

Text and Data Mining Limitation in Practice: Judgment GEMA vs. Open AI

Earlier this month, we wrote about the challenges of Text and Data Mining (TDM), which you can read about in the article [New Rules of the Game: How the Italian AI Framework Act Redefines the Boundaries of Text and Data Mining Limitations in Copyright Law](#). Italy is thus the first EU Member State to adopt concrete measures by formulating a legislative framework through the enactment of a comprehensive Artificial Intelligence Act, which also introduced changes in the field of TDM limitations. Unlike Italy, the **German court**, only one month later, in the judgment in the case **GEMA vs. OpenAI**, had the opportunity to set guidelines regarding the practical application of this limitation

GEMA vs. Open AI – Reasoning of the Judgment and the Court’s Legal Positions

The District Court of Munich I (Landgericht München I) published on 11 November 2025 a press release regarding the first-instance judgment rendered in the proceedings initiated by the collective management organization for musical authors, **GEMA**, against two companies of the **OpenAI group** (Az. 42 O 14139/24) for copyright infringement through **unlawful reproduction by memorizing works in and through a language model**.

The part of GEMA’s claim relating to **cessation and prohibition of further infringements, an obligation on the part of OpenAI to provide information on the extent of the use of works and revenues generated thereby, as well as damages for copyright infringement** in relation to nine musical compositions of well-known German authors (“Atemlos” (Kristina Bach), “36 Grad” (Thomas Eckart, Inga Humpe, Peter Plate, Ulf Leo Sommer), “Bochum” and “Männer” (Herbert Grönemeyer), “Über den Wolken” (Reinhard Mey), “Junge” (Jan Vetter), “Es schneit”, “In der Weihnachtsbäckerei” and “Wie schön, dass du geboren bist” (Rolf Zuckowski)) **was upheld**, whereas the part of the claim seeking **damages for violation of the general right of personality** due to modified lyrics being attributed to the authors **was dismissed**.

Below we provide an overview of the reasoning and legal positions adopted in the judgment published on the [official legal portal of Bavaria](#).

GEMA established its **active standing** to initiate and conduct these proceedings on the basis of specific agreements concluded with the lyricists of the nine compositions at the end of 2024, invoking previously existing authorizations and, for the purpose of removing any ambiguity, expressly regulating the transfer of exclusive rights and granting authorization for the exercise of moral rights.

Under these agreements, the authors, as original rightsholders of the disputed lyrics, “transferred” to GEMA “the exclusive right to use the song lyrics (with or without music), in whole or in part, in artificial intelligence models and systems and in applications based on them, in particular for training and fine-tuning, for processing contextual information and in the outputs of language models.” The agreements expressly include the use in chatbots and specific data-processing and data-search systems within the training processes of large language models.

With respect to the works represented by it, GEMA issued a **reservation (“opt-out”)**, prohibiting their use without authorization for the purposes of automated text and data mining, i.e., it expressly excluded the works in its repertoire from the application of the **TDM limitation** by publication on its website.

The defendants are operators of language models and chatbots of the **OpenAI group**, whose **passive standing** is based on the fact that the disputed lyrics of the nine compositions were contained in the training datasets and fully and unchanged memorized in the training models provided by the defendants, and that these lyrics were displayed as **outputs** in response to simple user prompts.

GEMA argued that the **infringement** was committed by acts of **reproduction** of the disputed lyrics **within the model, public communication** of the disputed lyrics when displayed through the chatbot, their **reproduction on users' devices and in chat histories**, violation of the right of **adaptation**, violation of the **integrity** of the work, and violation of the **general right of personality** by attributing modified song lyrics to the authors.

Regarding the finding of infringement through reproduction within the model, the court started from the position that interference with the reproduction right in large language models is a known issue, while emphasizing the need to distinguish between reproduction for the purposes of digitalization or analysis, and reproduction that remains embedded within the model. What is decisive is whether the works have been reproduced within the model by memorization.

In the present case, the court held that **memorizing the disputed lyrics in the defendants' models constitutes an act of reproduction of copyrighted works** both under German law and under Article 2 of the InfoSoc Directive, which grants authors the exclusive right to authorize or prohibit direct or indirect, temporary or permanent reproduction, by any means and in any form, in whole or in part, of their works. The fact that the disputed lyrics were incorporated as training data into the defendants' models and that these lyrics appeared in the outputs generated by simple prompts constitutes clear evidence that the works were reproduced within the language models themselves.

The court confirmed that reproduction is not limited to identical copies but includes any altered reproduction, including digitalization of analog works, MP3 compression, reduced-resolution images, etc., that the duration of the reproduction is irrelevant, and that reproduction includes any fixation of a work in material form which is suitable to be perceived, directly or indirectly, by human senses. Indirect perception exists when a work becomes perceptible through technical means. Applying these criteria, the court found that **reproduction occurred within the model** and that the disputed **lyrics were materially fixed** and could become **indirectly perceptible**.

Such permanent **reproduction** within the model is not justified, and memorization of training data **cannot fall within the TDM exception**, because the reproduction **does not serve further analysis of the data**. In this sense, the court held that although language models such as the disputed ones generally fall within the scope of the TDM limitation, in this situation the limitation cannot apply because in the analysis and

learning phase the model does not merely extract information from the training data but reproduces the works. Memorizing the disputed song lyrics exceeds the scope of analysis; they were not only analyzed but fully incorporated into the model parameters, thereby interfering with the authors' exploitation interests.

The research-organization and cultural heritage institutions TDM exception does not apply either. The defendants are not privileged research organizations, and even considering the parent entity, the purpose would remain commercial if the companies conduct research aimed at developing and marketing products or services.

The defendants challenged **GEMA's opt-out from the TDM exception**, claiming that the disputed lyrics were lawfully uploaded online with the consent of rightsholders and without protective measures, rendering each of the nine lyrics available for AI training. The court rejected this argument, holding that model training is not a usual or expected use of a work to which an author must consent.

The model outputs also **constitute copyright infringement** because, in response to certain prompts, the model generated content substantially identical to parts of the training material. The court held that the disputed outputs interfere with copyright, that the lyrics are recognizable in the outputs, and that the works are thereby reproduced and made publicly available. The outputs also constitute acts of public communication and adaptation carried out without authorization.

The act of communication by the defendants is direct because they themselves enable public enjoyment of the work, not merely indirectly by providing infrastructure for third-party acts. According to the court opinion, the defendants play a central role in enabling public communication, and chatbot users constitute a new public. A significant distinction exists compared to search engines: users no longer need to visit the original webpage, and the chatbot replaces the source.

The original songs were reproduced in recognizable form in the outputs. To establish copyright infringement through reproduction of protected parts, it is decisive only whether the protected parts are recognizable in the new work, not the overall impression that differentiates the works. The court found that the lyrics of "Atemlos", "Bochum", "Junge", "In der Weihnachtsbäckerei", and "Wie schön, dass du geboren bist" were reproduced to such an extent that recognition was beyond doubt, and that the outputs containing the other lyrics ("36 Grad", "Männer", "Über den Wolken", "Es schneit") clearly contained recognizable original elements and did not sufficiently differ from the original texts.

The defendants argued that there was no infringement because the user, by formulating the prompt, is the author of the reproduction, initiating the automated generation process. However, the court rejected this argument, stating that the defendants, as model operators, control the architecture, training, and selection of training data, as well as the structure of the training process leading to the memorization of specific content and the implementation or omission of protective filters. Therefore, they are responsible for the memorization and subsequent availability of the lyrics.

The justified part of the claim for injunction and prohibition resulted in a justified claim for damages and an obligation of defendants to provide information on the extent of the use of the works in the language models and revenues generated through such use.

The claim for damages for violation of the general right of personality due to attribution of modified lyrics to the authors was rejected. Such a claim would be justified only if misattributed authorship could cause confusion of identity. Relying on established case-law, the court held that the application of the general right of personality as a means of protection against misattribution is permissible only in situations involving the attribution of works created by third parties, which is not the case here. Personality protection would arise only if serious consequences, such as stigmatization or social exclusion, occurred - conditions not fulfilled by the disputed outputs.

Significance of the Legal Positions in the Proceedings

Although the judgment is not final, its legal positions are significant for defining acts constituting infringement, the application of TDM limitations, and the liability of AI-developing companies.

Acts constituting **infringement** include:

- **memorizing protected works within language model parameters**, which constitutes unlawful reproduction; the mere appearance of parts of the work in outputs is sufficient to prove memorization and therefore reproduction; and
- **model outputs** which, in response to a simple prompt, **reproduces parts of song lyrics** in a form almost identical to the original is considered as a new **act of reproduction, public communication, and adaptation of the work**. The output itself infringes the author's rights.

TDM limitations do not apply in this case. The court held that OpenAI is neither a research organization nor a cultural heritage institution within the meaning of the DSM Directive, even if it partially engages in research activities, because its primary and ultimate purpose is commercial. The TDM limitation permitting copying for analytical purposes also does not apply because training involving memorization of copyrighted works exceeds analysis and temporary copying; it constitutes permanent "writing" of works into the model. Moreover, GEMA validly excluded the application of TDM limitations, and the argument that "freely available on the internet" implies trainability was rejected with the reasoning that **training a model does not constitute an expected or customary use of the work from the author's perspective**.

Liability for infringement is on **OpenAI**, not on the user, because OpenAI is the one who has designed the model architecture and trained it on lyrics from repertoire of collective management organization GEMA's. OpenAI enables access to the content for a broad public, while the user does not "publish" anything but merely triggers the system.

It is expected that the German court's stance—that AI companies cannot freely use copyrighted works for training without licenses or agreements, and that exceptions such as TDM do not grant unlimited freedom to train models on copyrighted content—will significantly influence disputes concerning other categories of copyrighted works. Similar developments are expected beyond the domain of song lyrics.

The influence of the economic power of AI companies on legal practice will also be interesting to observe, particularly since the defendants argued during proceedings that establishing copyright infringement could force them to cease offering their product in Germany, potentially causing negative effects in areas of public interest—education, science, and research—where their technology plays a central social and economic role and transformative potential.

Rather than a Conclusion

The judgment of the District Court of Munich in the case **GEMA vs. OpenAI** represents the first clear judicial articulation of the boundaries between permissible TDM processing and impermissible permanent memorization of copyrighted works in large language models. The court has thereby set practical guidelines that will inevitably affect the European market for AI development and deployment. It is particularly significant that the court draws a clear distinction between technical, temporary processing operations and the permanent “inscription” of copyrighted works into the model’s parameters, considering the latter a form of reproduction that requires the author’s explicit authorization.

This interpretation significantly narrows the scope of TDM limitations under EU law and confirms that commercial AI systems cannot automatically rely on broad freedom to capture and process content available online. Solutions will therefore rely on licensing, contractual arrangements, or technical measures preventing memorization of protected works. This also reinforces the principle that the model operator—not the end-user—is responsible for how the model is trained and what it generates.

In the broader context, the judgment signals the future direction of European jurisprudence: artificial intelligence remains subject to copyright law, and exceptions such as TDM must be interpreted narrowly, balancing the protection of authors and rightsholders on the one hand and technological development on the other. As AI further evolves, this approach is likely to extend to other categories of content—visual works, literary texts, journalistic articles, and audiovisual productions—making the Munich case a key reference for future disputes at the intersection of AI and copyright.

This article is to be considered as exclusively informative, with no intention to provide legal advice. If you should need additional information, please contact us directly.