

STM SUBMISSION TO
UNITED STATES PATENT AND TRADE MARK OFFICE,
DEPARTMENT OF COMMERCE
REQUEST FOR COMMENTS, 84 FR 58141, pp 58141-58142:
INTELLECTUAL PROPERTY (IP) PROTECTION
FOR ARTIFICIAL INTELLIGENCE (AI) INNOVATION

January 10, 2010

TO: AIPartnership@uspto.gov

The International Association of Scientific, Technical and Medical ([‘STM’](#)) publishers welcomes the opportunity given by the United States Patent and Trade Mark Office, Department of Commerce to submit comments on IP protection of AI innovation (“IPP/AIP”).

STM’s members are at the forefront of digital innovation, providing stored and organized information, tagging and enriching content and creating ontologies. All of these advancements, together with the accuracy of the scientific record maintained by science and academic publishers, help to ensure that machine learning has both depth and accuracy. Publishers are already meeting the needs of the AI era, by developing tools, services and platforms that support and enhance machine learning. STM’s members will offer an ever-increasing number of products and services to make their content available, so that high quality and accurate content sets can be used as training data in machine learning under various licensing schemes. The availability and accessibility of high-quality training data is vital for empowering AI developers with the licensed materials required to achieve both deep learning and to unlock the great potential of AI. Likewise, the wide array of licenses offered by publishers ensures that there are ample, accessible materials available for the continued training of both people and machines.

This brief will under (A) provide input to Questions 1-13 set out in 84 FR 58141, and under (B) make additional observations that STM believes are relevant and as requested.

A. RESPONSES TO QUESTIONS 1-13 OF REQUEST FOR COMMENTS (RFC)

1. *Should a work produced by an AI algorithm or process, without the involvement of a natural person contributing expression to the resulting work, qualify as a work of authorship protectable under U.S. copyright law? Why or why not?*

STM agrees with the overall position reached in [AIPPI's resolution](#) of September 18th, 2019, namely:

Firstly, copyright, as understood in the Berne Convention, should only subsist in works that are original and created with human involvement

Secondly, copyright-like legislation, such as “related rights,” “neighboring rights,” “sui generis rights,” that may cover protected subject matter (non-original photographs and images, non-original databases, phonograms, videograms, broadcast, three-dimensional artistic works or works of craftsmanship) should continue to apply to such subject matter. Said subject matter should not be precluded from protection under these copyright-like statutes protecting for the most part investment in such subject matter, for the mere fact of having been derived or made by virtue of an AI algorithm or process.

Thirdly, it is too early to decide if certain kinds of non-original or non-human expressions should be protected by any newly devised AI legislation, merely because AI algorithms or processes were used in their making.

It should also be borne in mind that AI developers do not lack methods of protecting their investment:

For instance, an AI investor who developed a machine implementing artificial intelligence should not enjoy a related right in the artificially generated works solely because the investor funded or made the arrangements to develop and train the machine. Indeed, the intellectual property rights that such investor or developer may enjoy in the machine itself or potentially in the databases used in the context of the making, training, de-biasing or calibrating the machine, IP subsisting in materials embodied in any artificially generated works, as well as unfair competition law, will offer sufficient means to ensure protection and to allow recoupment of investments. In the absence of proprietary rights in an artificially-generated work, its designer-operator will not be without protection and can oppose or prevent the disclosure of the know-how underpinning its development or deployment, including (a) through the mechanism of trade secret protection, (b) through the general law of tort liability, to oppose or prevent any illegal uses, or (c) through technological protection measures.

For all the foregoing reasons and cognizant of the fact that AI is a fast-evolving field, STM concurs with AIPPI and reserves the right to reconsider this matter further in the light of new developments or future changes in technology.

2. Assuming involvement by a natural person is or should be required, what kind of involvement would or should be sufficient so that the work qualifies for copyright protection? For example, should it be sufficient if a person

(i) designed the AI algorithm or process that created the work;

(ii) contributed to the design of the algorithm or process;

(iii) chose data used by the algorithm for training or otherwise;

(iv) caused the AI algorithm or process to be used to yield the work; or

(v) engaged in some specific combination of the foregoing activities? Are there other contributions a person could make in a potentially copyrightable AI-generated work in order to be considered an “author”?

In the view expressed here, the question of “authorship” of any person involved in the making of any original work would be assessed according to a country’s law, including case law.

For example, under UK law “computer generated works” are protectable and, under case law, this protection does not create a new *category* of works.

The contours of protection depend on a qualification, under UK law, of the *category* of work in question (qualification of a work as, for instance, a literary work or dramatic work, artistic or audio-visual or musical). The author of a computer-generated work is the person or entity responsible for the making of the arrangements of generating the work in question.

From this perspective, the above questions frequently will translate into factual questions of (i) “who” (which human) should be regarded as the author, which in turn may lead to (ii) for what category of work is protection claimed. Depending on the facts, this question can become very intricate, but this is true also for any conventional, human-created work in which copyright subsists. In the STM sector, in fact, many original scientific works are multi-author works, sometimes authored by a very great number of authors but with an arithmetic average of 4.8 authors (about five) per work.

The UK case of [Nova Productions Ltd v Mazooma Games Ltd](#), confirmed on appeal, offers some guidance for what are “computer-generated” works: individual frames shown on a screen when playing a video or computer game were held to be computer-generated *artistic works*. The author of these frames was the person who had devised the rules and logic used to create them. The player of the game was not the author, not having contributed any artistic skill or labor. In this case it was held that *an artistic work* was authored by the person who devised the rules and logic used to create the frames (part of a computer or video game) and not the “player” of the game.

Conversely, CGI technology in the audio-visual sector and computer aided design “CAD” technology in architecture or in the creation of three-dimensional or virtual works of art tend to

attribute copyright to the architect or virtual visual artist, rather than to the owner or creator of the software used in the production of these art works.

3. To the extent an AI algorithm or process learns its function(s) by ingesting large volumes of copyrighted material, does the existing statutory language (e.g., the fair use doctrine) and related case law adequately address the legality of making such use? Should authors be recognized for this type of use of their works? If so, how?

Fair use is an affirmative defense and highly fact-specific. Frequently, users engaged in “novel” uses claim that such uses rate high on one of the four factors of fair use, namely, the character of the use being “transformative.” In this regard, STM would posit that for certain types of research, use of materials for AI and mining purposes is itself becoming a “primary use” of those materials. This is so for two reasons: firstly, deployment of AI tools and mining tools becomes commonplace in research intensive domains of human and business activity; secondly, publishers of STM materials increasingly publish copyright works and associated datasets with AI ingestion technologies in mind. In other words, copyright content of “look-up” type information will increasingly be published in ways that facilitate machine reading, learning, etc.

It follows that licensing, rather than reliance on fair use or other exceptions and limitations, should in most instances be the method of choice for enabling access to copyright works. Licensing will simultaneously be the method to credit authors and recognize owners of copyright works or of subject-matter protected under related rights legislation. The question of how to reward these owners is of eminent importance to publishers. Our members have particular insight into these questions, as STM publishers (alongside educational publishers), are already both users of AI technology and owners and licensors of copyright works and copyright datasets. Works and copyrightable or otherwise protected datasets are eminently licensable to users wishing to engage in machine learning, verification, de-biasing or other forms of tuning or calibration. The added benefit of licensing is that this assures the “provenance” of the data used for AI. In the long term, this will lead to higher quality and more trustworthy AI systems.

Humans and legal entities alike that are using copyright works by virtue of using AI algorithms or processes, or that otherwise would infringe copyright in the process of first training, assessing, refining or verifying the proper functioning of such algorithms, should remunerate owners of any pre-existing works as pure and simple matter of doing business.

Developing AI technology relies on high-quality unbiased or de-biased “data,” understood broadly as very often involving substantial parts of copyright works or having their AI-value arise in the context of other passages of a work in which copyright subsists. AI cannot be trained using “mere facts,” as data only has a meaning in context and the context will frequently be a copyright work. Where the goal is to encourage human-centered high-quality AI technology, the owner of datasets protected as copyright works or as protected subject-matter under a related right should not merely be acknowledged but should be rewarded consistent with the very aim of

copyright to encourage creativity and innovation, and in the case of related rights, incentivizing investment. Licensing will in most instances be the method of choice where mass extraction of content is needed to train, calibrate or verify the operation of any AI algorithm. Frequently copyright works will first have to be adapted and “normalized” in order to be meaningfully deployed in any AI calibration.

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

In STM’s view, existing laws frequently struggle already in relation to conventional copyright works to allocate liability and to determine the contours of infringements. Moreover, as the World Wide Web turns 50, some of the provisions in existing laws devised around the year 2000, e.g. relating to state immunity and safe harbors for online platforms, appear outdated and no longer adequate in a world of online platforms, social networks, big data, and IoT (Internet of Things). In particular, overly broad interpretations of safe harbors in relation to primary infringements, and requiring actual or narrowly construed “constructive knowledge” of specific instances of infringement, should be reconsidered, as infringements increasingly are bound to occur in mass and automated fashion. In other words, a person generating an AI process should have the responsibility to create a technology that is not “willfully blind” to copyright and IP laws, but copyright compliance should be “built-in” to any automated AI processes. The best way to achieve this outcome is to incentivize licensing.

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

STM is of the view that this is best left to agreements between parties and to case law. The above-cited UK case about video game authorship/ownership is instructive (see response to question 2). In the STM sub-sector, AI technology related to image generation and image classification, correction, identification, fraud detection is a very fast evolving field. It seems premature to legislate or regulate this market as it is actively evolving based on existing norms; STM believes that at this stage a determination case by case is most appropriate.

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

As AI becomes more wide-spread, industry self- and co-regulation and the dissemination of best practices will be necessary. Industry as well as cross-industry standards may need to be devised

as AI evolves. For instance, it is plausible that metadata and metadata standards will be developed to identify data and datasets to train certain machines as part of machine learning. It is also plausible that some data may be licensable either on a voluntary mass basis in general or in relation to certain applications of AI.

Where needed to address potential market failures, a broad global consensus should be sought for human-centered AI to move forward on common inter-operable standards and evolving best practices.

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

STM reserves comment on trademark related questions. In general, any AI technology should be consistent with trademark law. For example, to the extent that reference works contain trademark words and references to brands, AI technology indexing those works would need to seek additional permission from trademark rights holders, as appropriate.

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

No comment.

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

In STM's view, the US may wish to consider whether or not additional protection of datasets and databases could be beneficial to spur investment in high-quality data or datasets of assured provenance, or to prevent more adequately against free-riding. Laws of the EU, Mexico, and South Korea may provide models for an appropriate approach in the US. The US may also need to consider the advantage or disadvantage of such frameworks for US-based data or dataset producers compared to counter-parts elsewhere in the world.

While additional protection mechanisms may need to be considered by the US to avoid the potential that innovators move outside the US because of frameworks that make it easier to produce and curate high-quality data in other jurisdictions, it is important to consider also the need for an open science and open data policy framework: data should be made available as open as possible and remain as proprietary as needed. Thus, where raw data is produced for instance by government, any private sector entity should be able to re-use such data openly and freely. These policies are best addressed as such: policies and licensing conditions are more flexible and

adaptable and should be preferred over legislation that may be hard to change as the field of AI evolves.

In this regard, we concur with the French chapter of AIPPI, which AIPPI summarized as follows in its [Summary Report](#): “[The] *French Group is also concerned with a possible curtailing of the creative freedom enjoyed by authors who are natural persons. Indeed, the risk with such sui generis protection is that an author could become an infringer of a machine generated creation. Because of these hesitations, the French Group recognizes that it is still too soon to discuss the relevance of sui generis Protection.*” In other words, we agree that at this time it would be premature to extend investment protection to AI creations per se and that this could potentially unduly interfere with free human creation and re-use of what is already in the public domain.

10. How, if at all, does AI impact trade secret law? Is the Defend Trade Secrets Act (DTSA), [18 U.S.C. 1836](#) et seq., adequate to address the use of AI in the marketplace?

STM offers no comment on trade secret law. Naturally, much high quality data tends to be confidential or secret. Trade secret law remains one of the prime justifications, in addition to the law of privacy and personal data, to maintain a proprietary approach to raw data or to otherwise public or public sector data.

11. Do any laws, policies, or practices need to change in order to ensure an appropriate balance between maintaining trade secrets on the one hand and obtaining patents, copyrights, or other forms of intellectual property protection related to AI on the other?

The constitutional and societal contract and balance in patent and copyright law and in the law regarding trade secrets appears to STM to remain valid. Inventors and learned creators should not be unjustly deprived of the rewards from their inventions and creations. It should be their choice to maximize value through the application of patent law, allowing them to the exclusion of others to work an invention for a limited time; or to disseminate their creative works, in full reliance on copyright; or to maintain a trade secret and rely on their unique know-how for as long as it provides a legitimate competitive edge.

The foregoing is without prejudice to any eminent state interests in secrecy by virtue of state secrecy or security considerations.

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

In STM’s view the issue of Standard Essential Patents and FRAND (Fair, Reasonable and Non-Discriminatory terms) licensing of patents and copyright works in the context of AI could be further examined by USPTO and/or any appropriate US government agencies.

A further topic of study may be IP-related questions pertaining to block-chain technology and state law of escrow of software code and how escrow (in particular in relation to bankruptcy and re-organization) may be a useful tool also in the safeguarding of AI algorithms and processes.

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

STM has no specific comment at this time regarding this question.

B. Additional observations relevant to the topic

STM welcomes the opportunity to make some additional observations on the following three topics associated with AI policy and coordination:

(I) Examples of actual deployment of AI technology in STM publishing;

STM has organized a working group that brings together the expertise of its members on issues of publishing production, standards and technology, IP policy and protection, open science and open data policy, as well as public and government affairs as they relate to AI. We are happy to be a resource to the US PTO and others in the US government on these issues as AI policy development continues. In the interest of providing context for our engagement on these issues, STM provides the following non-exhaustive list of specific examples of areas of application of AI technology in STM publishing:

<p>Recommend articles to readers -- e.g., “more-like-this” or “you might also be interested in...”</p>	<ul style="list-style-type: none"> ● Some recommendation engines are popular and trusted, including some that suggest articles from all publishers based on recent user interactions.
<p>Improve search and browse with dynamic classification by topic</p>	<ul style="list-style-type: none"> ● Taxonomies and classifications provide uniform subject terms across all articles in certain scientific domains. ● Some such taxonomies are semi-automatically learned, maintained and applied with AI tools.

Offer enhanced search & browse based on entity search	<ul style="list-style-type: none"> ● Deployed on some publisher database offerings.
Match author-to-journal, paper-to-editor, or paper-to-peer reviewer	<ul style="list-style-type: none"> ● Some publishers use peer review engines to all or a subset of data from Web of Science and other indexes.
Enrich citations	<ul style="list-style-type: none"> ● Publishing tools identify different structural components of textual references, used in accepted manuscripts during production of final version of record of a published article.
Integrate primary publications with information in secondary databases and vice-versa	<ul style="list-style-type: none"> ● Use of AI tools to match ID of research funders against a registry and improve details about funder official name, ancestors/descendants, country, etc.
Detect and prevent fraud	<ul style="list-style-type: none"> ● Some publishers use Cross Check plagiarism-detection service on every submitted manuscript. ● Some publishers use external fraud detection services that in turn deploy AI tools.
Journal suggestions	<ul style="list-style-type: none"> ● During manuscript submission process, authors are may be automatically given the suggestion of submitting to another or a set of alternative journals.
Machine-generated books	<ul style="list-style-type: none"> ● In 2019, Springer Nature published the first machine-generated book about the latest research on lithium-ion batteries.
Author (name) disambiguation	<ul style="list-style-type: none"> ● Used by some publishers.
Experimental data checking	<ul style="list-style-type: none"> ● Some publishers offer tools that have been trained by machine learning to so that authors and editors can detect assignment errors before publication.
ML-powered research workflow tools	<ul style="list-style-type: none"> ● Chemical Abstracts Service and Wiley have developed ChemPlanner, which has “learned” the organic chemistry literature and predicts novel synthetic routes to compounds for chemists in pharma etc.

Research reliability	<ul style="list-style-type: none"> • Scite.ai uses ML and NLP to determine whether a paper is supported or contradicted by the papers that cite it.
----------------------	--

The above list is **by no means comprehensive or a reflection of the full potential the STM sector** sees in applying AI in an ethical, trustworthy, human-centered and value enhancing way.

- (II) The need to monitor convergence and possibly develop global consensus on industry or sector-specific frameworks for policy and ethical guidelines;

STM is aware of a great number of efforts in the policy space to define, categorize and devise sensible objectives and policies related to AI. For convenience, these efforts may be classified as efforts of public bodies, corporate perspectives and public perspectives.

In STM's view it is important to see to what extent these different perspectives converge and where they emphasize different attributes of AI.

By way of example, but without endorsing any particular framework, we refer to the EU Commission framework "Ethics Guidelines for Trustworthy Artificial Intelligence" (April 2019), the framework of IEEE's "Global Initiative on Ethics of Autonomous and Intelligent Systems" (2nd edition 2017) and the Public Voice's "Universal Guidelines for Artificial Intelligence" (October 2018).

- (III) The need for broad consultation on these issues;

Any state intervention, government policy or regulation should be preceded by a definition of broadly consulted and then defined desired outcomes and to choose an evidenced-based approach in order to drive sound policy formulation and regulatory impact assessments.

Finally, STM is grateful for this consultation not least because it demonstrates the USPTO's commitment to an evidence-based, open and transparent and inclusive consultative approach to approaching AI for the benefit of humanity, societal and planetary welfare.

STM Contacts

Ian Moss, Chief Executive Officer, STM, Oxford, UK

ian@stm-assoc.org

David Weinreich, STM Director Public Affairs, Americas, Washington, DC

weinreich@stm-assoc.org

Outside Counsel:

Carlo Scollo Lavizzari, Lenz Caemmerer Attorneys & Notaries, Basel, Switzerland

csl@lclaw.ch

About STM

STM supports members in their mission to advance research worldwide. Together as partners in science and policy, we serve society by developing standards and technology to ensure research is of high quality, trustworthy and easy to access. STM promotes the contribution that publishers make to innovation, openness and the sharing of knowledge. For us, supporting the growth and sustainability of the research ecosystem means embracing change. For the benefit of the community we provide data and analysis for all involved in the global activity of research.