

Kluwer Copyright Blog

Kneschke vs. LAION – Landmark Ruling on TDM exceptions for AI training data – Part 2

Paul Goldstein (Lillick Professor of Law, Stanford Law School) and Christiane Stuetzle, Susan Bischoff (Morrison & Foerster LLP) · Thursday, November 14th, 2024

In the [first part of this post](#) on the Kneschke vs. LAION decision by the German Hamburg Regional Court (“Court”), we explored the Court’s key findings regarding the operational step in a generative AI model, and the decision on the exceptions for scientific research text and data mining (“TDM”) and temporary reproductions. Now, in this second part, we turn our focus to the Court’s comprehensive *obiter dictum* addressing the commercial TDM exception.

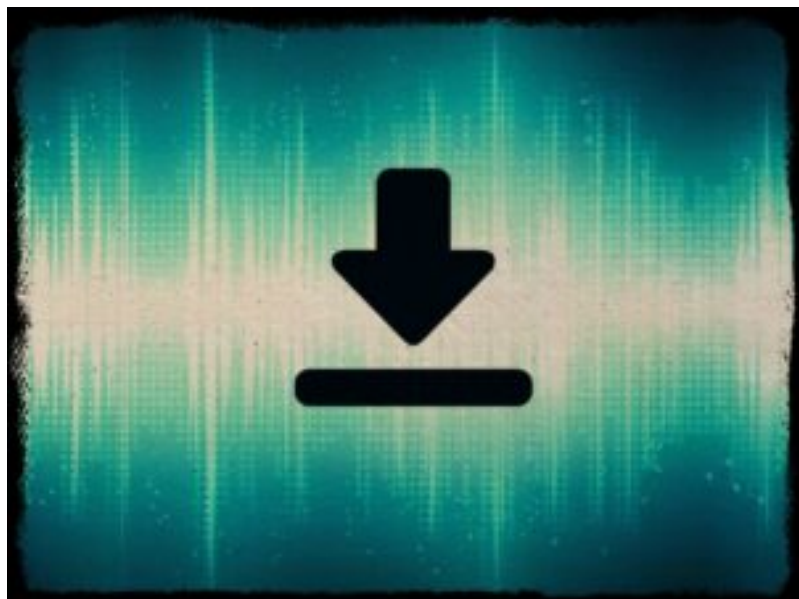


Image by [M. H.](#) from [Pixabay](#)

The case against LAION was dismissed on the grounds of the scientific research TDM exception. However, given the provision’s limited scope, the recent debate has focused on the broader TDM exception for other (commercial) purposes in Art. 4 DSM Directive (Sect. 44b UrhG). It is not surprising that the Court also considered the commercial TDM exception, but the extent is striking – the *obiter dictum* takes up almost half of the entire judgment.

Commercial TDM exception generally applicable to AI training data sets

The applicability of the commercial TDM exception to data collection for (generative) AI training has been widely debated. That Art. 53(1)(c) AI Act explicitly refers to the commercial TDM exception in the context of GPAI and thus generative AI is widely seen as a clear indication of the EU legislator’s intention that the exception covers AI data collection. Nonetheless, a recent [study conducted on behalf of the German Authors’ Initiative](#) still opposes an application. The Court rejected the study’s main arguments:

- TDM for AI training not distinct from other TDM: The study claimed that AI “uses” the very content of an intellectual creation which it has been trained on, rather than merely analyzing data for information. The Court rejected this idea, noting the unclear distinction between information and creation “hidden” in training data.
- Potential “creative” AI output irrelevant: The Court rejected the argument that the TDM exception should not apply to reproductions made for AI training because “AI web scraping” ultimately leads to competing creative products. As the Court pointed out, at the time of the TDM-relevant activity (reproduction when creating a data set), the training has not yet taken place, let alone the generation of (specific) AI output; therefore, the general intention to later obtain AI-generated output cannot be relevant for the legal assessment of the creation of a data set.
- No (presumed) contrary legislative intention: The Court emphasized that the relevant developments in AI since the introduction of the 2019 EU TDM exception relate less to the nature and scope of data mining, but rather to the performance of data-trained AI. The Court also found that by explicitly referencing the TDM exception in the 2024 AI Act, the EU legislator has “undoubtedly expressed” that it also covers the creation of data sets intended for AI training.
- Three-step test compliance: Under the overarching three-step interpretation standard (laid down in international and EU copyright law, see Art. 5(5) InfoSoc Directive, Art. 7(2) DSM Directive), copyright exceptions should only be applied in certain special cases that do not conflict with the normal exploitation of the copyrighted material and do not unreasonably prejudice the legitimate interests of the right holder. The Court focused on the potential conflicts between AI output and human creations but did not consider whether the application of the TDM exception to the creation of data sets for generative AI training also meets the three-step test at the input level, i.e., as colliding with the ability of rights holders to exploit their creations through licensing as training material.

The reproduction by LAION was found to have been made for the purpose of obtaining information on “correlations” within the meaning of TDM of Sect. 44b(1) UrhG, as the download facilitated the comparison of the image content and the description. LAION’s lack of “curating” (i.e., filtering) of the data set is irrelevant. The reproduced material was also lawfully accessible to LAION, as the photograph was publicly available on the photo stock agency’s website.

Requirements for an effective opt-out

Reproductions are permitted under the commercial TDM exception only if the rights holder has not reserved the right. For material available online, an effective opt-out must be expressed in a machine-readable form, Sect. 44b(3) UrhG, Art. 4(3) DSM Directive. The terms of use of the

crawled website stipulated:

- Opt-out can be issued by licensee and asserted by author: The Court stated that an opt-out can also be effectively declared by a subsequent rights holder, such as a legal successor or a licensee (here: the stock photo agency). Kneschke could also rely on the opt-out of his (non-exclusive) licensee in asserting his rights against LAION, because only the agency was able to implement an opt-out for the location where the photograph was available for web scraping (on the agency's website).
- Opt-out does not need to have specific law in mind: LAION argued that, since the terms were already implemented in January 2021, the reservation clause could not have been drafted in view of the current commercial TDM exception of Sect. 44b UrhG, which only came into force in June 2021. The Court clarified that an opt-out does not have to be declared in relation to a specific version of the law.
- Sufficiently clear wording: As the Court pointed out, the EU law model for the commercial TDM exception in Art. 4(3) DSM Directive requires that the use be “expressly” reserved. Although the wording of Sect. 44b UrhG does not include this criterion, the “expressiveness requirement” must still be taken into account to ensure conformity with EU law. The Court specified that the opt-out must (i) be explicitly declared (implied reservations are insufficient) and (ii.) be precise enough to unambiguously cover specific content (also satisfied by a reservation for *all* works on a website) and specific use (the Court found that the clause “easily” meets this requirement). An explicit mention of “text and data mining” or “reproductions” is therefore not required.
- Natural language opt-out may be machine-readable: The Court indicated that the opt-out in the terms of use met the requirement of being “machine-readable”. Whether an effective, i.e. machine-readable, opt-out is in place is of paramount importance when collecting data for AI training; disobedience results in a copyright infringement, and, due to the obligation of Art. 53(1)(c) AI Act, and exposes any provider doing business in the European Union to fines (up to 3% of the provider's total annual worldwide turnover or EUR 15 million, whichever is higher, Art. 101(1)(a) AI Act) and other enforcement measures, such as withdrawal of the model from the European market. These legal consequences under the AI Act are not relevant for LAION with regard to its data set (as it is not the provider within the meaning of Art. 3(3) AI Act of the GPAI models trained on the data set by third parties) but must be observed by any commercial actor conducting its own TDM for AI model training.

That an effective opt-out can be placed in a website's terms and conditions is already recognized in Recital 18 DSM Directive. But what constitutes machine-readability remains an open question. While the prevailing view holds that natural (human) language reservations are *not* machine-readable within the meaning of the TDM exception, favoring solutions like robots.txt and metadata, the Court takes the opposite view.

The Court justifies the natural language opt-out as “machine-understandable” with the AI Act. The Court’s reasoning is as follows: Art. 53(1)(c) AI Act requires GPAI model providers to put in place a policy for complying with TDM opt-outs “including through state-of-the-art technologies”. Although there is no such clarification in the law, the Court argued that these technologies “It asserts that the commercial TDM exception should not allow AI model providers to develop “increasingly powerful” text-understanding models without requiring them to use existing AI to detect natural language opt-outs. This novel argument’s assumed causality between the development of AI and a lowered threshold for machine-readability may be overly simplistic. Not all entities training AI under the TDM exception do so to develop text-proficient AI models. Nor is it apparent from the term *machine-readability* that it is sufficient for (highly specified) AI applications to comprehend the text, rather than that the declaration is technically coded and executable by a machine, i.e., crawler software. To avoid conflict with the narrower understanding of a machine-readable format in [Directive \(EU\) 2019/1024](#) (Recital 35: easily identifiable, recognizable, and extractable for software applications), the Court argued that a uniform definition across Directives is not required.

Although the plaintiff did not prove this, the Court saw indications that LAION had suitable technology in 2021 and was capable of automatically recognizing natural language opt-outs.

Outlook

The decision may be appealed to the Hamburg Higher Regional Court and then to the Federal Court of Justice (BGH). Given the fundamental legal issues involved and the ambiguity of the law, this case may indeed reach the BGH, which might refer to the ECJ for a preliminary ruling, particularly on a uniform interpretation of the machine-readability of opt-outs.

For entities that might be classified as GPAI model providers under the AI Act (which is not the case for LAION or other data set repositories, as they are not the providers of models trained with these data by third parties), such a copyright-specific clarification would come too late in terms of compliance with the AI Act, as their AI-product-related obligation to observe TDM opt-outs generally applies from August 2025 (per the non-binding Recital 106 AI Act even for training activities conducted outside the EU). Consequently, GPAI model providers will seek AI Act-specific clarification during the ongoing process of developing Codes of Practice under the leadership of the EU AI Office.

The authors wish to thank João Pedro Quintais for his most valuable feedback on this post.

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

Kluwer IP Law

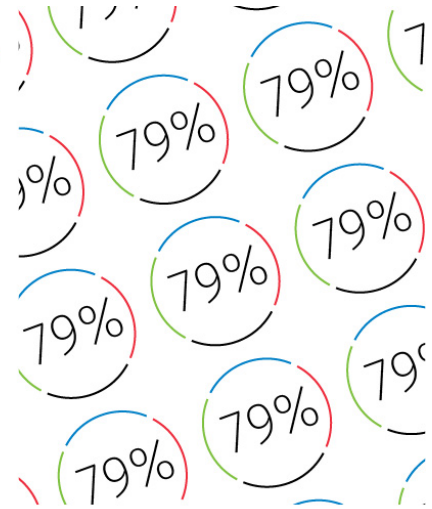
The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred 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 Thursday, November 14th, 2024 at 10:42 am and is filed under [Artificial Intelligence \(AI\)](#), [Case Law](#), [Exceptions and Limitations](#), [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.