

Kluwer Copyright Blog

Machine readable or not? – notes on the hearing in LAION e.v. vs Kneschke

Paul Keller (Institute for Information Law (IViR)) · Monday, July 22nd, 2024

Last week, the District Court of Hamburg, Germany, held a hearing in the first European case to examine the legality of using copyrighted works for the purpose of training generative AI models.

The case centers on [LAION e.V.'s](#) (a German non-profit organization that builds widely used training datasets) download of an image by German photographer [Robert Kneschke](#) for inclusion in the LAION 5B dataset. Neither party disputes that the image in question was downloaded, analyzed, and subsequently included in the training dataset, but LAION claims that this is legally permissible, while Kneschke disputes this. The disputed image was freely available without a paywall on the website [bigstock.com](#).



[Landgericht Hamburg Criminal Justice Building](#) by [Andrew Milligan](#) [sumo](#) from flickr [CC BY 2.0](#)

The first question was whether the reproductions made by LAION fell under the temporary copying exception of Article 5(1) of the InfoSoc Directive (implemented in Germany as § 44a UrhG). This approach was quickly rejected by the Court, which found that the copying was neither “transient or incidental” nor “an integral and essential part of a technical process”.

After rejecting the application of § 44a, the court turned to LAION’s next defense: that the reproductions were permitted under the text and data mining exception in Article 4 of the Digital Single Market Directive, transposed as § 44b UrhG.

Here it seems (from reports from [both sides](#) and [other observers](#)) that the court was inclined to take the position that making reproductions for the purpose of training AI systems falls within the scope of the TDM exception. This is in line with what [we have been arguing since early last year](#) and it is good to see that the Court seems to have a very similar understanding: that AI training is an automated analytical technique that generates correlations and thus falls within the scope of the definition of TDM in Article 2(2) of the CDSM Directive.

The court also held that the following passage in a [subsection of bigstock.com’s terms of service](#) constituted an opt-out from TDM within the meaning of Article 4(3) of the CDSM:

YOU MAY NOT [...] Use automated programs, applets, bots or the like to access the Bigstock.com website or any content thereon for any purpose, including, by way of example only, downloading Content, indexing, scraping or caching any content on the website.

The court pointed out that this passage clearly communicated an opt-out from the text and data mining use in question because it “[excluded the use of bots ‘for any purpose,’ including downloading](#)”. While this seems like a reasonable interpretation, it potentially raises questions down the road if all types of general statements (such as “for any purpose” or the much more commonly used “all rights reserved”) are to be interpreted as a reservation of rights under Article 4(3) of the CDSM Directive. Does such a statement really satisfy the “expressly reserved” condition for a reservation of rights? In the present case, the court seemed to find that the language in the ToS satisfied this requirement.

The main part of the hearing then revolved around the question of whether the above opt-out (expressed in English language and formatted in HTML in [a subsection of the website’s terms of use](#)) should be considered machine readable (as argued by the plaintiff) or not (as argued by LAION). In the discussion, LAION suggested that in order to be considered machine readable, an opt-out should be provided in a specific standardized format (in this case robots.txt) that can be easily understood by crawlers and other bots. The plaintiff argued that digital plain text is sufficiently readable and that requiring the use of specific formats is undesirable because most authors do not have the technical knowledge to effectively protect their works from being crawled in this way.

According to all observers, the court did not express an opinion on this issue, which seems to be the main factor in deciding the outcome of the case. The court set September 27 as the date for its decision, unless there is a need for further hearings.

For anyone who has been following the discussion of TDM opt-outs in the context of training generative AI models, the fact that the case appears to be resolving itself around the issue of machine readability can hardly come as a surprise.

As I have [argued elsewhere](#), the EU legal framework provides sufficient legal clarity regarding the use of copyrighted works for the purpose of AI training, but that without generally accepted standards for machine-readable opt-outs, this system is bound to fail. The hearing at the Hamburg District Court seems to confirm this thesis. Both sides raised legitimate concerns: LAION (channeling the concerns of AI model developers) points to the need for well-structured and standardized opt-out information that can be processed at scale. Kneschke (channeling the concerns of creators) pointed to the fact that the current situation, where there are no clear standards, is a barrier for anyone without a technical background – and control over the means to do so – to effectively exercise their rights.

As outlined in [this recent Open Future policy brief](#), creating more certainty for both sides of this debate will require building consensus around the following four distinct aspects of machine-readable opt-outs: the identifiers for works, the vocabulary for opt-outs, the infrastructure used to communicate and respect opt-outs, and the effect of an opt-out once recorded. Last week's hearing at the District Court of Hamburg is an important reminder that these issues need to be resolved urgently.

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

Kluwer IP Law

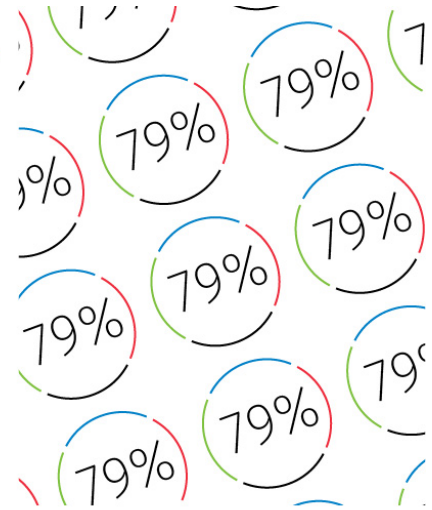
The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

Learn how **Kluwer IP Law** can support you.

79% of the lawyers think that the importance of legal technology will increase for next year.

Drive change with Kluwer IP Law.

The master resource for Intellectual Property rights and registration.



2022 SURVEY REPORT
The Wolters Kluwer Future Ready Lawyer
Leading change

This entry was posted on Monday, July 22nd, 2024 at 11:59 am and is filed under [Artificial Intelligence \(AI\)](#), [Germany](#), [Text and Data Mining \(TDM\)](#)

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