



According to a first-instance ruling by the Munich I Regional Court, the memorization of texts in language models infringes on the reproduction rights of the rights holder.

In its ruling of November 11, 2025, the Munich I Regional Court essentially upheld a lawsuit brought by the collecting society GEMA against two companies of the Open AI group for cease and desist, information, and damages (Ref. 42 O 14139/24). According to the court, the limitation provision for text and data mining under Section 44b of the German Copyright Act (dUrhG) is just as inapplicable as that under Section 57 dUrhG with regard to insignificant ancillary works.

The ruling is not yet final.

Facts

GEMA argued that the lyrics of nine well-known German authors (including “Atemlos” by Kristina Bach and “Wie schön, dass du geboren bist” by Rolf Zuckowski) were **memorized** in the defendants’ language models and, when the chatbot was used, were reproduced largely in the original version as responses (outputs) to simple prompts from users.

The defendants disputed this and replied that their language models do not store or copy specific training data, but rather reflect in their parameters what they have learned based on the entire training data set. Since the outputs are only generated as a result of user prompts, it is not the defendants but the respective user who is responsible as the producer of the output. Furthermore, any infringements of rights are covered by the free use for text and data mining; in addition, the song lyrics are to be regarded as incidental and dispensable alongside the entire training data set.

Legal assessment

According to the Munich I Regional Court, the song lyrics are **reproducible** in the defendants’ language models 4 and 4o. Information technology research shows that training data can be contained in language models and extracted as outputs – so-called “**memorization**.” According to the ruling, memorization occurs when the language models not only extract information from the training data set during training, but also **completely adopt the training data**. A comparison between the song lyrics used for training purposes and the outputs shows such memorization. Given the complexity and length of the song lyrics, the reproduction of the song lyrics cannot be based on chance.

Memorization constitutes a copyright **reproduction** of the song lyrics by means of data. According to Art. 2 of the InfoSoc Directive, reproduction occurs “*in any manner and in any form*.” New technologies such as language models are covered by the reproduction right under Article 2 of the InfoSoc Directive and Section 16 of dUrhG. According to the case law of the CJEU indirect perceptibility is sufficient for reproduction; this is the case if the work can be perceived using technical aids.

According to the Munich I Regional Court, the text and data mining exception does not apply. Although language models are also covered by the scope of the **text and data mining restrictions**, these restrictions cover necessary reproductions when **compiling the data corpus** for training, such as the reproduction of a work by converting it into another (digital) format or storing it in the working memory. The background to this is the idea that these reproductions are only created for subsequent analysis purposes and therefore do not affect the author’s exploitation interests in the work. However, if training involves not only extracting information from training data but also reproducing works, this does not constitute text and data mining. The premise that the freedom of text and data mining does not affect exploitation interests does not apply in this constellation.



An **analogous** application of the text and data mining exception is out of the question. Even if one assumes an unintended regulatory gap with regard to the lack of regulation on memorization, there is no comparable interest, since memorization has a lasting adverse effect on the exploitation of the work. The risk of memorization is solely attributable to the defendant.

In the absence of a main work, the reproductions of the song lyrics also do not constitute **an impermissible accessory** under Section 57 dUrhG. The entire training data set is not a copyright-protected work, which would be a prerequisite for the song lyrics to qualify as an accessory.

Nor can it be assumed that the **rights holders have given their consent**, as the training of models cannot be considered a normal and expected type of use that the rights holder must anticipate.

This eagerly awaited decision—should it ultimately become legally binding—could have a lasting impact on the business model of AI companies. It remains to be seen how the appeal courts will rule—it can be assumed that the defendants will file an appeal.





Durch die Memorisierung von Texten in Sprachmodellen wird nach einem erstinstanzlichen Urteil des Landgericht München I in das Vervielfältigungsrecht des Rechteinhabers eingegriffen.

Das Landgericht München I hat mit Urteil vom 11.11.2025 einer Klage der Verwertungsgesellschaft GEMA gegen zwei Unternehmen der Unternehmensgruppe Open AI auf Unterlassungs-, Auskunfts- und Schadensersatz im Wesentlichen stattgegeben (Az. 42 O 14139/24). Die Schrankenbestimmung des Text und Data Mining gemäß § 44b dUrhG ist nach dem Gericht ebenso wenig anwendbar wie jene nach § 57 dUrhG in Bezug auf unwesentliche Beiwerke.

Das Urteil ist nicht rechtskräftig.

Sachverhalt

GEMA brachte vor, dass die Liedtexte von neun bekannten deutschen Urheberinnen und Urhebern (darunter „Atemlos“ von Kristina Bach oder „Wie schön, dass du geboren bist“ von Rolf Zuckowski) in den Sprachmodellen der Beklagten **memorisiert** seien und bei Nutzung des Chatbots auf einfache Prompts der Nutzer als Antworten (Outputs) in weiten Teilen originalgetreu wiedergegeben werden.

Die Beklagten bestritten und erwiderten, dass ihre Sprachmodelle keine spezifischen Trainingsdaten speichern oder kopieren, sondern in ihren Parametern reflektieren, was sie basierend auf dem gesamten Trainingsdatensatz erlernt haben. Da die Outputs nur als Folge von Prompts der Nutzer generiert werden, seien nicht die Beklagten, sondern der jeweilige Nutzer als Hersteller des Outputs verantwortlich. Im Übrigen seien eventuelle Rechtseingriffe von der Schranke für das Text- und Data-Mining gedeckt; die Liedtexte seien darüber hinaus neben dem gesamten Trainingsdatensatz als nebenschließlich und verzichtbar anzusehen.

Rechtliche Beurteilung

Nach dem Landgericht München I sind die Liedtexte in den Sprachmodellen 4 und 4o der Beklagten **reproduzierbar** enthalten. Informationstechnische Forschung zeigt, dass Trainingsdaten in Sprachmodellen enthalten sein können und sich als Outputs extrahieren lassen – sogenannte „**Memorisierung**“. Eine Memorisierung liegt nach der Entscheidung vor, wenn die Sprachmodelle beim Training dem Trainingsdatensatz nicht nur Informationen entnehmen, sondern die Sprachmodelle die **Trainingsdaten vollständig übernehmen**. Der Vergleich zwischen den zu Trainingszwecken verwendeten Liedtexten und den Outputs zeigt eine solche Memorisierung. Angesichts der Komplexität und Länge der Liedtexte kann die Wiedergabe der Liedtexte auch nicht auf Zufall beruhen.

Die Memorisierung stellt eine urheberrechtliche **Vervielfältigung** der Liedtexte durch Daten dar. Gemäß Art. 2 InfoSoc-RL liegt eine Vervielfältigung „auf jede Art und Weise und in jeder Form“ vor. Neue Technologien wie Sprachmodelle sind vom Vervielfältigungsrecht nach Art. 2 InfoSoc-RL und § 16 dUrhG erfasst. Nach der Rechtsprechung des EuGH ist für die Vervielfältigung eine mittelbare Wahrnehmbarkeit ausreichend; eine solche liegt vor, wenn das Werk unter Einsatz technischer Hilfsmittel wahrgenommen werden kann.

Die Text und Data Mining Ausnahme kommt nach dem Landgericht München I nicht anwendbar. Sprachmodelle unterfallen zwar grundsätzlich auch dem Anwendungsbereich der **Text und Data Mining Schranken**, diese Schranken decken aber erforderliche Vervielfältigungen beim **Zusammenstellen des Datenkorpus** für das Training, wie etwa die Vervielfältigung eines Werks durch seine Überführung in ein anderes (digitales) Format oder Speicherungen im Arbeitsspeicher. Hintergrund ist der Gedanke, dass diese Vervielfältigungen lediglich zu nachfolgenden Analysezwecken erstellt werden und damit die Verwertungsinteressen des Urhebers am Werk nicht



beeinträchtigen. Wenn aber beim Training nicht nur Informationen aus Trainingsdaten extrahiert, sondern Werke vervielfältigt werden, liegt **kein Text und Data Mining** vor. Die Prämisse, dass für die Freiheit des Text und Data Minings keine Verwertungsinteressen berührt werden, greift in dieser Konstellation nicht.

Eine **analoge** Anwendung der Text und Date Mining Schranke kommt nicht in Frage. Selbst wenn man hinsichtlich der fehlenden Regelung zum Memorisieren eine planwidrige Regelungslücke annimmt, mangelt es an einer vergleichbaren Interessenlage, da bei einer Memorisierung die Werkverwertung nachhaltig beeinträchtigt wird. Das Risiko der Memorisierung ist allein der Sphäre der Beklagten zuzurechnen.

Mangels Vorliegens eines Hauptwerks stellen die Vervielfältigungen der Liedtexte auch **kein unzulässiges Beiwerk** nach § 57 dUrhG dar. Der gesamten Trainigsdatensatz ist kein urheberrechtlich geschütztes Werk, was aber Voraussetzung für die Qualifizierung der Liedtexte als Beiwerk wäre.

Man kann auch von keiner **Einwilligung der Rechteinhaber** ausgehen, da das Training von Modellen nicht als eine übliche und erwartbare Nutzungsart zu werten ist, mit der der Rechteinhaber rechnen muss.

Diese mit Spannung erwartete Entscheidung – sollte sie letztlich in Rechtskraft erwachsen – könnte das Geschäftsmodell von AI Unternehmen nachhaltig beeinträchtigen. Zunächst bleibt abzuwarten wie die Instanzgerichte entscheiden werden – von der Erhebung einer Berufung durch die Beklagten kann wohl ausgegangen werden.