

**fThe AIPLA/AIPPI/FICPI AI Colloquium Primer**  
**By Various Contributors<sup>1</sup>**

**V. AI Impact on the practice of Copyright Law<sup>2</sup>**

The copyright section will be broken up into two distinct areas, one for the potential IP issues derived during the learning phase (“AI System Training”) and the other dealing with IP issues when the AI system is in the “wild” and will produce anticipated output (“AI System Results”). The term “AI system” will be meant to include a computerized technology (i.e., algorithms and other computer software) with a purpose of independently performing an assigned task without human interaction, including the use of Big Data, Machine Learning and other AI tools.

**A. AI System Training**

In the *United States*, during the AI Learning phase, an AI system relies on copious input of information to “learn”, which will be called the “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 can include a large amount of art<sup>3</sup> or pictures, which may be (and typically are) obtained from the internet through bots. In other AI systems, music, formulas, chemical compounds, designs and even architecture may be obtained in the same manner.

In some cases, AI systems may write their own code during the AI Learning phase. But to do so, these AI systems generally scour the internet for other sources of code<sup>4</sup>.

More of the creative AI systems employ a GAN (“generative adversarial network”) system<sup>5</sup> 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 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.

In *Europe*, EU copyright law relating to AI is governed mainly by the InfoSoc Directive<sup>6</sup> and the Computer Software Directive<sup>7</sup>, wherein both Directives are implemented into national law by all the EU Member States. In terms of AI, the EU legislation has an impact on both the protection of the AI itself, as well as on works created by AI.

In *Japan*, for more than fifteen years, committees inside and outside of government have extensively discussed the possibility of having general provisions to restrict the rights provided by copyrights and allow free use of copyrighted works without approvals from copyright holders in certain circumstances, mechanisms known as “fair use” in the U.S. and as “fair dealing” in the U.K. Currently, the Copyright Act of Japan includes a lengthy and ever-extending list of what may be excluded from copyright protection. With the rapid emergence of new technologies, such as IoT and AI, new items needed to be added to the list one by one, but the statutory

amendments always lagged behind technological progress. The sharp differences of opinion between right holders and users have made it impossible to reach agreement on plausible general clauses with limiting the rights provided by copyrights.

Therefore, the amendments to the Copyright Act that became a law in 2018<sup>8</sup> were one step forward in an attempt to provide some room in which innovators can work without infringing copyrights, and strike a balance between innovators and right holders.

The amendments included three sets of relatively modest provisions that allow such free use and make flexible interpretations possible in view of innovative digital or AI services that are emerging and will emerge in the future. For example, without authorization of the right holder, the use of a work will be allowed if such use does not adversely affect the market value of the work. Also, the use of big data will be allowed for such services as Google Books and information analysis for detecting plagiarism among academic or school papers and displaying copied portions of the original.

In *China*, the current copyright law was enacted nearly 30 years ago. It originally addressed issues in connection with “normal” human works. However, with the development of AI, new issues have emerged.

#### **ISSUE #1-1 – Are there potentials for infringing use of copyrighted material or is this considered to be fair use?**

During this process, an AI system (typically the discriminator network) must make copies of that input, manipulate that material and create an intermediary dataset or database for the generator network.<sup>9</sup> In other cases, AI needs to have the input data “Labeled<sup>10</sup>.” In some cases the labeling is done manually, by humans<sup>11</sup> or through asking humans<sup>12</sup>, 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.

#### ***United States:***

Given the large amount of information required by an AI system to do its “work,” obtaining authorization from what literally could be hundreds of thousands of copyright owners would be an insurmountable hurdle. If the input is unstructured factual data or 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 unless authorized or excused by an exception. To avoid liability, the copying most likely would make use of the fair use exception. Fair use is a holistic case-by-case defense, so the answer depends upon the facts of each individual case. This would probably require a contextual, holistic analysis of four non-exclusive factors, which are discussed in more detail below.

There are no infringement cases involving AI input, so this discussion must look to analogous cases and assume certain facts. For instance, assume an AI system has to provide real-time recognition and understanding of objects it would encounter in the real

world – and so it inputs millions of photographs from many sources without authorization. An analogous case is Google Books. There, 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. In an opinion written by Judge Pierre Leval, the US Court of Appeals for the Second Circuit (the “Second Circuit”) found the use permitted under the fair use exception.<sup>13</sup>

**First Factor.** The purpose and character of the use. The Supreme Court has noted this factor favors secondary uses that are transformative, meaning the use 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.” Google Books quoted Campbell<sup>14</sup>, noting that “the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works.” The Second Circuit easily found Google’s search function to be “highly transformative.” Bearing this in mind, assume copyrighted images are not being used for their artistic value, but ultimately to teach the AI system to recognize objects in the real world. It is highly likely this would be considered a transformative use, and thus the first factor would favor fair use.

**Second Factor.** The nature of the copyrighted work. As the court said in Google Books, the second factor – the nature of the copyrighted work – “has rarely played a significant role in the determination of a fair use dispute.” Although the photographs are likely to be somewhat creative, given the transformative nature of the AI system’s use of them, the second factor is likely to be ignored or discounted.

**Third Factor.** The amount taken. Despite the fact that the AI system would be ingesting entire photographs, it is likely not to weigh against fair use. “[C]ourts have rejected any categorical rule that a copying of the entirety cannot be a fair use.” Transformative uses are permitted to take as much as is necessary to effect the use.

**Fourth Factor:** Effect on the market. This has often been called the most important factor, or at least, with the first factor, one of the two most important factors. The assumption here is that there is no collection of licensable photographs sufficient to create the AI data set. In Campbell, the Supreme Court noted the linkage between the first and fourth factor, i.e., “the more the copying is done to achieve a purpose that differs from the purpose of the original, the less likely it is that the copy will serve as a satisfactory substitute for the original.” Use of the photo dataset did not substitute for the purpose of the original photographs. Nor would any licensing revenue 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 likely that the use in creating a robust dataset would be a fair use under §107 of the Copyright Act.<sup>15</sup>

### ***China:***

In most cases, data used as input for AI systems is collected from other resources. The collection of data from other resources may infringe the copyright of the original owner if the use of data does not fall into the category of “fair use” according to article 22 of the copyright law.<sup>16</sup>

As to the raw material, it *per se* may involve copyright, such as a copyright of a drawing or a photo. According to the current Copyright Law, the use of this kind of raw material may infringe its copyright. Some professors and experts deem that the use of the raw material is not the original purpose of the raw material. Actually, the AI system does not care about the aesthetics of a photo or a drawing. Its use by an AI system is a secondary use and thus will not infringe a copyright. However, this secondary use opinion has not been extensively accepted.

### ***Korea:***

Data input for AI can be protected by copyright law. Copyright law protects the rights of the database producer. The rights of database producers are not protected by copyright but by similar rights as a database producer. The law states that database producers shall hold the rights to reproduce, distribute, broadcast, or transmit the whole or considerable parts of the relevant database. The term “database” means the compiled matters whose subject matters are systematically arranged or composed, so that they may be individually approached or retrieved.

## **ISSUE #1-2 – Is the intermediary database capable of having copyright protection and who is the owner?**

The AI System itself, whether or not it is a GAN system, is written as a computer program and executed on a system resembling a computer (it has a CPU and memory). It may or may not have displays or other connections. Therefore, it will be presumed that the entire AI system of program code and/or instructions is essentially computer software.

### ***United States:***

The intermediary database, like any other database, can receive copyright protection as a compilation so long as its selection and arrangement of input is minimally creative. As noted above, 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 very

valuable to the AI system provider. The impact of copyright law, however, is that it provides very little protection to this valuable AI asset. Other than the selection and arrangement of the works ingested, very little creative expression is contained in a non-expressive dataset. Despite the “sweat of the brow” effort to create the dataset, it is a combination 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.<sup>17</sup> Recently, in a case analogous to AI, the Ninth Circuit found *some* copyright protection for a database of customer information.<sup>18</sup> Thus, absent a change in the law, copyright provides little protection to the valuable AI dataset.<sup>19</sup>

Although authorship would need to be examined in light of the facts, and although the issue has yet to receive enough attention, it seems likely that the author of the database would be the party responsible for coding or otherwise training the AI system to select and arrange the data that eventually makes its way into the resulting database. See generally J. Ginsburg & L. Budiardjo, *Authors and Machines* (August 5, 2018 Submission Draft) (arguing that the programmer of “fully generative” machines would be the author of the output)<sup>20</sup>; *Digital Drilling Data Systems LLC v. Petrolink Services, Inc.*, 2018 WL 2267139 (S.D. Tex. May 16, 2018) (copyright owner of software program designed to make database schema was author of resulting schema).<sup>21</sup>

#### ***Europe:***

Databases used as input for the computer programs can also be protected by copyright. The same criteria applicable for other works are also applicable for AI technology.

#### ***China:***

Data used as input for the AI computer programs can also be protected by copyright, as long as they are original and reproducible according to rule 2 of “Regulation for the Implementation of the Copyright Law (2013)”.<sup>22</sup> As to the raw materials and the labelled materials, the Copyright Law and its Implementing Regulations will be applied directly.

#### ***Korea:***

The dataset for building the AI system is protected only if it falls under the copyright laws. The current AI dataset is not protected under other laws. The term “producer of database” means one who has made human or physical investments considerably in the production of the database, or in the renewal, verification or supplement of their subject matters. Therefore, if a large amount of manpower and time is spent in collecting data, it can be protected by a database under copyright law, but the level of protection is much lower than copyright.

## **ISSUE #1-3 – Is the AI system code used for training capable of protection by copyright?**

### ***United States:***

AI does not add or subtract anything to an already complex and unsettled area of copyright: the extent of protection afforded algorithms. Although copyright protection may extend to the way in which a process is expressed, it does not extend to the underlying process itself. This principle – called the idea/expression dichotomy – can be difficult to apply. That is true not only as a general matter, but especially in the context of computer programming. Without specifically examining an algorithm used in an AI system, it is difficult (if not impossible) to say whether the expression of that algorithm can receive copyright protection. In principle, however, AI algorithms could receive copyright protection, just like other any other computer algorithm or program. Indeed, as one court recently noted, “a set of commands to instruct a computer to carry out desired operations may contain expression that is eligible for copyright protection,” so long as the expression of that process is original and “the [programmer] had multiple ways to express the underlying idea.”<sup>23</sup>

So it seems likely the code is copyrightable if the commands for expressing the underlying functions can be expressed in at least several different ways, thus ensuring that the programmer would not receive what would amount to a monopoly.

### ***Europe:***

Before assessing possible protection for AI, we must agree on what AI is. If we, for the purposes of this discussion, can agree that AI is a computerized technology (i.e., algorithms and other computer software) with a purpose of independently performing an assigned task without human interaction, then AI may qualify for copyright protection under the Computer Software Directive.<sup>24</sup>

In terms of the originality criteria, it is important to keep in mind Article 9(2) of TRIPS, which also applies to EU copyright law<sup>25</sup>. Computer programs that are too simplistic and only reproduce purely functional factors can therefore be disqualified from protection.

As such, protection for AI itself can be afforded under copyright law to the author of the computer software.

### ***China:***

As to AI algorithms, when program codes are involved, the copyright is created. The related law is “Regulation on Computers Software Protection (2013).”<sup>26</sup>

***Korea:***

In Korea, 'THE INTELLIGENT ROBOTS DEVELOPMENT AND DISTRIBUTION PROMOTION ACT'<sup>27</sup> and 'THE SOFTWARE INDUSTRY PROMOTION ACT'<sup>28</sup> have been implemented to promote the development of science and technology related to AI and to develop related industries. However, there is no law that defines AI clearly or defines legal problems. Therefore, if AI itself corresponds to software, it will be judged as a computer program as defined in 'THE SOFTWARE INDUSTRY PROMOTION ACT' or 'Copyright Act', and it will receive the corresponding level of protection. Therefore, AI itself can be protected as a computer program under the copyright law of Korea (§4.9).<sup>29</sup>

**B. AI System Results**

In the aforementioned case of the AI system actually producing its own unique code<sup>30</sup>, the same issue arises, but for the actual software that is being created. While in many jurisdictions the answer may be the same as Issue #1-3, it is relevant to at least make certain there is no distinction in any jurisdiction.

**ISSUE #2-1 – Is the AI system generated code capable of receiving copyright protection and who is the owner?**

When the AI system produces results in some form of output, i.e., any form of media, a copyright protection decision may depend upon whether a “natural person” was involved, and if so, what that person’s involvement was in producing the output.

For the first case, the presumption is that the natural person is utilizing the AI system as a tool to help create the output.

***United States:***

Although this issue has yet to be addressed in the case law, some scholars have written articles suggesting that code created by the AI system should be copyrightable if the system was coded and trained to create output to that effect by a human programmer, who would be the author<sup>31</sup>. A more difficult question that may need to be addressed soon is whether the code (or other output) is copyrightable if its content is unexpected from the programmer’s perspective.

***China:***

Yes, if the code output satisfies the definition of copyright law, such as being original and reproducible, according to rule 2 of “Regulation for the Implementation of the Copyright Law (2013)”<sup>32</sup>. However, it may still have an issue of ownership under current law.

## **ISSUE #2-2 – Is the AI system generated output capable of having copyright protection and who is the owner?**

### ***United States:***

This should ideally be settled at the outset in an agreement between the AI system service provider and the entity requesting the AI service. If there is no agreement, then precedent suggests the programmer of the AI system is the author only if the AI system does the lion's share of the creative work. If not, then a court might find the human user is the author.

While there is no case precisely on point, the issue has arisen in analogous cases where there was no agreement. In one case, the Plaintiff's equipment and software used by the studios' contractor "to provide facial performance capture services and output files made with Plaintiff's system."<sup>33</sup> The contractor used the system without authorization. Plaintiff argued his company and not the studios owned the copyright to the output. The Judge cited the decision of the Ninth Circuit in *Design Data Corp. v. Unigate Enter., Inc.*<sup>34</sup> Pointing to all the actors' creative input, the judge held Plaintiff "has not alleged that the program 'does the lion's share of the work' ... or that the user's input is marginal," and dismissed the copyright claims.

### ***Japan:***

Copyright protection may be available for works created by application of AI technology, i.e., a new piece of artistic work such as a painting, music, a poem or a novel can be created by AI technology utilizing a database of the existing artistic works. In this case, a key issue is to decide who is the "author" that created the work (Copyright Act Article 2 (1)(ii)).<sup>35</sup> The decision depends on whether the case would fall in Situation (A) where a natural person instructed AI to create a piece of work and AI created the piece of work; or Situation (B) where a natural person utilized AI as a tool to create a piece of work intending to create it and contributing to its creation.

In Situation (A), the natural person is not entitled to a copyright because the person is a mere instructor and did not contribute to the creation of the work; and because the "work" must a production in which thoughts or sentiments are expressed in a creative way (Copyright Act Article 2 (1)(i)),<sup>36</sup> but AI is not a natural person and therefore AI cannot express thought or sentiments.

In Situation (B), a natural person can be entitled to a copyright since the person had his/her own intention of creation when using AI as a tool and contributed to the creation of the work during the production, for instance, by selecting one of multiple results or editing the result.

However, an issue may arise since the work created in Situation (A) and the work created in Situation (B) cannot be distinguished from each other. In addition, AI itself would automatically create works using big data without any instruction from a natural person.

***China:***

Yes, if the media output satisfies the definition of copyright law, such as being original and reproducible, according to rule 2 of “Regulation for the Implementation of the Copyright Law (2013)”<sup>37</sup>. However, the ownership of the media output is an issue, because it is created by running an AI algorithm on a database. The majority of the opinions in China deem the entity who creates the AI algorithm the author and owner of the media output based on the current AI technology because it is still used mainly to process the database and then send an output. However, this concept may need to be reconsidered in the near future.

**ISSUE #2-3 – Who would be responsible for potential infringement of the AI system generated output?**

For the second case, the presumption is that the natural person may have created or instructed the AI system, but was not involved after setting up the code and the training. For the infringement question, assume you are a company selling an AI system that generates output for a client that is found to be infringing the rights of a third party.

***United States:***

This may be unlikely to arise in many AI applications in that the output is not expressive. The output would likely be instructions to a device or a solution to a problem. As to expressive output, it would be a case-by-case determination.<sup>38</sup> As noted above, only the expression is protected, not the idea expressed.<sup>39</sup> One could imagine that the expressive AI system’s algorithms would be programmed to prevent any substantial similarity between the works input and the works output.

However, if we assume that the output is expressive, then existing precedent suggests it should depend on whether the program is responsible for the “lion’s share” of the work, as noted above. Caveat: this could become more complex if infringement is no longer tethered to a volitional act (see the majority and dissenting Supreme Court opinions in *Aereo*)<sup>40</sup>. However, almost every court faced with the issue still requires a volitional act in order to find infringement.

***Korea:***

While various arguments can be made, it is generally accepted that the person who uses AI as a "tool" has the rights to the creations and is responsible for the infringement of AI

intellectual property rights. Under the present law, only natural persons have such rights. This is similar to holding a pet owner liable if their pet has created an “accident.”

## **ISSUE #2-4 – Is the AI system generated output capable of having copyright protection and who is the owner?**

### ***United States:***

See #2-2 above, with the caveat that there is, so far, an academic discussion of whether there can be expressive AI output that would not be copyrightable – “emergent output.” While there are several definitions, one is “works of apparently creative expression that arise from the operation of a program but cannot be traced directly to a human source.”<sup>41</sup> This is a topic for further exploration, but is not an issue that has emerged as a problem so far.

Also, if there is no human apart from the programmer involved in the AI system’s creation of the output, existing commentary suggests the programmer is the author of the output so long as he coded or otherwise trained the AI system to create that output. If there is another human apart from the programmer who is involved in the creative process, like a user of the AI system, then the answer, as noted above in section #2-2, would depend on whether the AI system does the lion’s share of the work. If so, then the programmer may be the author. If not, then existing precedent suggests that the user of the program may be the author.

### ***Europe:***

EU copyright law, as stipulated in the above directives, has a highly humancentric approach in terms of what should be protected and who should benefit from such protection. This causes a number of difficulties when dealing with new types of works and creators.

When dealing with works created by AI, the humancentric approach of EU copyright law causes immediate obstacles. As stated above, copyright is only afforded where the work is original in the sense that it is the author’s own intellectual creation. This has been clarified by the EU Court of Justice as “an intellectual creation is an author’s own if it reflects the author’s personality” and the author’s “personal touch”<sup>42</sup>. Even though some may argue that some AI, for instance Amazon’s Alexa™ or Apple’s Siri™, has a personality of its own, it is highly doubtful that a court would consider works created by AI as having a “personal touch” in the traditional sense.

In fact, EU copyright law in many ways suggests that works created by AI are not eligible for copyright protection. Baseline rules such as the term of copyright protection being 70 years from the death of the author, to the author’s right to be named and not to have the work infringed are not adaptable to an AI counterpart. How would an AI

technology be able to enforce possible copyright infringements or enter into license agreements?

In the UK, for example, the above problem has been handled by giving authorship of a computer-generated work to the programmer. The effort made to program the AI is therefore rewarded with ownership of the work, even though the programmer's own personality or personal touch may not be reflected at all in the work. This solution is not without difficulties. First, who should really be rewarded with ownership of the works created by AI? Is it always the actual programmer or could the work be created by the AI based on user interaction? In many cases where AI is used to enhance a natural person's creation, it may seem more reasonable to award that natural person the full copyright to the creation. Second, should the programmer (and the employer?) then also be responsible for any infringement made by the AI in other people's copyright through the creation process? It is not inconceivable that the AI creates its work through the analysis and processing of data owned by someone else.

Therefore, it is possible that works created by AI in the EU in general would not be considered protectable under EU copyright law under the current legislation. As such works are becoming more and more sophisticated, a commercial need for a legal protection for such works will soon be needed to secure investments and developments. The UK solution may seem the most convenient and would help to create comfort to investors. However, as seen above, it is not without complications and there are still many unanswered questions.

### ***Japan:***

AI technology itself that produces copyrightable works cannot be protected by the Copyright Act, since the copyrightable "work" means a production in which thoughts or sentiments are expressed in a creative way, and which falls within the literary, scientific, artistic or musical domain (Copyright Act Article 2 (1)(i)).<sup>43</sup>

### ***China:***

Current copyright law requires that the author of works include Chinese citizens, legal persons or entities without legal personality and the works shall belong to the author if there is no exception. It excludes a machine that creates a new work (media output in this case) from being the owner of the copyright of the new work. In the recent pending case Beijing Film Law Firm v. Baidu Company before the Beijing Internet Court,<sup>44</sup> this issue was extensively discussed. Some of scholars have the opinion that certain parts of the rights of author, such as the right to claim authorship and to have the author's name indicated on the works, shall belong to the machine to reflect the fact of who completed the work. However, other rights, such as property rights, shall belong to the person or entity who actually controls the machine.

There is considerable uncertainty in this area, and thus more thoughtful consideration will be needed in the near future as more AI implementations are used to create works.

***Korea:***

Protection of copyright is a creative work expressing human thoughts and emotions.<sup>45</sup> Therefore, under the current law of the Republic of Korea, work by AI, i.e., not by a human, cannot be protected by copyright. Since it is doubtful whether the expression of AI corresponds to the expression of thoughts and emotions, it will be difficult to protect AI work by copyright. However, today, AI is studying human pictures and performances to create new works. Furthermore, it is expected that AI's creative activities will be done in various fields such as writing novels, making movies and designing clothes. Therefore, there is a need to protect AI's copyrightable works with copyright or other legal provisions. It can be divided into strong AI and weak AI. In reality, AI is being developed to reduce the efforts of humans in creative work, and the AI itself is becoming more and more creative, and will continue to do so in the future. The main problem is the latter, i.e., strong artificial intelligence. Under current copyright law it is not possible to determine the owner or discuss the protection of AI-related creations that are beyond human capacity to create. Therefore, new legal provisions (modified copyright law, trade secret protection law, etc.) will be required instead of the existing copyright law. In this case, it will be necessary to discuss the issue of responsibility for the work created by the AI in case of copyright infringement, and the scope of the AI-created work (protection period, author rights, etc.).

**ISSUE #2-5 – Who would be responsible for potential infringement of the AI system generated output?**

As discussed above in the AI Systems training section, the AI system may create its own labelled database that is used for a different application (not an AI system), which means this database becomes an output of the AI system. There are a number of ways that this could be done, and each may have different copyright protections depending upon the jurisdiction. Some labelled materials are “shared” via certain licensing allowances, for instance Creative Commons<sup>46</sup>.

***United States:***

The answer would likely depend on the facts. If there is no other human input in the creative process, then it would seem likely that the programmer (or the company employing that programmer) would be responsible for any infringement caused by the creation of infringing output. However, if there is other human input, like from the involvement of a user, then existing precedent suggests it should depend on whether the program is responsible for the “lion’s share” of the work, as noted above. Caveat: this could become more complex if infringement is no longer tethered to a volitional act (see the majority and dissenting Supreme Court opinions in *Aereo*)<sup>47</sup>. However, almost every court faced with the issue still requires a volitional act in order to find infringement.

**ISSUE #2-6 – Is the AI system database labelled by humans capable of copyright protection?**

***United States:***

A database, created by an AI system pre-programmed to do so by a human author, is a protectable work when labeled. The basic answer is that the database will be copyrightable so long as its selection or arrangement of data is minimally creative. That is a case-specific inquiry that would depend on the facts, viewed in light of analogous precedent.

**ISSUE #2-7 – Is the AI system shared database labelled by humans capable of copyright protection?**

***United States:***

This should ideally be settled at the outset in an agreement between the AI system service provider and the entity requesting the AI service based upon the licensing agreement of the shared database. At any rate, the capability to copyright the output still rests with the case-specific inquiry as stated above in at least #2-6.

***China:***

For labelled materials and a resultant creative output from an AI system, the Copyright Law and its Implementing Regulations would be applied directly.

**ISSUE #2-8 – Is the AI system database labelled by the AI system capable of copyright protection and who is the owner?**

***United States:***

For ownership, while the answer may be either similar to #2-6, or #2-7 above, it may be a factor of how the court construes the labelling done by the AI system. If it is merely using a trained system, there may be no factual distinction for any answers given above with human assistance. However, if the AI truly has developed its own labelling determinations, then this may become a similar pattern to #2-4 above.

Then the determination of copyright protection follows similar precedents as indicated above (if an AI is capable and allowed to be an owner of copyright).

## **ISSUE #2-9 – Who would be responsible for potential infringement of the AI system database labelled by the AI?**

### ***United States:***

The answer would likely depend on the facts and closely follow the discussion indicated in #2-5 above.

---

#### **<sup>1</sup> Contributors, in alphabetical order:**

Seong Tahk Ahn is a patent attorney at Lee & Ko IP in Seoul, South Korea. (ast@leekoip.com)

Takeshi Aoki is a partner in the electrical, computer-tech, and software group of PRIMEWORKS IP Attorneys (ata@primeworks-ip.com)

Vanessa Bailey, Global Director-Intellectual Property Policy, Intel Corporation

Raphael (“Ray”) Freiwirth – Owner RF IP Law, PLLC, JD, BSEE, BA International Relations, In-house Contract General & IP Counsel, mentoring startups, long time computer engineer, AI Legal Afficionado (Ray@rfiplaw.com).

Allison Gaul is a senior associate at Kilpatrick, Townsend, and Stockton LLP’s Washington, DC office and focuses her practice on patent prosecution for inventions related to artificial intelligence, machine vision, drones, and augmented reality. (agaul@kilpatricktownsend.com).

Dr Guillaume Henry, attorney at Law, AIPPI (g.henry@aippi.org)

Mincheol Kim is an intellectual property lawyer and global business adviser at Knobbe Martens. (mincheol.kim@knobbe.com)

Youping (Jennifer) Ma is a partner at Beijing Bridgeon Law Firm (mayouping@bridgeonlaw.com)

Amanda McAllister is an associate in Gray, Plant, Mooty’s Minneapolis Office, practicing in the area of data breach and cybersecurity law. (Amanda.McAllister@gpmlaw.com)

Dr. Shoichi Okuyama, is a Principal at Okuyama & Sasajima, practicing in electrical engineering, chemistry, and iotechnology fields. (Okuyama@aippi.org)

Lili Wu, is an attorney with the Han Kun Law Offices in Beijing, she provides whole-process patent portfolio management and patent enforcement. (13511006372@163.com)

Guoquan (Steven) Yang is a partner at Beijing Bridgeon Law Firm (steven.yang@bridgeonlaw.com)

#### **Special thanks to the editorial efforts of:**

R. Danny Huntington is a US patent attorney and partner in Rothwell Figg Ernst & Manbeck, P.C. in Washington, DC (dhuntington@rfem.com)

Claudia Ratner, J.D. candidate at the Maurice A. Deane School of Law at Hofstra University.

---

<sup>2</sup> Circulated March 2019. This primer is designed to orient attendees on the subject matters to be discussed at the forthcoming AIPLA-AIPPI-FICPI Colloquium (March 28-29, 2019 in Turin, Italy). This first edition of the primer is directed to patents, the primary focus of the Colloquium. Ideally it should be reviewed prior to the Colloquium. While not comprehensive, the contributors sought to provide an overview of the state of law as of the date of publication in at least the jurisdictions covered by the IP5 Offices (U.S. Patent and Trademark Office, European Patent Office, Japan Patent Office, Korean Intellectual Property Office, and National Intellectual Property Administration of China)

As with the previous Patent Section, this Copyright Draft is a work-in-progress. A further edition of the primer is already being prepared for circulation that include sections on trade secrets and trademarks. Those topics will also be briefly discussed during the Colloquium to provide a better understanding of the universe of possible IP protection of AI, as well as consideration of the interplay between the various types of protection.

The purpose of this document is to provide educational and informational content and is not intended to provide legal services or advice. The opinions, views and other statements expressed by the contributors are solely those of the contributors and do not necessarily represent those of AIPLA, AIPPI and FICPI.

<sup>3</sup>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."

<sup>4</sup> 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."

<sup>5</sup> *Generative Adversarial Network*, Wikipedia, [https://en.wikipedia.org/wiki/Generative\\_adversarial\\_network](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"

<sup>6</sup> DIRECTIVE 2001/29/EC of THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, *On the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society*, Official Journal of the European Communities, L167 (May 22, 2001) <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:167:0010:0019:EN:PDF>

---

<sup>7</sup> DIRECTIVE 2009/24/EC of THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, *On the Legal Protection of Computer Programs*, Codified Version, Official Journal of the European Communities, L111 (Apr. 23, 2009) <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:111:0016:0022:EN:PDF>

<sup>8</sup> Hitomi Iwase, *Revisions to Copyright Act and Unfair Competition Prevention Act*, International Law Office: Nishimura & Asahi (Nov. 5, 2018) <https://www.lexology.com/library/detail.aspx?g=8e59c176-ecd2-49c8-a5dd-8e4e9bbdbfc>

See Japanese Copyright Act Amendment, 2018

*Encompassing several existing provisions that restrict copyrights, the three generalized categories are now provided to allow free uses, and they are: (1) uses that do not involve personal enjoyment of ideas or emotions (Article 30-4), (2) uses associated with processing by a computer (Article 47-4), and (3) uses for computer processing that create new knowledge or information by data processing to promote the use of the work and that are considered insignificant in view of measures such as what proportion (whole v. partial) of a publicly provided work is used, how much of a work is used, how precisely expressions in the work are used, and other factors (Article 47-5).*

<sup>9</sup> See, for instance, *Big Data*, Wikipedia, [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data) (last modified Mar. 16, 2019, 01:14 UTC)

<sup>10</sup> *Labeled Data*, Wikipedia, [https://en.wikipedia.org/wiki/Labeled\\_data](https://en.wikipedia.org/wiki/Labeled_data) (last modified Aug. 29, 2018, 11:13 UTC)

"Labeled data is a group of samples that have been tagged with one or more labels"

<sup>11</sup> 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."

<sup>12</sup> *Labeled Data*, Wikipedia, [https://en.wikipedia.org/wiki/Labeled\\_data](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."

<sup>13</sup> Pierre N. Leval, *Toward A Fair Use Standard*, 103 Harv. L. Rev. 1105 (1990)

<sup>14</sup> *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015) (citing *Campbell v. Acuff-Rose Music Inc.*, 510 U.S. at 579 (1994)).

"Although transformative use is not "absolutely necessary" to a finding of fair use, "the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works" *Id.* at 290.

<sup>15</sup> 17 U.S.C.A. § 107

<sup>16</sup> Copyright Law of the People's Republic of China, Art. 22 (2010) <https://wipo.lex.wipo.int/en/text/186569>

<sup>17</sup> *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991)

---

<sup>18</sup> *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." Experian 893 F.3d 1176, 1180*

*"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." 893 F.3d 1176, 1180*

<sup>19</sup> *Experian Info. Sols., Inc. v. Nationwide Mktg. Servs. Inc.*, 893 F.3d 1176 (9th Cir. 2018)

*But the court only granted "thin" copyright protection. It found a competitor's copying 80 percent was not infringement. The court held copyright in a database is not infringed absent "a bodily appropriation of [Plaintiff's] work."*

*"Even assuming Natimark's pairings were exact copies of their counterparts in the Experian database, the match rate would only be 80% and insufficient to establish a bodily appropriation of Experian's work." Id. at 1187.*

*"Because Experian has not introduced the version of its database that it claimed was copied, it cannot establish infringement. Even if Experian could establish a triable issue as to copying by comparing later versions of its database with the alleged infringing database, however, the undisputed evidence shows that Experian \*1188 could not establish bodily appropriation of expression. It can at best show a match rate of 80%." Id. at 1187.*

<sup>20</sup> Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, Columbia Public Law Research Paper No. 14-597; Berkeley Technology Law Journal, Vol. 34, No. 2, 2019 (August 5, 2018).

<sup>21</sup> *Digital Drilling Data Sys. LLC v. Petrolink Servs. Inc.*, 2018 WL 2267139 (S.D. Tex. May 16, 2018)

*"Therefore, because the database is original, and based on the broad understanding of the nonliteral elements of computer programs, the Court finds that the schema generated by the DataLogger source code is covered by Digidrill's DataLogger copyright." Id. at 7.*

<sup>22</sup> Regulations for the Implementation of the Copyright Law of the People's Republic of China, Rule 2 (2013) <https://wipolex.wipo.int/en/text/456390>.

<sup>23</sup> *Oracle Am., Inc. v. Google Inc.*, 750 F.3d 1339 (Fed. Cir. 2014)

*"A set of commands to instruct a computer to carry out desired operations may contain expression that is eligible for copyright protection." Id. at 1367.*

*"We agree with Oracle that, under Ninth Circuit law, an original work—even one that serves a function—is entitled to copyright protection as long as the author had multiple ways to express the underlying idea." Id. at 1367.*

<sup>24</sup> Case C-5/08, *Infopaq International A/S v. Danske Dagblades Forening*, (2009) E.C.R. 465, ¶ 37

*Under the Computer Software Directive, a computer program shall be protected if it is original in the sense that it is the author's own intellectual creation. By author, only the natural persons who created the program can be considered authors (with the exception of those Member States where legal persons are also allowed to be designated as rightsholders). The originality criteria and the definition of authorship are also applicable to copyright in general throughout the EU:*

*"37 In those circumstances, copyright within the meaning of Article 2(a) of Directive 2001/29 is liable to apply only in relation to a subject-matter which is original in the sense that it is its author's own intellectual creation."*

---

"38 As regards the parts of a work, it should be borne in mind that there is nothing in Directive 2001/29 or any other relevant directive indicating that those parts are to be treated any differently from the work as a whole. It follows that they are protected by copyright since, as such, they share the originality of the whole work."

"39 In the light of the considerations referred to in paragraph 37 of this judgment, the various parts of a work thus enjoy protection under Article 2(a) of Directive 2001/29, provided that they contain elements which are the expression of the intellectual creation of the author of the work."

<sup>25</sup> TRIPS: Agreement on Trade-Related Aspects of Intellectual Property Rights, Art. 9, Sec. 2, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299 [hereinafter TRIPS Agreement].

"Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such."

<sup>26</sup> Regulations on Computer Software Protection of the People's Republic of China, (2013) <https://wipolex.wipo.int/en/text/455377>

<sup>27</sup> THE INTELLIGENT ROBOTS DEVELOPMENT AND DISTRIBUTION PROMOTION ACT of the Republic of Korea, Act No. 9014, Mar. 28, 2008 [http://elaw.klri.re.kr/eng\\_mobile/viewer.do?hseq=17399&type=sogan&key=13](http://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=17399&type=sogan&key=13)

<sup>28</sup> 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>

<sup>29</sup> 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."

<sup>30</sup> 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>

"Artificial intelligence has taught itself to create its own encryption and produced its own universal 'language'. Now it's writing its own code using similar techniques to humans."

<sup>31</sup> See generally Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, Columbia Public Law Research Paper No. 14-597; Berkeley Technology Law Journal, Vol. 34, No. 2, 2019 (August 5, 2018)

<sup>32</sup> Regulations for the Implementation of the Copyright Law of the People's Republic of China, Rule 2 (2013) <https://wipolex.wipo.int/en/text/456390>

<sup>33</sup> *Rearden LLC v. Walt Disney Co.*, 293 F. Supp. 3d 963 (N.D. Cal. 2018)

*The user input into the system two dimensional camera capture of Dan Stevens' (Beauty and the Beast) "facial expressions [in] all the scenes [shot] on previous days." The program processed those through computer graphics to create the animal-like face of the Beast on Stevens' body:*

"MOVA Contour's user inputs a two dimensional camera capture that may range from Dan Stevens' "facial expressions of all the scenes we had done on previous days" to the "subtle and dynamic motions performed by the actor [Josh Brolin playing Thanos in Guardians of the Galazy]" to "Brad Pitt's 44-year-old face." Disney Compl. ¶¶ 2, 4, 36." *Id.* at 971.

---

<sup>34</sup> *Design Data Corp. v. Unigate Enter., Inc.*, 847 F.3d 1169, (9th Cir. 2017)

"...the copyright protection afforded a computer program may extend to the program's output if the program "does the lion's share of the work" in creating the output and the user's role is so "marginal" that the output reflects the program's contents." *Id.* at 1173.

"Assuming, without deciding, that copyright protection does so extend, we nonetheless conclude that Design Data did not raise a question of material fact that the imported SDS/2-generated images and files reflected the contents of its program. Design Data did not present evidence establishing that SDS/2 "does the lion's share of the work" in creating the steel detailing files or that the user's input is "marginal." *Torah Soft*, 136 F.Supp.2d at 283. Thus, the district court correctly rejected Design Data's argument that the SDS/2 copyright protects the images and files that UE imported and distributed." *Id.* at 1173.

<sup>35</sup> Copyright Law of Japan, CRIC, Article 2(1)(ii)  
[http://www.cric.or.jp/english/clj/doc/20161018\\_October,2016\\_Copyright\\_Law\\_of\\_Japan.pdf](http://www.cric.or.jp/english/clj/doc/20161018_October,2016_Copyright_Law_of_Japan.pdf) (last modified Oct. 2016).

"author" means a person who creates a work;"

<sup>36</sup> Copyright Law of Japan, CRIC, Article 2(1)(i)  
[http://www.cric.or.jp/english/clj/doc/20161018\\_October,2016\\_Copyright\\_Law\\_of\\_Japan.pdf](http://www.cric.or.jp/english/clj/doc/20161018_October,2016_Copyright_Law_of_Japan.pdf) (last modified Oct. 2016).

"work" means a production in which thoughts or sentiments are expressed in a creative way and which falls within the literary, scientific, artistic or musical domain;"

<sup>37</sup> Regulations for the Implementation of the Copyright Law of the People's Republic of China, Rule 2 (2013)  
<https://wipolex.wipo.int/en/text/456390>

<sup>38</sup> See 17 U.S.C.A. § 501

*The normal infringement criteria would apply. (a) Was the alleged infringed work copied? That would be clear in that the work was input into the dataset. (b) Is the work registered and owned by the Plaintiff? Again, a binary question. Finally, (c) is the allegedly infringing work substantially similar. These are relatively straightforward questions, although the last question can be quite contentious in some cases.*

<sup>39</sup> See, e.g., *Hobbs v. John*, 722 F3d 1089, 1094 (7<sup>th</sup> Cir. 2013).

<sup>40</sup> *Am. Broad. Companies, Inc. v. Aereo, Inc.*, 573 U.S. 431 (2014).

<sup>41</sup> Bruce E. Boyden, *Emergent Works*, 39 COLUM. J.L. & ARTS 377 (2016)

"Second, the problem of computer-generated works is not a single problem, but rather a set of related problems, some of which are easier than others to resolve. The most difficult involve what might be called "emergent works"—works of apparently creative expression that arise from the operation of a program but cannot be traced directly to a human source." *Id.* at 379.

<sup>42</sup> Case C-145/10, *Eva-Maria Painer v Standard VerlagsGmbH*, (2009) E.C.R. 138, ¶ 88 and 92

"88 As stated in recital 17 in the preamble to Directive 93/98, an intellectual creation is an author's own if it reflects the author's personality."

---

"92 By making those various choices, the author of a portrait photograph can stamp the work created with his 'personal touch'."

<sup>43</sup> Copyright Law of Japan, CRIC, Article 2(1)(i)  
[http://www.cric.or.jp/english/clj/doc/20161018\\_October,2016\\_Copyright\\_Law\\_of\\_Japan.pdf](http://www.cric.or.jp/english/clj/doc/20161018_October,2016_Copyright_Law_of_Japan.pdf) (last modified Oct. 2016).

"“work” means a production in which thoughts or sentiments are expressed in a creative way and which falls within the literary, scientific, artistic or musical domain;"

<sup>44</sup> Chu Yi and Wang Yan, *China Focus: Up in the Air, Who Owns Copyright of AI-Generated Article?*, Xinhua (Dec. 10, 2018) [http://www.xinhuanet.com/english/2018-12/10/c\\_137663982.htm](http://www.xinhuanet.com/english/2018-12/10/c_137663982.htm)

<sup>45</sup> Copyright Law of Korea, Act No. 14634, Mar. 21, 2017, Article 2(1)  
[https://elaw.klri.re.kr/eng\\_service/lawView.do?hseq=42726&lang=ENG](https://elaw.klri.re.kr/eng_service/lawView.do?hseq=42726&lang=ENG)

"The term “work” means a creative production that expresses human thoughts and emotions;"

<sup>46</sup> See, for instance *Creative Commons*, Wikipedia, [https://en.wikipedia.org/wiki/Creative\\_Commons](https://en.wikipedia.org/wiki/Creative_Commons) (last modified Mar. 13, 2019, 08:10 UTC)

<sup>47</sup> *Am. Broad. Companies, Inc. v. Aereo, Inc.*, 573 U.S. 431 (2014).