

**European Law Blog**

# **Kneschke vs. LAION: Opening Fresh Perspectives on AI Training and TDM Exceptions**

**Maryna Manteghi**

**European Law Blog**

**Published on:** Jan 14, 2025

**DOI:** <https://doi.org/10.21428/9885764c.5f1ce2dd>

**License:** [Creative Commons Attribution-ShareAlike 4.0 International License \(CC-BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)

## Introduction

The ruling in [Kneschke vs. LAION](#) by the German Hamburg Regional Court (Court) represents the first landmark case clarifying the application of the [CDSM Directive's](#) text and data mining (TDM) exceptions to AI training. Kneschke, a photographer, sued [LAION](#) for using one of his original and creative works to build AI training datasets without authorisation, allegedly infringing copyright. In turn, LAION, a nonprofit organisation that promotes AI research by offering open datasets for training, argued that their actions are covered by specific TDM exceptions under the [German Copyright Act](#) and the EU CDSM Directive. The defendant downloaded the photo from a stock photo website, where it was freely available to the public, to verify that the image description created by a third-party provider actually describes the contents of the image. The final training datasets included only the hyperlinks to where the photos were accessible, along with the descriptions and other accompanying metadata. The Court dismissed the plaintiff's claim for the unauthorised reproduction, recognising that LAION could benefit from the TDM exception for scientific research (Art. 3 of the CDSM Directive and Sec. 60 d of the German Copyright Act). Even though the decision addresses only the issues related to the preparatory phase of AI training — such as downloading images — and not the training itself, some of the Court's key findings could shed light on the practicalities of the CDSM Directive's TDM exceptions. This is particularly significant in the time of AI growth when balancing powers in the digital environment is essential to facilitate the functionality of the EU AI sector. This post reflects on the most notable outcomes of the Court's decision.

## Overview of TDM Exceptions in Terms of AI Training

The building of generative AI (GenAI) models involves two main phases: training (the so-called 'input' phase), including a preparatory work step, and content generation (the so-called 'output' phase) (see [Verma](#)). AI developers use TDM to train their systems on vast amounts of digital materials by extracting valuable insights, patterns, or correlations from such data. The 'learning' process involves copying obtained data, which may include copyrighted materials, thus raising legal concerns regarding the use of this analytical technique. The rationale behind this is that when such copying is done without authorisation, it may potentially infringe the right to reproduction, one of the exclusive rights of copyright holders. Extensive research on the CDSM Directive's TDM exceptions highlights their inability to effectively address copyright-related issues in AI training (see, e.g., [Ducato & Strowel](#), [Geiger et al.](#), [Manteghi](#), etc.). Art. 3 of the CDSM Directive allows research organisations and cultural heritage institutions to carry out TDM on protected works to which they have 'lawful access' for the purpose of scientific research. The second exception in Art. 4 is much broader, as it benefits all types of users and covers any purposes of TDM analysis, provided that rightholders have not reserved the use of their works. Even though these exceptions have improved the regulation of TDM, clarifying the lawfulness of this analytical technique in principle, many practicalities remain unclear. The next section discusses the most controversial aspects raised in the *Kneschke vs. LAION* case, particularly concerning the

definition of ‘scientific research’ and the scope of beneficiaries of Art. 3 (including public-private partnerships), and the practical application of a machine-readable ‘opt-out’ mechanism of Art. 4 of the CDSM Directive.

### Case Analysis

In *Kneschke vs. LAION*, the Court ruled that the reproduction of protected works (through their download) by LAION in order to create training datasets falls within the meaning of ‘scientific research’ of Art. 3 of the CDSM Directive (Sect. 60 D of the German Copyright Act). Therefore, the preparatory work step in AI training should be covered by the specific TDM exception for scientific research as it constitutes a prerequisite for obtaining new knowledge during the training phase. Further, the Court stressed that there is no need to demonstrate that the research carried out at the preparatory phase of AI training leads to later research success to be considered ‘scientific research’. This point seems somewhat controversial, as the concept of ‘scientific research’ generally implies that the process contributes to the advancement of science in either theoretical or practical domains (for more discussions on ‘scientific research’, see [Peers et al.](#), pp.420-421). If the Court considered the preparatory phase as the essential and integral step of AI training, it would be the case, as the outcome of the whole process would matter. In the case of LAION, the training datasets are made publicly available free of charge, contributing to the advancement of science in the AI field. However, the Court considered the preparatory phase as a separate and independent process in relation to the later data analysis (the actual training). Logically, the process should contribute somewhat to the subsequent research success in the data training phase to qualify as ‘scientific research’. This seems crucial, as successful training would produce highly effective datasets for AI developers to train their systems on a wider range of patterns and correlations derived from diverse and adequate input data. The development of highly effective AI systems could facilitate the overall growth and advancement of the EU AI sector (see [Manteghi](#)).

Further, the Court dispelled all doubts that LAION falls within the scope of beneficiaries of Art. 3 of the CDSM Directive. The company makes training datasets publicly available free of charge, so it does not pursue commercial purposes. In this regard, it does not matter that LAION may actually benefit from creating such datasets for developing its commercial offerings. Moreover, it does not matter whether commercial actors then use such databases to train their models or not. The CDSM Directive does not limit the exception in Art. 3 to non-commercial research purposes. Indeed, both commercial and non-commercial research may generate valuable findings which would facilitate economic and technological developments. Additionally, it is argued that when knowledge is commercialised, it is often transformed into ‘practical’ forms (e.g., medication, apps, GPS systems, etc.) which could be directly accessible to the public (see [LIBER](#)). However, if a commercial company has a decisive influence on a not-for-profit research organisation (such as LAION)—for instance, because of the structural situations—this organisation should not be viewed as a beneficiary of Art. 3 (Recital 11, 12 and Art. 2 (1) (b) of the CDSM Directive). In this context, the plaintiff claimed that this was the case because of several issues. First, a co-founder of LAION is currently employed at the commercial (private)

company [Stability AI](#). Second, this commercial company provided computing resources to LAION in the early phase of its development. Third, this company, like many other commercial entities, now uses the LAION datasets to train its models, which subsequently become available for commercial use. However, the Court held that it did not demonstrate that LAION is dependent on Stability AI and that this commercial entity has a decisive influence on LAION's research (e.g., has preferential access to the research results). Nevertheless, the Court did not provide any additional clarification.

Further, in obiter dicta, the Court considered the potential application of a so-called 'commercial' TDM exception to AI training (Sec. 44b of the German Copyright Act, Art. 4 of the CDSM Directive). The exception is much broader than the exception for scientific research as it covers all types of users and any purpose. However, it can be diluted by a so-called 'opt-out' mechanism, which allows rightsholders to reserve the use of their works for TDM through contractual agreements or by using machine-readable means, including metadata and terms and conditions of a website or a service (Recital 18 of the CDSM Directive). The reservation right faced significant criticism for strengthening the position of authors and other rightsholders in relation to the EU AI sector (see, among others, Manteghi [here](#) and [here](#), [Tyner](#), [Hugenholtz](#), [Senftleben](#)). Many commentators argue that the 'opt-out' mechanism of Art. 4 of the CDSM Directive would require users to pay twice: first for 'lawful access' to works and then for reading and analysing training data, primarily benefiting large AI actors (see, e.g., [Ziaja](#), [Manteghi](#) p. 675). Moreover, it is still unclear how the 'opt-out' mechanism should work in practice as there are no generally recognised rules or protocols (for more on this, see, e.g., [Mezei](#), [Keller](#), [Manteghi](#)). In *Kneschke vs. LAION*, the Court recognised that the prohibition of scraping by automated tools without authorisation, written in natural language in the terms and conditions of the website where the photo in question was accessed, explicitly serves as a machine-readable 'opt-out' mechanism. This is because machines can read and understand the content of text written in natural language. The Court's consideration of the reservation right would provide more legal certainty to the use of this mechanism in future cases. However, the reasoning appears to be somewhat ambiguous, as it is not clear whether all machines are capable of interpreting natural language terms correctly and uniformly in different contexts, taking into account the complexity of legal texts.

## Conclusion

The ruling in *Kneschke vs. LAION* narrowly focused on AI training, limiting the analysis to the preparatory phase. However, the Court confirms that TDM exceptions cover AI training and clarifies (to some extent) the practicalities of reproducing protected works for building training datasets, as well as the reservation right under Art. 4 of the CDSM Directive. The judgement is significant for future cases as it, for instance, acknowledges that the creation of training datasets could be part of a broad process of 'scientific research' and strengthens the position of research organisations in their cooperation with private companies. However, some of the Court's considerations (e.g., the practical application of a machine-readable 'opt-out' mechanism) remain vague and may raise further questions in the future. The decision will guide the future use of protected

works for AI training, however, as AI continues to evolve, it is important to ensure that the EU regulation on TDM is properly adjusted to these changes.

---

**Maryna Manteghi is a doctoral researcher in the Faculty of Law at the University of Turku, Finland.**