AIPLA

American Intellectual Property Law Association

January 10, 2020

The Honorable. Andrei Iancu
Under Secretary of Commerce for Intellectual Property and
Director of the United States Patent and Trademark Office
United States Patent and Trademark Office
600 Dulany Street
Alexandria, Virginia 22314

Via email: AIPartnership@uspto.gov

Re: AIPLA Comments regarding "Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation," 84 Fed. Reg. 58141 (October 30, 2019) [Docket No. PTO–C–2019–0038]

Dear Under Secretary Iancu:

The American Intellectual Property Law Association (AIPLA) respectfully submits these comments in response to the Department of Commerce's Request regarding issues related to copyright, trademark, and other intellectual property rights impacted by artificial intelligence innovation raised in the October 30, 2019 Federal Register Notice (84 Fed. Reg. 58141).

AIPLA is a national bar association of approximately 12,000 members engaged in private or corporate practice, in government service, and in the academic community. AIPLA members represent a wide and diverse spectrum of individuals, companies, and institutions involved directly or indirectly in the practice of patent, trademark, copyright, trade secret, and unfair competition law, as well as other fields of law affecting intellectual property. Our members represent both owners and users of intellectual property. Our mission includes helping establish and maintain fair and effective laws and policies that stimulate and reward invention while balancing the public's interest in healthy competition, reasonable costs, and basic fairness.

I. 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?

No. A work produced without the involvement of a natural person should not be copyrightable. Under current U.S. copyright law, the hypothetical would result in an "authorless" work that is not protectable by copyright because no human being is involved in the resulting work.[1]

---

[1] The Patent and Copyright Clause of the U.S. Constitution, which authorizes Congress to enact the U.S. Copyright Act, provides: "The Congress shall have power . . . "[t]o 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." U. S. Const. Art. I § 8. At least

AIPLA believes that use of an AI algorithm or process should not change the fundamental requirement that authorship is limited to a natural person(s) or, in the case of a work made for hire, the employer of a natural person(s). A natural person may use technical tools (*e.g.,* a camera, a word processing program, a drafting program, image processing software, *etc.*) to produce a work and benefit from copyright protection if the natural person contributed protectable expression to the work. Similarly, a natural person may employ an AI algorithm or process as a tool to produce a work and benefit from copyright protection if – and only if – the natural person contributed protectable expression to the work. However, if an AI algorithm or process functions in a manner to produce a work to which no natural person contributed protectable expression, then the resulting should not be protectable by copyright.

II.   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 activities? Are there other contributions a person could make in a potentially copyrightable AI-generated work in order to be considered an "author"?

Contribution of protectable expression to the AI-created work should be required. Given the many different types of AI algorithms and processes and the many different applications thereof, evaluation of the existence and degree of protectable expression by a natural person to a work made using in whole or in part an AI algorithm will require case-by-case consideration. With this in mind, AIPLA provides the following general observations on the sub-questions:

(i)   Designing the AI algorithm or process that created the work.

It is unlikely a natural person who designs an AI algorithm or process that creates a work would be considered to have contributed protectable expression to the resultant work. Copyright may, of course, lie in the code that carries out the AI algorithm.

(ii)   Contributing to the design of the algorithm or process.

See (i) above.

---

one U.S. Court of Appeals has held, on a closely related issue, that under the Copyright Act, only humans (and not animals, like the monkey involved in the case) had statutory standing to sue for copyright infringement. *Naruto v. Slater*, 818 F.3d 418, 426 (9th Cir. 2018). Further, the U.S. Copyright Office interprets "authors" as human beings: [t]he U.S. Copyright Office will register an original work of authorship, provided that the work was created by a human being. *See* Copyright Compendium 306 The Human Authorship Requirement. "The copyright law only protects 'the fruits of intellectual labor' that 'are founded in the creative powers of the mind.' . . . Because copyright law is limited to 'original intellectual conceptions of the author,' the Office will refuse to register a claim if it determines that a human being did not create the work. . ..." *Id* (internal citations omitted).

(iii)   Choosing the data used by the algorithm for training or otherwise.

Choosing the data used to train an AI algorithm should generally be considered as part of the process of creating the AI process itself. As such, selection of training data would not be contribution of protectable expression to a work made using the AI process. However, selection of input data that dictates, at least in part, the output of an AI process may involve protectable expression, just as composing a scene with a camera suffices as contribution of protectable expression to the resulting photograph. This will depend on the application.

Consider, for example, a hypothetical AI process that has been trained to orchestrate melodies in the style of Georg Philipp Telemann. As the world's most prolific composer with more than 3000 works, an AI process trained with all of Telemann's works could be expected to be able to produce, from simple melody lines, orchestrations that are similar to those of Telemann. If a natural person were to use this hypothetical AI process to produce a Telemann-style work by inputting a melody that embodies the protectable expression of the natural person, the resultant work should be protectable by copyright. This is similar to an author using a tool such as a word processor, albeit more sophisticated.

Consider, on the other hand, a hypothetical AI process that has been trained to convert Baroque orchestrations into the style of George Gershwin. A natural person who used such a process by inputting the works of Telemann should not be considered to have contributed protectable expression to the resulting Gershwin-style work.

(iv)   Causing the AI algorithm or process to be used to yield the work.

Mere application of an AI algorithm or process to a certain task should not be considered contribution of protectable expression to the resultant work. However, as explained in (iii) above, where the use of the AI process by a natural person involves a contribution that does include protectable expression, copyright should lie in the resultant work.

(v)   Engaging in some specific combination of the foregoing activities.

In a given AI application, if any one of the foregoing activities involves contribution by a natural person of protectable expression to the resultant work, then so would a combination of the foregoing activities that includes such contribution. It also is possible, on a case-by-case basis, that contributions by a natural person to a combination of the foregoing activities could amount to contribution of protectable expression to the resultant work even if this is not true for any one of the foregoing activities taken alone.

III.   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?

AIPLA collaboratively contributed to a primer on AI and copyright, ("AI Copyright Primer,") published in conjunction with the Association Internationale pour la Protection de la Propriété Intellectuelle (AIPPI) and the International Federation of Intellectual Property Attorneys (FICPI).[2]

As discussed in the AI Copyright Primer, during the AI learning phase an AI system relies on copious input of information to "learn", which are referred to as "training materials." This is part of the necessary training experience before the AI system can be expected to produce any relevant output. The training materials are generally known as "data input" or "input database". For example, in an image recognition AI system, the training materials may include a large amount of art[3] or pictures, which may be (and typically are) obtained from the internet through bots[4]. In other AI systems, music, formulas, chemical compounds, designs and even architecture may be obtained in the same manner.

Some AI systems write their own code during the AI learning phase. To do so, these AI systems generally scour the internet for other sources of code.[5]

Some more creative AI systems employ a GAN ("generative adversarial network") system,[6] which typically uses two independent neural networks to critique each other. One of the systems is a generator network, which creates pieces of output (e.g., art) based on input from

---

[2] *See* https://ficpi.org/_/uploads/files/AIPLA-AIPPI-FICPI_Artificial_Intelligence_Colloquium_Copyright_ONLY_Primer.pdf (visited November 20, 2019).

[3] Jason Daley, *AI Project Produces New Styles of Art*, Smithsonian.com: Smart News (July 3, 2017) https://www.smithsonianmag.com/smart-news/ai-system-produces-new-styles-art-180963912/#vu8sZF2D4abpA0Ry.99 (Where AI is being used to create art: "Researchers from Rutgers University, the College of Charleston and Facebook's AI Lab collaborated on the system, which is a type of generative adversarial network or GAN, which uses two independent neural networks to critique each other. In this case, one of the systems is a generator network, which creates pieces of art. The other network is the 'discriminator' network, which is trained on 81,500 images from the WikiArt database, spanning centuries of painting. The algorithm learned how to tell the difference between a piece of art versus a photograph or diagram, and it also learned how to identify different styles of art, for instance impressionism versus pop art.").

[4] https://en.wikipedia.org/wiki/Internet_bot.

[5] Matt Burgess, *Microsoft's AI Writes Code by Looting Other Software*, Wired: Artificial Intelligence (Feb. 23, 2017) https://www.wired.co.uk/article/ai-write-code-microsoft; *see* Matej Balog et al., *DeepCoder: Learning to Write Programs*, ICLR (2017) https://openreview.net/pdf?id=ByldLrqlx (Wherein the system works by taking lines of code from existing programs and combining them: "A neural network, called DeepCoder, developed by Microsoft and University of Cambridge computer scientists, has learnt how to write programs without a prior knowledge of code. First reported by the New Scientist, the system works by taking lines of code from existing programs and combining them.").

[6] *Generative Adversarial Network*, Wikipedia, https://en.wikipedia.org/wiki/Generative_adversarial_network (last modified Mar. 13, 2019, 11:56 UTC) ("Generative adversarial networks (GANs) are a class of machine learning algorithms used in unsupervised learning, implemented by a system of two neural networks contesting with each other in a zero-sum game framework").

the discriminator trained in a specific genre (e.g., paintings by van Gogh). The other network is the "discriminator" network, which is trained on input from the internet to feed intermediate formats or images to the generator network.

During this process, an AI system (typically the discriminator network) makes copies of the input, manipulates the training materials, and creates an intermediary dataset or database for the generator network.[7] In other cases, AI needs to have the input data "labeled."[8] In some cases, the labeling is done manually, by humans[9] or through asking humans,[10] but in all cases this would be considered an intermediary database to be used by the AI system for the "generator network" or some similar output AI system entity.

There is no need for AI to employ copyrighted works as training materials. AI systems could also be coded to recognize and avoid using copyrighted materials based on digital rights management data associated with the copyrighted works. Nonetheless, many AI systems do use copyrighted materials as training materials.

If the AI were trained on copyrighted material, absent authorization or one of the exceptions to infringement in Section 107-122 of the Copyright Act, use of the copyrighted material would infringe. Given the large volume of information required to train a typical AI system, obtaining authorization from all of the copyright owners may present an insurmountable burden. If the input information is in the public domain, there should be no copyright problem. However, if the data is expressive and copyrighted (much of which is copyrighted in the U.S. upon fixation), that copying infringes the owner's exclusive right of reproduction or the exclusive right to make derivative works unless authorized or excused by an exception.

There is no guarantee that use of copyrighted training materials would constitute fair use of the copyrighted works. Fair use is a holistic, equitable defense. Whether a particular use is fair depends upon the facts of each individual case. This would require a contextual analysis of four non-exclusive fair use factors, which are discussed in more detail below.

AIPLA is aware of no copyright infringement case based on unauthorized use of copyrighted works as training materials. However, a similar situation was considered in *Authors*

---

[7] *See, e.g.*, *Big Data*, Wikipedia, https://en.wikipedia.org/wiki/Big_data (last modified Mar. 16, 2019, 01:14 UTC).

[8] *Labeled Data*, Wikipedia, https://en.wikipedia.org/wiki/Labeled_data (last modified Aug. 29, 2018, 11:13 UTC).

[9] Hope Reese, *Is 'Data Labeling' the New Blue-Collar Job of the AI Era?* TechRepublic: Innovation (Mar. 10, 2016, 10:00 AM PST) https://www.techrepublic.com/article/is-data-labeling-the-new-blue-collar-job-of-the-ai- era/ ("But Guru Banavar, the head of the team at IBM responsible for creating Watson, the AI system that mastered Jeopardy, told TechRepublic that this isn't necessarily the case. Banavar thinks that there will be "all kinds of jobs available" in the AI era. For workers at all skill levels. And for lower-skilled workers, data processing offers a new area of possibility. "Data labeling," is what Banavar calls it. "It will be the curation of data, where you take raw data and you clean it up and you have to kind of organize it for machines to ingest," he said. "If you look at any of the complicated analytical jobs we have today, 70% of that job is probably about the organizing and cleaning of data.").

[10] *Labeled Data*, Wikipedia, https://en.wikipedia.org/wiki/Labeled_data (last modified Aug. 29, 2018, 11:13 UTC) ("Labels can be obtained by asking humans to make judgments about a given piece of unlabeled data (e.g., 'Does this photo contain a horse or a cow?'),and are significantly more expensive to obtain than the raw unlabeled data.").

Guild, Inc. v. Google, Inc.[11] in which the U.S. Court of Appeals for the Second Circuit found the use permitted under the fair use exception. In that case, Google input over 20 million books, many copyrighted, into its servers. Researchers could enter a text inquiry, and the system would return snippets of the inquiry in context, as well as references to the books where the terms appeared. Applying the reasoning of *Authors Guild* to a typical AI use case, the following considerations are noted with respect to each of the four "fair use factors" set forth in 35 U.S.C. § 107:

(1) <u>The purpose and character of the use</u>. The United States Supreme Court in *Campbell v. Acuff-Rose Music, Inc.* noted that this factor favors transformative use, *i.e.*, a use that does not merely supersede the original work but, instead, "adds something new, with a further purpose or different character, altering the [original work] with new expression, meaning, or message."[12] *Authors Guild* quoted *Campbell*, noting that "the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works."[13] The Second Circuit found Google's search function to be "highly transformative." Assuming that copyrighted images are not being used for their artistic value but, rather, to train the AI system, the use may be considered transformative. Thus the first factor would favor fair use.

(2) <u>The nature of the copyrighted work</u>. The court in *Authors Guild* noted that the nature of the copyrighted work "has rarely played a significant role in the determination of a fair use dispute."[14] To the extent that training materials contain creative expression, the second factor would weigh against fair use.

(3) <u>The amount and substantiality of the portion used</u>. Typically, the AI would be intaking entire photographs, or substantial blocks of text. "[C]ourts have rejected any categorical rule that a copying of the entirety cannot be a fair use."[15] Transformative uses are permitted to take as much as is necessary to affect the use. In this case, intaking the entirety of certain works may be necessary to adequately train the AI. Thus, this factor may be relatively neutral.

(4) <u>Effect on the market</u>. This has often been called the most important factor, or at least, with the first factor, one of the two most important

---

[11] *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015) (*cert. denied*).

[12] *Campbell v. Acuff-Rose Music Inc.*, 510 U.S. 569, 579 (1994).

[13] *Authors Guild,* 804 F.3d at 214.

[14] *Id*. at 220.

[15] *Id*. at 221.

factors. In *Campbell*, the Supreme Court noted the linkage between the first and fourth factors: "when a commercial use amounts to mere duplication of the entirety of an original, it clearly "supersede[s] the objects" of the original and serves as a market replacement for it, making it likely that cognizable market harm to the original will occur. But when, on the contrary, the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred."[16] Use of the photo dataset did not substitute for the purpose of the original photographs. Nor would licensing revenue likely be lost by the copyright owners, because no potential market exists in light of the prohibitive transactional costs in finding and licensing from literally thousands (if not more) of copyright owners.

While, as noted, each AI input scenario would have to be analyzed contextually, it is possible that use of copyrighted material as training materials may constitute fair use depending on the nature of the materials used and how they are used. Fair use issues are not clearly addressed or resolved in the existing statutory framework. Additional guidance is needed whether unauthorized derivative works are being created by the AI, and, whether the use of copyrighted materials as training materials is "transformative."

Further even were the use fair, recognition of the authors of the pre-existing copyrighted works would probably be required under moral rights, which the U.S. has agreed to protect under its adherence to the Berne Convention.

IV.   Are current laws for assigning liability for copyright infringement adequate to address a situation in which an AI process creates a work that infringes a copyrighted work?

No. Current laws may rely on certain general law doctrines. Existing laws governing liability for copyright infringement and the fair use defense are probably not clear with respect to AI generated works.

V.   Should an entity or entities other than a natural person, or company to which a natural person assigns a copyrighted work, be able to own the copyright on the AI work? For example: Should a company who trains the artificial intelligence process that creates the work be able to be an owner?

No. Ownership should vest in the author (or employer of the author, in the case of works made for hire) and may then be assigned to another natural or juridical person.

---

[16] *Campbell,* 510 U.S. at 591 (*internal citations omitted*).

VI.   Are there other copyright issues that need to be addressed to promote the goals of copyright law in connection with the use of AI?

Existing U.S. law adequately addresses copyright protection for computer programs. Copyright protection for computer programs implementing AI routines are similarly protected. However, as the previous questions make clear, current U.S. law is subject to interpretation and may be unclear when applied to various scenarios raised by new applications of AI. Guidance, ideally with practical examples for various categories of AI applications, would be helpful to the public for clarity and predictability of copyright law.

VII.   Would the use of AI in trademark searching impact the registrability of trademarks? If so, how?

Yes, it could enhance searching, particularly for design marks that are difficult to search using current technology. AI searching might enable USPTO examining attorneys to find more potentially similar design marks than the current design code indexing method does, and this could lead to more refusals to register on the basis of section 2(d).

Various AI programs have been developed recently to enhance the efficiency of trademark searching.[17] These AI-powered solutions search and monitor words, phrases, and images using various government trademark offices' databases of marks and the internet.[18] However, the evaluation of search results is based on *human perception*. For example, "[f]ull of human-centric phrases such as imperfect recollection, average consumer, and confusion, human lawyers will be able to make those calls better than machines."[19] That said, humans inherently incorporate practical considerations into their arguments and decisions.

VIII.   How, if at all, does AI impact trademark law? Is the existing statutory language in the Lanham Act adequate to address the use of AI in the marketplace?

AI search tools have the potential to improve the quality of searching without otherwise impacting trademark law.

Certain AI algorithms and processes already commonly interact with consumers in many ways such as voice or text-based interfaced chatbots that suggest products. Issues will likely arise as to whether particular AI algorithms and processes are biased against certain

---

[17] For example, Trademark Now, TrademarkVision, MikeTM Suite and LawPanel's Aila.

[18] *See* "How AI Impacts Trademarks," Trademark Times (ManagingIP.com), May 19-23, 2018, https://www.managingip.com/pdfs/01-TrademarkTimes18Seattle.pdf.

[19] *Id.*

product selections or otherwise dilute (blur, whittle away, or tarnish) well-known marks. Existing case law[20] on liability issues will likely provide guidance.

AIPLA respectfully requests that the USPTO clarify whether services rendered by means of AI are properly classified as the underlying service itself, or as computer software. And, if properly classified as computer software, what specimen of use is sufficient if the mark is used only in connection with promoting the service, and the consumer or relevant market is never exposed to the back-end computer running the algorithm. AIPLA also requests that the USPTO clarify the definition of a specimen to take into account current ways that marks are used in commerce, including the way services are rendered by AI, that do not fit neatly into the categories currently stated in the statute.

IX.   How, if at all, does AI impact the need to protect databases and data sets? Are existing laws adequate to protect such data?

AI is a data-driven technology that could involve various types of databases and data sets. Therefore, further study is warranted on whether to protect, and if so what types of protection should be provided for, databases and data sets.

A.   Training issues

Aside from trade secret protection, there is little or no protection for data[21] and databases[22]. Patenting may be available for a training technique, but detecting infringement would be difficult. Creation of meaningful training data is a sunk cost associated with use of various machine learning techniques.

During training, a database is created called an "intermediary database." The intermediary database, like any other database, may be protected by copyright provided the selection and arrangement of input is minimally creative. To do its work the AI system must ingest and cull works to create a robust dataset. This process is expensive and time-consuming. The resulting dataset is valuable to the AI system provider. Copyright law, however, provides little protection. Other than the selection and arrangement of the works ingested, the dataset may contain no expression. Despite the "sweat of the brow" needed to produce it, it is a collection of facts. In *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, the Supreme Court held that a telephone book was not protected by copyright, despite the effort it took to create it.[23] Recently,

---

[20] *Cosmetic Warriors Limited and Lush Limited v Amazon.co.uk Limited and Amazon,* EU SARL [2014] EWHC 181 (Ch); *L'Oreal SA v eBay International AG*, (C-324/09) EU:C:2011:474 (12 July 2011).

[21] See 17 U.S.C § 103

[22] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.1996.077.01.0020.01.ENG&toc=OJ:L:1996:077:TOC, Also see https://en.wikipedia.org/wiki/Database_right.

[23] Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).

the Ninth Circuit found *some* copyright protection for a database of customer information.[24] Absent a change in the law, copyright provides little protection to the valuable AI dataset.[25]

Data and databases have been protected by various Government regulations for several decades. Beginning during the Second World War, when the Government coordinated production of war material between numerous former competitors, various forms of data protection were implemented by Department of Defense regulation. Similar regulatory scheme to protect proprietary data have been developed by EPA and FDA. Patient data has been protected by Health Information Patient Privacy Act (HIPPA). And general regulations regarding data privacy rights (GDPR) have been promulgated by the European Community. There remains a need for a *sui generis* form of data protection that would enable further development of AI systems.

B.   Characterization[26]

Data used for training a machine learning model can be characterized[27] in support of such training. Significant effort may be invested in such characterization, such as to avoid bias or to provide consistent predictive ability.[28] For IP protection, there are two aspects that should be considered, Structured or Data-Driven AI[29] and Unstructured or Model Driven AI:[30]

---

[24] *Experian Information Solutions, Inc. v. Nationwide Marketing Services, Inc*., 893 F.3d 1176 (9th Cir. 2018). The Plaintiff created culled data from some 2,200 public and proprietary sources and used methods to select only those customers it thought were valuable to its clients:

"The value, according to Experian, results from the process by which Experian determines the accuracy of its pairings and the utility of the selection of the pairings it includes in the CVD for its marketing clients." Ibid.

"For its database, Experian picks from roughly 2,200 public and proprietary sources that it believes have reliable, value-adding data. In determining whether to include a new source in its database, Experian runs the source through tests to measure the potential new data's quality and to identify the differences between the new source's data and existing data in the CVD." *Ibid*.

[25] *Ibid.*

[26] In fact, some have decided that the big 3 V's (Volume, Velocity, Variety) has become the 4 V's, where the fourth is now called Veracity, which refers to the data quality and the data value. The data quality of captured data can vary greatly, affecting the accurate analysis. https://en.wikipedia.org/wiki/Big_data

[27] Terms such as "tagging," "labeling," "characterization" are used synonymously herein.

[28] There are so many articles, but this is fairly recent and very disturbing https://www.nbcnews.com/mach/tech/playing-roulette-race-gender-data-your-face-ncna1056146

[29] The adjective data-driven means that progress in an activity is compelled by data, rather than by intuition or by personal experience. https://en.wikipedia.org/wiki/Data-driven

[30] In artificial intelligence, model-based reasoning refers to an inference method used in expert systems based on a model of the physical world. With this approach, the main focus of application development is developing the model. Then at run time, an "engine" combines this model knowledge with observed data to derive conclusions such as a diagnosis or a prediction. https://en.wikipedia.org/wiki/Model-based_reasoning

- Data-Driven is generally trained in a specific manner by categorization and review by people. Data-Driven, needs to have the input data "Labeled."[10] In some cases, the labeling is done manually, by humans[11] or through asking humans[12], but in all cases this would be considered an intermediary database. Data-Driven, uses labeling created by a process. For example, more of the creative AI systems employ a GAN ("generative adversarial network") system[31] which typically uses two independent neural networks to critique each other during training. One of the systems is a generator network, which creates pieces of output (e.g., art) based on input from the discriminator trained in a specific genre. During this process, the discriminator network must make copies of that input, manipulate that material and create an intermediary dataset or database for the generator network.[32]

- Model Driven is generally trained by indicating a model that the AI is to represent and allow it to find its way through data to derive that model.

Aside from keeping such data secret, there may be no appreciable patent protection available for the effort involved in curating and characterizing training data for the structured model. Copyright protection does not distinguish between the two different models. As indicated above for the data training section, absent a change in the law, copyright provides little protection to the valuable AI dataset. Absent some form of *sui generis* data rights protection, trade secret protection may be the strongest IP protection available for the characterization of training data.

C.   Volume issues

Although AI can find patterns and relationships in large volumes of data, it does not mean that these patterns and relationships are meaningful. Correlation is not causation. Noise in the data or "chance data"[33] can become meaningful in a way that could not be foreseen. [34]

---

[31] https://en.wikipedia.org/wiki/Generative_adversarial_network , Generative adversarial networks (GANs) are a class of machine learning algorithms used in unsupervised learning, implemented by a system of two neural networks contesting with each other in a zero-sum game framework.

[32] See, for instance, *Big Data,* Wikipedia, https://en.wikipedia.org/wiki/Big_data (last modified Mar. 16, 2019, 01:14 UTC)

[33]Data dredging (also data fishing, data snooping, data butchery, and p-hacking) is the misuse of data analysis to find patterns in data that can be presented as statistically significant, thus dramatically increasing and understating the risk of false positives. This is done by performing many statistical tests on the data and only reporting those that come back with significant results https://en.wikipedia.org/wiki/Data_dredging

[34]To see some of the fun being poked at curious relationships found - http://www.tylervigen.com/spurious-correlations

D.   Velocity issues

The volume of new data generated and used may or may not be important, depending upon its application. For instance, determining what is trending or what is not on a Twitter stream would require an ability to balance the velocity of data with the volume of data for analytics. A challenge for copyright may be that when here is high velocity and volume, intermediary databases are being created all the time as new data helps crate different paths for the model. Therefore, the intermediary and output database may have even more limited protection once created by the AI that it uses for its final output.

Providing for more flexibility with software methodology patents might help and having the data standards would improve reliability to assist with reliable outputs for patenting.

E.   Variety issues

It is easier to analyze data that is from similar sources such that correlation is quickly concluded. However, even when taking data from one source, it may not prove reliable due to contextual use of terms[35]. That big data now typically includes many different formats, i.e., pictures, emails, tweets, pdf, audio, spreadsheets, structured relational databases, amplified the problem. The ability to not only store this disparate data and analyze it in a contextual similar fashion may present new challenges.

F.   Privacy issues

A number of considerations related to privacy militate strongly for an effective system of data rights protection, that preserves proprietary value for compilers of datasets while not unduly inhibiting public access to public domain information.

Where and how data is obtained is not always well understood and may depend upon the mechanism built into the AI to obtain the data. Today, there are multiple tools[36] available to decipher web data (including hidden data), databases (structured or unstructured) or web crawlers to find those hidden sites not normally found in searches. Currently there may be some protection for the data (Europe)[37] with some data scrapped based upon HIPPA or GDRP similar laws such as in California.[38] However, those protections may not be available as the data may be scrapped without the knowledge of the owners of the data.

---

[35] Terms used for opioid users turned out to be problematic to determine using only twitter data, however finding users selling was easier. See https://www.scientificamerican.com/article/ai-scans-twitter-for-signs-of-opioid-abuse/

[36] https://www.ventureharbour.com/online-marketing-ai-machine-learning-tools-you-can-try-today/

[37] GDPR, https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

[38] California Consumer Privacy Act, https://en.wikipedia.org/wiki/California_Consumer_Privacy_Act

X.   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?

A.   Protecting AI inventions as trade secrets under the DTSA

The lack of effective protection for many of the elements of AI discussed above will likely lead to greater reliance on trade secrets protection. This in turn may deny the public the benefit of publication of and access to AI-related information that may support or induce the creation of further AI improvements.

A trade secret is any commercially valuable *information* that (1) derives economic value from not being publicly known; (2) is the subject of reasonable efforts to maintain its secrecy; and (3) provides economic benefit to its owner.

The Defend Trade Secrets Act (DTSA), 18 U.S.C. 1836, defines trade secrets broadly, covering a wide range of proprietary information to fall under its protection, including programs or codes. The DTSA "expressly contemplates protecting computer source code as a trade secret."[39] "[T]he broad scope of what may be a trade secret potentially allows for the protection of data sets that underlie the machine learning of AI."[40]

The DTSA defines a trade secret as:

> ***All forms and types of*** financial, business, scientific, technical, economic, or engineering ***information***, including patterns, plans, compilations, program devices, formulas, designs, prototypes, methods, techniques, processes, procedures, ***programs, or codes***, whether tangible or intangible, and whether or how stored, compiled, or memorialized physically, electronically, graphically, photographically, or in writing if:
>> (A) The owner thereof has taken reasonable measures to keep such information secret.
>> (B) The information ***derives independent economic value***, actual or potential, from ***not being generally known*** to, and not being readily ascertainable through proper means by, another person who can obtain economic value from the disclosure or use of the information.[41]

Notably, protection of a trade secret under this definition requires that "reasonable measures" are taken to maintain the secrecy of the relevant information. AI-specific "reasonable measures" may include "encrypting any AI software code; password protecting the AI software

---

[39] David A. Prange and Alyssa N. Lawson, *Re-evaluating Companies' AI Protection Strategies*, Robins Kaplan, (*available at,* https://www.robinskaplan.com/~/media/pdfs/reevaluating%20companies%20ai%20protection%20strategies.pdf?la=en).

[40] *Supra* note 1.

[41] 18 U.S.C. § 1839(3) (emphases added).

code; keeping the AI software or algorithm asset out of open source,"[42] among other things. However, even if the criteria are met, there are several recognized defenses to a claim of trade secret misappropriation, including legitimate reverse engineering, independent development, and the claim that the relevant information does not possess independent economic value, which are all possibilities when using an unpatented AI software or algorithm.[43]

Data used to train an AI system (training data) may be protected as a trade secret if the data meets the requirements for protection under the DTSA. Certain requirements include that the training data is properly within the scope of protectable subject matter; protection may not attach to data that falls outside those bounds.

Individual data sets provided to an AI system may be protected as trade secret, but potential challenges to independent economic value and public availability may make the protection of subparts of the training data inefficient and undesirable.

Regardless whether individual pieces of training data are publicly available or valuable, the entire set of information used to train a system could merit protection as a compilation. Compilations of public information (or mixed public and confidential information) can gain trade secret status so long as the compilation as a whole is not generally accessible or publicly available, has independent economic value, and is reasonably protected. Thus, trade secret law under the DTSA might protect that compilation of public data, as well as how an operator chooses what data to use to train an AI system.

B.   Issues with protecting AI-inventions as trade secrets under the DTSA

Precisely articulating what the exact trade secret is and the specific efforts taken to maintain its secrecy may be difficult when the *information* seeking to be protected is the product itself (e.g., the AI invention) or created by the AI system (i.e., so called "Inventive AI"), "particularly where it is within the black box of how the AI system performs its analysis."[44]

At the heart of AI lies the issue of the AI's ability to learn in an increasingly complex manner, and "essentially program[ing] itself."[45] Not only does this make articulating the exact trade secret difficult, but it also brings about issues of transparency, particularly as AI is increasingly used by government agencies in their decision-making processes.[46]

---

[42] 18 U.S.C. § 1839(3)(A).

[43] *Supra* note 1.

[44] Various Contributors, *Artificial Intelligence and Your Space Business: A Guide for Smart Navigation of the Challenges Ahead*, Hogan Lovells (February 2018) (*available at* https://www.hoganlovells.com/~/media/ai-article-space-09nl.pdf).

[45] Will Knight, *The Dark Secret at the Heart of AI*, MIT Tech. Rev. (Apr. 11, 2017) (*available at* https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/).

[46] Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, Univ. Cal. Davis 420 (2017).

Indeed, "banks, the military, employers, and others are now turning their attention to more complex machine-learning approaches that could make automated decision-making altogether inscrutable."[47] This issue is exacerbated by the gap in the AI system's creator or users' knowledge of the AI system and the analyses performed in the black box of the AI system.[48]

"As machine intelligence becomes more ubiquitous, there will be systems controlling online services and those that have an impact on the physical world." Those decisions may entail procedural due process rights to know: (1) whether a decision was made by an AI system; and (2) how the AI system came to its conclusion.[49]

Further, a lack of transparency makes it more likely that "biases . . . can slip through"[50] and there are serious challenges to redesigning AI systems if those biases are eventually identified.[51] As the AI Now Institute states, "[p]rocedural due process arguments [are] particularly adept at dismantling claims that these algorithmic decision systems are proprietary trade secrets, especially when the public's right to know . . . is a strong case against trade secrecy."[52] Further, the AI Now Institute advocates "[h]ighlighting the conflict between procedural due process rights and claims of proprietary or trade secrecy status for these systems up front, so that decision makers understand the inherent tension and incompatibility between them."[53]

Finally, the proliferation and rapid advancement of AI and machine learning begs the question: "[d]oes a human have to be involved to develop or utilize a trade secret?"[54] That question is not anticipated by the DTSA (nor any other U.S. laws relating to trade secrets) and presently remains unanswered.

---

[47] *Supra* note 8.

[48] Matt Burgess, *Holding AI to account: Will Algorithms Ever Be Free from Bias if They're Created by Humans?* Wired (Jan. 11, 2016) (*available at* https://www.wired.co.uk/article/creating-transparent-ai-algorithms-machine-learning).

[49] *Ibid.*

[50] Katyanna Quach, *US Government Use of AI Is Shoddy and Failing Citizens – Because No One Knows How it Works*, The Register (Sep. 26, 2018) (*available at* https://www.theregister.co.uk/2018/09/26/us_government_algorithms/).

[51] Various Contributors, *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems*, AI Now Institute (2018) (*available at* https://ainowinstitute.org/litigatingalgorithms.pdf).

[52] *Ibid.*

[53] *Ibid* at 17.

[54] Amanda G. Ciccatelli, *Making a Federal Case out of Trade Secrets*, IP Watchdog (Sep. 1, 2017) (*available at* https://www.ipwatchdog.com/2017/09/01/making-federal-case-trade-secrets/id=87234/).

XI.   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?

Different aspects of AI technologies can be protected through a variety of different intellectual property rights. These rights are not mutually exclusive. Generally speaking, aspects of AI inventions that cannot easily be reverse engineered or independently developed easily are best suited for trade secret protection over other forms of IP; however, the spectrum of inventions that fall into these categories may shift over time. Algorithms, standing alone, are generally ineligible for patent or copyright protection.

Under the current regime, trade secret laws could be useful for protecting certain elements of artificially intelligent systems, such as "neural networks, including modular network structure and individual modules; training sets, data output, and other data; software including underlying AI code and AI-generated code; and learning, backpropagation, and other algorithms."[55]   There are benefits to protecting AI innovations and the related underlying machine learning data as trade secrets as opposed to patents, such as preventing disclosure of the underlying algorithm or AI into the public domain, and allowing for protection for a potentially unlimited duration, as opposed to the time-limited scope of other forms of intellectual property.[56] However, relying solely on trade secret protection could tip the balance too far away from disclosure and thereby stifle innovation.

The increasing importance of AI technology highlights the need to further analyze whether and how to protect data, including a *sui generis* form of data protection is warranted.

XII. Are there any other AI-related issues pertinent to intellectual property rights (other than those related to patent rights) that the USPTO should examine?

Many scholars over the last two decades have argued that privacy is an intellectual property right and more recently argued that personal data sets should likewise be protectable as an intellectual property right.[57] In light of the passage of the European Union's General Data Protection Regulation (GDPR)[58] and the more recent passage of the California Consumer Privacy Act (CCPA),[59] the accountability of controllers of training data and their efforts to

---

[55] *Supra* note 1.

[56] 35 U.S.C. § 112 ("[t]he specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains…to make and use the same…").

[57] *See* Pamela Samuelson, "Privacy as Intellectual Property," 52 Stanford Law Review 1125 (2000), Trakman, L., Walters, R. & Zeller, "Is Privacy and Personal Data Set to Become the New Intellectual Property? 50 International Review of Intellectual Property and Competition Law 937 (2019).

[58] General Data Protection Regulation (EU) 2016/679 (GDPR)

[59] California Consumer Privacy Act of 2018, AB-375 (2017–2018 Session)

provide reasonable security of the personal data has become a global concern. One approach could be *sui generis* statutory protection of AI training data containing personal data.

XIII. Are there any relevant policies or practices from intellectual property agencies or legal systems in other countries that may help inform USPTO's policies and practices regarding intellectual property rights (other than those related to patent rights)?

Our response to the above question is provided under the following two topic headings Copyrights and Trade Secrets which will attempt to provide insight into how IP5 jurisdictions as well as standards bodies have developed (or are in the process of developing) policies and practices in response to the impact of artificial intelligence innovation. We did not address the topic of Trademark in response to this answer and refer back to our response to Question No. VII and VIII, *supra*.

A. Copyright

Below we provide recent significant copyright-related developments in other IP5 jurisdictions in response to the above question. Further, we direct the USPTO to a document we refer to as the "AI Copyright Primer" which we collaborative created with the Association Internationale pour la Protection de la Propriété Intellectuelle (AIPPI) and the International Federation of Intellectual Property Attorneys (FICPI).[60] The AI Copyright Primer examines how the IP5 examines AI copyright issues in two distinct areas; (1) potential IP issues derived during the AI algorithm's learning phase ("AI System Training") and (2) potential IP issues incurred when the AI system is in the "wild" and will produce anticipated output ("AI System Results").

1. The Chinese Approach

In April 2019, Beijing Internet Court in *Beijing Feilin Law Firm v. Beijing Baidu Wangxun Technology Co., Ltd.*,[61] issued a case of first impression on the issue of copyright protection of AI-generated content. More specifically, the Court addressed the question of whether artificial intelligence (AI) generated content is copyrightable and who is the author of such content. The plaintiff brought the case against Baidu claiming Baidu infringed the law firm's copyright article in Wolters Kluwer's legal information database. Similar to U.S. copyright law, the Beijing Internet Court found that the AI-generated content not protectable under the Chinese Copyright law because a natural person did not create the work. However, the Court did consider that the auto-generated work has originality and acknowledge that the investment in the generation of the production of the AI-generated content deserves some sort of protection.

---

[60] *See* https://ficpi.org/_/uploads/files/AIPLA-AIPPI-FICPI_Artificial_Intelligence_Colloquium_Copyright_ONLY_Primer.pdf (Visited November 20, 2019)

[61] *See* (2018) Beijing 0491 No. 239, Beijing Feilin Law Firm v. Beijing Baidu Wangxun Technology Co., Ltd. in a civil judgment on copyright infringement dispute (April 26, 2019) https://www.bjinternetcourt.gov.cn/cac/zw/1556272978673.html (Visited November 25, 2019)

A recent blog post on IPKAT[62] commented as follow:

> Even if the court denied the copyright protection could be available to purely AI-generated works, it also accepted that some sort of protection should be nonetheless available, on consideration that: (1) a work of this kind represents the investment of the software developer or the software user; (2) the content has the value of dissemination; (3) if there is no protection for the investment, the distribution of the content would be negatively affected. Regrettably, the court did not indicate what kind of protection should be granted. A neighbouring (sic) rights regime? A *sui generis* scheme? The judgment is silent in this regard.

In other words, at least one Chinese court is signaling a need for the PRC to create an alternative approach to the protection of software developers invest of time and sending a message that the current law does not encourage dissemination of AI-generated content. Whether that effort includes amending that jurisdictions copyright law or creating a new form of protection similar to the Japanese "protected data scheme" (discussed below with respect to trade secret policies) remains yet to be determined.

2. The European Approach

On April 17, 2019, the European Parliament and European Council enacted the "Directive on Copyright in the Digital Single Market" (Directive (EU) 2019/790)[63] [hereinafter "Copyright Directive"] to ensure "a well-functioning marketplace for copyright."[64] Article 3 "exempts acts of reproduction and extractions made by research organizations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access"[65] The Copyright Directive defines the term *text* and *data mining* (TDM) as "any automated analytical technique aimed at analyzing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations"[66] The definition of TDM also includes the "automated computational analysis of information in digital form, such as text, sounds, images or data"[67] In other words, AI training data sets would be considered TDM under the Copyright Directive. Article 4 of the Copyright Directive does in limited circumstances allow

---

[62] *See* November 9, 2019 post by Kan He "Feilin v. Baidu: Beijing Internet Court tackles protection of AI/software-generated work and holds that copyright only vests in works by human authors," http://ipkitten.blogspot.com/2019/11/feilin-v-baidu-beijing-internet-court.html (Visited November 24, 2019)

[63] *See* https://eur-lex.europa.eu/eli/dir/2019/790/oj.

[64] *See* Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on copyright in the Digital Single Market COM/2016/0593 final - 2016/0280 (COD).

[65] *Id.* Article 3.

[66] *Id.* Article 2 (2).

[67] *See* Recital (8).

orphan works if the TDM utilizes content in the public domain. If not then the TDM developer would likely need to, for example, approach a collective copyright management organization representing the creators of a particular category of work to obtain a license. Article 3 and Article 4 of Directive (EU) 2019/790 were controversial[68] and the subject of TDM was the subject of previous extensive study[69] by a consortium academics such as the University of Amsterdam and other industry interest group. It is unclear if a commercial entity could somehow bypass the copyright holders by strategically partner with a European research organization or cultural heritage institution.[70]

The implementation of European approach and subsequent decisional law of the Court of Justice of the European Union (CJEU) should be closely followed by US companies who deploy AI systems employing copyrighted training data.

3. The Japanese Approach

In contrast to the European Union approach discussed above, the Japanese National Diet amended its Copyright Act[71] in early 2018 to be "able to respond flexibly to changes in the times such as the advancement of information and communication technology, [and to provide] a right restriction provision with appropriate flexibility for the use of certain works that do not adversely affect the market of works."[72] The amended Japanese Copyright Act which took effect on January 1, 2019, provides an expansion to Japan's exclusive list of situations or types of actions in which third parties may freely use copyrighted works without obtaining the copyright owner's permission (refer to as "rights restriction provisions").[73]

The following three new situations or types of actions were determined by the Japanese Diet to be "less disadvantageous to the right holders"[74] to allow for appropriate flexibility:

---

[68] *See* http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ and http://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/ (Visited December 2, 2019).

[69] *See* Caspers, M.and Guibault, L, "Deliverable D3.3: Baseline report of policies and barriers of TDM in Europe," FutureTDM, https://dare.uva.nl/search?identifier=2e26ec5c-257c-4cfe-96eb-e44355dd2847 (Visited December 2, 2019).

[70] For example, it is unclear at this time if a commercial entity could sponsor research (i.e., design and train a particular AI algorithm or process) at a particular research organization or cultural heritage institution then acquire the finalized AI system thereafter commercializing the effort.

[71] See. Law amending part of the Copyright Law, enacted on May 18, 2018 at the 196th ordinary Diet session, and promulgated as Law No. 30 of 2018 on May 25, 2018, effective January 1, 2019.

English Translation http://cric.or.jp/english/clj/ocl.html (Visited November 22, 2019)

[72] See. http://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/ (Visited November 22, 2019)

[73] See. "Revisions-to-Copyright-Act-and-Unfair-Competition-Prevention-Act," by Hitomi Iwase, November 5, 2018 https://www.internationallawoffice.com/Newsletters/Intellectual-Property/Japan/Nishimura-Asahi/Revisions-to-Copyright-Act-and-Unfair-Competition-Prevention-Act (Visited November 22, 2019)

[74] *Id.*

(1)   Article 30-4 (Use not for the purpose of enjoying thoughts or feelings expressed in a work)

(2)   Article 47-4 (Uses associated with the use of copyrighted works on electronic computers)

(3)   Article 47-5 (Information processing by electronic computers and minor use accompanying the provision of the results)

The Japanese approach demonstrates a balancing of interests in which the harm to copyright owner (i.e., adverse effect the market of their work(s)) is weighed together with the broader national interest of advancing information and communication technology in the era of the Industrial Revolution 4.0.

4.   The South Korean Approach

To our knowledge the South Korea government does not have any current legislative initiatives or new legal precedence with respect to copyright law to help inform USPTO's policies and practices regarding intellectual property rights with respect to artificial intelligence innovation. That said, AI itself can be protected as a computer program under the 1987 Copyright Law of South Korea (§4.9).[75] Similarly, South Korea provides protection of databases defined as compiled matters whose subject matters are systematically arranged or composed, so that they may be individually approached or retrieved.[76]

In November 2018, President Moon formed the Presidential Committee on the Fourth Industrial Revolution (PCFIR) together with South Korea's "I-Korea 4.0" which is marketed as paralleling Japan's Society 5.0.[77] One of the four pillars of I-Korea 4.0 is to make preparations for future society which includes possible changes to regulations and policies affecting AI algorithms and processes to provide flexibility.

5.   Other Jurisdictional Approaches

Numerous other jurisdictions are grappling with determining whether their current Copyright Laws and practice meet the needs of protecting innovation related to artificial intelligence. AIPPI, a non-political association founded in 1897, devoted to harmonizing and improving IP laws around the world adopted a Resolution on "Copyright in Artificially

---

[75] See. THE SOFTWARE INDUSTRY PROMOTION ACT of the Republic of Korea, Act No. 3984 of December 4, 1987 http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN025684.pdf "9. Other matters necessary for developing the software industry."

[76] Article 2-19 of the South Korea Copyright Act

[77] https://www.asiapacific.ca/sites/default/files/filefield/ai_report_2019.pdf

Generated Works," at its Global IP Congress in September, 2019.[78] The AIPPI's Resolution was preceded by an international study project resulting in the creation of national reports covering the state of copyright regimes in light of AI in 32 countries[79] as well as a followed by a Summary Report.[80] Asa Kling formerly Director of the Israel Patent Office who recently hosted an AIPLA Delegations to Israel has provided a summary of Israel's copyright and trade secret regimes in light of artificial intelligence.[81]

       AIPLA's Harmonization Committee recently examined the approaches of numerous jurisdictions producing Table A and Table B below. Table A summarizes the status of copyright protection of AI created by a human programmer as well as output from an AI program in various countries (i.e., "protectability of the created work and protectability of the AI that created the work"). Table A attempts to summarize Table B by identifying where protection is and is not available for the work and the AI creating the work.

---

[78] See. , https://aippi.org/wp-content/uploads/2019/10/Resolution_Copyright_in_artificially_generated_works_English.pdf (Visited December 1, 2019)

[79] which are compiled at https://aippi.org/committee-publications/?committee-id=66327. (Visited December 1, 2019)

[80] https://aippi.org/wp-content/uploads/2019/07/SummaryReport_COPYRIGHT-DATA_London2019_final_160719.pdf.

[81] https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/israel

Table A: Copyright Protection for Human Program and AI Output among various Countries

Country	Protection of AI program created by human?	Protection of work created by AI?
Spain[82]	Possibly yes (not clear on copyrightability of software)	No
Germany[83]	Possibly yes (not clear on copyrightability of software)	No
Australia[84]	Possibly yes (not clear on copyrightability of software)	No
CJEU[85]	Possibly yes (not clear on copyrightability of software)	No
Portugal[86]	Possibly yes (not clear on copyrightability of software)	No
Hong Kong (SAR)[87]	Yes	Likely no
India	Yes	Likely no
Ireland	Yes	Likely no
New Zealand	Yes	Likely no
UK[88]	Yes	Likely no

[82] *See* https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html ("Preamble In turn, within the first set of rules are determined on the one hand, the rights belong to the author, who is who makes the purely human and personal task of creating the work and, therefore, constitute the core of the object of this Act and, secondly, the rights granted to certain natural or legal persons whose participation is essential for performance or for the dissemination of works created by the authors"

[83] *See* https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html.

[84] http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/FCAFC/2012/16.html (Renderings of MSDS (material safety data sheets) via HTML from a database of raw data are not protectable via copyright)

[85] *See* Directive 2001/29/EC (9) relates to "intellectual creation" and what appears to be only present in a person: "(9) Any harmonisation of copyright and related rights must take as a basis a high level of protection, since such rights are crucial to intellectual creation. Their protection helps to ensure the maintenance and development of creativity in the interests of authors, performers, producers, consumers, culture, industry and the public at large. Intellectual property has therefore been recognised as an integral part of property." (Emphasis added). See. Also. Infopaq International A/S v Danske Dagblades (Storing and printing 11 consecutive words of a document (e.g., a newspaper article) is an act of reproduction protectable under Directive 2001/29/EC), Originality must reflect the author's personality (suggesting a human author is necessary for a copyrightable work to exist). The CJEU left national courts of EU to decide whether the 11 word reproduction was the author's expression of intellectual creation. The understanding that is being upheld by CJEU also seems to prevent any similar solution to that of the United Kingdom, since the protection of a work by copyright required a "personal touch" of the author (v. Case C 145/10) as well as a margin for "creative freedom" (v. Case C 604/10).

[86] "Vitor Palmela Fidalgo, *Worldwide: Copyrights Protection Of Works Generated By Artificial Intelligence*, March 18, 2019 http://www.mondaq.com/x/790660/new+technology/Copyrights+Protection+Of+Works+Generated+By+Artificial+Intelligence ("Creation must result from effort of human intellect")

[87] *See* U.K. copyright law, section 9(3) of the Copyright, Designs and Patents Act (CDPA): ("In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.")

[88] In the U.K., the issue appears to focus on addressing question of who is causing the work to be generated. In a litigation relating to similarities of a user interface in a computer game, a court in England held a player's contribution is not significant

B.   Trade Secrets

IP5 jurisdictions have diverse trade secret laws, policies and practices unique to their jurisdiction providing various means of requiring how secrecy of a trade secret is maintained, how to articulate trade secrets as well as how and if training data sets are to be protected under trade secret law. With respect to laws, policies and practices related to the intersect of trade secreted AI algorithms and processes and transparency of the same; it is unclear at this time if jurisdictions will incorporate practical transparency requirements into their trade secret laws, policies or practice or will merely rely upon developing international standards[89] or await technical advances (e.g., explainable AI). That is, transparency of aspects of trade secreted training data for the purposes of identifying and avoiding algorithm bias in the AI algorithm and process as well as providing end users with trusted AI algorithms and processes.

1.   The Chinese Approach

On April 23, 2019 the PRC revised their Anti-Unfair Competition Law to strength trade secret protection (effective May 1, 2019) by (i) defining additional actions (Article 9) that infringe trade secret, (ii) identifying additional infringers beyond business operators to include natural person, legal entity, non-legal entity to be pursued for trade secret infringement; (iii) increasing compensation for trade secret infringement; (iv) enhancing administrative punishment against trade secret infringement; and (v) easing proof of trade secret infringement (Article 32). Despite this monumental change, no real guidance is found in this jurisdiction with respect to informing USPTO's policies and practices regarding intellectual property rights.

That said, in December 2018, Hangzhou Intermediate People's Court in *Anhui Jingmei Information Technology Co., Ltd. v. Taobao (China) Software Co., Ltd.* acknowledged that the data product for commercial sales can bring direct income and constitutes a competitive advantage for Taobao and should be protected by anti-unfair competition laws. The Court concluded that improper acquisition of the data product and getting profits thereby disrupted the market competition constituted unfair competition.

It is prescribed by Article 33 of Law of the People's Republic of China on the Promotion of Small and Medium-Sized Enterprises (2017 Revision) that the State supports small and medium-sized enterprises to apply modern technological means such as the Internet, cloud computing, big data, and artificial intelligence technology in research and development, design, manufacturing, and operation management to innovate production methods and improve production and operation efficiency. Article 33 appears to have a similar intent as the recent change in Korea Trade Secret law discussed below.

---

enough to contribute the generation of a copyrightable work, ("a player's input 'is not artistic in nature and he has contributed no skill or labour of an artistic kind'. So considering user action case by case could be one possible solution to the problem."). See Nova Productions v Mazooma Games [2007] EWCA Civ 219. *See also* https://www.slaughterandmay.com/media/251773/copyright_-_non-literal_software_copyright_in_the_wake_of_nova_v_mazooma.pdf.

[89] *See* "Ethically Aligned Design, Version 2 (EADv2), IEEE (December 2017), IEEE-SA P7001(Transparency of Autonomous Systems) and ISO/IEC JTC 1/SC 42/WG 3 (Trustworthiness).

2.   The European Approach

Directive 2016/943 (hereinafter the "Trade Secrets Directive") was introduced by the European Union's Legislation on June 8, 2016 as an attempt to harmonization trade secret laws among the 48 Member States. Article 2 (1) of the Trade Secrets Directive refers to a trade secret broadly as "information" which (i) is secret; (ii) has commercial value due to its secrecy; and (iii) is subject to reasonable steps to keep it secret.[90] Further, Article 1 (3) excludes "trivial information and the experience and skills gained by employees in the normal course of their employment, and also excludes information which is generally known among, or is readily accessible to, persons within the circles that normally deal with the kind of information in question."[91]

After the expiration of the transposition period (June 9, 2018) defined in Article 19 of the Trade Secrets Directive, Germany enacted its Trade secrets Protection Act. The German Act implemented Article 1(3) of the Directive by apparently incorporated prior German court rulings. That is, according to one German practitioner[92], absent an agreement between an employee and employer an employee in principle can use their experiences and knowledge gained at former employer at their new employer. For example, knowledge memorized in the normal course of work (even if a trade secret of former employer) may be transferred to the new employer. However, systematic memorizing would not be considered memorization in the course of normal work.

The European approach provides a broad enough definition of trade secrets such that artificial intelligence algorithms and processes can be protected. Further, the European approach definitely provides an illustrative example of how companies operating in the EU must proactively provide detailed employee agreements and limit employee's exposure to trade secreted training data and algorithms. The European approach should serve as a best practice guide for US-based companies as well as serve as informing future USPTO's policies and practices regarding intellectual property rights specific to AI innovation where at the present time employee job hopping is common.

3.   The Japanese Approach

The Japanese added a new category of data protection apart from trade secret yet similar in nature in its Unfair Competition Prevention Act (UCPA) adopted in May 2018 in order to improve utilization of "data."[93] That is, the revised UCPA took into account the reality that "useful data is easy to duplicate and, once illegally obtained, can be immediately redistributed

---

[90] Directive 2016/943 ("Trade Secrets Directive") https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0943&rid=2 (Visited November 24, 2019).

[91] *See id*. Recitation (14)

[92] *See* 2019 IPO Annual Meeting Presentation "Trade Secrets Trends – Comparative Law in the U.S., China, and Europe," Dr. Soenke Fock, Slide 18.

[93] *See* https://www.meti.go.jp/english/policy/economy/chizai/chiteki/index.html (visited November 23, 2019).

in its entirety, meaning that investors may lose the chance to recoup their investments."[94] Accordingly revised Article 2 (7) of the UCPA provides a new form of protection for data defined as follows:

> The term "protected data" as used in this Act means technical or business information that is accumulated in a reasonable amount by electronic or magnetic means (meaning an electronic form, magnetic form, or any other form that is impossible to perceive through the human senses alone; the same applies in the following paragraph) as information provided to specified persons on a regular basis and that is managed (excluding information that is kept secret).[95]

According to the "Overview Guidelines on Protected Data" provided by the METI (Ministry of Economy, Trade and Industry), this definition implies three requirements: (1) being technically controlled using, for example, an ID and a password, (2) having limited availability to third parties, and (3) having utility.

Available remedies include injunctions and damages as with other types of acts of unfair competition. These new provisions represent an entirely new type of protection for big data which may not be new or secret. How they are going to be implemented is another issue that we will have to deal with. For example, in a real-world setting, it is likely that data that fall in one of the above new categories may be mixed with data that are not protectable under the new provisions, and they may not be easily separable. Such a situation will certainly pose a significant challenge toward effective enforcement.

That said, the Japanese approach is worth investigating and could serve as informing future USPTO's policies and practices regarding intellectual property rights specific to training data sets for Artificial intelligence innovation.

4. The South Korean Approach

In July 2018, South Korea relaxed its trade secret law defined in their Unfair Competition Prevention and Trade Secret Protection Act ("UPCA"). That is, under the newly amended UPCA, there is no longer a need to take "reasonable efforts" to maintain the secrecy of the information. The purpose of this amendment was to significantly lowers the threshold for information to become a trade secret so as to provide more protection for small and mid-sized businesses that are typically not sophisticated (or undercapitalized) to meet the maintenance standard required under the former UPCA. In other words, so long as the training data is maintained as secret, there is no need to evaluate whether or not the measures taken were reasonable. However, this is approach is unlikely to be implemented in the US.[96]

---

[94] *See* Guidelines on Protection Data, Japanese Ministry of Economy, Trade and Industry, page 2 (January 23, 2019).

[95] *See* Unfair Competition Prevention Act (UCPA) adopted in May 2018, Article 2(7).

[96] *See* "What's Reasonable? – Protecting and Enforcing Trade Secrets in the Digital Age" AIPLA 2016 Spring Meeting, Jeanne M. Gills (Foley & Lardner),

5.   Other Jurisdictional Approaches

Numerous other jurisdictions are grappling with determining whether their current trade secret laws and practice meet the needs of protecting innovation related to artificial intelligence.

6.   Agreement on Trade-related Aspects of Intellectual Property Rights (TRIPS)

Output of an AI algorithm that is considered a trade secret is likely already protectable, both in the US and abroad under article 39 of the 1995 TRIPS (Trade-Related Aspects of Intellectual Property Rights) Agreement. That is, it can be argued that a trade secret can be almost anything that provides a "competitive advantage," something created by AI, whether it involves any creativity or not, is protectable. Article 39 provides:

> (1) In the course of ensuring effective protection against unfair competition as provided in Article 10bis of the Paris Convention (1967), Members shall protect undisclosed information in accordance with paragraph 2 and data submitted to governments or governmental agencies in accordance with paragraph 3. (2) Natural and legal persons shall have the possibility of preventing information lawfully within their control from being disclosed to, acquired by, or used by others without their consent in a manner contrary to honest commercial practices 1 so long as such information: (a) is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) has commercial value because it is secret; and (c) has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret. (3) Members, when requiring, as a condition of approving the marketing of pharmaceutical or of agricultural chemical products which utilize new chemical entities, the submission of undisclosed test or other data, the origination of which involves a considerable effort, shall protect such data against unfair commercial use. In addition, Members shall protect such data against disclosure, except where necessary to protect the public, or unless steps are taken to ensure that the data are protected against unfair commercial use. 1. For the purpose of this provision, "a manner contrary to honest commercial practices" shall mean at least practices such as breach of contract, breach of confidence and inducement to breach, and includes the acquisition of undisclosed information by third parties who knew, or were grossly negligent in failing to know, that such practices were involved in the acquisition.

---

https://www.foley.com/files/uploads/AIPLA%20Article%20on%20DTSA%20and%20Reasonable%20Efforts%20to%20Protect%20Trade%20Secrets%2048.pdf

We thank you for the opportunity to provide these comments. AIPLA supports the USPTO's efforts on improving the patent system, welcomes the opportunity to answer any questions these comments may raise, and looks forward to a continuing dialogue on this very important subject.

Sincerely,

Barbara A. Fiacco
President
American Intellectual Property Law Association