Kluwer Copyright Blog

The EU's false sense of isolationism in AI and copyright

Bertin Martens (Bruegel) · Thursday, May 29th, 2025

In October 2023, several music companies (Concord, ABKCO Music & Records, Universal Music) sued Anthropic AI in the US for alleged harm to their business interests because (a) its AI chatbot, Claude, was trained with unauthorized music lyrics data and (b) Claude's outputs in response to user queries contained copies of (parts of) these lyrics or derived lyrics.

On 25 March 2025, the US Northern district court of California dismissed the motion for an injunction. The judge argued that the plaintiffs had not demonstrated how using their lyrics as an input to train Claude could reduce the value of these lyrics in licensing markets or harm the reputation of the owners of rights to these lyrics. Moreover, the judge pointed out that these alleged harms relate to the unauthorized reproduction of lyrics in Claude's outputs. However, that complaint was resolved by an agreement between Anthropic and the plaintiffs whereby Anthropic agreed to put in place guardrails that prevent Claude from reproducing (parts of) copyright-protected lyrics in its outputs. That agreement effectively prevents harm in licensing markets for lyrics but does not prevent Anthropic from using lyrics as a training input for AI models. The judge thereby confirmed that in this case the use of lyrics for training AI models does not violate US copyright law.
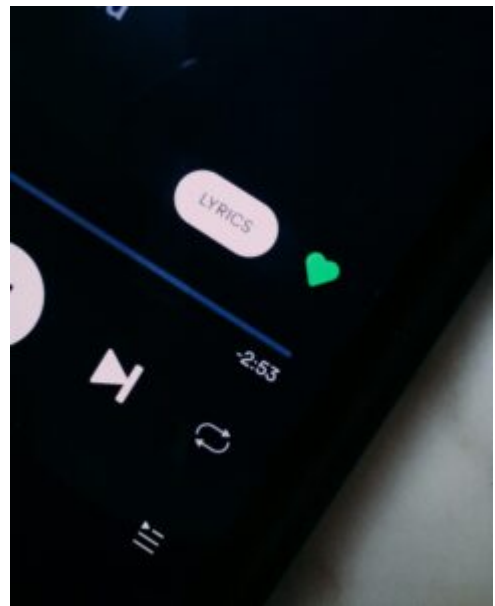

Photo by charlesdeluvio on Unsplash

Meanwhile, both the plaintiffs and the defendant are confident that their views will prevail in higher courts should the case go on appeal. Indeed, this judgment is just one among many judicial steps that dozens of AI-related copyright cases will have to go through in the US. It will take several years to resolve this legal uncertainty in court.

The case is of general importance however for all copyright industries, far beyond music and lyrics, because the court stresses the underlying economic reasoning in favour of unrestricted use of copyrighted material for AI training: it does not harm the market for copyrighted outputs, unless

AI models re-produce outputs that are nearly identical to training inputs. While early AI models were prone to doing so, the latest models have built-in guardrails to prevent regurgitationof inputs.

From an economic perspective, copyright grants an exclusive monopolistic right to the creator for the exploitation of the protected work. Monopolies are, almost by definition, inefficient from a societal point of view. However, in the case of copyright, they constitute an incentive to invest in new creative products that are beneficial for society. The scope of copyright protection should ensure that these dynamic innovation benefits exceed the welfare losses for society from granting an exclusive copyright. A key test for this economic criterion is the impact of an extension or reduction in scope on the supply of new creative products. The California judgment in the Anthropic case captured that economic reasoning very well.

There is no evidence so far that use of copyright protected materials as inputs for AI model training would harm the market for these materials. Authors can continue to sell their creative products in their usual outlet channels. Of course, AI may increase competition from AI-generated or hybrid products. But as long as these are not (nearly) identical reproductions of training inputs in model outputs, that is acceptable. In line with economic reasoning, the judge distinguished between possible harm from violation of reproduction rights in AI model inputs and outputs and accepted that only the latter could be harmful to the market for licensing of lyrics.

Just like firms, authors do not like competition. But consumers benefit from that competition. Evidence shows that authors tend to avoid publishing their copyrighted content on platforms that are known to use this for AI training. But that does not imply a reduction in production. However, there is evidence that limiting the scope of Text and Data Mining exception to copyright, is harmful for innovation in society. A third of all websites are already implementing copyright opt-out protocols, such as robot.txt, in line with the EU Copyright in the Digital Single Market Directive (CDSM Directive) Art 4(3). That reduces the volume of AI training data and the quality of AI models. It has a negative economic impact on AI-driven innovation across the entire economy, far beyond media industries, which account for less than four percent of GDP.

Copyright holders may of course claim that whatever US courts decide with regard to copyright and AI has no bearing on EU AI markets. They may feel protected by extraterritoriality claims in the AI Act that, though controversial, may give European authors a feeling of protection against liberal US interpretations of copyright in AI model applications. This may result in a false sense of economic isolationism. Foundational generative AI or Large Language Models, such as OpenAI's ChatGPT, Meta's Llama and Google's Gemini, are all developed and trained in the US, but widely used in the EU. Very few foundational models, and certainly no current technology frontier models, are trained in the EU. The absence of suitable computing infrastructure in the EU, and the fact that all big tech firms are US-based, has a lot to do with this. But uncertainty and risks surrounding the EU AI copyright regime are an important factor too. If the EU wants to strengthen its sovereignty in the AI supply chain and promote EU home-grown AI models, it should take copyright economics seriously.

Rightsholders may reserve their rights with opt-outs in the hope that it generates licensing revenue. This may work for a few very large publishers who have indeed signed licensing agreements with some AI model developers. It is unlikely to work for smaller publishers. Trying to do so with millions of unidentified web publishers would run into unsurmountable licensing transaction costs. Collective licensing only displaces that problem from AI developers to intermediary agencies; it does not solve that problem. This results in biased AI training datasets, limited to a few large

publishers only and goes against the anti-bias requirements imposed by Recitals 70, 110 and 156, and Art 10 § 5, of the AI Act.

Creative industries already benefit substantially from the use of AI models for the production of content. At the same time, they claim that these models harm them. Their schizophrenic mindset is perhaps best captured in copyright measure 2.3(5) in the proposed AI Code of Practice. It reiterates that AI model developers should respect publishers' opt-outs from the use of their webpage content for AI model training. However, they should ensure that the same content is picked up by their search engines. Webpage publishers want their content to be found by users of AI algorithms in search engines, but not by AI model algorithms that deliver a more reasoned response to user queries. In other words, EU copyright allows humans to learn from their content, but not more efficient machines that work for humans.

This policy stance on (the economics of) AI copyright risks marginalizing the EU even further in global AI markets and services, both with regard to model training and use, especially if US jurisprudence would move towards fair use and transformative use of copyrighted content in AI model training. Some big tech firms are already withholding their most advanced AI models and services from the EU market, so far mostly related to legal uncertainty about the use of personal data for training of AI models. Even if, miraculously, EU extra-territoriality claims with regard to copyright in AI would be admissible, they will not protect it from falling into the trap of low-quality and underperforming AI services.

———————————————

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe here.

This entry was posted on Thursday, May 29th, 2025 at 9:06 am and is filed under Artificial Intelligence (AI), USA

You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.