovations will receive adequate intellectual property rights in exchange for the extent of disclosure required under patent and copyright laws, innovators may increasingly elect to maintain their AI innovations as trade secrets (when possible).

Innovation was such a priority to the founding fathers that they expressly gave Congress in Article 1, Section 8, Clause 8 of the U.S. Constitution the power to draft laws to address it: [The Congress shall have power] . . . "To promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The scriveners' choice of terms is telling: The language suggests that (1) copyright ("authors") and patent rights ("inventors") could serve to promote the progress of science and useful arts, and (2) these

rights should be exhausted by time ("for limited times"). One of the many benefits of these two statutory rights is that it encourages disclosure of innovations to the public in exchange for certain benefits to the owners.

If owners of AI innovations perceive copyright and patent protection as uncertain and therefore an inadequate exchange for the disclosure of their innovations they may look elsewhere, namely trade secret law. A few of the challenges in relying on trade secret law as a strategy for the protection of innovation stem from the requirement that the owner must establish the innovation is "the subject of efforts that are reasonable under the circumstances to maintain its secrecy." Efforts to maintain the secrecy of an AI innovation may be inconsistent with:

a. **Collaborations between entities.** Collaboration can be an incredibly important engine for spurring scientific and technological advancement. To collaborate effectively, it may be desirable or even necessary for the collaborators to disclose valuable details regarding their AI innovations to each other.

b. **Disclosure or publication.** The public greatly benefits from the inventors' obligation to disclose in order to receive intellectual property rights. If AI innovators increasingly elect to maintain their algorithms, datasets, and results as trade secret, fewer contributions may be made to the scientific community. These disclosures enable scientists to learn from the advancements (and failures) of others, which may reduce unnecessary repeating of experiments We predict that a system which protected AI innovator rights resulting in greater disclosure will increase the pace of research advancements. The commercial market relies upon disclosures from the academic community to derisk new products and ventures.

II. **Protecting AI innovations using trade secret strategies does not align with the ethos of university culture.** Universities largely do not maintain information as trade secret for at least one or more of the following reasons:

a. **Open Dissemination of Research Results & Collaboration is Encouraged.** Trade secret strategies are incongruent with collaboration. If trade secret information constitutes an increasing share of private companies' IP assets, fewer collaborations between the private and public sectors are likely to occur. Many U.S. academic institutions (in particular public universities) have policies that prohibit agreeing to provisions in contracts with third parties (e.g., collaborative or sponsored research agreements) that restrain their academic researchers' ability to disseminate their research methods and results. Also, a fundamental tenet of university culture is the freedom to interpret, publish and otherwise disseminate research results in order to support the transfer of knowledge to others and maintain an open academic environment that fosters intellectual creativity. Trade secrets do not easily align with this mission.

b. **Commitment to Students:** University professors teach their students so that they can necessarily go out into the world with the knowledge they have gained while at the university. Graduate students nearly always need to publish the results of their research in order to graduate. As such, it may be challenging for a university to enter into research or collaborative relationships with external parties that are consistent with the university's commitment to the education of its students.

c. **Importance of Building upon Prior Research:** Academic researchers are often among the leading experts in their fields, which expertise tends to be the result of a foundation of prior research built over the course of years. Agreeing to maintain some of this knowledge as trade secret may negatively impact their ability to publish and conduct further research on the secret subject matter, which their careers and the advancement of research and knowledge for the betterment of society in general necessarily depends upon.

d. **Public Benefit of Verification and Efficiency:** Academia contributes to the public store of knowledge, and in particular to the public store of testable information due to the requirements of peer reviewed journals to disclose the materials and methods used. Published results which are subject to falsification lead to new, useful knowledge. Trade secrets are, by their nature, not subject to falsification and thus do not efficiently advance knowledge within a technical field. For example, when an algorithm is weakly predictive, data maintained as a trade secret may motivate investigators to simply to acquire more data in an attempt to salvage the causal inference through a larger sample size. The transparency offered by an intellectual property right to data should mitigate that motivation. The better approach motivated by data transparency supplied by an intellectual property right would be for the investigator to identify the not-yet-appreciated hidden or confounding variables, or combinations of such variables. This second approach is only possible when the data and proxy variables used by the algorithm are known. Excessive reliance on trade secrets may ironically foster excessive reliance on ever larger-and-harder-to-access datasets instead of ever more insightful use and curation of data.

e. **Public Benefit of Transparency and Ethics**: In addition to the challenges presented when results or other information is not subject to falsification, there is also the problem of bias resulting in potentially harmful proxy classifiers. As recently publicized (see *Science* 25 Oct 2019:Vol. 366, Issue 6464, pp. 447-453DOI: 10.1126/science.aax2342 "Dissecting racial bias in an algorithm used to manage the health of populations"), AI can deliver misleading results – when health costs were used as a proxy for health needs, without appreciating that patients with significant needs often simply do not use the health system, and thus have lower costs. This paper is notable for a few reasons. First, the academic researchers knew how to properly gain access to protected health information "PHI", which shows that privacy protection and peer reviewed research, including verifying and/or falsifying prior results which themselves used PHI, are possible. Secondly, the authors provided a synthetic data set to enable yet other researchers to verify their results ("Because the data used in this analysis are protected health information, they cannot be made public available. We provide instead a synthetic dataset … and all code necessary to reproduce our analyses.").

f. **Public Benefit of Interoperability:** In due course, arriving at consensus classifiers and electronic relational diagrams ("ERDs") has enormous potential to facilitate research. As an aspirational example, imagine if electronic health records had standardized fields, ERDs, and consensus predictive biomarkers – this would greatly increase the speed (and efficiency) at which insights into health care could be developed and tested. For example, recent studies (https://globalforum.diaglobal.org/issue/may-2019/what-are-the-chances-of-getting-a-cancer-drug-approved/, and https://academic.oup.com/biostatistics/article/20/2/273/4817524 report that biomarkers used to select patients for cancer clinical trials significantly increase the likelihood of approval of the treatment. Examples of other areas which may similarly benefit from data interoperability (such as consensus classifiers and ERDs) include agriculture and marine resource management.

III. Artificial Intelligence-enabled inventions are often created using vast training datasets. Clarifying what intellectual property protection is available to owners to protect their investments in creating these databases may encourage disclosure and interoperability of datasets in this space.

a. The European Union's Database Directive recognizes property rights for databases.

The European Union has been leading the way in both ownership, and privacy rights, of data since the 1990s and 2010s, respectively. In 1996, the European Commission introduced the Database Directive which carved out *sui generis* property rights for databases. That is to say, under the EU regime, databases are granted *bona fide* property rights (as opposed to IP rights). In jurisdictions wherein such right is observed, once a database is established, the creator is afforded 15 years of protection. This exclusive right specifically targets the extraction and re-utilization of substantial portions of databases and archives. Beyond this threshold, authorization via a license is mandated.

Since 1996 the European Court of Justice "ECJ" has adjudicated several cases referred to it (*see* British Horseracing Board v. William Hill Organization LtD (UK), Fixtures Marketing v. Oy Veikkaus Ab (Finland), Fixtures Marketing v. Svenska Spel Ab (Sweden), and Fixtures Marketing v. OPAP (Greece)). The ECJ's conclusions underline that the "*sui generis* right stemmed from investing in the creation of the database, mainly the act of seeking, discovering, and collecting disparate data from independently existing sources, − in contrast to the mechanical collection or creation of data…". Importantly, this designation is entirely removed from Copyright Law. Unlike in the Copyright system, which protects database structure, originality does not factor into the determination of eligibility for the *sui generis* right. Instead, a substantial investment standard, akin to the old American "sweat of the brow" requirement, is employed. The core stated purposes of the Database Directive were to "harmonize protection of databases, stimulate investment in them, and safeguard the balance between rights and interests of database producers and users."

The Database Directive and the protections it affords have gone through two rounds of extensive multifactorial analysis. The European Commission has concluded that the Database Directive has indeed had a harmonizing effect, and set the right balance of interests. However, apparently it has had no tangible impact on the production of databases or the competitiveness of the industry. Nonetheless, discontent for the current policy has grown, and there have been calls by the European Commission to create a so-called "Data Producer's Right."

b. The EU's General Data Protection Regulation (GDPR)

Consistent with its leadership in database protection rights, the EU pioneered data privacy protection through dramatic new regulation. Implemented in 2018, the General Data Protection Regulation (GDPR) established standards for the handling of data and the norms governing its use. The overarching theme of the GDPR is an increased emphasis on consumer control when it comes to their data. As per the legislation: (i) consent requests must not be couched with vague terms, (ii) consent to various terms may not be bundled, (iii) consent must be readily revocable, (iv) companies bear a burden of promptly reporting breaches, and (v) consumers will have access to their personal data and information surrounding its use by the companies that house it. The penalty for violation of GDPR provisions is the greater of 20,000,000 euros, or 4% of global annual revenue.

c. Similarities between databases and tangible biomaterials.

While the U.S. has rejected a "sweat of the brow" doctrine, universities have been transferring biological materials between themselves and industry for decades. Through such experience, universities have generated several template material transfer templates that reflect common understanding universities have as to ownership rights in such biological materials.

It is generally appreciated that these biological materials, such as engineered cell lines, reagents, or mouse models, were made with significant effort and resources, often over a period of years. These biological materials are transferred to other researchers for several reasons, including: i) they are needed for peers and colleagues to confirm or revise the published results, ii) making them available is often a condition of receiving funding, and iii) sharing materials saves time and effort for future researchers. It is also understood that it is reasonable to put some conditions on the transfer. Examples of these conditions may be i) according credit to the source of the biological materials and ii) using it only under certain conditions.

AUTM member institutions are reporting increasing work on curating data sets to make them useful for research and for use as training datasets. This additional work includes combining data from more than one source, and creating the taxonomies by which the data are organized. Trial classification systems are particularly important when the categories that matter are not known to begin with. The Linnean classification system, for example, did not come from collection alone. The table below illustrates some similarities and differences between biological materials and curated datasets:

Attribute	Biological Materials	Curated datasets
Tangible	Yes.	No.
Effort is required to make and maintain	Yes "Making" means taking certain actions in a lab with tangible materials, cells, animals, etc…, performing, and documenting quality control tests to assure that the material has been produced as expected, and storing or maintaining the material securely.	Yes "Making" means gathering, curating (sorting according to a potential property of interest), performing quality control tests, such as relational integrity, indexing of records, etc…, documenting all of the above, including the sorting criteria, the fields, and the electronic relational diagram "ERD" and storing securely
Useful to share, transfer, and make available to others	Yes	Yes
Interest in transferring with conditions on the recipient	Yes. May wish to receive credit when the material is used, for example. May wish to limit use and further distribution.	Yes. May wish to receive credit when the database is used, for example. May wish to limit use and further distribution.

Despite these similarities, because biological materials are tangible, they fall easily into the concept of property rights, while curated datasets do not. AUTM is interested in exploring the possible expansion of *sui generis* property rights to curated datasets in view of the:

1) reported increasing efforts toward data curation,

2) fact that curation may not, in some circumstances, meet the criteria required for copyright protection,

3) benefits of sharing, improving, and incentivizing the creation of such datasets, including in the context of public-private partnerships,

4) disadvantages of trade secrets, particularly in an academic environment, and

5) different infringement protections needed for proprietary data sets that are used in a "once and done" manner to train AI algorithms.

IV. Helpful clarifications with respect to data management and use.

a. Secure storage

Standards for secure electronic data transfer and storage will be appreciated and help facilitate controlled sharing of such data for public benefit. How many and what type of electronic defensive measures are reasonably expected of today's data managers?

b. Malicious re-identification

A law expressly prohibiting malicious re-identification of de-identified data will be helpful, as it would shift some of the defensive burden away from the data custodians.

c. Synthetic datasets

A clear expression from federal sponsors on when and if they will require researchers who study PHI and other confidential information to providing synthetic datasets which can then be used to verify their published results. Would such synthetic datasets be of interest to the USPTO in showing compliance with aspects of 35 U.S.C. § 112?

d. Use of copyrighted material for machine learning

It would also be interesting to explore electronic approaches to "fair use" of copyrighted material by an AI system. Should owners of copyrighted material receive an electronic watermark to enable such owner to either i) opt in for unlimited with no compensation use by AI systems, ii) reject outright the use of their copyrighted material in a machine learning context, or iii) potentially specify a fee for such use?

These are some of the complexities and new ways of working with AI datasets and AI related inventions that technology transfer offices struggle to balance. While AI offers the potential for significant value, the current intellectual property framework is inadequate to fairly reward the "curator" or "owner" of the datasets. Additional discussion and resources to build effective policies and intellectual property protections are needed so that AI can effectively contribute to advances across many fields of study.

Sincerely,

Stephen J Susalka

Stephen J. Susalka, PhD, CLP, RTTP

AUTM CEO