Kluwer Copyright Blog

Kneschke vs. LAION – Landmark Ruling on TDM exceptions for AI training data – Part 1

Paul Goldstein (Lillick Professor of Law, Stanford Law School) and Christiane Stuetzle, Susan Bischoff (Morrison & Foerster LLP) · Wednesday, November 13th, 2024

Image by M. H. from Pixabay

On September 27, 2024, the German Hamburg Regional Court ("**Court**") issued the first ruling on reproductions of copyrighted content from the Internet made during the creation of an AI training data set – and on whether the copyright exceptions for text and data mining ("**TDM**") provide statutory permission for such use (Landgericht Hamburg, 310 O 227/23). The facts of the case and the practical implications of the decision were covered in the blog post by Jonathan Pukas and Jan Bernd Nordemann. Building on that, this post delves further into the Court's key findings in two parts. This first part identifies the operational step in a generative AI model that the Court ruled on and reviews the decision on the exceptions for scientific research TDM and temporary reproductions. The second part examines the Court's extensive obiter dictum on the commercial TDM exception.

The Court dismissed photographer Robert Kneschke's injunction claim against the German non-profit organization LAION for the unauthorized reproduction of Kneschke's photograph through a download from a stock photo agency website in connection with the creation of an open-source

dataset for AI training purposes. The decision provides a valuable first judicial insight, well beyond the specific case, into the TDM exceptions for commercial and scientific research purposes under the 2019 Copyright in the Digital Single Market (DSM) Directive (EU 2019/790) and, if upheld, may influence the interpretation of the law far beyond German borders. The ruling is also relevant beyond "just" copyright law, as it sheds light on the uncertainty among providers of general-purpose AI ("**GPAI**") models as to how to meet their obligation under the EU AI Act (Regulation (EU) 2024/1689) to implement a policy to comply with TDM opt-outs for their training data.

Operational stages of generative AI (not) covered by the Court ruling

For the legal assessment of generative AI, three operational levels can be roughly distinguished: the input level, the "Black box" (machine/deep learning), and the output level. LAION creates and provides open-source data sets that are used by third parties free of charge as training data for generative AI models, e.g., by Stability AI for Stable Diffusion. LAION's 5B data set consists of descriptions of almost 6 billion images and the links to the images' publicly accessible locations on the Internet. The data sets do not contain the images themselves (or reproductions thereof), but in order to ensure through a software check that the image descriptions match the linked images, LAION had to download the images. And it was only such a download of Kneschke's photo for that purpose that was the subject of the Court proceedings. As a result, the Court did not address the copyright implications of the "black box" stage of a neural network or the output level.

The decision relates only to the input level, and only to a limited extent. A training process consists of several individual steps and may differ from model to model, with individual copyright implications for each step and each process. The Court's decision concerns a preparatory step to the actual training of AI and only a very specific one, the downloading of material during the setting up of a training data set. The decision therefore does not concern what we understand to be the actual training of an AI model, i.e., the analysis of the material by the neural network for patterns, relations, correlations, and probabilities.

Reproduction permitted under TDM exception for scientific research

The Court held that LAION's reproduction by downloading the photo was legally permitted under the TDM exception for scientific research (Sect. 60d German Copyright Act, "**UrhG**"). As of 2021, such an exception must be provided in all EU Member States through the implementation of Art. 3 DSM Directive, which allows for "*reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access*.". The Court based its decision on the following general findings:

- Preparatory steps as part of "scientific research": The Court defined scientific research as the "methodical and systemic pursuit of new knowledge", explicitly including not only the work steps *directly* related to the acquisition of knowledge, but also the preparatory steps, provided that they are aimed at the (subsequent) acquisition of knowledge. Accordingly, the Court confirmed that the creation of a data set can constitute scientific research; it may not yet be associated with the acquisition of knowledge, but it is a fundamental work step with the aim that the data set will be used for this purpose at a later stage.

- Third-Party use of the data set is sufficient: The ruling clarifies that for the creation of the data set to be considered as scientific research by LAION, it does not have to be used by LAION itself for AI training. It is sufficient that the data set is published free of charge, making it available (also) to researchers of artificial neural networks. Even the use by commercial entities for AI training is acceptable, as their commercial research is still a relevant pursuit of knowledge.

- Research success not required: The Court emphasized that actual subsequent research *success* is not necessary for the work step to qualify as scientific research.

- Non-commercial nature of the research: The Court stated that the only decisive factor is that the activity is non-commercial, making LAION's organization and financing irrelevant in this regard. This is the case because LAION makes the data set publicly available free of charge. It was not apparent to the Court that the creation of the data set (also) served the development of LAION's own commercial offerings. The use of the data set by commercial entities to train their commercialized AI models is irrelevant for the assessment of the non-commercial nature of the research activity, i.e., the creation of the data set by LAION. It is also irrelevant that individual members of LAION pursue paid activities with such companies, as this does not attribute the activities of these companies to LAION.

- No relevant cooperation with private companies: The TDM exception for scientific research does not apply when the researching organization cooperates with a private enterprise that exerts a certain degree of influence on the organization and has preferential access to the research results. The plaintiff had unsuccessfully argued that this was the case, e.g., by pointing out that a co-founding member of LAION now works at Stability AI. However, the Court did not consider this to be evidence of Stability AI's decisive influence on LAION's research. The Court also found that the plaintiff's argument regarding the alleged financing of the LAION data set by a private company against early access was without substance. LAION itself had stated that a company had provided computing resources during the start-up phase, but the Court did not evaluate this.

Copyright exception for temporary reproductions: Not applicable here

Unsurprisingly, the Court found that the reproduction of the photograph made by LAION through downloading was not covered by the exception for temporary acts of reproduction (Sect. 44a UrhG; Art. 5(1) InfoSoc Directive. LAION had argued that the image was not stored permanently but was only used briefly for the analysis of the description-image conformity and was then automatically and irrevocably deleted immediately.

- No mere transient reproduction if deletion is only due to specific programming: The Court rejected the argument that the reproduction was transient in nature based on the clarification of this criterion by the European Court of Justice (ECJ), which in particular requires that, to qualify,

the technological process must be automated in such a way that it deletes the reproduction automatically without human intervention (case C-5/08, para. 64). The Court found that the deletion of the downloaded image was not "user-independent", but the result of a deliberate programming of the analysis process by LAION. Furthermore, LAION had not provided any information on the actual duration of the storage.

- No mere incidental reproduction if download is preparatory step for analysis: The ECJ defines an incidental reproduction as one which neither exists independently of the technological process of which it forms part nor has a purpose independent of that process (case C-360/13, para. 43). The Court did not consider the downloading of the photograph by LAION to be an incidental reproduction because it was an actively controlled procurement process and not merely an ancillary step to the subsequent analysis.

Part II of this post will examine the Court's extensive obiter dictum on the commercial TDM exception.

_____

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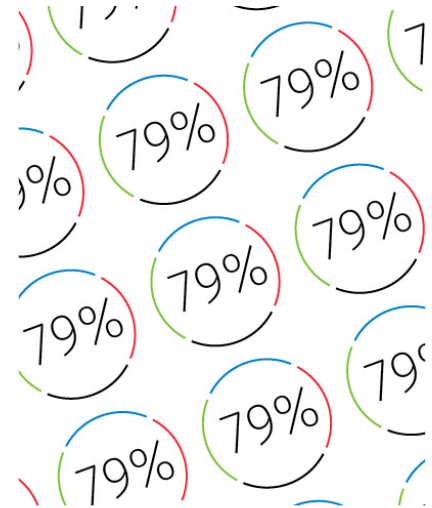
Learn how **Kluwer IP Law** can support you.

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.

Wolters Kluwer

2022 SURVEY REPORT
The Wolters Kluwer Future Ready Lawyer
Leading change

This entry was posted on Wednesday, November 13th, 2024 at 12:27 pm 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.