

Copyright Infringement by Generative AI Tools Under US and UK Law: Common Threads and Contrasting Approaches.

February 13, 2025

This article was originally published in the [New York Law Journal](#) on February 10, 2025

This article was co-authored by Jack Griem, Chair of Carter Ledyard's [Intellectual Property practice](#), and Robert Lands, Head of IP and Commercial at Howard Kennedy.

Introduction

Generative artificial intelligence is transforming the way we do business, make art, and learn new information. Every day, another company launches or updates a product or service that integrates generative AI. Many of these products and services are available on the internet, available in many countries at once.

Intellectual property law is famously country-specific. Each country has developed its copyright law over many years, through a mixture of statutory provisions and decisional case law. At the same time, most countries are now members of the Berne Convention, which enables enforcement of the copyright in a work created in one Berne signatory country to be enforced in any other signatory country. The law governing enforcement of copyright, however, remains country-specific.

Two countries at the heart of the development of generative AI are the United Kingdom and the United States. Many copyright infringement actions have been filed in both countries by owners of copyrights in works used to train generative AI models, against companies that developed and released AI tools. This article explores how the courts in the US and the UK are addressing the key copyright infringement issues as they relate to generative AI models and output, and highlights the differences, particularly in the area of “fair use”/“fair dealing” and statutory provisions unique to each country.^[1] Fair Dealing is often confused with the US notion of Fair Use. They are both concerned with whether the unauthorised use of a copyright work is fair to the copyright owner, but despite this shared ethos, they are not the same.

If an original work is used to train a “large language model” (“LLM”), like ChatGPT or Google’s Bard, or to create training sets or algorithms for an AI image generator, or if a copyrighted work is used as an input prompt, does the resulting data set, algorithm, or expression infringe the copyright on the original works?

At a very high level, LLMs and AI image generators take apart the works they are trained on, transforming them into component parts of a neural network that are then weighted using mathematical principles. These AI-powered engines can then create new expression by breaking an input prompt into weighted tokens that are run through the engine.

Companies will have to use their judgment in balancing the risk and reward of adopting generative technology, given the mismatch between the speed of legal decision making and generative AI technology advances. Law is slow, but technology is fast, and generative AI is especially fast – the world’s largest technology companies are competing with each other to release improved versions of generative AI

models. Decisional law is made slowly, as cases are brought and judges issue decisions, which are then appealed. Major changes in statutory law often take years to be enacted, particularly where there are opposing economic interests.

US Approach

In the United States, copyright law protects original works of authorship, including literary, dramatic, musical, and artistic works. The United States Copyright Act defines a “derivative work” as “a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted.” (17 U.S.C. § 101). Copyright holders have the right to create, control and license “derivative works” based on their copyrighted work, and the “derivative works” are also copyrightable, to the extent that they add new, copyrightable expression. (17 U.S.C. § 106).

Historically, derivative works were created directly by a person writing or drawing a new work, perhaps with the use of a tool or technique that created some predictable effect or change. The question of whether AI data sets, generators, and AI-generated content are derivative works of copyrighted works is at the core of the class-action lawsuits filed by writers and visual artists against companies that provide LLMs and AI art generators like OpenAI (creator of ChatGPT), Meta and Stability AI, Midjourney and DeviantArt.

To be considered a derivative work, the new work must “copy”, in some way, from an original work. However, a work can be copied from one medium to another, and be adapted for the new medium in the process, and still be a “copy.” *ABS Entm’t, Inc. v. CBS Corp.*, 908 F.3d 405, 416, 418 (9th Cir. 2018) (citing the treatise Nimmer on Copyright). The amount of incorporation required to create a derivative work is not always clear. Existing court decisions have generally held that the new work must incorporate a substantial amount of the original work in order to be considered a derivative work. See, e.g. *Caffey v. Cook*, 409 F. Supp. 2d 484, 496 (S.D.N.Y. 2006) (citing Nimmer on Copyright). In order to be an infringement, a “substantial similarity”

must exist between the defendant’s work and the protectible elements of plaintiff’s. *Peter F. Gaito Architecture, LLC v. Simone Dev. Corp.*, 602 F.3d 57, 63 (2d Cir. 2010). Because ideas, facts, and concepts are not copyrightable, see 17 U.S.C. § 102(b), a copyright owner must prove that the defendant copied the owner’s particular means of expressing an idea, not merely that the defendant expressed the same idea. *Abdin v. CBS Broadcasting, Inc.*, 971 F.3d 57, 67 (2d Cir. 2020).

How US courts will apply these infringement principles to an unpredictable technology like LLMs is still unknown. There is no district court decision definitively ruling on whether any particular LLM plausibly creates a derivative work, although several dispositive motions are pending in different courts. District courts have denied motions to dismiss claims for direct infringement based on the use of a training set that includes copyrighted works to train the LLM, suggesting that claim might have merit.

LLMs are trained on a huge number of works. Their output is fundamentally unpredictable in its details, but will always include elements of that input. What output is created depends on the prompt used, which interacts unpredictably with the trained LLM. Does an LLM create a new type of derivative work, like a movie adaptation of a book, or does it merely abstract facts and use new words or images to express those facts?

Many cases are ongoing between content owners and GenAI companies, but none have reached a definitive conclusion. Most are in discovery now. Some cases have been pared down to the question of whether use of copyrighted content to train a generative AI system is copyright infringement, and whether that use is excused under fair use principles. Each system has used different training sets, and these are generally not publicly known. Similarly, whether the training sets were stored by the generative AI company, and how they were adapted for training, is generally not publicly known, so discovery is required.

Assuming that use of copyrighted materials to train a generative AI system is infringement, in order to prevail copyright owners will have to overcome defendants' fair use arguments. In the US, the fair use defense is statutory and turns on four factors: "(1) the purpose and character of the use . . . ; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of" the work. 17 U.S.C. § 107; see *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994). Defendants argue that using copyrighted works to train a generative AI system adds a "different purpose or character" than the original works, and so is transformative and protected as a fair use. *Andy Warhol Found. for the Visual Arts, Inc. v.*

Goldsmith, 598 U.S. 508, 525 (2023). Defendants also point to a line of cases that find that "intermediate copying" is a fair use, so long as the copying is used to create a different product or service.^[2] Plaintiff copyright owners, on the other hand, point to the admitted wholesale copying of their works, the fact that the generative AI models can often be induced to output copies or near-copies of the works, and the loss of any license value to generative AI systems once the works have been used for training.^[3]

We are far from any authoritative resolution of the question of infringement by AI systems. Any final district court decision will be subject to appeal, which will likely focus on the legal and statutory interpretation questions around the meaning of "derivative work" in the Copyright Act and whether the use of copyrighted content to train a commercial generative AI system is a fair use. In addition, the US Copyright Office has been charged with studying this issue and issuing reports with recommendations for legislative changes. The first report has been issued, recommending a law that protects all individuals from the knowing distribution of unauthorized digital replicas. Forthcoming reports, now overdue, are expected to address the copyrightability of materials created in whole or in part by generative AI, the legal implications of training AI models on copyrighted works, licensing considerations, and the allocation of any potential liability.

UK Approach

As in the United States, copyright in the UK subsists in original literary, dramatic, musical and artistic works, as well as in sound recordings and films. The copyright owner has the exclusive right to do certain acts, including copying the copyright work, communicating the work to the public, and making adaptations of the work. Copyright is infringed by a person doing any of the acts restricted to the copyright owner, without the permission of the copyright owner.^[4]

Applying this basic proposition to Generative AI raises a number of questions:

- To what extent does training an AI model by feeding copyright works into it, infringe the rights of copyright owners?
- Is the generated output of the AI an infringement if it is based on an earlier copyright work?
- Is the AI model itself an "infringing article" once it is trained on copyright works, in the way that a book containing unlicensed reproductions of pictures would be an infringing article?
- If the AI generated output is a copyright infringement, who is liable for that infringement? Is it the "prompt engineer" who requested the output? Is it the developer of the AI, or whoever trained it?
- If there is an infringement- are there any applicable exceptions or defences that might apply?

Training the AI

The training of AI by exposing it to copyright works necessitates reproduction by the AI of those works, by copying them into the model. Whether that form of reproduction is permitted will broadly depend on the source of the works.

It will also depend on whether one of the statutory exceptions from infringement applies. UK copyright law contains a limited exception for the creation of temporary copies as part of a technical process, although the copies made by the AI are unlikely to be temporary if it is to

remember the work. The UK Government has announced an intention to legislate to clarify the scope of this exemption in the context of AI training. Further, there is an exception for “text and data mining” which allows the extraction and reproduction of copyright works, but only under certain, limited circumstances. In particular, the data mining must be for non-commercial research.

The previous UK Government proposed widening the text and data mining exception in an effort to bolster the tech industry. It met resistance from the UK’s creative industries who were less keen to see their work exploited in this way, and the proposal was shelved. However, the idea is now back on the table. The present Government published a consultation in December 2024 on the reform of copyright law, including a proposal to allow text and data mining exemption for commercial, as well as non-commercial, purposes. This would be similar to the exception already in place in the European Union. Like the EU regime, rightsholders would have a right to prevent text and data mining of their work by reserving their rights. Further, AI developers would be required to be more transparent around their use of existing works for training, although the level of detail that they will be required to disclose is not yet known.

Outputs from Generative AI

Whether or not the exceptions for temporary copying or for text and data mining discussed above apply to the input, those exceptions will not in any event apply to the output of the generative AI.

Under UK law, an article which copies the whole or a “substantial part” of an original copyright work will infringe. The test is not whether a substantial part of the new work replicates the earlier work, but the degree to which the original has been copied. Further, whether a part is “substantial” is a qualitative test, not a quantitative test. In other words, the infringing copying could make up a small part of the new work, and may also be a small part of the original work, provided it was an important part of the original.

It is therefore quite possible that an AI will reproduce a substantial part of one or more original copyright works when generating a new work in response to a user’s prompt.

Fair Dealing Defences

Is it possible that the generative output could benefit from one of the Fair Dealing defences under UK law?

To qualify for the defence of Fair Dealing, the use must fall within one of the purposes specified in the Copyright Designs and Patents Act 1988 before there is any consideration of fairness. The permitted purposes are specific and limited, and some come with conditions attached. For example, certain fair dealing purposes require that a sufficient acknowledgement is given to the original source.

If an appropriate fair dealing purpose is identified, the use of the original work must still be fair in all the circumstances. The UK court would consider various factors to determine fairness, including the amount of the work used and the impact on the rights holder.

Fair Dealing purposes include research and private study, criticism and review, and news reporting. However, the Fair Dealing purposes which are most relevant (and debated) when it comes to generative AI are quotation, parody, pastiche and caricature. Does the output from generative AI quote from the source material? Is it a parody of the original? Or perhaps a pastiche of someone’s work?

There have been very few decided cases on the meaning of “parody” for fair dealing purposes, and even fewer for “pastiche”, which is likely to be the purpose most relevant to generative AI produced “in the style of” a particular author or artist.

This may change in the near future due to the rash of AI related copyright infringement cases currently working their way through the Courts. One of the most keenly followed will be *Getty Images v Stability AI*.

Getty Images is an image library. It accuses Stability AI of infringing copyright in its content, both because its content was used to train their Stable Diffusion image generating AI, and also because the outputs from that system are alleged to infringe Getty Images' rights. Getty Images have pointed out that in some images generated by the AI, their watermark (being their brand names added to the image to deter copying) has been reproduced.

Amongst other defences, Stability AI argues that the training activities did not occur in the UK, and that the generated images are a pastiche and fall within the UK's fair dealing exception.

If fair dealing for pastiche does not save the day for Stability AI, they could be liable for infringement. Under UK law, AI developers could potentially be liable as secondary infringers, accessories/joint tortfeasors, or primary infringers in relation to the output of a generative AI.

The person who writes the prompts that leads to the output being generated could also be liable. Particularly if the prompt in some way directed the AI to copy an existing work. For example, by asking it to produce something similar to or "in the style of" the original work. Notably, the image generating AI, Dall-E will not reproduce images in the style of a living artist. Further, OpenAI, the company that built both Dall-E and ChatGPT, also allows artists and copyright owners to direct that their works are removed from training.

The Way Forward

The judgement in the Getty Images case will be hugely important, but with the trial not expected before the Summer, a final judgment (particularly if the case is appealed) is some way off.

In the meantime, there remains an uncertainty around whether the development and use of generative AI infringes the copyright in input works.

The Government's consultation on the introduction of a broader text and data mining exception seeks to address this. Although their preferred solution is to increase freedom to operate for AI developers by widening the exception, the consultation acknowledges that this is not the only possible solution. An alternative to a broader exception would be to tip the scales in the other direction- to strengthen copyright law, making it clear that an express licence is required to use a copyright work for training AI.

Recognising that an infringement would only occur if the use of the work was not authorised, some AI developers have already taken steps to obtain that authorisation by agreeing licences with copyright holders. If the Government's proposed text and data mining exception does not become law, or if it does not go far enough to shield AI developers and users from copyright infringement, then licensing could be the way forward, either individually or collectively. The latter would involve a statutory licensing regime under which, AI developers could obtain a licence use copyright works on payment of a fee or royalties to a collecting society.

But would collective licensing work on an opt-in or opt-out model? If it was the latter (as is proposed for the expanded text and data mining exception), any copyright work could be used by the AI unless the owner has opted out. This is likely to be resisted by the creative community, particularly given the administrative burden it would place on rightsowners to opt-out their works. How to make opt-outs work effectively is an issue discussed in the Government's current consultation, particularly as there is some uncertainty over how rightsholders effectively reserve their rights under the equivalent text and data mining exception in EU law.

Conclusion

The question of whether AI-generated content is a derivative work will depend largely on the technical details of its creation, as well as how the courts interpret the relevant statutory language and decisional precedent.

There are significant copyright cases before the courts at the moment and it is likely that these will in due course provide at least some of the answers. We may also see governments legislate on AI, either by introducing governance rules such as transparency requirements on AI

developers, by amending copyright laws to introduce new exceptions, or perhaps by introducing new statutory royalties and collective licensing arrangements.

When any significant new technology brushes up against intellectual property laws, all that is certain is that there will be some uncertainty. However, the core principles of copyright remain applicable, and infringements and compensation to rights holders will no doubt continue to be pursued.

[1] There are also fascinating open questions regarding what level of human input is necessary in order to make a work created with the assistance of an AI tool copyrightable. Those questions are beyond the scope of this article.

[2] See, e.g., ANTHROPIC'S OPP. TO RENEWED MOT. FOR PRELIM. INJ., *Concord Music Group, Inc. v. Anthropic PBC*, N.D. Cal., 5:24-CV-03811-EKL, pp. 17-24.

[3] See e.g., PLAINTIFFS' REPLY IN SUPPORT OF MOTION FOR PRELIMINARY INJUNCTION, *Concord Music Group, Inc. v. Anthropic PBC*, N.D. Cal., 5:24-CV-03811-EKL, pp. 9-12.

[4] Copyright Designs and Patents Act 1988, Section 16(2)

related professionals

John M. Griem, Jr. / Partner

D 212-238-8659

griem@clm.com