

GEMA versus OpenAI: The Key Takeaways

Mohamed Ataul Karim

Introduction

The claimant, GEMA, is a German music collecting society that manages the rights to musical works transferred to it by composers, lyricists, and music publishers.¹ The defendants, OpenAI Ireland Ltd (operator of ChatGPT in Europe) and OpenAI L.L.C. (parent company), are a group of companies that develop, operate, and license generative artificial intelligence.² The dispute concerns the use of nine (9) German song lyrics to train GPT large-language models and chatbots (Models 4 and 4o), without obtaining a licence from GEMA.³ The claimant sued the defendants in the 42nd Civil Chamber of the Munich I Regional Court for copyright infringement arising from the unauthorised use of those song lyrics (case no. 42 O 14139/24).⁴ On November 11, 2025, the Court ruled in favour of the claimant. This paper outlines two key takeaways of this decision.

The Key Takeaways of the Decision

The first key takeaway is that the Court determined that the reproduction of copyrighted song lyrics occurred through fixation, storage, and memorisation in the AI model.

The claimant argued that copyright infringement occurred in the model (by fixation and storage of lyrics). This is a reproduction in the sense of Article 2 of the InfoSoc Directive,⁵ and Section 16 of the German Copyright Act (UrhG).⁶ The defendants submitted that models do not store

¹ Certified copy of the decision (English translation), Case no. 42 O 14139/24, Munich Regional Court I, delivered on 11 November 2025, p.3 (hereinafter the Court's Decision).

² *ibid*

³ *ibid*

⁴ *ibid*

⁵ Article 2: Reproduction right

Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part:

- (a) for authors, of their works;
- (b) for performers, of fixations of their performances;
- (c) for phonogram producers, of their phonograms;
- (d) for the producers of the first fixations of films, in respect of the original and copies of their films;
- (e) for broadcasting organisations, of fixations of their broadcasts, whether those broadcasts are transmitted by wire or over the air, including by cable or satellite

⁶ Section 16: Right of reproduction

(1) 'Right of reproduction' means the right to produce copies of the work, whether on a temporary or on a permanent basis and regardless of by which means of procedure or in which quantity they are made.

(2) The transfer of the work to devices for the purposes of repeated communication of video and sound sequences (video and audio recordings), regardless of whether this is the recording of a communication of the work on a video or audio recording medium or the transfer of the work from one video or audio recording medium to another, also constitutes reproduction.

or copy the training data (in this case, song lyrics).⁷ Instead, they learn statistical relationships and patterns from the entire training dataset. Moreover, the outputs are generated based on statistical probabilities: prediction of the following words in a large language model (LLM) in response to a user's prompt. Eventually, a model does not serve as a database of training data by storing it and reproducing it when searched with the user's prompts.

The Court decided that the texts of the song lyrics at issue were physically fixed in the models, which constitutes a reproduction.⁸ The defendants contested this in their submissions. According to the defendants, there was no physical determination of the training data in the model (a requirement for reproduction).⁹ They submitted that the claimant could not prove the exact fixation of song lyrics in the Model.¹⁰ Instead, the claimants only referred to memorising training data, which is insufficient to substantiate the fixation.¹¹ The Court rejected this argument. According to the Court, a concrete physical determination of the dataset (lyrics) in the models is not required.¹² It is sufficient to show that lyrics exist within the model's overall parameters. This view was based on decomposing song lyrics into parameters and then identifying those parameters within the Model as a whole.¹³

Regarding the memorisation of training data by models, the Court found that song lyrics are reproducibly defined in the models.¹⁴ The memorisation of lyrics in the Models effectively results in regurgitation: the reproduction of outputs verbatim or with slight derivations of the training inputs (song lyrics). A simple prompt can reproduce the song lyrics as the model's outputs. In this case, the claimant used the following prompts: "What is the text of [song title]", "Who wrote the lyrics", "What is the refrain of [song title]", "Please also name me the 1st verse", and "Please also name me the 2nd verse".¹⁵ The chatbots produced the exact verbatim or slightly modified outputs of the protected song lyrics.¹⁶

The defendants tried to avoid liability. They argued that models are not designed to generate training data again.¹⁷ They have adopted measures to prevent the generation of outputs wholly

⁷ The Court's Decision, p. 19

⁸ *ibid*, 37

⁹ Under section 16 of the German Copyright Act.

¹⁰ The Court's Decision, p. 22

¹¹ *ibid*, 22

¹² *ibid*, 38

¹³ *ibid*

¹⁴ *Ibid*, 37

¹⁵ *Ibid*, 8

¹⁶ An example, the chat history for the song lyrics "Men" and "36 Degrees" is submitted in Exhibit K.

¹⁷ The Court's Decision, p. 19

or partly mimicking training data or undesirable content.¹⁸ However, current state-of-the-art methods cannot entirely prevent the memorisation problem in AI models.¹⁹ This is a problem that machine learning scientists are now trying to solve. Moreover, the claimant provoked the model through several well-designed prompts. Prompt engineering disrupted the model's neutral path, turning it into an “expert in song lyrics”.²⁰ In essence, the model will not operate this way in everyday use.

The Court viewed that the model's outputs, produced by memorisation, are reproductions within the meaning of EU and German Copyright laws. In summary, the analysis of fixation, storage, and memorisation of song lyrics in the AI Models is anchored in the broad notion of reproduction under Article 2 of the InfoSoc Directive and Section 16 of the German Copyright Act.

The second key takeaway is that the Court held the text and data mining (TDM) exception inapplicable to the above reproductions. TDM means “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.”²¹ The defendants argued that the reproductions were covered by sections 44b(2)²² and 60d(1)²³ of the German Copyright Act. In brief, the former permits reproduction for any TDM purposes, while the latter allows reproduction for non-commercial TDM by privileged research institutions.

¹⁸ *ibid*, 20

¹⁹ *ibid*

²⁰ *ibid*

²¹ Article 2(2), the Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance.) PE/51/2019/REV/1, OJ L 130, 17.5.2019, pp. 92–125 (the DSM Directive)

²² Section 44b: Text and data mining

(1) ‘Text and data mining’ means the automated analysis of individual or several digital or digitised works for the purpose of gathering information, in particular regarding patterns, trends and correlations.

(2) It is permitted to reproduce lawfully accessible works in order to carry out text and data mining. Copies are to be deleted when they are no longer needed to carry out text and data mining.

(3) Uses in accordance with subsection (2) sentence 1 are permitted only if they have not been reserved by the rightholder. A reservation of use in the case of works which are available online is effective only if it is made in a machine-readable format.

²³ Section 60d: Text and data mining for scientific research purposes

(1) It is permitted to make reproductions to carry out text and data mining (section 44b (1) and (2) sentence 1) for scientific research purposes in accordance with the following provisions.

(2) Research organisations are authorised to make reproductions. ‘Research organisations’ means universities, research institutes and other establishments conducting scientific research if they

1. pursue non-commercial purposes,

2. reinvest all their profits in scientific research or

3. act in the public interest based on a state-approved mandate.

The authorisation under sentence 1 does not extend to research organisations cooperating with a private enterprise which exerts a certain degree of influence on the research organisation and has preferential access to the findings of its scientific research.

The defendants advanced three key arguments in support of the TDM exception.²⁴ First, reproductions for AI training are not limited to creating a training corpus; they also extend to subsequent training activities. Second, the parent companies of the defendants' groups were not for-profit oriented. The sub-companies of the groups are for-profit. The defendants argued that the sub-companies' profit-based business model does not change the legal status of the group as a whole. Thus, the defendants should be treated as a privileged research organisation. Third, the claimant did not effectively opt out of TDM. The defendants submitted that they duly consulted through the robots.txt protocol. This is the market standard for machine-readable opt-out from TDM. But the claimant failed to opt out via robots.txt. Instead, the claimant made a declaration of opt-out only in 2024, which was not in machine-readable format.

The claimant, on the other hand, argued that the TDM exception is not maintainable.²⁵ First, the reproductions in the model were not exempted under section 44b of the German Copyright Act. They were not used for data or computational analysis. Second, the defendants cannot resort to Section 60d since they are not privileged research organisations. Third, the claimant had opted out of TDM by effective reservation. The claimant argued that the defendants have the burden to disclose all sources of the song lyrics.²⁶ Moreover, the robots.txt Protocol has some limitations. It is not entirely suitable for opting out of TDM. The reservation can be made for broad directories, subdirectories or URLs, but not for individual, specific content or works.²⁷ Effective reservations cannot be made for conditions of use or licensing terms via the robots.txt Protocol.²⁸

The Court recognised that there may be three phases of AI training: pre-training, training and post-training phases:²⁹

1. Pre-training: creation of training data by extraction and conversion of training data into a machine-readable format
2. Training: the analysis of data for training an AI model
3. Post-training deployment: the use of the trained model through prompts and outputs

According to the Court, the TDM exception may cover some pre-training activities, such as the reproduction and extraction of lawfully accessed works. But it does not cover reproductions in

²⁴ The Court's Decision, pp.22-23

²⁵ *ibid*, 14

²⁶ *ibid*

²⁷ *ibid*

²⁸ *ibid*

²⁹ *ibid*, 34

phase 2. In this phase, the training of an AI model not only extracts information from the training data but also encodes (reproduces) it into the model.³⁰ These reproductions are not for TDM but for memorising training data to generate outputs. Therefore, they are not excused. Moreover, the defendants are not research organisations, as they neither reinvest all profits in scientific research nor act in the public interest, as recognised by the state mission.³¹ The Court, however, did not address the issue of opt-out from TDM.

Conclusion

This is a significant decision. It clarifies that AI models can contain and reproduce training data. This can be an actionable claim under German and EU copyright law. The TDM exception does not apply to such an actionable claim. The fixation and reproduction of training data in AI models are not TDM exceptions. Furthermore, the commercial nature of AI companies plays a decisive role in this regard. The decision is not final. It is likely to be appealed or even referred to the Court of Justice of the European Union (CJEU). It will be interesting to see how things unfold in the future.

³⁰ *ibid*,40-42

³¹ *ibid*,44-45