



## **Stellungnahme**

### **für das öffentliche Fachgespräch des Ausschusses für Kultur und Medien des Deutschen Bundestages zu „KI und Urheberrecht – Herausforderungen und Handlungsoptionen für den Schutz kreativer Inhalte“ am 24. Februar 2026**

#### **Zusammenfassung**

1. Das Training generativer KI erfordert die Nutzung sehr vieler urheberrechtlich geschützter Daten. Rechteinhaber:innen haben Sorge, dass ihre Werke genutzt werden, um KI-Anwendungen zu erstellen, die sie ersetzen werden. Eine Herausforderung für den Gesetzgeber besteht darin, dass die Phasen der Wertschöpfung von KI, bei denen die meisten urheberrechtlich relevanten Handlungen anfallen, nicht diejenigen sind, bei denen die größte Substitutionsgefahr besteht.
2. Der Gesetzgeber sollte sich die KI-Wertschöpfungskette zunutze machen, um zu gesellschaftlich wertvoller, offener Innovation anzureizen. Wo Substitutionseffekte drohen, sollte er durch regulatorische Eingriffe sicherstellen, dass Rechteinhaber:innen an dem Wert, der durch KI geschaffen wird, beteiligt werden.
3. Die Qualität der Trainingsdaten ist essenziell für die Qualität von KI-Modellen und KI-Anwendungen. Das gesellschaftliche Interesse, dass KI-Unternehmen möglichst ausgewogene, hochwertige Datensätze auch aus Europa verwenden, ist groß. Nur so lässt sich sicherstellen, dass KI-Modelle europäische Werte repräsentieren.
4. Das Urheberrecht kann und sollte Anreize dafür setzen, dass Datensätze frei zur Verfügung gestellt werden. (Forschungs-)Projekte, die Datensätze frei verfügbar machen, dürfen Nutzungen unabhängig von der Zustimmung der Rechteinhaber:innen vergütungsfrei vornehmen (§ 60d UrhG). Unternehmen, die Datensätze proprietär vertreiben, müssen Lizenzvereinbarungen mit Rechteinhaber:innen abschließen, wenn diese sich die Nutzung für das KI-Training vorbehalten haben (§ 44b UrhG).
5. Das gesellschaftliche Interesse an größtmöglicher Offenheit von Basismodellen ist ebenfalls sehr groß. Offenheit erhöht die Sicherheit der Modelle und fördert den Wettbewerb. Basismodelle sind zentrale Bausteine unterschiedlichster KI-Anwendungen. Sie ersetzen menschlich geschaffene Werke nicht. Auch hier kann schon das bestehende Recht mit §§ 44b, 60d UrhG Anreize setzen, damit sich Unternehmen für offene Entwicklung entscheiden.
6. KI-Anwendungen werden meist nicht offen zur Verfügung gestellt. Ihr Training erfordert die Verwendung kleiner, spezialisierter Datensätze. Das macht die Lizenzierung für Rechteinhaber:innen besonders attraktiv. Größere Rechteinhaber:innen können auf diesem Wege derzeit schon erhebliche Einnahmen generieren. Kleineren Rechteinhaber:innen gelingt das (noch) nicht.
7. Der Substitutionsdruck entsteht vor allem beim Einsatz generativer KI-Anwendungen. Dabei finden allerdings oft keine urheberrechtlich relevanten Handlungen statt. Der Gesetzgeber sollte KI-Unternehmen mit einer Abgabe belegen, deren Höhe sich nach den Umsätzen richtet, die die Unternehmen erwirtschaften. Die so gewonnenen Einnahmen sollten über Verwertungsgesellschaften an Rechteinhaber:innen ausgeschüttet werden. Durch die Abgabe würde sichergestellt, dass ein Teil des Gewinns, der mit generativer KI erwirtschaftet wird, an Künstler:innen zurückfließt.

## 1. Einleitung

Seit OpenAI Ende November 2022 ChatGPT auf den Markt gebracht hat, erscheinen nahezu täglich neue Anwendungen, die Text-, Bild-, Audio- und/oder Videodateien generieren können (generative KI). Die Kreativbranche ist deswegen stark verunsichert – in Europa ebenso wie in anderen Ländern, etwa in den USA. Befürchtet wird, dass kreative Leistungen zunehmend durch KI-Output ersetzt werden. Zudem wird beklagt, dass beim Training generativer KI riesige Mengen an Daten benötigt werden, von denen viele urheberrechtlich geschützt sind. Die Sorge lautet: Unsere Werke werden genutzt, um Anwendungen zu erstellen, die uns letztlich ersetzen werden.

Der Gesetzgeber kann das Urheberrecht als einen Hebel einsetzen, um zu KI-Innovation anzureizen, die mit europäischen Werten vereinbar ist und die Interessen von Rechteinhaber:innen achtet. Überdies kann er mittels des Urheberrechts die bestehende Marktmacht von KI-Unternehmen brechen den Wettbewerb sowie europäische Unternehmen stärken.<sup>1</sup> Dafür muss er sich die KI-Wertschöpfungskette zunutze machen und unterschiedliche Regeln für unterschiedliche Entwicklungsphasen bereithalten. Das geltende Recht tut dies teilweise, aber noch nicht durchgehend.

Regulatorisch herausfordernd ist vor allem, dass urheberrechtlich relevante Handlungen vor allem während der Trainingsphase anfallen. Diese Phase gefährdet das Geschäftsmodell von Rechteinhaber:innen nicht unmittelbar, Substitutionseffekte können vor allem durch den Betrieb von KI entstehen. Dabei finden aber oftmals keine urheberrechtlich relevanten Handlungen statt.

## 2. Technischer Hintergrund

Durch das Training lernen KI-Modelle, Muster und Strukturen zu erkennen. Textmodelle etwa müssen abbilden können, wie Wörter und Sätze zusammenhängen und wie Wörter in einem bestimmten Kontext interpretiert werden. Die Trainingsdaten speichert das Modell nicht in Form von Text- oder Bilddaten. Es leitet aus den Inhalten Parameter ab, die als Zahlenwerte gespeichert werden. Auf Basis dieser Zahlenwerte kann ein Textmodell beispielsweise berechnen, dass auf die Wortfolge „Der Baum ist“ häufiger das Wort „grün“ als das Wort „lila“ folgt, und dass das Wort „Bank“ in einem Kontext ein Sitzmöbel bezeichnet und in einem anderen ein Finanzinstitut. Das Training versetzt das Modell in die Lage, Wahrscheinlichkeiten zu berechnen. Diese Berechnungen führt es durch, wenn eine KI-Anwendung eine Anfrage von einer Nutzerin erhält, einen sogenannten Prompt. Auf Grundlage dieser Wahrscheinlichkeitsrechnungen generiert es ein Ergebnis (Output).

Die Qualität der Trainingsdaten ist essenziell für die Qualität des Outputs: Ist das Datenmaterial von schlechter Qualität, etwa weil es Fehler enthält oder einseitig ist, so ist auch der Output meist fehlerhaft und/oder biased. Es besteht daher ein großes gesellschaftliches Interesse daran, dass KI-Unternehmen möglichst ausgewogene, qualitativ hochwertige Datensätze verwenden.

## 3. Urheberrechtlicher Hintergrund

Werke und sonstige Schutzgegenstände sind zwar nur implizit im KI-Modell enthalten. Während des Trainingsprozesses müssen die Daten aber wenigstens kurzfristig im Arbeitsspeicher vervielfältigt werden. Zumeist werden sie sogar längerfristig (außerhalb des Modells) abgespeichert und für das Training kuratiert und verändert. Beim Trainingsprozess fallen also urheberrechtlich relevante Handlungen an. Das Recht, ein Werk zu vervielfältigen und zu bearbeiten, steht prinzipiell der Rechteinhaberin zu. Nutzer:innen dürfen ein Werk nur nutzen, wenn die Rechteinhaberin ihr dies erlaubt hat oder wenn die Nutzung gesetzlich erlaubt ist.

Das deutsche Recht verfügt über zwei gesetzliche Erlaubnistatbestände, die Nutzungen zum Zwecke des Text und Data Mining erlauben (§§ 44b, 60d UrhG). Text und Data Mining ist die automatisierte Analyse von digitalen Daten, um Informationen zu gewinnen, etwa um Sprachmuster zu

---

<sup>1</sup> Daneben erfordert KI-Innovation Investitionsbereitschaft und Unternehmertum, qualifizierte Arbeitskräfte und große Rechenkapazitäten.

erkennen. Beide Erlaubnistatbestände basieren auf unionsrechtlichen Vorgaben (Art. 3 und 4 DSM-Richtlinie). Mit der KI-VO hat der Unionsgesetzgeber verdeutlicht, dass diese Vorschriften seines Erachtens unter bestimmten Voraussetzungen auf das Training generativer KI-Anwendungen anwendbar sind. Im weltweiten Vergleich bietet die EU daher relativ viel Rechtssicherheit. Im Wettbewerb der Rechtsordnungen ist dies von Vorteil sowohl für KI-Unternehmen als auch für Rechteinhaber:innen.

§ 60d UrhG erlaubt es zum Zweck des Text und Data Mining, Schutzgegenstände für Forschungszwecke zu vervielfältigen und Datenbanken zu entnehmen. Auch begleitende Handlungen, etwa Veränderungen der Werke, die zur Aufbereitung der Daten erforderlich sind, sind vom Anwendungsbereich der Norm erfasst. Nutzer:innen müssen rechtmäßigen Zugang zu den Schutzgegenständen haben. Die Nutzung ist auch dann gestattet, wenn die Rechteinhaberin kommuniziert, dass sie dies nicht wünscht.

§ 44b UrhG gestattet Nutzungen auch für kommerzielles Text und Data Mining. Eine Nutzung ist nur erlaubt, wenn sich die Rechteinhaberin die Nutzung nicht ausdrücklich in angemessener Weise vorbehalten haben. Bei Werken, die frei im Internet verfügbar sind, muss die Rechteinhaberin den Vorbehalt in maschinenlesbarer Form kommunizieren. Der Nutzungsvorbehalt ist das Instrument, um die Entstehung eines Lizenzmarktes zu fördern: Hat eine Rechteinhaberin kommuniziert, dass sie die Nutzung für Trainingszwecke nicht wünscht, darf die Nutzung nicht auf gesetzlicher Basis erfolgen, sondern nur, wenn die Nutzerin, also das KI-Unternehmen, einen Lizenzvertrag mit der Rechteinhaberin abgeschlossen hat. Für gewisse Stufen der KI-Wertschöpfungskette (dazu unten) bildet sich derzeit ein Lizenzmarkt heraus. Vertragspartner:innen sind in aller Regel große KI-Unternehmen und große Verwerterinnen. Kleinere Rechteinhaber:innen profitieren hingegen zumeist nicht. Dadurch werden bestehende Machtverhältnisse in der Tech- und Kreativbranche weiter zementiert.

Wichtig zu wissen ist: Beim Betrieb von KI-Anwendungen fallen nicht unbedingt urheberrechtlich relevante Handlungen an. Zudem treten die Substitutionseffekte, die Rechteinhaber:innen befürchten, nicht in erster Linie dadurch ein, dass Nutzer:innen geschützte Werke mittels KI reproduzieren. Gefährlich für die Kreativen sind vor allem die vielen Outputs, die Werken stilistisch ähneln und diese daher ersetzen können. Stilistische Nachahmungen sind aber keine Urheberrechtsverletzungen.

#### **4. KI-Wertschöpfungskette**

Die KI-Wertschöpfungskette ist komplex. Sie besteht aus mehreren Phasen, in denen unterschiedliche Akteur:innen tätig werden. Dieser Tatsache sollte das Recht Rechnung tragen. Mehr noch: Der Gesetzgeber sollte sich die Wertschöpfungskette zunutze machen, um zu größtmöglicher, qualitativ hochwertiger, offener KI-Innovation anzureizen und zugleich dort, wo Substitutionseffekte drohen, durch regulatorische Eingriffe sicherzustellen, dass Rechteinhaber:innen an dem Wert, der durch KI geschaffen wird, beteiligt werden.

##### **a) Erstellung von (offenen) Trainingsdatensätzen**

Datensätze bilden das Fundament der KI-Entwicklung, die Qualität einer KI-Anwendung wird insbesondere durch das Datenmaterial bestimmt, das beim Training zum Einsatz kommt. Die großen US-amerikanischen Unternehmen trainieren ihre Modelle insbesondere mit englisch-sprachigen Daten; damit sind auch die Werte, die in diesen Daten zum Ausdruck kommen, besonders stark in den Modellen repräsentiert. Wir Europäerinnen sollten ein Interesse daran haben, dass die Modelle, die auch hier in Europa zum Einsatz kommen, auch unsere Werte repräsentieren.

Trainingsdatensätze werden nicht selten von Forschungsgruppen erstellt und dann jedermann für alle Zwecke frei zur Verfügung gestellt (vgl. nur die Datensätze des deutschen LAION e.V.). Diese Datensätze dürfen große Unternehmen nutzen, besonders wertvoll sind sie aber für kleinere

Unternehmen sowie für gemeinnützige Projekte, die nicht über die erforderlichen finanziellen Mittel verfügen, um große Datensätze zu lizenzieren. Sowohl für den Wettbewerbs- als auch für den Forschungsstandort Deutschland ist es mithin gut, wenn Datensätze möglichst frei zur Verfügung stehen.

Das Urheberrecht kann und sollte diese freie Verfügbarkeit fördern. § 60d UrhG bietet dazu bereits jetzt eine gute Grundlage. Die Vorschrift erlaubt die Vervielfältigung unabhängig von der Zustimmung der Rechteinhaberin, soweit die Nutzung zu Forschungszwecken geschieht. Die ersten Urteile des LG und des OLG Hamburg belegen: Wer Datensätze offen und frei zur Verfügung stellt, kann sich auf § 60d UrhG berufen.

Die Ersteller von Datensätzen, welche nicht frei zur Verfügung gestellt werden, profitieren hingegen zurecht nicht von § 60d UrhG, sondern müssen den Voraussetzungen des § 44b UrhG Rechnung tragen. Sie handeln in kommerzieller Absicht, ihr Handeln ist nicht auf Erkenntnisgewinn für die Allgemeinheit gerichtet, sondern darauf, einen Vorteil im Vergleich zu Wettbewerbern zu erlangen. Dass sie Inhalte, für die Nutzungsvorbehalte erklärt wurden, nur nutzen dürfen, wenn sie eine Lizenz erworben (und Rechteinhaber:innen vergütet) haben, ist daher interessengerecht. KI-Unternehmen müssen sich also entscheiden: Wollen sie einen Wettbewerbsvorteil für sich, so müssen sie zahlen. Wollen sie einen Beitrag erbringen, der allen Marktteilnehmer:innen sowie der Allgemeinheit zugutekommt, so dürfen sie kostenlos nutzen.

## **b) Training von (offenen) Basismodellen**

An die Erstellung von Trainingsdatensätzen schließt sich das Training von Basismodellen an. Basismodelle wie GPT oder Llama werden vor allem in den USA und in China entwickelt. Sie erfordern ganz erheblichen finanziellen Einsatz und gewaltige Rechenleistung; diese Modelle sind zentrale Bausteine verschiedenster KI-Anwendungen. Um ein Modell derartig breit zu machen, dass es in unterschiedliche Richtungen weiterentwickelt werden kann, muss es mit sehr vielen, unterschiedlichen Daten trainiert werden.

Gesellschaftlich wünschenswert ist, dass auch möglichst viele Basismodelle offen sind. Dies ermöglicht es der Allgemeinheit, die Modelle sowie die damit verbundenen Risiken zu überprüfen – Offenheit erhöht die Sicherheit von KI. Vor allem aber können kleinere Unternehmen sowie nicht kommerzielle Projekte lizenzfrei auf offenen Basismodellen aufbauen und unterschiedlichste Anwendungen konzipieren. Basismodelle ersetzen nicht unmittelbar menschlich geschaffene Werke. Auch viele KI-Anwendungen, die auf Basismodellen aufbauen, stellen keine Konkurrenz für Rechteinhaber:innen dar; man denke nur an diagnostische Anwendungen im Medizinbereich. Weil die Entwicklung konkreter KI-Anwendungen weniger teuer und aufwendig ist als jene von Basismodellen, haben auch europäische Unternehmen eine realistische Chance, in Wettbewerb zu Unternehmen aus den USA und aus China zu treten.

Auch hier ist also das gesellschaftliche und volkswirtschaftliche Interesse groß, KI-Unternehmen zur Offenlegung ihrer Modelle anzureizen. Das Urheberrecht kann dazu beitragen, indem es auch das Training offener Basismodelle nach § 60d UrhG lizenzfrei ermöglicht. Damit steht ein KI-Unternehmen auch auf dieser Entwicklungsstufe vor der Frage, ob, es lieber ein Forschungsvorhaben finanzieren möchte, das das Basismodell entwickelt und frei zur Verfügung stellt, oder ob es lieber ein proprietäres Modell entwickeln und die dafür erforderlichen Daten wenigstens zum Teil lizenzieren möchte – mit der Gefahr, dass die Qualität des Datensatzes geringer ist.

## **c) Training von KI-Anwendungen**

KI-Unternehmen nutzen Basismodelle, um KI-Anwendungen zu entwickeln. Diese Anwendungen werden meist proprietär, also nicht offen vertrieben. Sie müssen ebenfalls trainiert werden. Das Training erfolgt unter Verwendung spezialisierter Datensätze. Je kleiner der Datensatz, desto wertvoller ist das einzelne Datum, und desto schwerer lässt es sich durch ein anderes Datum ersetzen.

Das bedeutet: Für Rechteinhaber:innen ist die Lizenzierung in dieser Phase der Wertschöpfung potenziell lukrativer als in den vorherigen Phasen. Gerade für das Training von KI-Anwendungen hat sich bereits ein Lizenzmarkt entwickelt.

Auch die Größe dieses Lizenzmarkts kann durch offene Basismodelle gefördert werden. Je mehr Unternehmen auf den offenen Fundamenten aufbauen können, desto mehr potenzielle Lizenznehmer:innen haben Rechteinhaber:innen. Der stärkere Wettbewerb, den offene Basismodelle bei KI-Anwendungen zur Folge haben, kommt also nicht nur der deutschen Volkswirtschaft, sondern auch Rechteinhaber:innen zugute.

#### **d) Einsatz von KI-Anwendungen**

Den Substitutionsdruck spüren Rechteinhaber:innen (erst bzw. insbesondere) dann, wenn generative KI-Anwendungen eingesetzt werden. Aus urheberrechtlicher Sicht ist dabei zu beachten: Solange der Output einer Anwendung keinen geschützten Inhalt vervielfältigt oder bearbeitet, findet keine urheberrechtlich relevante Handlung statt. Paradoxe Weise ist der Schritt, der urheberrechtlich am unproblematischsten ist, derjenige, der wirtschaftlich die größten Gefahren birgt.

Diese wirtschaftliche Schieflage korrigiert das geltende Recht nur unzureichend. Lizenzen für Vervielfältigungen beim Training können zwar einen gewissen Ausgleich schaffen. Der Staat sollte den Kultursektor aber nicht allein auf Lizenzoptionen verweisen, zumal nur die Inhalte großer Rechteinhaber:innen für KI-Unternehmen interessant sind, kleinere Rechteinhaber:innen aber allenfalls über Verwertungsgesellschaften überhaupt am Lizenzmarkt partizipieren können. Die Sorgen von Werkschaffenden sollte der Gesetzgeber ernst nehmen.

Der Gesetzgeber kann hier Abhilfe schaffen und sicherstellen, dass menschliche Kreativität allgemein finanziell unterstützt wird, indem er KI-Unternehmen mit einer Abgabe belegt, deren Höhe sich nach den Einnahmen richten sollte, die sie innerhalb des Territoriums mit ihren Diensten erwirtschaften. Die so generierten Einnahmen sollten über Verwertungsgesellschaften an Rechteinhaber:innen ausgeschüttet werden. Dabei sollte berücksichtigt werden, ob Rechteinhaber:innen bereits auf Lizenzbasis entlohnt wurden. Durch die Abgabe würde sichergestellt, dass ein Teil des Gewinns, der mit generativer KI erwirtschaftet wird, an Künstler:innen zurückfließt.

Berlin, den 13. Februar 2026

Weitere Informationen:

- de la Durantaye, Garbage in, garbage out – Die Regulierung generativer KI durch das Urheberrecht, ZUM 2023, 645-660
- de la Durantaye, Akkommodation statt Assimilation. Warum die EU bei der KI-Regulierung nicht auf den Brussels Effect setzen sollte – und was stattdessen sinnvoll wäre, ZUM 2025, 165-174
- de la Durantaye, American Beauty: KI-Training und die fair use doctrine, GRUR 2025, 1550-1557
- de la Durantaye, Control and Compensation. A Comparative Analysis of Copyright Exceptions for Training Generative AI, IIC 2025, 737-770 DOI: 10.1007/s40319-025-01569-6