

Alliance of Independent Authors Commentary and Suggestions

Artificial intelligence and machine learning systems (“AI” or “AI systems”) today autonomously generate creative works which would, were they created by humans, be eligible for copyright protection. AI systems are creative and communicative, rational and analytical, independent and autonomous, efficient and accurate, and capable of assigning free choice from alternatives.

The use of AI systems in the production of works grows ever more common in what is being called “the 3A era” of advanced technology and the “self-publishing 3.0” era of independent author-publishing. Authors, publishers and other creators now use, and want to expand their use, of AI tools to generate text, as working drafts and completed works.

This will foster creative expression (among humans and machines) and is to be encouraged—but many questions remain to be answered, as a matter of urgency.

To be effective, a contemporary copyright policy must be framed around how author-publishers actually work, trade and negotiate today, and how today’s readers actually discover, buy and read books. This submission addresses what we at the Alliance of Independent Authors consider to be the two most pressing questions around the copyright of text, artworks and other content generated by AI systems: ownership and accountability.

There are three routes that copyright courts can take when considering creative works generated by an AI system

1. deny copyright protection for works that have been generated by an AI system
2. attribute copyright of such works to the creator of the AI system
3. attribute copyright of such works to the creator of the AI system and the (already copyrighted) creators of the works used to feed the system

We favour option 3, while acknowledging the challenges involved.

The ability of authors and publishers to make a living rests on copyright law. We argue that traditional copyright law has become outdated with respect to works created by AI systems and call on policy makers to rethink current copyright law governing AI systems and introduce an extension to the copyright framework which is applicable to the new (3A and self-publishing 3.0) era of advanced, automated, autonomous AI systems.

Such an extension needs to consider the rights of creator of the AI system as well as the copyright of the creators of the works used to feed the system.

The question for us is not so much whether a finished work created by AI is worthy of copyright. It seems to us that it indubitably is and a (Chinese) court has already deemed it so.

The issue is ownership of that finished work, which is generated by a system that has human creator(s) and is fed by content created by humans.

Any extension to copyright law must include a new right related to data (copyrighted works) used to train an AI system.

Two core challenges in the era of global self-publishing are: There is no such thing as international copyright. International law is based on a combination of domestic legal systems, regional and international regimes, as well as bilateral and multilateral treaties and agreements. And a myriad of social, technological, economic and legal developments now sees copyright legislators and policy makers in various jurisdictions—especially the EU—increasingly in conflict with writers and artists, critics and scholars and free expression activists.

We need not just balanced, robust and flexible copyright law but the education to understand what's at stake and the means to interpret and assert it, taking into account the shifting roles each individual creator experiences as author, reader, student, consumer and researcher.

Our concern is to see a healthy, supportive, and functioning global copyright environment that balances the benefits of copyright ownership with the flexibility to run a successful and sustainable online author business, in which AI systems already figure significantly, and in which they will play an increasing part in the coming years.

Issue 6: Authorship and Ownership

12. AI applications are capable of producing literary and artistic works autonomously. This capacity raises major policy questions for the copyright system, which has always been intimately associated with the human creative spirit and with respect and reward for, and the WIPO/IP/AI/2/GE/20/1 page 5 encouragement of, the expression of human creativity. The policy positions adopted in relation to the attribution of copyright to AI-generated works will go to the heart of the social purpose for which the copyright system exists.

If AI-generated works were excluded from eligibility for copyright protection, the copyright system would be seen as an instrument for encouraging and favoring the dignity of human creativity over machine creativity.

This is **not** the role of copyright law or policy

If copyright protection were accorded to AI-generated works, the copyright system would tend to be seen as an instrument favoring the availability for the consumer of the largest number of creative works and of placing an equal value on human and machine creativity.

This **is** the role of copyright law and policy, in line with the values of free expression, democratisation of publishing, and spread of literacy that has long united copyright activists and policy makers.

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

This has already been decided in one jurisdiction. In Jan 2020, a Chinese court assigned copyright to an AI, Tencent's Dreamwriter, which has been producing articles for some years. The court "found that the article's articulation and expression had a "certain originality" and met the legal requirements to be classed as a written work — thus it qualified for copyright protection" and awarded copyright ownership to Tencent for all stories its AI produces.

The question of how the AI was created, whether creator copyright was respected in the creation of the AI has not been made clear.

Does a creator need to own the AI for their copyright to be protected? Clearly this is not possible for individual creators

(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?

The AI system is a creation in itself, yes and copyright has been attributed to AI-generated works in the Chinese courts.

In giving consideration to the concept of a "legal personality" we must ask how useful that concept is to the creators of the work. Where is the original creator, on whose work the system was created, to be credited? Their rights must be balanced with the rights of the owner of the AI system.

We say this in recognition that it is challenging for any platform or service to vet all the material claimed as copyright by putative rights holders. Copyright clearance is undoubtedly a challenge not just in negotiating licenses to works, but figuring out who the works belong to in order to begin those negotiations. It is, however, a challenge that must be met if copyright is to be meaningful in an AI context.

(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?

Considering AI generated work as a performance is an interesting option. How would that work in practice? Where do prior creators figure in this treatment of AI generated works as performances?

Issue 7: Infringement and Exceptions

13. An AI application can produce creative works by learning from data with AI techniques such as machine learning. The data used for training the AI application may represent creative works that are subject to copyright (see also Issue 10). A number of issues arise in this regard, specifically,

(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?

Yes, use of copyright works to train machine learning models should be considered a breach of copyright. However, we recommend that a new right be created related to data (copyrighted works) used to train a model. Those wishing to use an AI system to create must license the works on which the system is trained.

(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?

This is a very important question, particularly around inherent bias in machine learning models. If an AI model can only learn on out of copyright works, or those explicitly licensed, it is likely to be trained on those works that are only written by middle/upper-class white men in English-language, or religious texts, or other works that entrench the current lack of diversity in publishing.

(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?

Yes, in the same way as fair use allows for research purposes currently. Creative Commons licenses could also be assigned.

(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?

The exceptions for text and data mining may provide a model for consideration of AI systems.

Text and data mining often involves computer programs copying large amounts of copyright material with the intention of allowing scientists and other researchers to discover patterns, trends and other information that cannot be detected through human reading. Scientific publishers have long recognised the importance of text mining and offer various applications

for use by researchers. UK copyright law provides an exception that allows researchers to make copies of works for such text and data analysis.

The exception stipulates that where a user has lawful access to a work they can make a copy of it for the purpose of computational analysis under the following conditions:

1. The computational analysis must be for the purpose of non-commercial research
2. Wherever practically possible, the copy must acknowledge the source

Copyright is infringed if copies made for text and data analysis are sold, let for hire, transferred to another person, or used for purposes different than those permitted by the exception (although the researcher could receive explicit permission to do any of these).

Although text mining is mainly concerned with literary works, the exception covers all categories of copyright works, and a parallel exception applies to recordings of performances. This may provide a model for consideration of AI systems more generally.

(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?

Unauthorized use of data in copyright works for machine learning is an infringement of copyright and such authorisation needs to be sought. the necessary legal and policy framework to ensure compliance is needed.

Again, text mining may provide a model. Researchers rely on the exception to allow text and data analysis but collaboration between database owners and researchers remains a fundamental part of the process, and the respect for copyright a fundamental principle.

(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? WIPO/IP/AI/2/GE/20/1 page 6

The Alliance of Independent Authors' guide to blockchain *Authors and the Blockchain: Towards a Creator Centered Business Model* explores how the new technology of blockchain allows a work to be tracked through the chain of use and how acknowledgement and compensation for creators (copyright holders) can be given precedence.

Services are coming into the publishing and self-publishing space with the intention of providing such solutions e.g. Access copyright, StreetLib.

The use of blockchain to monitor, attribute and recompense copyright is worth considering in relation to AI systems. Blockchain provides the ability to track micro-percentages so that, for example, 100,000 books read into an AI system could each have their copyright acknowledged and a micro-percentage assigned for use of the text.

Issue 8: Deep Fakes

14. The technology for deep fakes, or the generation of simulated likenesses of persons and their attributes, such as voice and appearance, exists and is being deployed. Considerable controversy surrounds deep fakes, especially when they have been created without the authorization of a person depicted in the deep fake and when the representation creates actions or attributes views that are not authentic. Some call for the use of deep fake technology to be specifically banned or limited. Others point to the possibility of creating audiovisual works that might allow the deployment of popular or famous performers after their demise in a continuing manner; indeed, it might be possible for a person to authorize such use.

15. Should the copyright system take cognizance of deep fakes and, specifically,

(i) Since deep fakes are created on the basis of data that may be the subject of copyright, to whom should the copyright in a deep fake belong? Should there be a system of equitable remuneration for persons whose likenesses and “performances” are used in a deep fake?

There needs to be a careful distinction between “deep fakes” created without the person’s permission, and those creators/artists/actors etc who license their voice and/or image to be used in further creation.

Actor Samuel L Jackson [licensed his voice to Amazon Alexa](#) for their first voice synth, described by The Verge as a “lightweight deep fake”.

Joanna Penn, author, audiobook narrator and podcaster at The Creative Penn Podcast, is working with Descript.com to create a “voice double” with the intention of licensing her voice once the technology is mature enough.

The likelihood is that a market for licensable voices will emerge, so the subsidiary rights that allow this to happen need to be in place.

Issue 10: Further Rights in Relation to Data

(i) Should IP policy consider the creation of new rights in relation to data or are current IP rights, unfair competition laws and similar protection regimes, contractual arrangements and technological measures sufficient to protect data?

Yes, IP policy should create a new right for the use of data to train AI models.

(ii) If new IP rights were to be considered for data, what types of data would be the subject of protection?

All books licensed by authors to publishers which have not got a specific clause around AI. For example, a large romance publisher who has a contract for the term of copyright (e.g.

Harlequin) may “own” thousands of books that they could use to train an AI system to write romance novels according to their framework. But the original license has not licensed the publishers to use their works to train an AI system.

(iii) If new IP rights were to be considered for data, what would be the policy reasons for considering the creation of any such rights?

Publishers with a large amount of data already licensed from authors e.g. newspapers, book publishers, could use that data to train models and replace the humans who created the original work.

(iv) If new IP rights were to be considered for data, what IP rights would be appropriate, exclusive rights or rights of remuneration or both?

Both.

See Clause 3 of the Alliance of Independent Authors *Copyright Bill of Rights: Eight Rights that Underwrite Author Income*:

“Authors have the right to fair remuneration from publishing and self-publishing companies who trade in their copyright.”

Copyright that does not include a right of remuneration is of little use to a working author.

(v) Would any new rights be based on the inherent qualities of data (such as its commercial value) or on protection against certain forms of competition or activity in relation to certain classes of data that are deemed to be inappropriate or unfair, or on both?

(vi) How would any such rights affect the free flow of data that may be necessary for the improvement of AI, science, technology or business applications of AI?

The existing exceptions in UK law provide for such

(vii) How would any new IP rights affect or interact with other policy frameworks in relation to data, such as privacy or security?

The existing exceptions provide for such

(viii) How would any new IP rights be effectively enforced?

The enforcement of rights is notoriously difficult in publishing. We look to new technology, particularly blockchain for books, to help solve some of these challenges

That copyright has real and significant value can be seen in how it is ever more vehemently contested by three competing interests regularly contesting each other across a digital, machine learning and AI battleground:

- **Big Tech:** Internet publishing platforms like Google, Facebook and Amazon that authors use to publish and/or promote books.
- **Big Content:** global media corporations like Penguin Random House, News International and Hollywood to whom authors licence publishing rights. Here we also find large self-publishing services like Author Solutions, some of whom grab rights as well as charging service fees.
- **Big Legal:** Large legislating territories and blocks like the USA or the European Union who aim to modify the power of big tech and big content.

All of these affect how independent authors work and earn but none speak directly to, or for, such authors. Neither do trade publishers, literary agents nor, sadly, many authors' organizations.

The interests of an Internet tech giant like Google, Facebook, or YouTube; a social media company like Twitter or LinkedIn, a news organization like The Guardian or New York Times, a traditional publishing house like Random House or Penguin, a literary agency, a legislative body, a trade-published author, or even an authors' representative body, may not align with those of the independent, self-publishing author. All are investigating the use of AI in their work, as are some advanced indie authors.

This is taking place against a background of heated debate about existing copyright law.

Early in 2019, opposition to a proposed EU Copyright Directive (now passed) saw, among other protests, an online petition¹ gather more than five million signatures, a Polish newspaper printing a blank front page², Italian Wikipedia blacking out, and concerned German citizens³ taking to the streets.

Many of the free access protesters against Article 13 (now Article 17) of the controversial EU Directive were authors and other creators. They do not want the changes Big Content is asking for, purportedly on their behalf. These authors want the free and flexible Internet we already have and believe that the foundational copyright law we have is already robust and flexible enough to meet today's needs.

They are concerned that some changes which Big Content depicts as updating and improving copyright law for the digital age do not work well for independent authors and their readers.

And within the author community itself, indie authors take many different approaches and hold widely varying opinions about copyright issues like piracy, plagiarism and digital rights management.

Increased author confidence is now showing in the variety of business models they are adopting, in the variety of outlets and platforms they use, and in the variety of content (not just text but audio and video) that many now publish, not least on their own websites.

Understanding this changed landscape is challenging for many authors and for the government bodies that support them: ministries responsible for business, entrepreneurship, culture, intellectual property, the knowledge economy, and related industry organizations.

It's challenging, yes, but to be effective, contemporary copyright policy must be framed around how author-publishers actually work, trade and negotiate today, and how today's readers actually discover, buy and read books.

We must all get better at identifying, tracking, quantifying, and supporting independent authors as creative digital micro-businesses. This includes considering the impact on the independent author when reviewing and updating policy and legislation around issues such as copyright.

We urge interested parties from all sides—including authors—to better understand the trading environment and copyright challenges of the self-publishing author in a digital environment and how copyright policy, and its implementation, directly affects their ability to earn a living and contribute to the fabric of society.

We need not just balanced, robust and flexible copyright law but the education to understand what's at stake and the means to interpret and assert it, taking into account citizen's shifting roles as authors, readers, scholars and consumers.

The Alliance of Independent Author's *Copyright Bill of Rights* paints a picture of a healthy, supportive, and functioning global copyright environment that balances the benefits of ownership with the flexibility to run a successful and sustainable online author business.

New technologies like AI, the blockchain, cryptocurrency and other tools are fast coming on track, with the power to further shift economies, businesses and behaviors.

More than ever, authors' advancement depends on their ability to think critically about how copyright supports their work as author business models evolve in the digital, machine learning and AI age of Self-Publishing 3.0.

1 [Change.org](https://www.change.org/p/european-parliament-stop-the-censorship-machinery-save-the-internet). *Stop the censorship machinery! Save the Internet!* <https://www.change.org/p/european-parliament-stop-the-censorship-machinery-save-the-internet>

2 France24. *Polish dailies print blank front page in EU copyright appeal*. March 25, 2019. <https://www.france24.com/en/20190325-polish-dailies-print-blank-front-page-eu-copyright-appeal>

3 CNBC. *Thousands protest against controversial EU internet law claiming it will enable online censorship*. March 25, 2019. <https://www.cnbc.com/2019/03/25/protesters-in-germany-say-new-eu-law-will-enable-online-censorship.html>