

Munich court rules ChatGPT's memorised song lyrics infringe copyright (GEMA v OpenAI)

Source copy kept by wheresthe.ai for the story "Munich court rules ChatGPT's memorised song lyrics infringe copyright (GEMA v OpenAI)" (<https://wheresthe.ai/d/de-gema-v-openai-munich-judgment-2025-11/>).

Source: <https://www.taylorwessing.com/de/insights-and-events/insights/2025/11/gema-v-openai>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 15b00179f1658172b1d564467b684a8d43207d25aaa7221d4de6c471e952bf49

Rightsholders Take the Lead: GEMA v. OpenAI

Home

News & Insights

Insights

Rightsholders Take t...

Diese Seite teilen

Diese Seite teilen

Auf dieser Seite

Autor

Dr. Christian Frank, Licencié en droit (Paris II / Panthéon-Assas)

Partner

München

Read More

Autor

Dr. Christian Frank, Licencié en droit (Paris II / Panthéon-Assas)

Partner

München

Read More

Auf dieser Seite

Diese Seite teilen

Diese Seite teilen

11. November 2025

Rightsholders Take the Lead: GEMA v. OpenAI

In-depth analysis

Regional Court of Munich I — 11 November 2025, Case No. 42 O 14139/24

Overview

In today's ruling (11.11.2025), the Regional Court of Munich I essentially upheld GEMA's claims for injunctive relief, information, and damages against two companies in the OpenAI group (Case No. 42 O 14139/24). The court dismissed claims based on a violation of general personality rights arising from incorrect attribution of modified song lyrics. GEMA—the German collecting society for musical performing and mechanical reproduction rights—sued OpenAI for copyright infringement concerning the lyrics of nine well-known German authors, arguing

that the lyrics were memorized by OpenAI's language models and, when the chatbot was used, were reproduced in large parts verbatim in response to simple user queries. OpenAI argued that its language models do not store or copy specific training data, but reflect—within their parameters—patterns learned from the entire training dataset. Because outputs are generated in response to user prompts, responsibility for any output lay with the user as the creator. In any event, any legal infringements were said to be covered by copyright limitations, in particular the text-and-data mining (TDM) exception.

The court held that GEMA is entitled to the asserted claims both for (i) reproduction in the language models and (ii) reproduction in outputs. In the court's view, memorization within the models and the subsequent reproduction of lyrics via the chatbot each infringed exclusive rights of exploitation. None of the relevant exceptions applied—in particular the TDM exception in § 44b UrhG / Art. 4 DSM Directive. The lyrics were found to be reproducible using ChatGPT-4 and ChatGPT-4o. The court observed that “memorization” is recognized in information-technology research: large language models do not merely extract information from the training dataset during training; they may also adopt training data in their post-training parameters. Here, memorization was established by comparing the lyrics in the training data with the model's outputs. Given the complexity and length of the songs, the court ruled out coincidence as the cause of the reproductions.

To situate the court's reasoning, it is helpful to compare two recent decisions abroad: *Getty Images v. Stability AI* (High Court of Justice, London, 4 Nov 2025, [2025] EWHC 2863 (Ch)) and *Kadrey v. Meta* (N.D. Cal., 25 Jun 2025, No. 23-cv-03417-VC).

Getty Images v. Stability AI (UK) — Core Reasoning

The Getty court has adopted the experts' account that diffusion models learn patterns/statistics rather than store training images:

"Stable Diffusion does not itself store the data on which it was trained ... Rather than storing their training data, diffusion models learn the statistics of patterns... It is impossible to store all training images in the weights... LAION-5B ~220TB vs 3.44GB model weights."

[items 552–554]

The court then articulated the legal test for an “infringing copy” (i.e., whether anything in the model amounts to a stored reproduction). Relying on CDPA s.17 (copying includes electronic storage in any medium) and *Sony v Ball* (RAM can be an infringing copy while it contains the data), the judge wrote:

"An infringing copy must be a copy... I cannot see how an article can be an infringing copy if it has never consisted of/stored/contained a copy. In *Sony v Ball* the RAM chip was only an infringing copy while it contained the copy..."

[items 584, 587]

The court confirmed that storage in intangible media (e.g., cloud) is still “storage in any medium” under s.17. Applying this test to model weights and “memorization,” the court accepted as a technical matter that memorization can occur (e.g., watermarks with high duplication), but emphasized that Getty did not allege or prove that any copyrighted work was memorized or stored in the weights:

"There is no evidence of any Copyright Work having been ‘memorized’... and no evidence of any image having been derived from a Copyright Work."

[item 559]

Doctrinally, therefore, even though training elsewhere involved reproductions, weights that never contain a copy are not an ‘infringing copy’:

"Is an AI model which derives from a training process itself an infringing copy? In my judgment, it is not... by the end of that process the model does not store any of those works... The model weights... have never contained or stored an infringing copy."

[items 599–600]

Bottom line (Getty): The court distinguishes learned parameters from stored reproductions. “Memorization,” without proof that the article contains a copy at some point, does not make weights an infringing copy.

Kadrey v. Meta (US) — Core Reasoning

The Kadrey court does not analyze “weights” as copies. It addresses training-stage copying under fair use and treats “output regurgitation” as relevant to market harm, not as a weights-as-copies theory.

Assessing memorization/regurgitation, the court asked whether Llama regurgitates plaintiffs’ books (as output-level reproduction). It found that record evidence showed mitigations and only de minimis regurgitation under adversarial prompts:

“Meta... post-trained its models to prevent them from ‘memorizing’ and outputting certain text... Even using that method [adversarial prompting], the expert could get no model to generate more than fifty word and punctuation marks ... Llama cannot currently be used to read or otherwise meaningfully access the plaintiffs’ books.”

[p. 12 (bottom)–13 (top)]

That output evidence fed Factor 4 (market effects). The court repeatedly stressed that market substitution is “undoubtedly the single most important” factor, and held that plaintiffs failed to offer empirical proof of harm on this record.

Bottom line (Kadrey): No weights-as-copies analysis. “Memorization” appears only as regurgitation risk relevant to market harm; the court found little proof of such regurgitation here.

GEMA v. OpenAI — Reasoning in More Detail

The disputed lyrics were found to be reproducibly embedded in the models; no deceptive prompting was required.

“Due to memorization, embodiment as a prerequisite for copyright reproduction of the song lyrics at issue is given by data in the specified parameters of the model. The song lyrics at issue are reproducibly fixed in the models. ... For the purposes of copyright reproduction, it can remain open how memorization works in detail. It is irrelevant whether we are talking about storing or copying the training data or, as the defendants put it, whether the model reflects in its parameters what it has learned based on the entire training data set, namely relationships and patterns of all words or tokens that represent the diversity of human language and its contexts. The decisive factor is that the song lyrics that served as training data are reproducibly contained in the model and thus embodied in it.”

Item 3 b bb (1)

The fact that the model represents content as probabilities is irrelevant. New technologies such as language models fall within the right of reproduction (§ 16 UrhG / Art. 2 InfoSoc Directive) and the right of making available to the public (§ 19a UrhG / Art. 3 InfoSoc Directive). Under CJEU case law, indirect perceptibility suffices for reproduction where the work can be perceived with technical aids.

The court held that reproduction within the models is not covered by the TDM exception in § 44b UrhG.

In language models, text and data mining aims to evaluate information such as abstract syntactic rules, common terms, and semantic relationships. This means that levels of expression such as word choice, range of expression, and repetitions are also evaluated This raw information from training data can be converted into model parameters. However, memorizing the song lyrics at issue exceeds such evaluation and is therefore not merely text and data mining. The song lyrics as training data were not evaluated alone but were incorporated in their entirety into the parameters of the model, which in turn interferes with the exploitation interests of the authors.

Item 4a. bb. (2) cc.

The court also confirmed OpenAI’s responsibility and rejected an attempt to shift responsibility to users. The outputs at issue were generated by simple prompts. The defendants operated the language models, chose and used the training data (including the lyrics), and were responsible for the model architecture and any memorization of training data. The models therefore had a significant influence on the outputs, and the specific content was generated by the models.

Bottom Line

The GEMA decision accepts in principle the distinction between learned parameters and stored reproductions, but —on the facts—finds memorization and reproduction in the models and infringing outputs. In Getty, the evidentiary

record did not reveal comparable reproductions, which appears decisive to the different outcomes; the approaches are analytically comparable but part ways on proof. Kadrey likewise turns on evidence, weighing market effects (Factor Four) against the purposes of fair use; analogously, the GEMA court weighs the scope and rationale of the TDM exception against rightsholders' exploitation interests.

Unlike Getty and Kadrey, the GEMA court squarely addresses the relevance of infringing outputs and holds OpenAI—as provider—responsible for the generation of reproductions in outputs.

The judgment will likely be appealed and could ultimately reach the Court of Justice of the European Union (CJEU)—so there is more to come.

Branchen Technologie, Medien & Kommunikation

Newsletter-Anmeldung

Wählen Sie aus unserem Angebot Ihre Interessen aus!

Jetzt abonnieren

Jetzt abonnieren

Related Insights

Technologie, Medien & Kommunikation Vom Opt-out zur Marktordnung: Die Entschließung des Europäischen Parlaments zu GenAI und Urheberrecht vom 10. März 2026

16. März 2026

Briefing

von Dr. Christian Frank, Licencié en droit (Paris II / Panthéon-Assas)

Klicken Sie hier für Details

Artificial intelligence KI und digitale Daten im Koalitionsvertrag 2025 – Eckpunkte und europarechtlicher Kontext

11. April 2025

In-depth analysis

von mehreren Autoren

Klicken Sie hier für Details

Trade secrets Transparenz vs. Rechtssicherheit – Spezifizierungslast in Geheimhaltungsklauseln

8. April 2024

Briefing

von Dr. Christian Frank, Licencié en droit (Paris II / Panthéon-Assas) und Dr. Julia Freifrau von Imhoff

Klicken Sie hier für Details

Newsletter-Anmeldung

Über Taylor Wessing

Campaigns and online tools

News

Unsere Anwälte

Expertise

International

Insights

Events

[Presse & News](#)

[Kontakt](#)

[Karriere](#)

[Datenschutzhinweise](#)

[Datenschutzbestimmungen & Cookie Hinweise](#)

[Rechtliche und regulatorische Informationen](#)

[Nutzungsbedingungen](#)

[Scam E-Mails](#)

[© Taylor Wessing](#)