

**Max-Planck-Institut
für Innovation und Wettbewerb**

Copyright in the Spotlight - 30th Anniversary of the Hellenic Copyright Organization

Athens, 7 March 2025

**Article 4 DSM Directive and genAI:
Where do we stand?**

Prof. Dr. Silke von Lewinski



Applicability of Art. 4 DSM Directive to training of genAI models

Art. 4 (1) DSM-Directive, combined with Art. 2 no (2) DSM-Directive (Definition)

1. Member States shall provide for an **exception or limitation** to the rights provided for in [...] for reproductions and extractions of lawfully accessible works and other subject matter **for the purposes of** [text and data mining, ie,] any **automated analytical technique aimed at analysing** text and data in digital form in order to **generate information** which includes but is not limited to **patterns, trends and correlations**;

Recital 8 DSM-Directive

(8) New technologies enable the **automated computational analysis** of information in digital form, such as text, sounds, images or data, generally known as text and data mining. Text and data mining makes the processing of large amounts of information **with a view to gaining new knowledge and discovering new trends** possible [...]



Applicability of Art. 4 DSM Directive to training of genAI models

- Acts of reproduction occur:
 - When works etc are collected and stored in databases as a basis for the data corpus
 - When AI outputs correspond to works contained in such input databases
 - But also **during the training of generative AI models?**



Applicability of Art. 4 DSM Directive to training of genAI models

Classical TDM :

During training, **only semantic information** of works is accessed, not syntax (idea vs expression) (eg, number of female vs male names in literature of 1970s) – thus, TDM itself not in conflict with copyright

Is training of AI models TDM? No!

- technological reasons:

- not only semantic, but also **syntax information is copied**;
- extensive use also of expression/syntax is the **very purpose** of genAI training,
and **indispensable** to achieve AI outputs similar to human works
- much larger datasets for genAI needed than for TDM



Applicability of Art. 4 DSM Directive to training of genAI models

Is training of AI models TDM? No!

- legal reasons:

Wording and purpose of Art. 4 and 2 no. (2):

- purpose of reproduction must only be ‚analysis‘ of data, not generation of new products, and
- this only to gain patterns, trends, correlations, information/new knowledge, but not to produce images, texts, music etc, similar to input

Historical interpretation (2019!); also AI Act: no copyright instrument, no discussion of genAI (with its severe consequences for copyright holders), only reminder to respect law where TDM is applied



Applicability of Art. 4 DSM Directive to training of genAI models

For more detail, see common study of computer scientist Prof. Stober and law professor T. Dornis:

Study in German: T. Dornis and S. Stober, Urheberrecht und Training generativer KI-Modelle - technologische und juristische Grundlagen (September 4, 2024). Available at SSRN:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4946214



Applicability of Art. 4 DSM Directive to training of genAI models

For more detail, see english articles of Prof. Dornis:

Dornis, Tim W., The Training of Generative AI is not Text and Data Mining (October 19, 2024). Available at **SSRN:**

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4993782

Dornis, Tim W., Generative AI, reproductions inside the model, and the making available to the public (November 27, 2024). Available at **SSRN:**

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5036008



Consequences?

Exception does not apply (thus, opt out irrelevant), **use must be licensed.**

However, given **different views** on applicability of Art. 4 to training of genAI models: **legal uncertainty**

Practice (examples from Germany):

- Litigation: GEMA claim against OpenAI and SunoAI (Nov. 24, Jan. 25)
and
- Negotiation and licensing
 - CMO VG Wort: AI licence for internal use of companies
 - Collective agreement between union of actors and film producers, 1.3.25
 - Individual agreements between authors and publishers



Thank you for your attention!
svl@ip.mpg.de

