



Creative Commons response to the UK Consultation on Copyright and Artificial Intelligence

20 February, 2025

Submitted by email: copyrightconsultation@ipo.gov.uk

Creative Commons is pleased to respond to the United Kingdom Intellectual Property Office, Department for Science, Innovation and Technology, and Department for Culture, Media and Sport's open consultation on "[Copyright and Artificial Intelligence](#)." In our comments, we explain our overall approach to creativity, copyright, text and data mining (TDM), and artificial intelligence (AI), and then respond to the consultation's questions.

Contents

[About Creative Commons](#)

[Summary](#)

[About Creative Commons' approach to copyright, TDM, and AI](#)

[Responses related to copyrighted works as "inputs" to TDM and AI](#)

[Responses related to "opt-outs," and Option 3](#)

[Responses related to licensing](#)

[Responses on training data transparency](#)

[Responses related to questions on copyrightability of computer generated works \(e.g. "AI outputs"\)](#)

[Responses related to potential infringement in the outputs](#)

About Creative Commons

Creative Commons is a nonprofit organization that helps overcome legal obstacles to the sharing of knowledge and creativity. To that end, we publish a suite of licenses and legal tools designed to give every person and organisation in the world a free, simple, and standardised way to grant copyright permissions for creative and academic works, ensure proper attribution, and allow others to copy, distribute, and make use of those works. In addition, we work closely with major institutions and governments to create, adopt and implement open licensing and ensure the correct use of CC licenses and CC-licensed content. We also engage in public policy, contributing to shaping the laws and rules governing sharing and the commons, including any technological impacts, with a particular emphasis on copyright reform in the public interest.

Summary

In short, we believe copyright should be appropriately balanced in order to facilitate TDM. We are grateful that the UK is now taking steps to modernise its copyright laws, and discourage the government from treating this as an “all or nothing” choice or a zero sum game, where one “side” wins and another loses. Instead, we believe that a clear, fair exception can be carefully crafted to serve the public interest. An exception can be complemented by ways for rightsholders to express their preferences about use of their works in AI training, and we highlight key principles to ensure such “opt-outs” or “preference signals” are workable and socially beneficial.

To be clear, we do have concerns about AI and its development and use: for example the impact these tools will have on artists and creators’ jobs and compensation, the use of these tools to develop harmful misinformation, to exploit people’s privacy (e.g., their biometric data), create safety and security risks, or perpetuate biases, as well as the sustainability of the information commons. These issues particularly impact the open sharing community, which often is producing information, scholarship, and creative output in the public interest, without the same commercial incentives for creation as many other content creators. That said, it is important to recognise that copyright is not the only lens to address these problems, and our response here cautions against using copyright law as the primary tool to address all social concerns in place of other, potentially more appropriate remedies.

About Creative Commons’ approach to copyright, TDM, and AI

A founding insight of Creative Commons is that all creativity builds on the past. Creators necessarily learn from, and train their own skills by, engaging with pre-existing works and artists — for instance, noticing the style in which musicians arrange notes, or building on surrealist styles initiated by visual artists. Likewise, scientists and researchers build on past discoveries and the existing literature to gain a better understanding of how the world works and to progress ideas.

TDM is a way to study and analyse existing works, using machines, in order to create new insights and materials. TDM typically consists of computers analyzing huge amounts of text or data, and has the potential to unlock huge swaths of interesting connections between textual and other types of content. Understanding these new connections can help organize our shared knowledge, enable new research capabilities that result in novel scholarly discoveries and critical scientific breakthroughs. Because of this, text and data mining is increasingly important for research. While much of the discourse around TDM as applied to generative AI has focused



on the creation of artistic works, TDM and AI have uses that can help generate advances across science, education, healthcare, and other domains that are of significant importance to society.¹

One of the motivations for founding Creative Commons was also offering more choices for people who wish to share their works. Through engagement with a wide variety of stakeholders, we heard frustrations with the inflexibility they seemed to face with copyright's default framework, i.e. all rights reserved. Instead they wanted to let the public share and reuse their works in some ways but not others. Our engagement revealed that people were motivated to share not merely to serve their own individual interests, but rather because of a sense of societal interest. Many wanted to support and expand the body of knowledge and creativity that people could access and build upon — that is, the commons. Creativity depends on a thriving commons, and expanding the array of choices was a means to that end.

Similar themes have come through as we have consulted with our community about generative AI. Obviously, the details of AI and technology in society today are different from 2002 when the first CC license was published. But the challenges of an all-or-nothing system where works are either open to all uses, including AI training, or entirely closed, are a through-line. So, too, is the desire to do so in a way that supports creativity, collaboration, and the commons.

In turn, we have been engaging actively with stakeholders to avoid “all-or-nothing” choices in the age of AI. Specifically, we have been exploring the development of “preference signals” for AI training – a way for copyright holders to make requests to potential users about some uses, not enforceable through the licenses, but as an indication of their wishes. Preference signals raise a number of tricky questions, including how to ensure they are a part of a comprehensive approach to supporting a thriving commons — as opposed to merely a way to limit particular ways people build on existing works, and whether that approach is compatible with the intent of open licensing. At the same time, we do see potential for them to help facilitate sharing in the public interest. We have heard from many different people, including artists and rightsholders, as well as developers of AI technologies, and we believe that it is possible for policy to guide the development of AI in ways that benefit all.

Responses related to copyrighted works as “inputs” to TDM and AI

A TDM exception is consistent with copyright's purpose and function – providing sufficient incentives for creativity in order to benefit the public. No creativity happens in a vacuum, purely original and separate from what's come before; people observe the ideas, styles, genres, and other tropes of past creativity, and use what they learn to create anew. Copyright law has always reflected this fact – by differentiating between protectable expression and uncopyrightable ideas

¹ See examples here:

<https://www.knowledgerights21.org/news-story/kr21-principles-on-artificial-intelligence-science-and-research/>

and facts, through appropriate exceptions, and by ensuring broad freedoms associated with the public domain, among other limitations to copyright.

Today, people are using TDM, including in its application to develop AI, including general purpose and generative AI, to study, analyse, and learn from past works. Just as people learn from past works by studying them directly, they can use computers to study works – for example, “training” generative AI by analyzing past materials in order to extract uncopyrightable information and make non-infringing outputs.

Given how digital technologies function, engaging in TDM may involve making intermediate copies of copyrighted works in order to analyse them. Unfortunately, today, UK law encumbers this sort of analysis in overly broad ways by limiting non-expressive uses – that is, uses that involve copying, but do not communicate the protectable expressive aspects of the underlying work so that it can be read or otherwise enjoyed. Treating this sort of copying as per se infringing copyright in effect shrinks the commons and impedes others’ creativity, by restricting legitimate uses only because they are machine-enabled and involve copying as an intermediate step. It extends creators a monopoly over ideas, genres, and other concepts not limited to a specific creative expression, as well as over new tools for creativity.

In general, we think that analysis of existing works in order to derive uncopyrightable elements or make otherwise non-infringing uses should be permissible under copyright law, even if it involves making a copy of a whole work as an intermediate step, such as through TDM. The UK government is correct to revise its copyright law to this end.

In so doing, the UK can build on and learn from examples around the world on how to best craft such an exception in ways that are clear and fair. As one recent paper notes,

“Globally, the binary policy debate that assumes that text and data mining and AI training must be categorically condemned or applauded has been eclipsed by a more granular debate about the specific circumstances in which the unauthorised use of copyrighted works for AI training should be allowed or prohibited. Countries that have hesitated until now to modernise their copyright laws in the area of AI training have several templates open to them and little reason for hesitation.”²

² Sag, Matthew and Yu, Peter K., The Globalization of Copyright Exceptions for AI Training (October 04, 2024). Emory Law Journal, Vol. 74, 2025, Forthcoming, Texas A&M University School of Law Legal Studies Research Paper No. 24-75, Emory Legal Studies Research Paper Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4976393> or <http://dx.doi.org/10.2139/ssrn.4976393>, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4976393

Responses related to “opt-outs,” and Option 3

We are also supportive of preference signals that help people state how they wish for their works to be used. It is critical that such signals are narrowly tailored to serve the public interest, and to recognise that the development of these signals is also not a binary choice. They could be crafted to differentiate different types of uses and users. Like robots.txt, they could be developed as voluntary standards and norms, providing flexibility and adaptability for the signals to evolve over time and be interpreted contextually. Alternatively, as Option 3 imagines, they could be part of a legally binding “rights reservation” or “opt-out” scheme, although a more rigid, prescriptive approach instantiated in law makes it even more essential to carefully address implementation and operational challenges at the outset.

More specifically, any “opt-out” system would need to address key challenges and in turn meet important criteria, including:

- Differentiate and broadly protect research uses: In the EU’s [2019 copyright directive](#) there are two prongs to the TDM exception. Article 3 provides a TDM exception for certain research uses, and this protection is not subject to the rights reservation also referred to as “opt-out.” Meanwhile, Article 4 provides an exception for all users, except that respect for the rights reservation is required. The UK should follow this basic structure, and consider a broader, clearer exception for research, such that, in addition to research and cultural heritage institutions, nonprofits and other research labs unaffiliated with research institutions can also easily avail themselves of the exception. As researcher advocacy organization KnowledgeRights21 has argued, “public research institutions and knowledge valorisation is held back by the artificial and unworkable distinction that EU / UK copyright law makes between commercial and non-commercial research. Despite European governments’ claimed strong support for public-private partnerships in research, the reality is that the distinction between commercial and non-commercial research that European copyright law makes means that working with and sharing information between partners in the context of knowledge valorisation projects is impossible.”³
- Require machine readable rights reservations by rightsholders: Opt-outs should be provided via clearly understandable and machine readable formats, such that it is easy for TDM users to readily determine the copyright status of a work. Moreover, it should only be rightsholders of a particular copyright-protected work that can exercise the reservation as, after all, they are the ones who control the pertinent rights.
- Ensure flexibility for new technologies: TDM and its uses will continue to evolve. The exception should be able to accommodate TDM not just in today’s context of generative AI but also in future contexts. As such, the legislative framework the UK pursues should

3

<https://www.knowledgerights21.org/news-story/kr21-principles-on-artificial-intelligence-science-and-research/>

be crafted in such a way that it is as future proof as possible or at the very least includes a mechanism for review, without undermining legal certainty.

- Ensure feasibility: in line with the above, it is crucial that rights reservations be readily identifiable and available. They must not be highly difficult or costly to implement for rightsholders or users. To that end, developers should not be held liable where the validity of a rightsholder claim is uncertain, such as when multiple entities offer conflicting preferences or when someone fraudulently or inadvertently or mistakenly claims to be the rightsholder.
- Facilitate interoperability: where appropriate, rights reservation systems should build on and complement existing tools like robots.txt. Where practical, it is also important to develop systems that work across media types and across jurisdictional borders.

Responses related to licensing

The consultation asks about how to support effective licensing. Here, it is important to recognise that the creation of a TDM exception is not antithetical to licensing as a whole. Rather, TDM is already facilitating the creation of new types of tools and services, which in turn are opening up new licensing opportunities in *other* uses beyond protected TDM. For instance, while generative AI developers may train their base models on copyrighted works under national copyright limitations and exceptions, many are still working to obtain licenses over works for uses that go above and beyond permitted uses, such as displaying significant expressive text from a particular news article or expressing images with a high degree of similarity to protected works. This is not a new phenomenon, but rather consistent with the history of copyright – for instance, uses of the video camera and video cassette recorder (VCR) were built on copyright’s limitations and exceptions, but gave rise to a huge market for film rentals and sales.

To be clear, a new TDM exception should not be conditioned on any affirmative requirement to seek licenses. To the extent the government pursues Option 3, licensing would not be necessary in situations where the relevant rights have not been clearly reserved by the rightsholders. Moreover, it should be possible for rightsholders to use granular preference signals to reserve their rights in more specific, less binary ways (i.e., deciding not to reserve rights for some uses and/or some users, while reserving them for other contexts). This system should also support open or otherly permissive licensing.

Responses on training data transparency

CC generally supports measures that provide the public with transparency into AI models and the data used to train them. Transparency can help build trust in AI models by allowing others to “look under the bonnet” and investigate how those models work. We believe disclosure of training data in the development of AI models is desirable for numerous reasons, including

suitability for academic and research purposes, supporting the sustainability of communities and institutions where training data is created, crediting sources, and supporting fundamental rights. However, on balance, these measures should not be so onerous as to make it impractical or impossible for developers to develop socially-beneficial technology. Moreover, transparency should not be enforced through existing or newly created copyright measures, as it is in fact a requirement that is fully independent from the copyright status of the training data.

Responses related to questions on copyrightability of computer generated works (e.g. “AI outputs”)

Works not authored by humans should not be eligible for copyright protection - i.e., they should be in the public domain.

Expressions of human creativity should be eligible for copyright, even where realised with AI-generated assistance. The current standards for copyrightability of a creative output should remain as they are in the context of AI, where human authorship is a prerequisite for copyright protection and machine authorship is a non sequitur. We recognise that determining which parts of a work are authored by a human in such situations will not always be clear — indeed, just as determining authorship of individual parts of a work is not always clear in works created without AI assistance.⁴

There is a tremendous variety of tools and processes lumped together under “generative AI” which involve varying degrees of human involvement in the creative process. We consider the approach of granting copyright to the human-authored parts of a work while exempting the AI-generated parts to be the correct approach, even where drawing the distinction in practice may be difficult. This problem is not unique to AI: authors have always created work building upon and remixing uncopyrightable and other public domain works, by remixing old songs, translating or editing old texts, incorporating government-authored works, and other such artistic components from the public domain.

Responses related to potential infringement in the outputs

It is possible for AI-generated works to implicate the exclusive rights of rightsholders in some cases. We cannot enumerate all cases specifically; however, there are many possible ways this can happen, and we urge caution and flexibility in determining when an output has infringed exclusive rights. It is possible for an output to be infringing; it is also possible that an output including material under copyright may be using it in a way that is fair, legitimate, and lawful.

⁴ Notably, the US Copyright Office drew similar distinctions in its report. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>

Liability for any infringement in this case should depend heavily on the facts of the situation, as the nature of generative AI tools and usage of those tools can vary greatly. For example, where a user has made a knowing attempt to get copyrighted material they have previously had access to as output from an AI tool (by directing prompts specifically toward that output, for example), that user has simply used the tool to reproduce the material, in the same way that one might use other tools to create a copy of an existing work. This may be non-infringing.

We would caution against considering a simple attempt to make reference to a rightsholder's work as an infringement—for example, a user requesting artwork in the style of an existing artist or a user quoting from a literary work in order to comment or parody. Copyright law already contemplates such cases, and already deals with some ambiguity that depends on facts: was a creator simply being inspired by the style of an artist, which is not an infringement and is critical to free expression and the development of artistic genres? Or was a creator attempting to reproduce copyrightable elements in an infringing manner? Likewise, quotation for purposes of criticism, commentary, parody or pastiche is largely considered fair and legitimate, even though it involves use of works protected by exclusive rights.

The debate about how copyright should apply to generative AI has often been cast in all-or-nothing terms—does it infringe on pre-existing copyrights or not? The answer to this question is certainly that generative AI *can* infringe on other works, but just as easily it may not. The criteria that already exist in copyright to determine substantial similarity between works are apposite to address these questions with regard to AI outputs. Our concern is that overregulation of generative AI could easily tip the carefully crafted balance achieved within the copyright system, undermining AI tools' ability to help people realise their creative visions, advance research, improve productivity, and positively transform society.

Thank you for the opportunity to respond. We remain at your disposal should you require any further information.

[End of document]
