Computer & Communications
Industry Association
Tech Advocacy Since 1972

The Computer & Communications Industry Association ("CCIA")[1] submits the following comments in response to Issues 1-7 in the World Intellectual Property Organization ("WIPO")'s draft issues paper on artificial intelligence ("AI") and intellectual property ("IP").[2] Recognizing that additional time may be available to furnish answers, CCIA is submitting this presently to assist WIPO's further consideration. Insofar as WIPO decides to revise the questions to obtain additional information, it should take into account and elicit information that will expand Member States' understanding of the importance of balanced intellectual property protections for AI innovations.

PATENTS

Issue 1: Inventorship and Ownership

It is not yet clear that AI-generated inventions have actually been created, though there are several instances in which a human has applied for a patent and claimed that the invention was generated by AI. In understanding AI-related inventions, it is important to distinguish between three classes of inventions: inventions that improve the functioning of AI, inventions that use AI, and inventions created by AI. Each presents its own set of concerns with respect to the questions asked.

CCIA believes that existing patent rules are mostly sufficient for inventions of the first two types, while there is no reason that patent law can or should provide patent protection to inventions created by AI.

In the case of inventions autonomously generated by AI:

> **(i) Should the law permit or require that the AI application be named as the inventor or should it be required that a human being be named as the inventor? In the event that a human inventor is required to be named, should the law give indications of the way in which the human inventor should be determined, or should this decision be left to private arrangements, such as corporate policy, with the**

[1] CCIA is an international nonprofit membership organization representing companies in the computer, Internet, information technology, and telecommunications industries. Together, CCIA's members employ nearly one million workers and generate approximately a quarter of a trillion dollars in annual revenue. CCIA promotes open markets, open systems, open networks, and full, fair, and open competition in the computer, telecommunications, and Internet industries. A complete list of CCIA members is available at http://www.ccianet.org/members.

[2] Available at https://www.wipo.int/about-ip/en/artificial_intelligence/call_for_comments/index.html.

possibility of judicial review by appeal in accordance with existing laws concerning disputes over inventorship?

(ii) The inventorship issue also raises the question of who should be recorded as the owner of a patent involving an AI application. Do specific legal provisions need to be introduced to govern the ownership of autonomously generated AI inventions, or should ownership follow from inventorship and any relevant private arrangements, such as corporate policy, concerning attribution of inventorship and ownership?

CCIA opposes allowing an AI to be named as an inventor. The law does not presently permit such a situation, as evidenced by the European Patent Office's recent decision to refuse a patent application filed naming an AI as an inventor. Nor is there any policy justification for naming an AI as an inventor. However, neither should a human be named as an inventor of technology created by an AI—in such circumstances, there is no inventor who could be properly named on the patent.

This does not mean that there may not be a need to differentiate between AI-contributed and human-contributed aspect of an invention when a human utilizes autonomous AI as part of the process used by the human to create an invention—in such circumstances, it would be appropriate to name the human as the inventor and owner under the ordinary course of patent law. That ownership would be solely with respect to the contribution the human made to the invention—the portion contributed autonomously by AI would not itself be patentable, as it would represent nothing more than what is available to any artisan of ordinary skill using the available prior art. However, that human contribution must itself provide novelty and inventive step over the prior art and the skill in the art—including the output of autonomous AIs.

(iii) Should the law exclude from the availability of patent protection any invention that has been generated autonomously by an AI application? See also Issue 2, below.

CCIA submits that the law should exclude from patent protection inventions that are generated autonomously by an AI application. Such inventions lack any justification for the grant of a patent monopoly, as well as being nothing more than the exercise of ordinary skill.

There is no reason to provide a patent to an AI-generated invention. There are no moral rights considerations, as the only potential holder of such a right—the AI itself—lacks any sense that would justify a moral right in the invention. And on an economic basis, the reward of the patent does not incentivize the AI to operate. The AI is not motivated by the prospect of a reward of exclusivity—it is not motivated at all. It would produce the idea upon being run

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

regardless of the availability of a patent, suggesting that the essential value of a patent as an encouragement to invent is absent and suggesting that the inefficiency and anti-competitive harm of an exclusive right is unnecessary to cause the idea to be created.

Further, an invention autonomously generated by an AI application is inherently lacking in inventive step. An invention lacks inventive step when it would have been obvious to a person skilled in the art. Once an AI is made available, it becomes part of the ordinary skill in the art. While that inventing AI might itself be patentable, its output would not be patentable absent additional human intervention that goes beyond the ordinary skill. Ordinary creativity—the output of a skilled person—exceeds the creativity of an 'automaton'. But an automaton is exactly what an AI is. In the event that an AI is capable of the needed creativity to create an idea, it is definitionally within the creativity of a person skilled in the art.

Much like the availability of computation and computer-aided design tools has affected what is reasonable to treat as ordinary skill, the availability of AI tools will affect what is reasonable to treat as the ordinary skill in the art. An ordinary artisan, relying on the output of an AI tool, has not created anything beyond the ordinary skill unless they contribute something to the combination that rises above what any ordinary artisan could do with the same AI tool. We would not find patentable the set of logic gates created on a field-programmable gate array by a VHDL compiler. It is simply the output of an ordinary tool—though the tool itself might contain patentable inventions. Similarly, the output of an AI is within the scope of ordinary creativity and presumptively unpatentable, even though the AI itself might be patentable.

To the extent that technology reaches a point at which it would be correct to say that an AI is self-motivated and contains sufficient indicia of consciousness to justify treatment as a person, such a change would implicate issues far beyond the expertise and scope of patent law and should be addressed at that time.

CCIA thus opposes granting ownership of inventions created by an AI to anyone, as such inventions should not be patentable.

Issue 2: Patentable Subject Matter and Patentability Guidelines

In the case of AI-generated or -assisted inventions:

(i) Should the law exclude from patent eligibility inventions that are autonomously generated by an AI application? See also Issue 1(iii), above.

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

The law should not exclude from patent eligibility inventions that are autonomously generated by an AI application on the basis of patentable subject matter unless the invention would itself be excluded on that basis if it had been invented by a human. The fact of AI-inventorship is not relevant to the patentable subject matter inquiry. However, as set forth above, such inventions should not be patentable as they lack any inventive step, being simply the result of ordinary creativity and the use of the ordinary tools available to a skilled artisan.

(ii) Should specific provisions be introduced for inventions assisted by AI or should such inventions be treated in the same way as other computer-assisted inventions?

(iii) Do amendments need to be introduced in patent examination guidelines for AI-assisted inventions? If so, please identify which parts or provisions of patent examination guidelines need to be reviewed.

Inventions where a human inventor is assisted by AI should be treated as if the AI-contributed portion is part of the prior art. Once this rule is set in place, existing provisions with respect to patentability and inventorship are sufficient to address other considerations of patentability.

Issue 3: Inventive Step or Non-Obviousness

(i) In the context of AI inventions, what art does the standard refer to? Should the art be the field of technology of the product or service that emerges as the invention from the AI application?

In the context of an AI-generated or -assisted invention, the relevant art is that of the field of technology of the product or service that emerges as the invention.

When examining a mechanical invention, the field of art is that of the product, even though the invention might itself have been created using simulation and CAD software. An ordinary artisan creates using the tools available to them, including software and AI, even if they could not necessarily themselves create those resources. There is no reason to diverge from this rule for AI-generated or -assisted inventions.

If the invention itself incorporates AI, or uses AI to implement the technology, then the relevant art may include the AI arts in addition to the field of technology of the product or service—for example, a product that applies machine learning computer vision to the fruit-sorting arts would involve relevant art in both fruit-sorting and AI arts.

(ii) Should the standard of a person skilled in the art be maintained where the invention is autonomously generated by an AI application or should consideration

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

be given to replacing the person by an algorithm trained with data from a designated field of art?

As described above with respect to Issue 1(iii), while the standard of a person skilled in the art should be maintained, the standard of skill in the art becomes the abilities of a person skilled in that art who can use a known AI tool for inventing, in the same way that the standard of skill in the art today is that of a person skilled in the art with access to known computing technologies.

An autonomously-generated AI invention is thus always lacking in inventive step, representing nothing more than the ordinary output of a person skilled in the art equipped with their ordinary tools.

(iii) What implications will having an AI replacing a person skilled in the art have on the determination of the prior art base?

Adding an AI to the person of skill, rather than replacing that person, is the appropriate course. Neither would impact the prior art base, which is presumed to be all relevant prior art, but would impact the ordinary skill in the art in terms of a general increase in skill as ordinary artisans would have more sophisticated tools with which to build upon the prior art.

(iv) Should AI-generated content qualify as prior art?

AI-generated content should generally qualify as prior art in the same way that any other publicly available information qualifies as prior art. Further, the output of publicly available AI tools is presumptively part of the prior art sphere, even if the specific output is not. This presumption would not apply if the AI itself is not publicly available or described, or if the output is based on a known AI architecture with non-public application-specific training, as it is unreasonable to think that an ordinary artisan would have access to such a non-public tool. However, if the AI itself is available to the public, either directly or via replication of a described system, then any output of that AI forms part of the sphere of prior art, even if that specific output item is itself not publicly available.

Issue 4: Disclosure

(i) What are the issues that AI-assisted or AI-generated inventions present for the disclosure requirement?

Sufficient disclosure for an AI invention is dependent on how AI is integrated into the invention and the type of AI in question. Much of current AI is based on machine learning

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

techniques, and these comments focus on the disclosures that are most relevant to machine learning. However, other forms of AI exist and will likely be developed, and distinct disclosure considerations may apply to other AI architectures.

Autonomous AI-generated inventions would not create distinct disclosure considerations beyond those typical to the subject matter of the invention—*e.g.*, a novel compound identified by an AI would need to comply with typical written description constraints for chemical compounds.

Inventions that relate to improving the functionality of AI inventions (AI-improving inventions), such as new algorithms for machine learning or training, will generally implicate the same disclosure concerns as computer-implemented inventions. The written description must describe not just the desired aspect of AI, but how it is achieved—typically, by disclosing the algorithm or implementing structure and how to utilize it. For example, a patent application for a new deep learning structure could describe the number of layers, the number of units per layer, data flows between units or layers, and activation functions. A patent application for a new training algorithm would likely describe the structure to be trained, the starting point and data used in training, and other relevant information necessary to reproduce the invention.

Inventions that apply AI to a problem outside of the AI sphere (AI-application inventions), such as AI machine vision or AI drug discovery, present additional unique considerations with respect to disclosure. Beyond the computer-implemented concerns described above with respect to AI inventions, which would also apply to most AI-application inventions, AI-application inventions are particularly likely to be described in functional terms and would require disclosure of a suitable AI structure in order to avoid invalidity concerns. Such a structural description in the specification would need to ensure replicability by a reader by requiring that the inventor provide access to sufficient information to replicate the invention. That could be in the form of describing the type of training data and the training methodology, or in the form of the trained model itself, such as the hidden layer weights post-training. Applications relying on modern AI techniques may implement a well-known structure (*e.g.*, a deep learning network) with the novelty coming from how the structure is adapted to solve a specific problem. The necessary disclosure might be description of the types of training data used, the pre-processing of data used, or how outputs are analyzed and applied to solve the problem.

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

In both circumstances, the nature of AI systems can raise additional concerns for disclosure and enablement. In particular, AI systems can be unpredictable. Seemingly minor changes in training data or algorithm can produce wildly varying output models.

Given this, it is crucial—particularly for AI-application inventions—that the specification enable the full scope of the claim. This means that the disclosure of a single species would not generally be sufficient to support a generic claim—*e.g.*, disclosure of a deep learning model that recognizes images of dogs would not be sufficient to support a claim to "using a deep learning model to recognize images by type" without additional support in the specification to provide modeling for generic image recognition. Claims to "do a function with AI" should only be considered enabled if they teach how to do the function with respect to any form of AI, not just a particular model of AI. To treat such claims otherwise would effectively, and unjustifiably, imply that the patent could predictably enable the use of novel AI architectures to perform the claimed function.

(ii) In the case of machine learning, where the algorithm changes over time with access to data, is the disclosure of the initial algorithm sufficient?

As described above, disclosure of the algorithm alone will often be insufficient. The model alone does not implement the invention—access to the training data and methodology, or the trained model itself, is necessary for a patent to enable others to replicate it.

(iii) Would a system of deposit for algorithms, similar to the deposit of microorganisms, be useful?

It is unclear why a system of deposit would prove useful. The algorithms, training methodology, and model details useful to enabling and describing existing AI tools are generally amenable to textual or digital description and should be submitted as part of the patent application. Any additional data necessary for enabling and disclosing an invention that is more voluminous than is appropriate for inclusion in the patent document could be submitted as an appendix to the application, similar to the permitted appendices for computer program listings.

To the extent that disclosure of future AI architectures would be well-suited for a depository system, such a system could be created at that time.

(iv) How should data used to train an algorithm be treated for the purposes of disclosure? Should the data used to train an algorithm be disclosed or described in the patent application?

Sufficient information must be provided as part of the patent application to enable a skilled person to replicate the invention without undue experiment. This can be achieved in a

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

variety of ways. The exact way in which written description is provided for an AI-improving or AI-application invention could likely be left to the inventor's discretion, so long as the inventor provides sufficient description to enable replication of the claimed AI. This could be in the form of a description of the trained model itself and the algorithm with which to apply it, a description of appropriate training data and training methodology, the provision of the actual training data and training methodology, or some other combination of data—so long as the totality of the provided information is sufficient to enable replication.

Non-replicability is *prima facie* evidence of insufficient disclosure, and replicability of an AI-improving or AI-application invention essentially requires disclosure of either sufficient training data to produce the model or else of the trained model itself. An inventor's assertion that such data is unnecessary to replication would need to be supported by additional evidence before it could be accepted. Conversely, provision of the trained model would essentially always allow replication, serving as a "safe harbor" for enablement and description—sufficient, though not required.

(v) Should the human expertise used to select data and to train the algorithm be required to be disclosed?

CCIA generally believes that the human expertise used to select data and train the algorithm should be disclosed. If the trained model is itself provided, and the claim is sufficiently narrow that the trained model is sufficient to enable the practice of the full scope of the claim, this human expertise might be omitted; however, given the importance of data selection and training to successful application of AI technology, it should generally be presumed that the human expertise necessary to train the AI should be disclosed absent a showing by the applicant that it is unnecessary.

Issue 5: General Policy Considerations for the Patent System

(i) Should consideration be given to a sui generis system of IP rights for AI-generated inventions in order to adjust innovation incentives for AI?

CCIA opposes a *sui generis* system of IP rights for AI-generated inventions.

There is no reason to provide a patent to an AI-generated invention. There are no moral rights considerations, as the only potential holder of such a right—the AI itself—lacks any sense that would justify a moral right in the invention. And on an economic basis, the reward of the patent does not incentivize the AI to operate. The AI is not motivated by the prospect of a

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

reward of exclusivity—it is not motivated at all. It would produce the idea upon being run regardless of the availability of a patent, suggesting that the essential value of a patent as an encouragement to invent is absent and suggesting that the inefficiency and anti-competitive harm of an exclusive right is unnecessary to cause the idea to be created.

Further, as described above with respect to Issue 1(iii), AI-generated inventions are no more than the output of an automaton—not the type of non-obvious innovation the patent system is intended to reward. As the level of skill rises, the standard for non-obviousness rises with it. And as the tools available to an ordinary artisan become more sophisticated, the level of skill rises. The availability of patents for invention-generating AI systems is sufficient to incentivize the creation of those systems. There is no need to further incentivize the operation of those systems to create AI-generated inventions.

(ii) Is it too early to consider these questions because the impact of AI on both science and technology is still unfolding at a rapid rate and there is, at this stage, insufficient understanding of that impact or of what policy measures, if any, might be appropriate in the circumstances?

AI is a rapidly developing field. The uncertainty of a rapidly developing technology is best served by limiting the constraint that current choices place on future developments.

It is possible that future developments might require changes to the positions set forth in this document. But it is particularly important not to take actions that might significantly delay or divert AI development. In particular, the drastic reward of a 20-year monopoly on an AI technology, difficult to remove once granted, must be balanced against the potential of a small delay in development if that reward is less available. Widespread availability of patents on AI-generated inventions would lead to less innovation by placing ordinary creativity into the realm of monopoly and chilling the rationale to pursue such creativity. Such an outcome should be avoided by ensuring robust disclosure of AI inventions and determining that the output of autonomous AI-generating systems is presumptively lacking in inventive step.

COPYRIGHT AND RELATED RIGHTS

Issue 6: Authorship and Ownership

(i) Should copyright be attributed to original literary and artistic works that are autonomously generated by AI or should a human creator be required?

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

A work produced by an AI algorithm or process, without the involvement of a natural person contributing to the resulting work, should not qualify as a work of authorship protectable under copyright law.

The U.S. Copyright Office currently refuses to register a work that was not created by a human being. It explains:

> The copyright law only protects "the fruits of intellectual labor" that "are founded in the creative powers of the mind." *Trade-Mark Cases*, 100 U.S. 82, 94 (1879). 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. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 58 (1884).[3]

The U.S. Copyright Office adds that it "will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author."[4]

Accordingly, if an AI algorithm or process creates a work that lacks expression from a natural person in the resulting work, the U.S. Copyright Office correctly would refuse to register the work. A court would uphold this refusal under the U.S. Supreme Court precedent cited in the Compendium.

As a matter of policy, the result should be the same globally. Withholding copyright protection from a work resulting from an AI process for which there was no expressive contribution by a natural person is justifiable from a policy perspective; the AI algorithm, and the computer that runs it, does not require the economic incentive provided by copyright in order to create works. Indeed, AI is capable of quickly producing an enormous array of works. Recognizing copyright in such output could quickly create a minefield of legal issues, leading to litigation and uncertainty.[5]

To be sure, the human creator of the software that runs the AI algorithm or process would receive a copyright in the expressive aspects of the AI software (and perhaps a patent for inventions in the AI software). We do not need copyright to incentivize the creation of AI-generated works by giving copyright to each generated work because copyright already incentivizes the creation of AI software that generates all manner of content.

---

[3] Compendium of U.S. Copyright Office Practices III, § 306 (2017).

[4] *Id*. at § 313.2.

[5] *See* Daniel Gervais, *The Machine As Author*, 105 IOWA L. REV. (forthcoming 2020), *available at* https://ssrn.com/abstract=3359524.

(ii) In the event copyright can be attributed to AI-generated works, in whom should the copyright vest? Should consideration be given to according a legal personality to an AI application where it creates original works autonomously, so that the copyright would vest in the personality and the personality could be governed and sold in a manner similar to a corporation?

As noted in response to Issue 6(i), AI-generated works should not receive copyright protection. If a work resulting from an AI process is modified or enhanced by a natural person, that modification or enhancement may reflect sufficient originality to rise to the level of authorship. *See Feist v. Rural Telephone*, 499 U.S. 340 (1991). A court applying traditional copyright principles to the facts of such a case would be able to determine whether there was enough expressive contribution by a human to rise to the level of authorship and whether the portions of a work that are reproduced in a particular case are protectable human creativity or unprotectable, AI-generated output. Of course, any copyright would extend only to the original elements contributed by the natural person.

CCIA opposes the concept of according a legal personality to an AI application that creates original works autonomously.

(iii) Should a separate sui generis system of protection (for example, one offering a reduced term of protection and other limitations, or one treating AI-generated works as performances) be envisaged for original literary and artistic works autonomously generated by AI?

CCIA opposes the adoption of the adoption of a *sui generis* form of protection for original literary and artistic works autonomously generated by AI for the policy reasons discussed above.

Issue 7: Infringement and Exceptions

(i) Should the use of the data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright? If not, should an explicit exception be made under copyright law or other relevant laws for the use of such data to train AI applications?

The use of the data subsisting in copyright works without authorization for machine learning should not constitute an infringement of copyright. In the United States, the existing statutory framework and related case law concerning the fair use right, 17 U.S.C. § 107, clearly permit the ingestion of large amounts of copyrightable material for the purpose of an AI algorithm or process learning its function. In jurisdictions without a fair use provision, an

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

explicit exception may be necessary. Because of the importance of the lawfulness of ingestion to this inquiry and AI more generally, we will explain in detail how fair use permits this activity.

AI algorithms and other processes often require the ingestion of large amounts of material. Assembling that material may entail converting it into a more usable format, *e.g.*, translating image files into machine-readable files.[6] In addition, backup copies of the materials will be necessary to protect against loss of data in the event of system failure.[7] Temporary reproductions of portions of the material in a computer's random access memory are a normal part of the process of training and AI algorithm.[8] All these copies are not viewable or consumable by the outside world.[9] Because these non-expressive copies are not consumable by the public, they do not function as market substitutes for copies of the ingested works.[10]

Numerous U.S. appellate courts have correctly found the mass copying of raw material to build databases for uses by AI processes to be fair use under 17 U.S.C. § 107. *See, e.g.*, *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015); *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014); *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 640 (4th Cir. 2009); *Perfect 10 v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2003). Judge Leval's opinion in *Google* provides the clearest analysis of why the creation of an AI database, and its subsequent uses, are fair uses.

---

[6] In the Google Library Project, Google made a digital scan of each book it borrowed from a research library, then used optical character recognition software to convert the scan machine readable text. Google retained both the scanned image and machine-readable text. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 208 (2d Cir. 2015).

[7] For example, HathiTrust created and maintained four text-only copies of its entire database (one on the primary server at the University of Michigan, another at the mirror server at the University of Indiana, and two encrypted backup tapes at two secure locations on the University of Michigan campus) for the purpose of balancing the load of user web traffic and serving as back-up in the case of a disaster. *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).

[8] These temporary reproductions may not constitute copies under the Copyright Act. *See Cartoon Networks LP v. CSC Holdings, Inc*., 536 F.3d 121 (2d Cir. 2008).

[9] Professor Matthew Sag characterizes acts of copying which do not communicate the author's original expression to the public as "nonexpressive uses." *See, e.g*., Matthew Sag, *Copyright and Copy-reliant Technology*, 103 NW. U. L. REV. 1607, 1624 (2009). Professor Edward Lee describes three kinds of uses: creational uses (uses of copyrighted works to create a technology); operational uses (uses that occur during the operation of the technology once it has been created); and output uses (the distribution or display of works as an output of the technology). Edward Lee, *Technological Fair Use*, 83 SO. CAL. L. REV. 797, 842-44 (2010).

[10] The non-expressive uses of works for the creation of AI algorithm are analogous to the "intermediate copies" made during the course of software reverse engineering. Courts have found that fair use permitted the translation of machine-readable object code into human readable source code as an essential step in the development of noninfringing interoperable computer programs. In these cases, the source code was used internally and was never distributed to the public. *See Sony Computer Entm't v. Connectix Corp*., 203 F.3d 596 (9th Cir. 2000); *Sega Enters. Ltd. v. Accolade, Inc*., 977 F.2d 1510 (9th Cir. 1992); *Atari Games Corp. v. Nintendo of Am., Inc*., 975 F.2d 832 (Fed. Cir. 1992).

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

1. The Purpose and Character of the Use

The *Google* court stated that it "had no difficulty concluding that Google's making of a digital copy of Plaintiff's books for the purpose of enabling a search for identification of books containing a term of interest to the searcher involves a highly transformative purpose. . . ." *Google*, 804 F.3d at 216. In reaching this conclusion regarding these copies, the court relied on *HathiTrust*, where the court found that "both the making of the digital copies and the use of those copies to offer the search tool were fair uses." *HathiTrust*, 755 F.3d at 105. The *Google* court noted that the *HathiTrust* court had found that the downloading and storing of complete digital copies of books "was essential to permit searchers to identify and locate the books in which words or phrases of interest to them appeared." *Google*, 804 F.3d at 217. The *Google* court quoted the *HathiTrust* court's conclusion that "the creation of a full-text searchable database is a quintessentially transformative use . . . [a]s the result of a word search is different in purpose, character, expression, meaning, and message for the page (and the book) from which it is drawn." *Google*, 804 F.3d at 217 (quoting *HathiTrust*, 755 F.3d at 97).

The *Google* court then cited *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 640 (4th Cir. 2009); *Perfect 10 v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007); and *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2003), as "examples of cases in which courts had similarly found the creation of complete digital copies of copyrighted works to be transformative uses when the copies served at different function from the original." *Google*, 804 F.3d at 217 (quotations omitted). All three of these cases involved the creation of a database to which AI algorithms were applied. *Kelly* and *Perfect 10* involved search engines designed to find images on the World Wide Web. *iParadigms* involved a plagiarism detection service that enabled an instructor to find works from which a student paper may have been copied. The *Google* court explained that "[a]s with *HathiTrust* (and *iParadigms*), the purpose of Google's copying of the original copyrighted books is to make available significant information *about those books*, permitting a searcher to identify those that contain a word or term of interest. . . ." *Google*, 804 F.3d at 217. *See also iParadigms*, 562 F.3d at 639-40 ("iParadigms' use of plaintiffs' works had an entirely different function and purpose than the original works. . . . iParadigms' use of these works was completely unrelated to the expressive content and instead aimed at detecting and discouraging plagiarism.").

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

The *Google* court also made clear that the commercial motivation of a provider of the AI database should not tilt the first factor against the provider: "[o]ur court has . . . repeatedly rejected the contention that commercial motivation should outweigh a convincing transformative purpose and absence of significant substitutive competition of the original." *Google*, 804 F.3d at 219. *See also iParadigms*, 562 F.3d at 639. In short, the first fair use factor should always weigh in favor of the creator of an AI database.

2. The Nature of the Copyrighted Work

Neither the *Google* nor *HathiTrust* courts found the second fair use factor, the nature of the copyrighted work, to be dispositive. This is "because the secondary use transformatively provides information about the original, rather than replicating protected expression in a manner that provides a meaningful substitute for the original." *Google*, 804 F.3d at 220. In other words, because the copying involved in the creation of a search database does not provide a meaningful substitute of the original, the nature of the original has little relevance. *See also iParadigms*, 562 F.3d at 641-42 (second factor does not weigh against iParadigms because its "use of the works in this case—as part of a digitized database from which to compare the similarity of typewritten characters used in other student works—is . . . unrelated to any creative component."). There could be cases where the nature of the works used, such as more factual works, might also support the finding of fair use. In any event, the second fair use factor does not tilt against non-expressive uses by AI database creators.

3. The Amount and Substantiality of the Portion Used

Uses in the creation of AI algorithms often require the copying of entire works. The *HathiTrust* court concluded its discussion of the third fair use factor by noting that "[b]ecause it was reasonably necessary for the [HathiTrust Digital Library] to make use of the entirety of works in order to enable the full-text search function, we do not believe the copying was excessive." *HathiTrust*, 755 F.3d at 98. Likewise, the *Google* court found that "not only is the copying of the totality of the original reasonably appropriate to Google's transformative purpose, it is literally necessary to achieve that purpose. If Google copied less than the totality of the originals, its search function could not advise searchers reliably whether the searched term appears in a book (or how many times)." *Google*, 804 F.3d at 221. *See also iParadigms*, 562 F.3d at 642 (endorsing the district court's conclusion that iParadigms' use of the entirety of original works was limited in purpose and scope as a digitized record for electronic comparison

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

purposes only). In other cases, an AI database might be made up of individual words or inconsequential portions of larger copyrighted works and thus the third factor would be expected to weigh in favor of such uses as well. In general, this factor and the case law interpreting it have well-served innovation in AI.

4. The Effect of the Use on the Market for the Original

The *HathiTrust* court found that the fourth fair use factor supported a finding of fair use because the ability to search the text of a book to determine whether it includes a search term "does not serve as a substitute for the books that are being searched." *HathiTrust*, 755 F.3d at 100. The *HathiTrust* court rejected the plaintiffs' suggestion that HathiTrust impaired the emergence of a market for licensing books for digital search: "Lost licensing revenue counts under Factor Four only when the use serves as a substitute for the original and full-text-search does not." *Id*. The *Google* court cited with approval *HathiTrust*'s conclusion that the search function does not substitute for the books being searched, *Google*, 804 F.3d at 223, and devoted the rest of its discussion of the fourth factor to snippet view. *See also iParadigms*, 562 F.3d at 644 ("Clearly no market substitute was created by iParadigms, whose archived student works do not supplant the plaintiffs' work in the 'paper mill' market so much as merely suppress demand for them, by keeping record of the fact that such works have previously been submitted.").

The *Google* court dismissed plaintiffs' assertion that the search function usurped their market for derivative uses. The court explained that "the copyright that protects Plaintiffs' works does not include an exclusive derivative right to supply . . . information [contained in the works] through query of a digitized copy." *Google*, 804 F.3d at 225. The court underscored this point: "Nothing in the statutory definition of a derivative work, or of the logic that underlies it, suggests that the author of an original work enjoys an exclusive right to supply information about that work of the sort communicated by Google's search functions." *Google*, 804 F.3d at 226. Given the absence of harm the Copyright Act is intended to prevent, the fourth fair use factor favors uses by AI database providers when those uses cause little or no harm to the legitimate markets of the originals.

5. The Four Factors Weighed Together

At the end of its fair use analysis, the *Google* court stated that "considering the four fair use factors in light of the goals of copyright, we conclude that Google's making of a complete digital copy of Plaintiffs' works for the purpose of providing the public with its search . . .

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

functions . . . is a fair use and does not infringe Plaintiff's copyrights in their books." *Google*, 804 F.3d at 225. The reasoning of *HathiTrust*, *Google*, and *iParadigms* compels the conclusion that the uses necessary to make any AI database are fair use. Regardless of the nature of the content, providing AI functionality *always* has a different purpose and function from the content itself; copying entire works is *often* necessary to provide accurate operation of AI algorithms and processes; and such copying *never* substitutes for the original works.

To help prevent this issue from being relitigated in every case involving an AI database, there should be a bright line permitting uses related to the creation and operation of an AI database. Such clear guidance not only would conserve judicial resources, it would prevent erroneous decisions. Drawing a bright line permitting the copying necessary to enable an AI process would benefit innovators and the courts.

(ii) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, what would be the impact on the development of AI and on the free flow of data to improve innovation in AI?

Treating the unauthorized use of data subsisting in copyrighted work for machine learning as an infringement would have a significant adverse impact on the development of AI and the free flow of data to improve innovation in AI. Companies and research institutions alike would become hesitant to ingest copyrighted works for machine learning because of the potential exposure to infringement liability. Such a result would be counterproductive to copyright law's ultimate goal of promoting the creation and dissemination of works.

Data-dependent research has already yielded significant achievements in science, medicine, and the humanities. For example, digital humanities scholars—in fields ranging from literature and linguistics to history and philosophy—use AI to gain insights into history by making linguistic connections between texts over time.[11] The data necessary to fuel these discoveries often depends on digitizing entire books, while the finished product does not contain protected expression from those works. Similarly, medical historians have used AI algorithms to learn about harmful drug reactions and make connections among documents in the historical

[11] *See generally* Brief for Digital Humanities and Law Scholars as *Amici Curiae* Supporting Defendants-Appellees, *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) (No. 12-4547-cv) (describing researchers' non-expressive use of metadata for socially beneficial purposes).

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

record.[12] U.S. case law recognizes and enables these uses, due to the benefits this type of research yields, as explained in response to Issue 7(i).

(iii) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an exception be made for at least certain acts for limited purposes, such as the use in non-commercial user-generated works or the use for research?

An exception permitting unauthorized uses of works for the purpose of machine learning should not be limited to non-commercial user generated works or non-commercial research. The U.S. fair use jurisprudence described above does not contain such limitations, with no adverse impact on rights holders. The Digital Single Market ("DSM") Directive currently treats text- and data-mining ("TDM") by commercial entities differently from TDM by noncommercial entities, and it should be broadened to permit AI innovation to develop and thrive in Europe, so as not to discriminate against business stakeholders.

(iv) If the use of the data subsisting of copyright works without authorization for machine learning is considered to constitute an infringement of copyright, how would existing exceptions for text and data mining interact with such infringement?

Existing text- and data-mining exceptions typically are worded broadly enough to encompass the ingestion of copyright works for the purpose of enabling TDM. For example, Article 3(1) of the Digital Single Market Directive requires adoption of an exception "for reproductions and extractions made . . . in order to carry out . . . text and data mining of works. . . ."[13]

(v) Would any policy intervention be necessary to facilitate licensing if the unauthorized use of data subsisting in copyright works for machine learning were to be considered an infringement of copyright?

Licensing is the wrong approach to copyright issues related to TDM. The enormous number and variety of works, the difficulty of identifying and locating rights holders, and the small economic value derived from the use of any one work mean that the administrative costs would exceed the fees distributed to rights holders. In other words, a licensing regime would benefit collective management organizations, not rights holders.

---

[12] *See generally* Brief for Medical Historians as *Amicus Curiae* Supporting Defendants-Appellees, *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) (No. 12-4547-cv) (describing researchers' use of pre-1966 medical literature not found in conventional digital databases).

[13] The DSM Directive defines TDM as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations."

25 Massachusetts Avenue NW Suite 300C Washington, DC 20001 P: (202) 783-0070 www.ccianet.org

(vi) How would the unauthorized use of data subsisting in copyright works for machine learning be detected and enforced, in particular when a large number of copyright works are created by AI?

As noted above, CCIA believes that copyright should not prevent the use of data subsisting in copyright works for machine learning, regardless of whether the rights holder authorized the use. In any event, software enables a website or online database operator to detect if information is being harvested. Moreover, the operator could employ technological means to prevent the harvesting of data. The circumvention of those technological measures may be unlawful.

February 14, 2020

Respectfully submitted,

Jonathan Band
policybandwidth
21 Dupont Circle NW, Suite 800
Washington, DC 20036

Joshua Landau
Ali Sternburg
Computer & Communications Industry Association
25 Massachusetts Ave NW, Suite 300C
Washington, DC 20001