Florent Thouvenin | Lena A. Jäger | Joël Donzé | Viviane Ammann | Lennart Deutschmann | Lena Hänni | Maimuna Jobarteh | Aleksandra Marchewka

Training von KI-Modellen und Urheberrecht: Technische Grundlagen, Stand der Diskussion und Lösungsansätze

Werke der Literatur und Kunst werden im grossen Stil für das Training von Modellen der Künstlichen Intelligenz (KI) verwendet. Ob und unter welchen Voraussetzungen diese Werknutzungen urheberrechtlich zulässig sind, wird im In- und Ausland intensiv diskutiert. Dieser Beitrag vermittelt die technischen Grundlagen, deren Verständnis für die rechtliche Beurteilung erforderlich ist, skizziert die Rechtsentwicklung in ausgewählten Rechtsordnungen und stellt den Stand der Diskussion in der Schweiz dar. Auf dieser Grundlage werden mögliche Lösungsansätze für das Schweizer Recht aufgezeigt.

Les œuvres littéraires et artistiques sont largement utilisées pour l'entraînement des modèles d'intelligence artificielle (IA). La question de savoir si et dans quelles conditions cette utilisation des œuvres est autorisée au regard du droit d'auteur fait l'objet d'un débat intense en Suisse et à l'étranger. La présente contribution introduit les bases techniques dont la compréhension est nécessaire pour l'évaluation juridique, décrit l'évolution du droit dans certains ordres juridiques et présente l'état actuel du débat en Suisse. Sur cette base, elle propose des solutions possibles pour le droit suisse.

Florent Thouvenin, Prof. Dr. iur., Rechtsanwalt, Ordinarius für Informations- und Kommunikationsrecht und Vorsitzender des Leitungsausschusses des Center für Information Technology, Society, and Law (ITSL) an der Universität Zürich.

Lena A. Jäger, Prof. Dr., ausserordentliche Professorin für Digitale Linguistik Institut für Computerlinguistik und Institut für Informatik, Leiterin des Instituts für Computerlinguistik und Co-Direktorin des Kompetenzzentrums Language & Medicine, Universität Zürich.

Joël Donzé, MLaw, LL.M., Wissenschaftlicher Assistent am Lehrstuhl für Informations- und Kommunikationsrecht, Universität Zürich.

Viviane Ammann, MLaw, LL.M., Wissenschaftliche Assistentin am Lehrstuhl für Informations- und Kommunikationsrecht, Universität Zürich.

Lennart Deutschmann, Wissenschaftlicher Assistent am Lehrstuhl für Informations- und Kommunikationsrecht, Universität Zürich.

Lena Hänni, MLaw, Wissenschaftliche Assistentin am Lehrstuhl für Informations- und Kommunikationsrecht, Universität Zürich.

Maimuna Jobarteh, MLaw, Wissenschaftliche Assistentin ITSL, Universität Zürich.

Aleksandra Marchewka, MLaw, Wissenschaftliche Assistentin ITSL, Universität Zürich.

I. Einleitung

Die Entwicklung und Verwendung von Modellen und Systemen der Künstlichen Intelligenz (KI) werfen eine Vielzahl von Rechtsfragen auf.[1] Diese Fragen werden seit einigen Jahren intensiv diskutiert – nicht nur, aber auch im Urheberrecht. Die Diskussion konzentrierte sich hier zunächst auf die Frage, *ob von KI generierte Werke urheberrechtlich geschützt* sind. Das ist mittlerweile im Grundsatz geklärt: Autonom von KI-Systemen generierte Werke sind keine geistigen Schöpfungen und damit nicht urheberrechtlich geschützt.[2] Offen bleibt allerdings, welches Mass an menschlicher Beteiligung erforderlich ist, damit KI-Systeme als blosse Werkzeuge angesehen werden und mit ihrer Hilfe geschaffene Werke als geistige Schöpfungen gelten sowie urheberrechtlich geschützt sein können.

Seit einiger Zeit fokussiert die Diskussion auf die Frage, wie die *Nutzung von urheberrechtlich geschützten Werken beim Training von KI-Modellen* urheberrechtlich zu qualifizieren ist. Diese Frage wird in vielen Rechtsordnungen untersucht, und sie ist Gegenstand zahlreicher Gerichtsverfahren, insb. in den USA.[3] Anders als der urheberrechtliche Schutz des Outputs von KI-Systemen kann die Nutzung von Werken beim Training von KI-Modellen nur auf der Grundlage eines hinreichenden technischen Verständnisses beurteilt werden. Die Vorgänge beim Training sind komplex, und die Komplexität nimmt mit der technischen Entwicklung weiter zu. Zudem bestehen in technischer Hinsicht Unterschiede zwischen den Werkarten und den für deren Generierung verwendeten Modellen. Die rechtswissenschaftliche Literatur bemüht sich zwar um ein hinreichendes technisches Verständnis, fokussiert aber bisher auf Sprachmodelle und bleibt notwendigerweise hinter der technischen Entwicklung zurück. Das ist oft unproblematisch, weil gewisse technische Unterschiede und Entwicklungen die rechtliche Beurteilung unberührt lassen. Bisweilen führen die technischen Entwicklungen und das verbesserte Verständnis der Vorgänge aber dazu, dass die rechtliche Beurteilung überdacht werden muss.

Von der Frage des Trainings von KI-Modellen ist diejenige der *Verwendung der trainierten Modelle in KI-Systemen* zu unterscheiden. Urheberrechtliche Fragen stellen sich dabei in aller Regel nur bei generativer KI, also bei KI-Systemen, die Texte, Bilder, Musik, Videos, Source Code und andere Inhalte generieren. Im Gegensatz dazu erzeugen diskriminative KI-Systeme keine neuen Inhalte, sondern treffen Vorhersagen oder nehmen Klassifikationen vor. Solche Systeme werden bspw. für das Erkennen von Spam-Mails, die Klassifikation von Objekten in der Medizin (gutartiger oder bösartiger Tumor?), die Bilderkennung (Hund oder Katze?), die Gesichtserkennung, das Kreditscoring von Webshops (Zahlung nur mit Kreditkarte oder auch auf Rechnung?) oder für Vorhersagen (Ist die Bewerberin für die Stelle geeignet?) verwendet. Der Einsatz von generativen KI-Systemen kann Urheberrechte verletzen, wenn der Output ein vorbestehendes Werk (oder Teile davon) enthält oder einem vorbestehenden Werk (oder Teilen davon) so ähnlich ist, dass er in dessen Schutzbereich fällt. Diese Frage ist zwar mit derjenigen des Trainings verknüpft, weil selbst bei einem Eingreifen in den Schutzbereich nur eine Urheberrechtsverletzung vorliegt, wenn ein bestehendes Werk bei der Erzeugung eines neuen Werks tatsächlich genutzt worden ist.[4] Eine sinnvolle urheberrechtliche Beurteilung ist aber nur möglich, wenn die Frage der Verletzung von Urheberrechten bei der Verwendung von generativen KI-Systemen von derjenigen der Nutzung von Werken beim Training der in diesen Systemen verwendeten KI-Modelle unterschieden wird. Dieser Beitrag untersucht nur die letztere Frage.

Bei der Beurteilung der Nutzung von urheberrechtlich geschützten Werken beim Training von KI-Modellen muss nicht zwischen generativer und diskriminativer KI unterschieden werden. In beiden Fällen stellen sich dieselben Rechtsfragen, und diese können auch gleich beurteilt werden, sofern die Verletzung von Urheberrechten bei der Verwendung von KI-Systemen als eigenständiger Vorgang verstanden und beurteilt wird. Aus Sicht der Rechteinhaber sind generative KI-Systeme allerdings ungleich problematischer als diskriminative Systeme, weil deren Output von Menschen geschaffene Werke substituieren und das Funktionieren etablierter Geschäftsmodelle in Frage stellen kann.

Vor diesem Hintergrund verfolgt der vorliegende Beitrag drei Ziele: Zum einen werden die technischen Grundlagen dargestellt, soweit sie für die urheberrechtliche Beurteilung des Trainings von KI-Modellen erforderlich sind. Der Fokus liegt dabei auf den urheberrechtlich besonders herausfordernden Modellen der generativen KI, die Erkenntnisse lassen sich aber auf diskriminative KI übertragen (II.). Zum anderen werden die Rechtslage in ausgewählten ausländischen Rechtsordnungen skizziert (III.) und der Stand der Diskussion in der Schweiz dargestellt (IV.). Auf dieser Grundlage

1 In diesem Beitrag werden die Begriffe «KI-System» und «KI-Modell» im Sinn der KI-Verordnung der EU (Verordnung EU 2024/1689 vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz) verwendet. Nach Art. 3 Nr. 1 KI-Verordnung ist ein KI-System «ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können». Ein KI-System ist damit eine konkrete Anwendung, die von Nutzern eingesetzt werden kann (bspw. ein Chatbot) und das ein (oder mehrere) KI-Modell(e) nutzt. Der Begriff des KI-Modells wird in der KI-Verordnung nicht definiert, sondern vorausgesetzt. Gemeint ist damit ein mathematisches Modell, das typischerweise durch maschinelles Lernen mit Daten trainiert worden ist und das Vorhersagen trifft oder Inhalte generiert. KI-Modelle sind die grundlegenden Bausteine von KI-Systemen.

2 Siehe dazu statt vieler: F. Thouvenin/P. Picht, AI & IP: Empfehlungen für Rechtsetzung, Rechtsanwendung und Forschung zu den Herausforderungen an den Schnittstellen von Artificial Intelligence (AI) und Intellectual Property (IP), sic! 2023, 510 f.; C. Semmelmann, Künstliche Intelligenz und Urheberrecht: Stand zu Training und Output 2024, sic! 2024, 637 ff.; I. Cherpillod, Intelligence artificielle et droit d'auteur, sic! 2023, 445 ff.

3 Siehe dazu hinten III.6.b).

4 F. Thouvenin, Retrieval-Augmented Generation (RAG) – ein urheberrechtliches 101, sic! 2025, 387.

werden in einem dritten Schritt mögliche Lösungsansätze für das Schweizer Recht aufgezeigt (V.).

II. Technische Grundlagen

1. Maschinelles Lernen

Die generativen KI-Modelle der jüngsten Generation basieren auf sog. *Deep Learning*. *Deep Learning* ist ein Teilbereich des maschinellen Lernens, bei dem komplexe künstliche neuronale Netze verwendet werden.[5] Das maschinelle Lernen selbst ist ein Teilgebiet von KI.[6]

a) Künstliche Neuronale Netze (KNN)

Aktuelle KI-Systeme verwenden häufig eine neuronale Netzwerkarchitektur als zentralen Baustein.[7] Der Begriff «neuronales Netzwerk» stammt aus den Neurowissenschaften und verweist auf die Verbindungen zwischen Neuronen im menschlichen Nervensystem, was historisch die Entwicklung von künstlichen neuronalen Netzen inspiriert hat.[8]

Die Neuronen eines künstlichen neuronalen Netzwerks sind in aufeinanderfolgenden Ebenen organisiert.[9] In jeder Ebene befinden sich mehrere Neuronen, die jeweils mit einigen oder allen Neuronen der vorhergehenden Ebene verbunden sind.[10] Neuronen sind einfache Recheneinheiten, die Informationen verarbeiten und anschliessend weitergeben.[11] Die Verbindungen zwischen den Neuronen werden durch Gewichte dargestellt. Diese Gewichte sind numerische Werte, welche die Richtung und Stärke der Signalweitergabe bestimmen.[12] Zu Beginn des Trainingsprozesses werden die Gewichte zufällig initialisiert und anschliessend durch Optimierungsverfahren schrittweise angepasst.[13] Die Gewichte sind die trainierbaren Parameter des Modells.[14] Die Architektur des Netzwerks – also etwa die Anzahl der Ebenen oder die Art der Verbindungen – wird vor dem Trainingsprozess festgelegt und verändert sich während diesem in der Regel nicht.[15]

Die erste Ebene, die Eingabeebene, verarbeitet die Rohdaten.[16] Jede weitere Ebene erhält in der Regel die Ausgaben der jeweils vorhergehenden Ebene.[17] Ein Modell kann zahlreiche sog. verborgene Ebenen *(Hidden Layers)* haben.[18] In der letzten Ebene, der Ausgabeebene, erzeugt das System die finalen Resultate und gibt sie aus.[19] Bei einer grossen Anzahl von Ebenen spricht man von einem tiefen neuronalen Netzwerk *(Deep Neural Network)*[20] oder auch von *Deep Learning*.[21]

b) Trainingslogik und Lernparadigmen

Ziel des maschinellen Lernens ist es, ein Modell so zu trainieren, dass es eine bestimmte Aufgabe möglichst gut löst.[22] Das «Modell» implementiert dabei eine mathematische Funktion, die anhand einer Eingabe eine Ausgabe erzeugt.[23] Das Training eines KI-Modells kann deshalb als Optimierungsproblem verstanden werden.[24] Man versucht die Parameter (die Verbindungsgewichte) des Modells so anzupassen, dass der Trainingsfehler minimiert wird.[25] Dieser wird durch die Fehlerfunktion (häufig auch als Kostenfunktion, Verlustfunktion oder «*loss function*» bezeichnet) beschrieben, welche die Differenz zwischen der gegebenen und der gewünschten Ausgabe darstellt.[26] Das Minimieren des Fehlers erfolgt in der Regel durch iterative Optimierungsverfahren, allen voran den sog. *Gradientenabstieg*.[27] Dabei wird die Fehlerfunktion ausgewertet, und es werden deren Ableitungen mit Bezug auf die Modellparameter berechnet.[28] Diese Ab-

5 M. Kuipers/R. Prasad, Journey of Artificial Intelligence, Wireless Pers Commun 2022, 384; R. T. Kreutzer, Künstliche Intelligenz verstehen, Grundlagen – Use-Cases – unternehmenseigene KI-Journey, 2. Aufl., Wiesbaden 2023, 10, 13.

6 T. W. Dornis/S. Stober, Urheberrecht und Training generativer KI-Modelle, Technologische und juristische Grundlagen, Baden-Baden 2024, 23; A. Man-Cho So, Technical Elements of Machine Learning for Intellectual Property Law, in: Jyh-An Lee/Hilty Reto/Liu Kung-Chung (Hg.), Artificial Intelligence and Intellectual Property, Oxford 2021, 11.

7 S. Yanisky-Ravid, Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era – The Human-Like Authors Are Already Here – A New Model, Mich. St. L. Rev. 2017, 675.

8 F. Rosenblatt, The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Psychological Review 1958, 386 ff.; S. Papastefanou, Smart Grids and Machine Learning in Chinese and Western Intellectual Property Law IIC 2021, 994; R. H. Chen/C. Chen, Artificial Intelligence: An Introduction to the Big Ideas and their Development, 2. Aufl., 2024, 75 ff.; Kreutzer (Fn. 5), 10.

9 Chen/Chen (Fn. 8), 80 ff.

10 Man-Cho So (Fn. 6), 15.

11 Dornis/Stober (Fn. 6), 28.

12 S. Shalev-Shwartz/S. Ben-David, Understanding Machine Learning, From Theory to Algorithms, New York 2014, 269 ff.

13 J. Drexl/R. Hilty/F. Beneke/L. Desaunettes-Barbero/M. Finck/J. Globocnik/B. Gonzalez Otero/J. Hoffmann/L. Hollander/D. Kim/H. Richter/S. Scheuerer/P. R. Slowinski/J. Thonemann, Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective, Max Planck Institute for Innovation and Competition Research Paper Series 13/2019, 1–14, 6. Zu den Optimierungsverfahren siehe hinten, II.1.b) ff.

14 D. Cohn/L. Atlas/R. Ladner, Improving Generalization with Active Learning, Machine Learning Nr. 21994, 207; Drexl / Hilty / Beneke / Desaunettes-Barbero / Finck / Globocnik / Gonzalez Otero / Hoffmann / Hollander / Kim / Richter / Scheuerer / Slowinski / Thonemann (Fn. 13), 6.

15 Cohn/Atlas/Ladner (Fn. 14), 207; Drexl/Hilty/Beneke/Desaunettes-Barbero/Finck/Globocnik/Gonzalez Otero/Hoffmann/Hollander/Kim/Richter/Scheuerer/Slowinski/Thonemann (Fn. 13), 6. Es gibt aber dynamische Netzwerkarchitekturen: Siehe Y. Han/G. Huang/S. Song/L. Yang/H. Wang/Y. Wang, Dynamic Neural Networks: A Survey, IEEE Transactions on Pattern Analysis & Machine Learning Intelligence 2022, 7436 ff.; siehe auch M Looks/M. Herreshoff/D. Hutchins/P. Norvig, Deep Learning with Dynamic Computation Graphs, ICLR Conference Paper 2017, 1 ff.

16 Kreutzer (Fn. 5), 11.

17 Kreutzer (Fn. 5), 11.

18 Kreutzer (Fn. 5), 11.

19 Kreutzer (Fn. 5), 11.

20 Drexl/Hilty/Beneke/Desaunettes-Barbero/Finck/Globocnik/Gonzalez Otero/Hoffmann/Hollander/Kim/Richter/Scheuerer/Slowinski/Thonemann (Fn. 13), 6; Kreutzer (Fn. 5), 13.

21 Kuipers/Prasad (Fn. 5), 3284 ff; Kreutzer (Fn. 5), 13.

22 Dornis/Stober (Fn. 6), 23.

23 Dornis/Stober (Fn. 6), 23 f.

24 I. Goodfellow/Y. Bengio/A. Courville, Deep Learning, MIT Press 2016, ‹www.deeplearningbook.org/› (abgerufen am 5. Januar 2026), 271 ff.

25 Man-Cho So (Fn. 6), 14.

26 Man-Cho So (Fn. 6), 14.

27 Goodfellow/Bengio/Courville (Fn. 24), 296 f.

28 Dornis/Stober (Fn. 6), 33 f.

leitungen (der Gradient) geben die Richtung des steilsten Anstiegs der Fehlerfunktion an.[29] Durch die Aktualisierung der Gewichte in entgegengesetzter Richtung bewegt sich das Modell schrittweise hin zu einer Konfiguration mit geringerem Fehler.[30] Die konkrete Ausgestaltung der Fehlerfunktion und der Trainingsprozess hängen jedoch stark davon ab, in welchem Lernparadigma (überwachtes, unüberwachtes oder bestärkendes Lernen; siehe dazu sogleich) das Modell trainiert und für welche Aufgabe es eingesetzt wird, also welche Problemstellung adressiert wird und welche Zielgrössen dem Modell während des Trainings vorgegeben werden.[31] Ziel aller Trainingsmethoden ist, dass das Modell nach dem Training möglichst gut generalisieren kann.[32]

Grundsätzlich lässt sich das Training von KI-Modellen in drei Hauptparadigmen einordnen: überwacht *(Supervised Learning)*, unüberwacht *(Unsupervised Learning)* und bestärkend *(Reinforcement Learning)*:

Überwachtes Lernen *(Supervised Learning)* basiert auf gelabelten Daten, d.h. jedes Trainingsbeispiel enthält die Eingabemerkmale und die gewünschte Ausgabe (Zielwerte).[33] Das Modell soll dabei eine Funktion erlernen, die Eingaben korrekt den Ausgaben zuordnet, sodass es auf neue, unbekannte Daten generalisieren kann,[34] also zu einer beliebigen, während des Trainings nicht gesehenen Eingabe eine korrekte Ausgabe berechnet. Ziel ist, dass das Modell unbekannte Daten mit minimalem Fehler erkennen kann.[35] Eine zu enge Annäherung an die Trainingsdaten kann jedoch zu einem sog. *Overfitting* führen; in diesem Fall ist das Modell nicht gut darin, für unbekannte Eingabedaten korrekte Ausgaben zu erzeugen.[36]

Beim unüberwachten Lernen *(Unsupervised Learning)* wird nicht von einer externen «*ground truth*», also einer bekannten, «richtigen» Ausgabe, ausgegangen, mit der die Ausgaben des trainierten Modells verglichen werden.[37] Anders als beim überwachten Lernen werden die Modelle hier mit einer Menge unstrukturierter, nicht gelabelter Trainingsdaten trainiert.[38] Im selbstüberwachten Lernen *(Self-Supervised Learning)*, einem sehr häufig verwendeten Unterparadigma des unüberwachten Lernens, das insb. für generative KI-Modelle verwendet wird,[39] werden die für das Training benötigten «richtigen» Ausgaben aus den jeweiligen Eingaben erstellt.[40] Typischerweise wird ein Teil der Eingabe maskiert, und das Modell wird trainiert, diesen fehlenden Teil aus dem nicht maskierten Teil der Eingabe (dem Kontext) zu rekonstruieren. Dadurch lernt das Modell Muster und Zusammenhänge in den Trainingsdaten. Ein Beispiel sind generative Sprachmodelle *(Language Models)*, die durch selbstüberwachtes Lernen auf grossen Textmengen die Regelmässigkeiten der jeweiligen Sprache lernen.[41] Namentlich bei grossen Modellen *(Large Language Models, LLMs)* gehen die erlernten Regelmässigkeiten über die rein sprachlichen Eigenschaften wie lexikalische oder syntaktische Regeln hinaus. Heutige Modelle abstrahieren aus den Trainingsdaten auch Fakten und Zusammenhänge.[42] Ein anderes Beispiel sind Modelle zur Generierung von Bildern. Ein verbreiteter Ansatz besteht bei diesen Modellen darin, die Trainingsbilder (stark) zu verrauschen und aus diesen verrauschten Versionen die Originale zu rekonstruieren.[43] Sobald das Modell Muster und Zusammenhänge innerhalb der Trainingsdaten gelernt hat, ist der Trainingsprozess abgeschlossen.[44] Das ermöglicht den Modellen, neue Daten zu generieren, die sich von den Trainingsdaten unterscheiden, aber denselben Regelmässigkeiten folgen.[45] Bei Sprachmodellen ist dies Text in der jeweiligen Sprache,[46] bei Bildmodellen sind es Bilder, die in ihrer Art den in den Trainingsdaten enthaltenen Bildern gleichen, aber neue Motive wiedergeben können.[47]

Beim verstärkenden Lernen (Reinforcement Learning) handelt ein Agent («Entscheidungsinstanz des Systems») in einer Umgebung, indem er aus einer endlichen oder unendlichen Menge an möglichen Handlungen die jeweils nächste Handlung auswählt, die (langfristig) zu Belohnungen oder Strafen führt (z.B. den nächsten Zug beim Schach).[48] Ziel des Agenten ist es, eine Strategie zu entwickeln, die ihm

29 Dornis/Stober (Fn. 6), 34.
30 M. Kaulartz, Technische Hintergründe, in: M. Kaulartz/T. Braegelmann (Hg.), Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 2.2 Rz. 19.
31 Umfassend: J. Terven/D.-M. Cordova-Esparza/J.-A. Romero-González/A. Ramírez-Pedraza/E. A. Chávez-Urbiola, A comprehensive survey of loss functions and metrics in deep learning, Artif Intell Rev Nr. 195 2025, 1 ff.
32 Dornis/Stober (Fn. 6), 26.
33 R. S. Sutton/A. G. Barto, Reinforcement Learning, An Introduction, 2. Aufl., Cambridge MA 2018, 2; A. Lauterbach, Introduction to Artificial Intelligence and Machine Learning, in: T. F. Claypoole (Hg.), Law of Artificial Intelligence and Smart Machines, Understanding A.I. and the Legal Impact, Chicago 2019, 36.
34 I. Sarker, Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Nature 2021, 159 f.; D. Bergmann/C. Stryker, What is model training?, IBM Think, ‹www.ibm.com/think/topics/model-training› (abgerufen am 5. Januar 2026).
35 Kaulartz (Fn. 30), Kap. 2.2 Rz. 19.
36 S. Tsimenidis, Limitations of Deep Neural Networks: a discussion of G. Marcus' critical appraisal of deep learning, arXiv 2020, 3 f.
37 H. Ugochi Dike/Y. Zhou/K. Kumar Deveerasetty/Q. Wu, Unsupervised Learning Based On Artificial Neural Network: A Review, Proceedings of the 2018 IEEE International Conference on Cyborg and Bionic Systems Shenzhen 2018, 324.
38 Sutton/Barto (Fn. 33), 2; Kaulartz (Fn. 30), Kap. 2.2 Rz. 25.
39 Zu generativen Modellen im Detail siehe hinten, II.2.
40 J. Zhang/L. Yang/S. Mahmoud Sajjadi Mohammadabadi/F. Yan, A survey on self-supervised learning: Recent advances and open problems, Neurocomputing Nr. 131409 2025, 1 ff.
41 Zum Ganzen J. Devlin, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2019, 4171 ff.
42 Siehe dazu im Detail hinten, II.2.b) aa).
43 Siehe dazu im Detail hinten, II.2.b) bb).
44 S. Lapuschkin/S. Wäldchen/A. Binder/G. Montavon/W. Samek/K.-R. Müller, Unmasking Clever Hans predictors and assessing what machines really learn, Nature Communications Nr. 1096 2019, 2 ff.
45 G. M. Harshvardhan/M. Kumar Gourisaria/M. Pandey/S. Swarup Rautaray, A comprehensive survey and analysis of generative models in machine learning, Computer Science Review Nr. 100285 2020, 2; M. Chen/S. Mei/J. Fan/M. Wang, Opportunities and challenges of diffusion models for generative AI, National Science Review Nr. 12 2024, 2.
46 Siehe dazu im Detail hinten, II.2.b) aa).
47 Siehe dazu im Detail hinten, II.2.b) bb).
48 Kreutzer (Fn. 5), 19 f.; Sutton/Barto (Fn. 33), 2; Man-Cho So (Fn. 6), 23.

langfristig eine möglichst hohe Belohnung einbringt.[49] Die Belohnungen müssen also so definiert werden, dass der Agent lernen kann, welche Handlungen und Handlungssequenzen zielführend sind (beim Schach könnte das eine Belohnung von +1 für einen Sieg und eine Strafe von -1 für eine Niederlage sein). Der Agent lernt durch systematisches Ausprobieren und muss dabei *Exploration* (Ausprobieren neuer Handlungen) und *Exploitation* (Ausnutzen bekannter, lohnender Handlungen) ausbalancieren.[50] Auch beim *Reinforcement Learning* kann es zu einem *Overfitting* kommen, wenn der Agent sich zu sehr an seine Umgebung anpasst.[51] Wird bspw. ein Agent in einem Rennspiel auf einer bestimmten Strecke trainiert, wird er diese optimal fahren, aber versagen, sobald sie ein wenig verändert wird.

Grundsätzlich gilt, dass eine grössere Datenmenge die Generalisierungsfähigkeit von Modellen verbessert, weil das Risiko des *Overfitting* sinkt und mehr Muster erfasst werden können. Allerdings unterliegt dieser Effekt dem Prinzip des abnehmenden Grenznutzens: Jeder zusätzliche Datensatz trägt zwar in der Regel zu einer Leistungssteigerung bei, aber mit immer geringerer Wirkung.[52] Grundsätzlich verhält es sich so, dass für ein Modell mit mehr trainierbaren Parametern auch mehr Daten zum Training benötigt werden.[53] Sind nicht genügend Daten verfügbar, können Menge und Vielfalt durch Datenaugmentierung erhöht werden.[54] Dazu werden bspw. von den vorhandenen Daten leicht veränderte Kopien oder synthetische Daten erzeugt.[55]

2. Training generativer KI-Modelle

Generative KI-Modelle werden darauf trainiert, neue Daten zu erzeugen, die den Trainingsdaten in Struktur und Charakter ähnlich sind. Anders als bei diskriminativen Modellen ist es nicht das Ziel, bestehende Daten zu klassifizieren oder zu erkennen, sondern die zugrunde liegende Wahrscheinlichkeitsverteilung der Trainingsdaten zu erfassen und auf dieser Basis eigenständige neue Inhalte zu generieren.[56] Der Prozess lässt sich dabei in zwei Schritte aufteilen: In einem ersten Schritt werden die Daten gesammelt und kuratiert (a) in einem zweiten Schritt folgt das eigentliche Training der Modelle (b).[57]

a) Datensammlung und Datenkuratierung

Für das Training komplexer KI-Modelle werden in der Regel sehr grosse Datenmengen benötigt.[58] Für das Training von generativen Modellen werden die Daten regelmässig mittels *Webscraping* beschafft.[59] Dabei werden Daten automatisiert von Websites extrahiert.[60] Entsprechende Programme suchen dazu nach vorgegebenen Regeln das Internet ab und speichern die gewünschten Daten (Texte, Bilder etc.).[61] Ein bekanntes Beispiel ist *Common Crawl*.[62] Dieses Webarchiv enthält den HTML-Quelltext, Links, Metadaten und Textinhalte von ca. 250 Milliarden Websites.[63]

Die Datensätze des *Common Crawl* sind öffentlich zugänglich, und viele KI-Anbieter erstellen daraus ihre eigenen vorgefilterten Datenkorpora, um ihre Modelle zu trainieren.[64] Viele der so entstandenen grossen Datensätze enthalten nicht nur Daten aus dem *Common Crawl*, sondern auch Daten aus anderen, nicht deklarierten Quellen.[65] In vielen Fällen ist es daher für Dritte nahezu unmöglich herauszufinden, welche Daten im Trainingsset enthalten sind, und aus welchen Quellen sie stammen.[66]

Um die Qualität des Datensatzes zu gewährleisten, werden die Daten in einem zweiten Schritt kuratiert; dabei werden namentlich Duplikate und unerwünschte Informationen entfernt.[67] Oft werden die Daten auch normiert, um sie einfacher verarbeiten zu können.[68] So werden bspw. die für das Training gebrauchten Bilddaten hochskaliert oder verworfen, wenn sie eine bestimmte Mindestauflösung nicht erreichen.[69] Text wird meist von Vollbreite auf ASCII-

49 Kreutzer (Fn. 5), 19 f.; Sutton/Barto (Fn. 33), 2; Man-Cho So (Fn. 6), 23.
50 Sutton/Barto (Fn. 33), 2.
51 Y. Jiang/J. Z. Kolter/R. Raileanu, On the Importance of Exploration for Generalization in Reinforcement Learning, 37th Conference on Neural Information Processing Systems NIPS 2023, 1; C. Zhang/O. Vinyals/R. Munos/S. Bengio, A Study on Overfitting in Deep Reinforcement Learning, 2018, 1 ff.
52 J. Hestness/S. Narang/N. Ardalani/G. Diamos/H. Jun/H. Kianinejad/Md. M. Ali Patwary/Y. Yang/Y. Zhou, Deep Learning Scaling is Predictable, Empirically, arXiv 2022, 13.
53 J. Hoffmann et al., Training compute-optimal large language models, Proceedings of the 36th International Conference on Neural Information Processing Systems 2022, 1 ff.
54 Dornis/Stober (Fn. 6), 27.
55 A. Mumuni/F. Mumuni, Data augmentation: A comprehensive survey of modern approaches, Array Nr. 16 2022, 3 ff.
56 Zum Ganzen Harshvardhan/Kumar Gourisaria/Pandey/Swarup Rautaray (Fn. 45), 2; Chen/Mei/Fan/Wang (Fn. 45), 1.
57 K. Lee/A. F. Cooper/J. Grimmelmann, Talkin' 'Bout AI Generation: Copyright And The Generative-AI Supply Chain, J. Copyright Soc'y U.S.A. 2025, 290 ff.
58 Lee/Cooper/Grimmelmann (Fn. 57), 290.
59 Dornis/Stober (Fn. 6), 54; Lee/Cooper/Grimmelmann (Fn. 57), 290.
60 Dornis/Stober (Fn. 6), 54.
61 Dornis/Stober (Fn. 6), 54.
62 ‹https://commoncrawl.org/› (abgerufen am 5. Januar 2026).
63 I. Ilyankou/M. Wang/S. Cavazzi/J. Haworth, Quantifying Geospatial in the Common Crawl Corpus, Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems 2024, 2.
64 Siehe dazu z.B. für GPT-3 von OpenAI T. B. Brown et al., Language models are few-shot learners, 34th Conference on Neural Information Processing Systems 2020, 6.
65 Siehe z.B. für GPT-5 von OpenAI OpenAI, GPT-5 System Card, ‹https://openai.com/de-DE/index/gpt-5-system-card/› (abgerufen am 5. Januar 2026), 5.
66 Siehe dazu C. Schuhmann/R. Beaumont/R. Vencu/C. Gordon/R. Wightman/M. Cherti/T. Coombes/A. Katta/C. Mullis/M. Wortsman/P. Schramowski/S. Kundurthy/K. Crowson/L. Schmidt/R. Kaczmarczyk/J. Jitsev, LAION-5B: An open large-scale dataset for training next generation image-text models, Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track 2022, 2.
67 Brown et al. (Fn. 64), 8; Schuhmann/Beaumont/Vencu/Gordon/Wightman/Cherti/Coombes/Katta/Mullis/Wortsman/Schramowski/Kundurthy/Crowson/Schmidt/Kaczmarczyk/Jitsev (Fn. 66), 5 f.
68 Dornis/Stober (Fn. 6), 59.
69 Siehe D. Podell/Z. English/K. Lacey/A. Blattmann/T. Dockhorn/J. Müller/J. Penna/R. Rombach, SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis, ICLR Conference Paper 2024, 3 f., die allerdings eine Methode vorschlagen, die auf eine Normierung verzichten.

Normalbreite und zusätzlich auf Kleinbuchstaben normalisiert.[70] Zudem werden alle Daten in eine einheitliche, maschinenlesbare Form überführt.[71]

Ein bekanntes Beispiel für einen öffentlich verfügbaren Datensatz ist der LAION-Datensatz. Dieser wurde auf Grundlage von Websitedaten aus dem *Common Crawl* erstellt.[72] Dafür wurden aus den *Common-Crawl*-Beständen Bild-URLs und zugehörige Textbeschreibungen extrahiert.[73] Zur Qualitätssicherung wurden die verlinkten Bilder heruntergeladen und nach inhaltlicher Übereinstimmung zwischen Text und Bild gefiltert.[74] Der LAION-Datensatz enthält allerdings nicht die Bilder und die dazugehörigen Textbeschreibungen selbst, sondern nur die Metadaten, also die Links zu den Bildern und den entsprechenden Beschreibungen.[75] Die Nutzerinnen und Nutzer des LAION-Datensatzes laden also Daten immer selbst herunter und kreieren und kuratieren meist eigene *Subsets*, um ihre Modelle zu trainieren.[76]

b) Trainingsprozess

aa) Textgenerierung

Grosse Sprachmodelle *Large (Language Models, LLMs)* werden mit sehr grossen Textkorpora trainiert[77] *(*; so enthält z.B. der *Common-Crawl*-Datensatz fast eine Billion Wörter.[78] Für die Textgenerierung können verschiedene Modellarchitekturen verwendet werden. Die meisten Sprachmodelle, einschliesslich der grossen Sprachmodelle, verwenden eine sog. Transformerarchitektur; es können aber auch mit anderen Architekturen wie SSMs *(State Space Models)* oder mit traditionelleren RNNs *(Recurrent Neural Networks)* kompetitive Ergebnisse erzielt werden.[79] Der zentrale Baustein der Transformerarchitektur sind die sogenannten *Self-Attention Layers*. *Self-Attention* ist ein Mechanismus, bei dem zur Berechnung der Repräsentation eines Wortes für die nächsthöhere Ebene eines neuronalen Netzes der Kontext, in dem dieses Wort im konkreten Eingabesatz oder -text auftaucht, einbezogen wird, indem die unterschiedlichen Positionen innerhalb dieses einen Satzes (oder Textes) miteinander in Beziehung gesetzt werden.[80] Im Wesentlichen bewertet das Modell für jedes Wort (bzw. *Token*, siehe sogleich) im Text, wie wichtig oder relevant alle anderen Wörter in derselben Sequenz für dessen Bedeutung sind.[81]

Als Eingabe erhält das Modell eine Sequenz von Wörtern oder kleineren sprachlichen Einheiten, bestehend aus einzelnen Silben oder Buchstabenkombinationen (*Word tokens* oder *Subword tokens*).[82] Die genaue Aufteilung hängt vom verwendeten *Tokenizer* ab. Ein konkretes Beispiel ist die Aufteilung des aus sieben Wörtern bestehenden Satzes «Der Apfel fällt nicht weit vom Stamm» in neun *Tokens* durch den *OpenAI Tokenizer*:[83]

Der	Ap	fel	fällt	nicht	weit	vom	Stamm	.

Damit das Modell sprachliche Zusammenhänge erfassen kann, werden die *Tokens* typischerweise mithilfe einer sog. *Embedding-Matrix* in mehrdimensionale Vektoren aus reellen Zahlen übersetzt, etwa [0.12, –0.83, 0.45, …].[84] Diese Vektoren bilden gemeinsam einen hochdimensionalen Raum, in dem *Tokens* mit ähnlicher Bedeutung oder Verwendung räumlich nahe beieinanderliegen, während solche mit unterschiedlicher Bedeutung oder Verwendung weiter voneinander entfernt sind.[85] Auf diese Weise entsteht eine geometrische Struktur, in der semantische und syntaktische Beziehungen nicht durch Regeln, sondern durch Abstände und Richtungen im Vektorraum ausgedrückt werden.[86] Semantische und syntaktische Beziehungen entstehen aber nicht alleine aus statischen Abständen zwischen Wortvektoren, sondern aus deren Verarbeitung durch die kontext-

70 T. Kudo/J. Richardson, SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations 2018, 68.

71 L. Käde, Kreative Maschinen und Urheberrecht, Die Machine Learning-Werkschöpfungskette vom Training über Modellschutz bis zu Computational Creativity, Baden-Baden 2021, 65.

72 Schuhmann/Beaumont/Vencu/Gordon/Wightman/Cherti/Coombes/Katta/Mullis/Wortsman/Schramowski/Kundurthy/Crowson/Schmidt/Kaczmarczyk/Jitsev (Fn. 66), 4.

73 Schuhmann/Beaumont/Vencu/Gordon/Wightman/Cherti/Coombes/Katta/Mullis/Wortsman/Schramowski/Kundurthy/Crowson/Schmidt/Kaczmarczyk/Jitsev (Fn. 66), 4 ff.

74 Schuhmann/Beaumont/Vencu/Gordon/Wightman/Cherti/Coombes/Katta/Mullis/Wortsman/Schramowski/Kundurthy/Crowson/Schmidt/Kaczmarczyk/Jitsev (Fn. 66), 4 ff.

75 Schuhmann/Beaumont/Vencu/Gordon/Wightman/Cherti/Coombes/Katta/Mullis/Wortsman/Schramowski/Kundurthy/Crowson/Schmidt/Kaczmarczyk/Jitsev (Fn. 66), 7.

76 Stability AI hat z.B. das generative Modell «Stable Diffusion» auf einem Subset von LAION-5B trainiert. Siehe dazu R. Rombach/A. Blattmann/D. Lorenz/P. Esser/B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, IEEE 2022/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2022, 10675 f.

77 Für eine Einführung in die Funktionsweise des Trainings von Sprachmodellen: ‹https://huggingface.co/learn/llm-course/chapter1/1› (abgerufen am 5. Januar 2026).

78 Brown et al. (Fn. 64), 8.

79 W. X. Zhao et al., Survey of Large Language Models, arXiv 2025, 1 ff.

80 Siehe dazu im Detail A. Vaswani/N. Shazeer/N. Parmar/J. Uszkoreit/L. Jones/A. N. Gomez/L. Kaiser/I. Polosukhin, Attention is All You Need, Proceedings of the 31st International Conference on Neural Information Processing Systems 2017, 2.

81 Vaswani/Shazeer/Parmar/Uszkoreit/Jones/Gomez/Kaiser/Polosukhin (Fn. 80), 4 ff.

82 R. Sennrich/B. Haddow/A. Birch, Neural Machine Translation of Rare Words with Subword Units, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics 2016, 1715 ff.; Zhao et al. (Fn. 79), 19; Dornis/Stober (Fn. 6), 42; D. Rosenthal, Teil 17: Was in einem KI-Modell steckt und wie es funktioniert, ‹www.vischer.com/kuenstliche-intelligenz/› (abgerufen am 5. Januar 2026).

83 ‹https://platform.openai.com/tokenizer› (abgerufen am 5. Januar 2026).

84 Siehe Vaswani/Shazeer/Parmar/Uszkoreit/Jones/Gomez/Kaiser/Polosukhin (Fn. 80), 5.

85 J. Pennington/R. Socher/C. D. Manning, GloVe: Global Vectors for Word Representation, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing 2014, 1532; T. Mikolov/K. Chen/G. Corrado/J. Dean, Efficient Estimation of Word Representations in Vector Space, arXiv 2013, 4 ff.

86 T. Mikolov/I. Sutskever/K. Chen/G. Corrado/J. Dean, Distributed Representations of Words and Phrases and their Compositionality, Advances in Neural Information Processing Systems 2013, 1 ff.

sensitiven *Self-Attention*-Ebenen.[87] Während des Trainings werden insb. die Gewichte der *Self-Attention*-Komponenten optimiert, die Beziehungen zwischen den *Tokens* abhängig vom Kontext modellieren. Damit das Modell zudem Reihenfolgeinformationen nutzen kann, werden die *Token Embeddings* um Positionsinformationen ergänzt (*Positional Encodings* oder *Positional Embeddings*). Dadurch kann das Modell die Reihenfolge der Wörter berücksichtigen.[88]

So wird die Beziehung zwischen den *Tokens* [Ap] und [fel] sowie [Stamm] nicht über explizit oder manuell kodiertes Bedeutungswissen, sondern über Ähnlichkeiten in ihrer statistischen Nutzung abgebildet. Das Trainingsziel besteht darin, für jedes *Token* in einer Sequenz die Wahrscheinlichkeit seines Auftretens auf Grundlage des Kontexts zu maximieren.[89] In der Anwendung generiert das Modell dann Sätze oder Texte, deren Wahrscheinlichkeit im gegebenen Kontext hoch ist. Um ein nichtdeterministisches Modellverhalten (also sprachliche Varianz) zu erreichen, sagen die meisten Modelle nicht immer das wahrscheinlichste *Token* voraus *(greedy decoding)*,[90] stattdessen wird ein *Token* wird bspw. aus einer festen Anzahl mehrerer möglicher *Tokens* über einer bestimmten Wahrscheinlichkeit zufällig gezogen *(Top-k Sampling)*[91], oder es werden so viele der wahrscheinlichsten *Tokens* ausgewählt, bis ihre aufsummierte Wahrscheinlichkeit einen gewissen Wert erreicht, und anschliessend wird aus diesem Pool ein *Token* zufällig gezogen *(Top-p Sampling)*.[92]

Ziel des Trainings ist nicht das Erlernen des Inhalts der für das Training verwendeten Texte, sondern lediglich die statistische Wahrscheinlichkeit, wie *Tokens* in diesen Texten in Relation zueinander auftreten; gespeichert werden sollen also nur die Wahrscheinlichkeitsbeziehungen zwischen den *Tokens*.[93] Dies ermöglicht dem Modell, statistisch wahrscheinliche Kombinationen von *Tokens* zu generieren.[94] Neben den sprachlichen Mustern können die trainierten Modelle jedoch, wie erwähnt, auch relationale Informationen wiedergeben, weil nicht nur die lexikalischen und syntaktischen Regeln, sondern auch die statistischen Beziehungen zwischen den in den Texten repräsentierten Informationen erlernt werden.[95]

Die grundlegenden Prinzipien zum Training von Texten in natürlicher Sprache gelten auch für das Lernen und Generieren von *Source Code*, auch wenn im Einzelnen Unterschiede bestehen können, bspw. bei der sinnvollen Tokenisierung von *Code*.[96]

bb) Bildgenerierung

Für die Text-zu-Bild- und die Bild-zu-Bild-Generierung sind heute drei Modelltypen üblich: *Generative Adversarial Networks (GANs)*, *Variational Autoencoders (VAEs)* und Diffusionsmodelle.[97]

Aktuell werden vor allem Diffusionsmodelle verwendet.[98] Diese kombinieren zwei komplementäre Prozesse. Im Vorwärtsprozess wird eine Probe aus der realen Datenverteilung schrittweise durch das Hinzufügen von Rauschen verfälscht, bis sie nicht mehr von reinem Rauschen zu unterscheiden ist. Im Rückwärtsprozess wird ein neuronales Netz darauf trainiert, diese Verfälschung rückgängig zu machen, indem es das Rauschen schrittweise entfernt (sog. *Denoising*) und so aus zufälligem Rauschen wieder Bilder erzeugt, die der ursprünglichen Datenverteilung entsprechen.[99]

Ein digitales Bild liegt auf einem Computer immer als numerisches Raster von Pixelwerten vor. In diesem Raster sind jedem Pixel jeweils eine Position und eine Farbe zugewiesen. Die Position im Raster wird durch Koordinaten repräsentiert (z.B. (1, 2)). Die zugewiesenen Farben werden im RGB-Farbmodell aus den Lichtfarben Rot, Grün und Blau gebildet. Jeder der drei Kanäle kann eine Intensität zwischen 0 und 255 annehmen. Auf dieser Grundlage entsteht die sichtbare Farbe eines Pixels durch additive Mischung der Kanalwerte. Ein Pixel mit den Werten R: 255, G: 0, B: 0 erscheint als reines Rot, während Gelb durch die Kombination von Rot und Grün mit den Werten R: 255, G: 255, B: 0 entsteht. Durch Einstellung der Farbkanäle kann so jede Farbe gemischt werden.[100] Für Diffusions-

87 K. Ethayarajh, How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing 2019, 55 ff.

88 Zum Ganzen Vaswani/Shazeer/Parmar/Uszkoreit/Jones/Gomez/Kaiser/Polosukhin (Fn. 80), 1 ff.

89 Zhao et al. (Fn. 79), 26.

90 Zhao et al. (Fn. 79), 26.

91 Zhao et al. (Fn. 79), 27.

92 Zhao et al. (Fn. 79), 27.

93 Zhao et al. (Fn. 79), 26; H. Cho/Y. Sakai/K. Tanaka/M. Kato/N. Inoue, Understanding Token Probability Encoding in Output Embeddings, Proceedings of the 31st International Conference on Computational Linguistics 2025, 10618 ff.; D. Rosenthal/L. Veraldi, Das Training von KI-Sprachmodellen mit fremden Inhalten und Daten aus rechtlicher Sicht, in: Jusletter 3. Februar 2025, Rz. 27 ff. Zur Problematik der Memorisierung siehe hinten, II.3).

94 Dornis/Stober (Fn. 6), 42.

95 F. Petroni/T. Rocktäschel/P. Lewis/A. Bakhtin/Y. Wu/A. H. Miller/S. Riedel, Language Models as Knowledge Bases?, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing 2019, 2463 ff.

96 D. Fried, InCoder: A Generative Model for Code Infilling and Synthesis, The Eleventh International Conference on Learning Representations 2023, 1 ff.; X. Du, Evaluating Large Language Models in Class-Level Code Generation, 2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE) 2024, 982 ff.

97 Z. Sordo/E. Chagnon/Z. Hu/J. J. Donatelli/P. Andeer/P. S. Nico/T. Northen/D. Ushizima, Synthetic Scientific Image Generation with VAE, GAN, and Diffusion Model Architectures Journal of Imaging Nr. 252 2025, 7 ff.

98 Chen/Mei/Fan/Wang (Fn. 45), 5; Rombach/Blattmann/Lorenz/Esser/Ommer (Fn. 76), 10675 f.; C. Chen/E. Liu/D. Liu/M. Shah/C. Xu, Investigating Memorization in Video Diffusion Models, ICLR 2025 Workshop on Navigating and Addressing Data Problems for Foundation Models, 1.

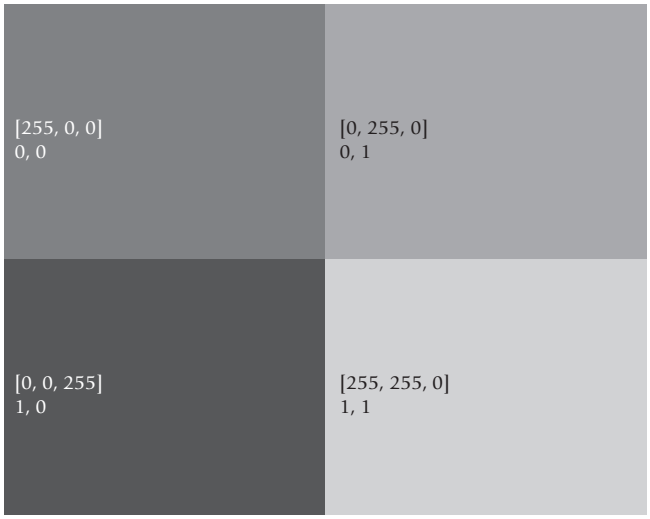
99 J. Ho/A. Jain/P. Abbeel, Denoising Diffusion Probabilistic Models, Advances in Neural Information Processing Systems 2020, 2; Chen/Mei/Fan/Wang (Fn. 45), 3 ff.; P. Esser et al., Scaling Rectified Flow Transformers for High-Resolution Image Synthesis, Forty-first International Conference on Machine Learning 2024, 2.

100 Zum Ganzen siehe M. Muratbekova, Color Models in Image Processing: A Review and Experimental Comparison, arXiv 2025, 1 ff.

modelle werden diese Werte in der Praxis als *Tensor*[101] verarbeitet.

Ein *Tensor* eines RGB-Bildes enthält die Dimensionen Pixelhöhe *(Height)*, Pixelbreite *(Width)* und Anzahl Farbkanäle *(Channels)*. Für ein Bild mit vier Pixeln hätte ein *Tensor* also die Form [2, 2, 3] (zwei Pixel hoch, zwei Pixel breit und drei Farbkanäle).

Jede Pixelposition enthält die jeweiligen Werte für die Intensität von Rot, Grün und Blau:[102]



Um neue Bilder zu generieren, beginnt das Modell mit reinem Rauschen (zufälliger *Tensor* ohne Struktur) und wandelt dieses schrittweise in einen neuen *Tensor* um. Dieser *Tensor* besteht aus numerischen Werten, die die RGB-Kanäle eines späteren Bildes beschreiben. In vielen Diffusionsverfahren enthält dieser Rückwärtsprozess stochastische Komponenten, wodurch bei gleicher Modellkonfiguration unterschiedliche Bilder aus der gelernten Verteilung erzeugt werden können. Dass die resultierenden Bilder in der Regel keine identischen Kopien von Trainingsbeispielen sind, liegt primär daran, dass das Modell beim Training Muster über viele Daten hinweg lernt und so eine Verteilung approximiert, statt einzelne Bilder exakt zu speichern.[103]

Diffusionsmodelle zur Generierung von Bildern aus einem Eingabetext werden mit Bild-Text-Paaren trainiert. Dabei lernt das Modell, welche Bildkomposition bei einer gegebenen Textbeschreibung wahrscheinlich ist, sodass es anschliessend Bilder generieren kann, die einem Textprompt entsprechen.[104]

Auch bei diesen Modellen werden typischerweise die Bildbeschreibungen zunächst in eine Folge von *Tokens* umgewandelt und anschliessend mithilfe eines Sprachmodells in numerische Repräsentationen *(Embeddings)* transformiert.[105] Diese Textembeddings dienen dem Diffusionsmodell als zusätzlicher konditionierender Input während des *Denoising*-Prozesses und beeinflussen so, welche Inhalte im Bild erzeugt werden.[106] Auf dieser Grundlage kann das Modell, bspw. über sog. *Cross-Attention*, lernen, welche Teile der Textbeschreibung für welche Bildinhalte relevant sind, und diese entsprechend gewichten.[107] Ziel des Trainings ist, dass das Modell durch die enorme Anzahl an Trainingsbeispielen[108] stabile statistische Zusammenhänge zwischen Sprache und Bildinhalt aufbaut, und generalisierbare Muster lernt, ohne einzelne Bilder zu speichern.

cc) Audiogenerierung

Die Audiogenerierung kombiniert verschiedene technische Ansätze, die sich grundlegend darin unterscheiden, welche Form von Daten erzeugt wird. Während symbolische Modelle musikalische Strukturen wie Noten, Akkorde und zeitliche Ereignisse in Form von Befehlen generieren,[109] erzeugen audiobasierte Modelle direkt hörbare Klangsignale.[110] Für die Erzeugung von Audio hat sich in den letzten Jahren ein Methodenmix etabliert.[111] Es werden unter anderem autoregressive Modelle und Diffusionsmodelle eingesetzt.[112]

Für das Training werden die Audiodateien je nach Modell entweder – ähnlich wie bei Sprachmodellen[113] – *tokenisiert*[114] oder für Diffusionsmodelle in eine kontinuierliche Darstellung, z.B. in ein Spektrogramm überführt.[115] Dabei

101 Ein Tensor ist ein mathematisches Objekt, das Skalare, Vektoren, Matrizen und Arrays noch höherer Dimensionen umfasst. Tensoren werden verwendet, um Daten wie zweidimensionale Bilder (Höhe und Breite), dreidimensionale Videos (Höhe, Breite und Zeit) oder vierdimensionale Tensoren für Sequenzen dreidimensionaler Bilder (Höhe, Breite, Tiefe und Zeit) darzustellen.

102 J. Kausthub, How do images get converted to tensors in PyTorch?, Kausthub's Substack, ‹https://kausthub.substack.com/p/how-do-images-get-converted-to-tensors› (abgerufen am 5. Januar 2026).

103 Zum Ganzen grundlegend J. Ho/A. Jain/P. Abbeel, Denoising Diffusion Probabilistic Models, Proceedings of the 34th International Conference on Neural Information Processing Systems 2020, 1 ff.; vereinfacht D. Bergmann/C. Stryker, What are Diffusion Models?, IBM, ‹www.ibm.com/think/topics/diffusion-models› (abgerufen am 5. Januar 2026). Aufgrund ihres deutlich geringeren Rechenaufwandes werden heute mehrheitlich latente Diffusionsmodelle eingesetzt. Diese arbeiten nicht im Pixelraum, sondern mit stark informationsverdichtetem Bilddaten. Das Grundprinzip des Diffusionsprozesses bleibt jedoch identisch, siehe dazu im Detail Rombach/Blattmann/Lorenz/Esser/Ommer (Fn. 76), 10674 ff. Zur Problematik der Memorisierung siehe hinten, II.3.b).

104 Chen/Mei/Fan/Wang (Fn. 45), 5.

105 Siehe dazu oben, II.2.b) aa).

106 Rombach/Blattmann/Lorenz/Esser/Ommer (Fn. 76), 10677 ff.

107 Rombach/Blattmann/Lorenz/Esser/Ommer (Fn. 76), 10677 f.; Chen/Mei/Fan/Wang (Fn. 45), 5.

108 Der LAION-5B-Datensatz enthält z.B. 5 Milliarden Text-Bild-Paare: Schuhmann/Beaumont/Vencu/Gordon/Wightman/Cherti/Coombes/Katta/Mullis/Wortsman/Schramowski/Kundurthy/Crowson/Schmidt/Kaczmarczyk/Jitsev (Fn. 66), 13.

109 A. Muhamed et al., Symbolic Music Generation with Transformer-GANs, Proceedings of the AAAI Conference on Artificial Intelligence 2021, 408.

110 Für Text zu Audiogeneration siehe z.B. ‹https://suno.com/› (abgerufen am 5. Januar 2026).

111 W. Lin/C. He, Continuous Autoregressive Modeling with Stochastic Monotonic Alignment for Speech Synthesis, The Thirteenth International Conference on Learning Representations 2025, 3.

112 Lin/He (Fn. 111), 3 ff.

113 Siehe dazu oben, II.2.b) aa).

114 Die *Tokens* enthalten dabei Informationen wie Tonlänge und -höhe sowie deren Platzierung in der Audiodatei. Siehe dazu W. X. Hsiao/J. Liu/Y. Yeh/Y. Yang, Compound Word Transformer: Learning to Compose Full-Song Music over Dynamic Directed Hypergraphs, Proceedings of the AAAI Conference on Artificial Intelligence 2021, 178.

115 F. Schneider/O. Kamal/Z. Jin/B. Schölkopf, *Moûsai*: Efficient Text-to-Music Diffusion Models, Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics 2024, 8052.

handelt es sich um Abbilder von Audiosignalen, die über drei Dimensionen Zeit, Frequenz und Intensität wiedergeben.[116]

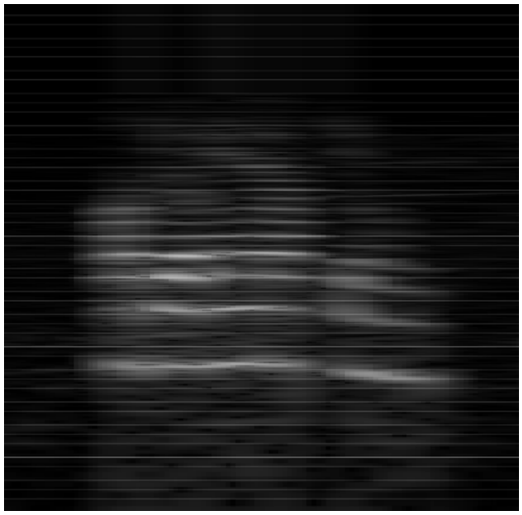


Abbildung 1: Spektrogramm des gesprochenen Wortes «Audiogenerierung».

Für das Training werden grosse Datensätze aus Musik und dazugehörigen Textbeschreibungen genutzt. Für das Modell «Moûsai» wurden zur Erstellung des Datensatzes bspw. über Spotify[117] 49'000 Audiodateien gesucht und anschliessend über YouTube heruntergeladen. Anschliessend werden die zu den Songs gehörenden Metadaten dazu benutzt, um die Textbeschreibungen der Audiodateien zu erstellen.[118]

Das Training von Diffusionsmodellen zur Audiogenerierung gleicht im Kern demjenigen von bildbasierten Diffusionsmodellen.[119] Eine kompakte Darstellung des Audiomaterials wird schrittweise mit Rauschen überlagert. Ein neuronales Netzwerk lernt dann, aus diesem verrauschten Zustand wieder sinnvolle (d.h. der Wahrscheinlichkeitsverteilung der Trainingsdaten entsprechende) Audiodaten zu generieren.[120] Zudem lernt das Modell, wie sich Textbeschreibungen in typische Muster der Audiodaten übersetzen lassen, sodass im Rückwärtsprozess aus Texteingaben der Nutzerinnen und Nutzer neue Audiodaten generiert werden können, die mit der gelernten Datenverteilung und mit dem Prompt konsistent sind.[121]

dd) Videogenerierung

Auch bei der Videogenerierung dominieren derzeit Diffusionsmodelle.[122] Sie funktionieren prinzipiell gleich wie Diffusionsmodelle zur Bildgenerierung,[123] mit dem Unterschied, dass bei Text-zu-Video-Modellen eine zeitliche Komponente hinzukommt.[124] Viele Modelle bauen auf bestehenden Text-zu-Bild-Ansätzen auf und werden durch zeitbezogene Module erweitert, die Objektbewegungen, Kamerabewegungen und andere Veränderungen zwischen den einzelnen *Frames* (Einzelbild eines Videos) modellieren.[125]

Für das Training werden die Modelle mit grossen Mengen an Videodaten trainiert. Für das Lumiere-Modell wurden bspw. dreissig Millionen Videos inklusive Beschreibung verwendet. Vor dem Training werden die Videos normiert, d.h. die Anzahl Frames pro Sekunde wird vereinheitlicht (bspw. pro Video achtzig Frames bei sechzehn Frames pro Sekunde), und die räumliche Auflösung wird auf ein einheitliches Format gebracht (bspw. 128 × 128 Pixel).[126] So entstehen einheitliche fünf Sekunden lange Videos, unabhängig davon, mit welchen Frames pro Sekunde das ursprüngliche Video aufgenommen wurde.

Vor dem Training werden die Videosequenzen in vierdimensionale *Tensoren* mit der Struktur Frames, Höhe, Breite und Kanäle überführt. Anschliessend wird dem gesamten *Tensor* ein gemeinsames Rauschmuster hinzugefügt, sodass die zeitliche Dimension nicht unabhängig pro Frame verrauscht wird, sondern als zusammenhängende Einheit. Analog zur Bildgenerierung lernt das Modell im Rückwärtsprozess der Diffusion, den verrauschten *Videotensor* schrittweise zu entrauschen.[127]

Für die Generierung werden üblicherweise zunächst einige *Keyframes* erzeugt, die als zeitliche Fixpunkte dienen. Anschliessend berechnet das System die fehlenden Übergangsbilder zwischen diesen Fixpunkten, sodass aus wenigen Ankerpunkten eine flüssige und konsistente Videosequenz entsteht.[128]

Damit das Modell nach Anweisungen der Nutzerinnen und Nutzer neue Videos erstellen kann, werden alle Videos mit einer Beschreibung versehen, die miterlernt wird.[129]

116 Siehe dazu ‹https://listeningtowaves.com/sound-exploration› (abgerufen am 5. Januar 2026); Auf ‹https://spectrogram.sciencemusic.org/› (abgerufen am 5. Januar 2026) kann man ganz einfach eigene Spektrogramme erstellen.

117 Bei Spotify handelt es sich um eine der grössten Musikstreamingplattformen: ‹https://open.spotify.com/intl-de/› (abgerufen am 5. Januar 2026).

118 Schneider/Kamal/Jin/Schölkopf (Fn. 115), 8055 f.

119 Siehe dazu oben, II.2.b) bb).

120 Siehe dazu im Detail Schneider/Kamal/Jin/Schölkopf (Fn. 115), 8052 ff.

121 H. Liu et al., AudioLDM: Text-to-Audio Generation with Latent Diffusion Models, AudioLDM: Text-to-Audio Generation with Latent Diffusion Models, Proceedings of the 40th International Conference on Machine Learning 2023, 1 ff.

122 Z. Xing et al., A Survey on Video Diffusion Models, ACM Comput. Surv. 2024, 1 f.; Y. Wang et al., Survey of Video Diffusion Models: Foundations, Implementations, and Applications, Transactions on Machine Learning 2025, 4, 6.

123 Siehe dazu oben, II.2.b) bb).

124 A. Blattmann/R. Rombach/H. Ling/T. Dockhorn/S. Wook Kim/S. Fidler/K. Kreis, Align Your Latents: High-Resolution Video Synthesis with Latent Diffusion Models, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023, 22565.

125 Blattmann / Rombach / Ling / Dockhorn / Wook Kim / Fidler / Kreis (Fn. 124), 22563 ff.

126 O. Bar-Tal et al., Lumiere: A Space-Time Diffusion Model for Video Generation, SIGGRAPH Asia 2024 Conference Papers 2024, 5.

127 Zum Ganzen siehe J. Ho et al., Video Diffusion Models, Video Diffusion Models, Advances in Neural Information Processing Systems 2022,1 ff.

128 Blattmann / Rombach / Ling / Dockhorn / Wook Kim / Fidler / Kreis (Fn. 124), 22567.

129 Liu et al., Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models, arXiv 2024, 16.

3. Memorisierung von Trainingsdaten

Das Ziel des Trainings von KI-Modellen besteht darin, dass die Modelle statistische Muster aus dem Trainingsdatensatz extrahieren, und nicht darin, dass sie die im Trainingsdatensatz enthaltenen Werke als solche speichern.[130] Verschiedene Studien zeigen jedoch, dass sich Modelle an Text-,[131] Bild-,[132] Musik-[133] und Videodaten[134] exakt oder nahezu exakt «erinnern» können.

a) Memorisierung von Textdaten

Die Informatik kennt verschiedene, leicht unterschiedliche Definitionen der «Memorisierung». Verwendet werden Begriffe wie *«verbatim memorization»*, *«extractable memorization»* oder *«eidetic memorization»*. Den verschiedenen Begriffen ist allerdings gemeinsam, dass sie die Eigenschaft grosser Sprachmodelle bezeichnen, längere Textpassagen zu erzeugen, die exakte oder nahezu exakte Übereinstimmungen mit Sequenzen aus ihren Trainingsdaten darstellen.[135]

Um Sprachmodelle auf Memorisierung zu prüfen, verwendet man Trainingsdaten-Extraktionsangriffe *(extraction attacks)*, die die Reproduktion von Trainingsinhalten provozieren. Dabei werden dem Modell gezielte Prompts oder Textsequenzen aus den möglicherweise memorisierten Werken eingegeben, und diese Eingaben werden automatisiert angepasst und vielfach wiederholt; anschliessend wird unter den Ausgaben nach Sequenzen gesucht, die mit hoher Wahrscheinlichkeit wörtlich aus den Trainingsdaten stammen.[136] Generiert das Modell lange und spezifische *(non trivial)* Textsequenzen, die wörtlich einem Werk entsprechen, so spricht dies gegen eine zufällige Neugenerierung und dafür, dass das Modell diese Passage tatsächlich memorisiert hat.[137] Reproduziert werden dabei vor allem Textpassagen, die in den Trainingsdaten oft vorkommen,[138] bspw. der gesamte Text der *MIT-*, *Creative-Commons-* oder *Project-Gutenberg*-License.[139] In anderen Experimenten ist es sogar gelungen, den ganzen Text des Buches *«Harry Potter and the Sorcerer's Stone»* zu reproduzieren.[140]

Mit zunehmender Grösse des LLM nimmt auch der Anteil der extrahierbaren Trainingsdaten zu.[141] Diese Erkenntnis stützt sich auf Vergleiche zwischen den Modellen *GPT-2* und *GPT-Neo*, die unter ähnlichen Bedingungen, aber mit unterschiedlichen Datenmengen (40 GB und 800 GB) trainiert wurden. In den durchgeführten Experimenten «erinnerte» sich das auf 800 GB Textdaten trainierte Modell *GPT-Neo* an 40% der Daten aus dem Evaluationsset, während das kleinere *GPT-2* lediglich 6% des Sets wiederherstellen konnte. Diese Experimente legen nahe, dass grössere Sprachmodelle nicht nur besser generalisieren und Textpassagen «neu» erzeugen, sondern Texte oder Textpassagen, die sie im Training «gesehen» haben, bei geeigneten Prompts auch wörtlich oder nahezu wörtlich reproduzieren können.[142] Die Texte oder Textpassagen können zwar nicht deterministisch abgefragt werden, wie dies bei einer herkömmlichen Datenbank der Fall wäre, es ist aber unter Umständen möglich, Werkteile oder nahezu ganze Werke exakt oder nahezu exakt zu reproduzieren.

Der Befund legt nahe, dass mit wachsender Modellkapazität nicht nur die Mustererkennung, sondern auch die Fähigkeit zur wörtlichen Wiedergabe steigt. Dabei gilt: Je öfter ein Text im Training vorkam, desto höher ist die Wahrscheinlichkeit, dass das Modell ihn memorisiert.[143] Ein Risiko der Memorisierung besteht bereits, wenn ein Text in relevanter Häufigkeit wiederholt wird. So hat eine Studie gezeigt, dass eine 33-fache Wiederholung eines Textes zu dessen Memorisierung in einem grossen Sprachmodell führen kann.[144]

b) Memorisierung von Bilddaten

Ähnlich wie bei den Tests zur Prüfung der Memorisierung von Textdaten versucht man auch Bildmodelle durch *attacks*

130 H. Brown et al., What Does it Mean for a Language Model to Preserve Privacy?, FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2282 f.; N. Carlini/C. Liu/Ú. Erlingsson/J. Kos/D. Song, The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks, 28th USENIX Security Symposium 2019, 269.

131 N. Carlini/F. Tramèr/E. Wallace/M. Jagielski/A. Herbert-Voss/K. Lee/A. Roberts/T. Brown/D. Song/U. Erlingsson/A. Oprea/C. Raffel, Extracting Training Data from Large Language Models, 30th USENIX Security Symposium 2021, 2633 ff.; M. Nasr et al., Scalable Extraction of Training Data from Aligned, Production Language Models, The Thirteenth International Conference on Learning Representations 2025, 1 ff.

132 N. Carlini et al., Extracting Training Data from Diffusion Models, Proceedings of the 32nd USENIX Conference on Security Symposium 2021, 5253 ff.

133 J. Roh et al., Bob's Confetti: Phonetic Memorization Attacks in Music and Video Generation, submitted to The Fourteenth International Conference on Learning Representations *(erscheint 2026)*, 1 ff.; D. Bralios et al., Generation or Replication: Auscultating Audio Latent Diffusion Models, ICASSP 2024 – IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2024, 1156 ff.

134 Ceh/Liu/Liu/Sha/Xu (Fn. 98), 1 ff.

135 J. Huang/D. Yang/C. Potts, Demystifying Verbatim Memorization in Large Language Models, in: The Stanford AI Lab Blog, ‹https://ai.stanford.edu/blog/verbatim-memorization/› (abgerufen am 5. Januar 2026).

136 Nasr et al. (Fn. 131), 3; Carlini/Tramèr/Wallace/Jagielski/Herbert-Voss / Lee / Roberts / Brown / Song / Erlingsson / Oprea / Raffel (Fn. 131), 2637.

137 A. Feder Cooper et al., Extracting memorized pieces of (copyrighted) books from open-weight language models, ICML 2025 Workshop on Reliable and Responsible Foundation Models 2025, 1 ff.; Carlini / Tramèr / Wallace / Jagielski / Herbert-Voss / Lee / Roberts / Brown / Song/Erlingsson/Oprea/Raffel (Fn. 131), 2639 ff.

138 Carlini / Tramèr / Wallace / Jagielski / Herbert-Voss / Lee / Roberts / Brown / Song / Erlingsson / Oprea / Raffel (Fn. 131), 2638 f.; N. Carlini et al., Quantifying Memorization Across Neural Language Models, The Eleventh International Conference on Learning Representations 2023, 5.

139 Carlini / Tramèr / Wallace / Jagielski / Herbert-Voss / Lee / Roberts / Brown/Song/Erlingsson/Oprea/Raffel (Fn. 131), 2638 f.

140 Feder Cooper et al. (Fn. 137), 10.

141 Carlini / Tramèr / Wallace / Jagielski / Herbert-Voss / Lee / Roberts / Brown / Song / Erlingsson / Oprea / Raffel (Fn. 131), 2644; Brown et al. (Fn. 130), 4.

142 Carlini et al. (Fn. 138), 4 f.

143 Carlini et al. (Fn. 138), 1.

144 Zum Ganzen Carlini/Tramèr/Wallace/Jagielski/Herbert-Voss/Lee/Roberts/Brown/Song/Erlingsson/Oprea/Raffel (Fn. 131), 2644.

dazu zu bringen, in den Trainingsdaten enthaltene Bilder zu generieren.[145] Experimente haben bspw. gezeigt, dass *Stable Diffusion*[146] Bildinhalte aus dem Trainingsdatensatz reproduzieren kann. In den Versuchen wurden unter anderem Bildbeschreibungen *(Captions)* aus dem LAION-Datensatz als *Prompts* verwendet, um gezielt Reproduktionen von Bildern zu provozieren. Im Durchschnitt führen diese *Captions* nicht zu einer eigentlichen Reproduktion des zugehörigen Trainingsbildes. Für die Bewertung als Memorisierung ist jedoch entscheidend, dass in einzelnen Fällen (nahezu) identische Reproduktionen auftreten können. So führen z.B. bestimmte Schlüsselphrasen wie etwa Künstlernamen oder häufig vorkommende stilistische Begriffe eher zu Reproduktionen. Ein Prompt wie «*A painting of The Great Wave off Kanagawa by Katsushika Hokusai*» kann zur Ausgabe einer nahezu identischen Kopie des Originals führen. Häufige Wiederholungen eines Bildes im Trainingsdatensatz begünstigen Reproduktionen zusätzlich, wenn die passenden Prompts eingegeben werden. Es ist allerdings sehr unwahrscheinlich, dass bei zufälligen Texteingaben ein in den Trainingsdaten enthaltenes Bild reproduziert wird.[147]

Generell werden viel eher Bilder reproduziert, die «einfach» sind und deren Beschreibung nahe an der Bildkomposition liegt.[148] Komplexere Bilder lassen sich selbst mit detaillierten Prompts und vielen Iterationen kaum reproduzieren.[149]

c) Memorisierung von Audiodaten

Die Forschung zur Memorisierung bei Audiomodellen ist limitiert.[150] *Attacks* auf SUNO[151] haben gezeigt, dass sehr ähnliche Songs hinsichtlich Rhythmus und Kadenz reproduziert werden können, wenn man Textprompts verwendet, die phonetisch sehr nahe am Original sind. Für den Prompt verwendeten die Autoren einer Studie anstatt der ersten Zeile von Eminems «*Lose Yourself*» – «*His palms are sweaty, knees weak, arms are heavy*» – die phonetisch sehr ähnliche Variante: «*His pants are sweaty, cheese weak, cars are heavy*». Das Modell generierte daraufhin einen Song, der dem Original in Rhythmus und Kadenz sehr stark ähnlich war.[152]

d) Memorisierung von Videodaten

Dass Videodaten im Unterschied zu Bilddaten eine zeitliche Dimension haben, erhöht die Komplexität bei der Analyse der Memorisierung. Neben dem statischen Bildinhalt eines Frames können Modelle auch Bewegungsdynamiken erfassen und unter Umständen reproduzieren.[153] Das wirft die Frage auf, ob bereits eine Memorisierung vorliegt, wenn einzelne Frames reproduziert werden, oder ob erst die getreue Reproduktion von Inhalt und zeitlicher Abfolge als Memorisierung anzusehen ist.[154]

Unabhängig von dieser Definitionsfrage neigen Videodiffusionsmodelle insgesamt weniger dazu, Trainingsdaten zu reproduzieren. Der Grund dürfte darin liegen, dass viele Videodiffusionsmodelle auf Bilddiffusionsmodellen basieren,[155] die mit grossen Mengen an Videodaten trainiert werden, und deshalb variantenreichere und kreativere Outputs generieren.[156] Wie bei der Memorisierung von Bilddaten[157] gilt, dass die Wahrscheinlichkeit einer exakten Reproduktion mit zunehmender Grösse und Vielfalt des Trainingsdatensatzes abnimmt.[158]

4. Technische Massnahmen gegen Memorisierung

Da generative KI-Modelle neue Inhalte generieren und nicht bestehende Inhalte wiedergeben sollen, ist die Memorisierung von Trainingsdaten grundsätzlich unerwünscht.[159] Bei der Entwicklung von KI-Modellen werden verschiedene Ansätze verwendet, um der Memorisierung entgegenzuwirken. Allerdings löst keiner dieser Ansätze das Problem vollständig; in der Praxis lässt sich Memorisierung meist nur reduzieren, nicht vollständig ausschliessen.

Ein weit verbreiteter Ansatz ist die sog. De-Duplizierung von Trainingsdaten, mit der erreicht werden soll, dass jedes Quelldatum im Training nur einmal benutzt wird.[160] Damit sinkt die statistische Wahrscheinlichkeit, dass ein Text reproduziert wird.[161]

Ein anderer Ansatz besteht darin, mit sog. *Soft-Prompts*, die vom Modell erzeugt und dem vom Benutzer eingegebenen Prompt vorangestellt werden, das Modellverhalten so zu steuern, dass die Wahrscheinlichkeit der Ausgabe memorisierter Trainingsdaten reduziert wird. Für die Konstruktion solcher *Soft-Prompts* werden die Gewichte eines Modells ein-

145 G. Somepalli et al., Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023, 6048 ff.; Carlini et al. (Fn. 132), 5253 ff.

146 *Stable Diffusion* ist ein Text-zu-Bild-Diffusionsmodell: ‹https://stablediffusionweb.com/de› (abgerufen am 5. Januar 2026).

147 Zum Ganzen Somepalli et al. (Fn. 145), 6053 ff.

148 M. Sag, Copyright Safety for Generative AI, Hous L Rev (2023), 328 ff.

149 Sag (Fn. 148), 328 f.

150 Messina et al., Mitigating Data Replication In Text-to-Audio Generative Diffusion Models Through Anti-Memorization Guidance, arXiv 2025, 1.

151 ‹https://suno.com/› (abgerufen am 5. Januar 2026).

152 Roh et al. (Fn. 133), 1 ff. Das Original und die Reproduktion können unter: ‹https://bobsconfetti.github.io/bobsconfetti/› (abgerufen am 5. Januar 2026) abgerufen werden.

153 A. Rahman/M. V. Perera/V. M. Patel, Frame by Familiar Frame: Understanding Replication in Video Diffusion Models, IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2025, 2768 ff.; Chen/Liu/Liu/Sha/Xu (Fn. 98), 3.

154 Chen/Liu/Liu/Sha/Xu (Fn. 98), 4 ff.; Rahman/Perera/Patel (Fn. 153), 2768 f.

155 Siehe dazu oben, II.2.b) bb) und cc).

156 Rahman/Perera/Patel (Fn. 153), 2771.

157 Siehe dazu oben, II.3.b).

158 Rahman/Perera/Patel (Fn. 153), 2771.

159 Carlini/Tramèr/Wallace/Jagielski/Herbert-Voss/Lee/Roberts/Brown/Song/Erlingsson/Oprea/Raffel (Fn. 131), 2635.

160 Für Textdaten K. Lee et al., Deduplicating Training Data Makes Language Models Better, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics 2022, 8424 ff.; für Bilddaten C. Chen/D. Liu/C. Xu, Towards Memorization-Free Diffusion Models, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2024, 8426; für Audiodaten Bralios et al. (Fn. 133); für Videodaten Chen/Liu/Liu/Sha/Xu (Fn. 98), 8.

161 Siehe dazu Lee et al. (Fn. 160), 8424 ff.; Carlini et al. (Fn. 138), 5 f., 9.

gefroren und in einem *Attack*-Szenario mit Prompts trainiert, welche die Ausgabe memorisierter Inhalte maximieren; auf diese Weise können Schwachstellen identifiziert und defensive *Soft-Prompts* optimiert werden, die das Risiko einer Reproduktion von Trainingsdaten verringern.[162]

Eine weitere Möglichkeit besteht darin, die von Nutzern eingegebenen Prompts im Modell abzuwandeln, bspw. wenn diese auf spezifische Stile von Künstlerinnen und Künstlern oder auf andere markante Merkmale von Trainingsdaten abzielen. Die Abwandlung der Prompts soll verhindern, dass das Modell die charakteristischen Merkmale der Trainingsdaten (bspw. den Stil einer Künstlerin) reproduziert *(Concept Ablation)*.[163] Mit sog. *Machine Unlearning* kann zudem der Einfluss einzelner Datenpunkte im Modell gezielt reduziert werden, um Reproduktionen von Trainingsdaten zu verhindern, ohne das gesamte Modell vollständig neu trainieren zu müssen.[164]

Bei Diffusionsmodellen kann durch *Anti-Memorization Guidance* in die Datengenerierung eingegriffen werden. Während der Schritte für das *Denoising* wird dabei fortlaufend geprüft, ob eine aktuelle Zwischenvorhersage den Trainingsdaten zu ähnlich ist; trifft dies zu, wird von dieser Zwischenvorhersage weggesteuert.[165] Bei Diffusionsmodellen tritt zudem das Problem auf, dass beim *Cross-Attention*-Mechanismus[166] auf einige *Trigger-Tokens* zu grosses Gewicht gelegt wird und deshalb Trainingsdaten reproduziert werden können. Dem kann entgegengewirkt werden, indem man die entsprechenden *Tokens* maskiert.[167]

Die Memorisierung lässt sich auch auf der Ebene der Modellparameter betrachten. Oft sind es einzelne Gruppen von Neuronen eines Netzwerks, die für die Memorisierung verantwortlich sind. Wenn sich diese (Gruppen von) Neuronen identifizieren lassen, können sie deaktiviert werden. Dadurch lassen sich Reproduktionen nachhaltig verhindern, weil diese Massnahme nicht durch «*attacks*» umgangen werden kann.[168]

5. Erkenntnisse

Die Erarbeitung der technischen Grundlagen vermittelt im Wesentlichen drei Erkenntnisse, die für die urheberrechtliche Beurteilung der Nutzung von Werken beim Training von generativen KI-Modellen zentral sind:

Erstens müssen die für das Training verwendeten Text-, Bild-, Audio- und Videodateien in einem ersten Schritt gesammelt und kuratiert sowie in einer Datenbank gespeichert werden. Dieser Vorgang führt zu Vervielfältigungen dieser Dateien in den Datenbanken.

Zweitens funktioniert das Training von generativen KI-Modellen für alle Werkarten (Text, Bild, Ton, Video und *Source Code*) im Grundsatz gleich. Es geht immer darum, dass die Modelle die statistische Verteilung in den Trainingsdaten lernen, um auf dieser Grundlage ähnliche neue Werke generieren zu können.

Drittens hat sich gezeigt, dass die Modelle nicht nur neue und statistisch wahrscheinliche Inhalte generieren können, sondern auch in der Lage sind, exakte und nahezu exakte Kopien von in den Trainingsdaten enthaltenen Werken zu generieren. Dieser Memorisierung kann zwar durch verschiedene Massnahmen entgegengewirkt werden, sie lässt sich aber (bisher) nicht vollständig verhindern. Trainierte Modelle unterscheiden sich zwar von klassischen Datenbanken, bei denen sich alle gespeicherten Werke ohne Weiteres jederzeit abrufen lassen. Die in KI-Modellen gespeicherten Repräsentationen von Werken ermöglichen es Nutzerinnen und Nutzern aber teilweise, dem Modell die beim Training verwendeten Werke exakt oder nahezu exakt durch geeignete Prompts wieder zu entnehmen. Diese Problematik besteht bei allen bekannten Modellen und bei allen Werkarten. Die Memorisierung zeigt, dass die beim Training verwendeten Werke in den Modellen in gewisser Weise «gespeichert» sein können. Ob die Repräsentationen von Werken in den Parametern der trainierten Modelle urheberrechtlich als Vervielfältigungen anzusehen sind, ist allerdings eine Rechtsfrage, die das Urheberrecht autonom entscheiden kann. Die Entscheidung muss aber auf der Grundlage eines hinreichenden Verständnisses der technischen Vorgänge erfolgen.

III. Rechtslage in Europa und den USA

1. Vorbemerkungen

Wie die Nutzung urheberrechtlich geschützter Werke beim Training von KI-Modellen zu beurteilen ist, wird in vielen Rechtsordnungen intensiv diskutiert. Besonders relevant sind aus Schweizer Sicht die gesetzgeberischen Entwicklungen und die Gerichtsverfahren in der EU sowie in Deutschland und Frankreich, im Vereinigten Königreich (UK) und in den USA. Diese werden nachfolgend skizziert.

162 Für Textdaten M. Ozdayi, Controlling the Extraction of Memorized Data from Large Language Models via Prompt-Tuning, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics 2023, 1512 ff.

163 Für Bilddaten N. Kumari et al., Ablating Concepts in Text-to-Image Diffusion Models, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2023, 22691 ff.

164 Für Textdaten L. Bourtoule, Machine Unlearning, IEEE Symposium on Security and Privacy (SP) 2021, 141 ff.; für Bilddaten S. Alberti/K. Hasanaliyev/M. Shah/S. Ermon, Data Unlearning in Diffusion Models, The Thirteenth International Conference on Learning Representations 2025, 1 ff.

165 Für Bilddaten Chen/Liu/Xu (Fn. 160), 8425 ff; für Audiodaten Messina et al. (Fn. 150), 1 ff.

166 Siehe dazu oben, II.2.b) bb).

167 Für Bilddaten J. Ren et al., Unveiling and Mitigating Memorization in Text-to-Image Diffusion Models Through Cross Attention, Computer Vision – ECCV 2024, 340 ff.

168 Für Bilddaten D. Hintersdorf et al., Finding NeMo: Localizing Neurons Responsible for Memorization in Diffusion Models, Proceedings of the 38th International Conference on Neural Information Processing Systems 2024, 1 ff.

2. EU

a) Schranken für Text und Data Mining

Die EU hat 2019 die *Digital Single Market-Richtlinie (DSM-RL)*[169] erlassen, um das Urheberrecht ans digitale Umfeld anzupassen.[170] Die DSM-RL sieht zwei Schranken für Text und Data Mining vor, eine allgemeine Schranke für Text und Data Mining (Art. 4 DSM-RL) und eine spezifische Schranke für Text und Data Mining zum Zweck der wissenschaftlichen Forschung (Art. 3 DSM-RL). Die Mitgliedstaaten der EU mussten diese Vorgaben bis 2021 im nationalen Recht umsetzen.

Der Begriff des Text und Data Mining wird in Art. 2 Nr. 2 der DSM-RL definiert. Nach dieser Definition ist Text und Data Mining «eine Technik für die automatisierte Analyse von Texten und Daten in digitaler Form, mit deren Hilfe Informationen unter anderem – aber nicht ausschliesslich – über Muster, Trends und Korrelationen gewonnen werden können».

Die Schranke für *Text und Data Mining zum Zweck der wissenschaftlichen Forschung* (Art. 3 DSM-RL) stellt Vervielfältigungen von Werken und anderen Schutzgegenständen für Text und Data Mining frei, die durch Forschungsorganisationen und Einrichtungen des Kulturerbes zum Zweck der wissenschaftlichen Forschung vorgenommen werden, sofern diese Organisationen zu den Werken und Schutzgegenständen rechtmässig Zugang haben. Einrichtungen des Kulturerbes sind namentlich öffentlich zugängliche Bibliotheken, Museen oder Archive (Art. 2 Nr. 3 DSM-RL). Als Forschungsorganisationen gelten Hochschulen (inkl. ihrer Bibliotheken), Forschungsinstitute und andere Einrichtungen wie Forschungskliniken, deren vorrangiges Ziel die wissenschaftliche Forschung oder die Lehrtätigkeit ist (Art. 2 Nr. 1 DSM-RL). Die Organisation darf nicht gewinnorientiert sein, oder sie muss sämtliche Gewinne in die Forschung reinvestieren (Art. 2 Nr. 1 Bst. a DSM-RL); alternativ muss sie im Rahmen eines von einem Mitgliedstaat anerkannten Auftrags im öffentlichen Interesse tätig sein (Art. 2 Nr. 1 Bst. b DSM-RL). Nicht auf die Schranke von Art. 3 DSM-RL berufen können sich Forschungsorganisation, die einem Unternehmen, das einen bestimmenden Einfluss auf die Organisation hat, bevorzugten Zugang zu den Forschungsergebnissen gewähren (Art. 2 Nr. 1 DSM-RL).

Die *allgemeine Schranke für Text und Data Mining* (Art. 4 DSM-RL) stellt alle für Text und Data Mining erstellten Vervielfältigungen von Werken und anderen Schutzobjekten frei, unabhängig vom Zweck des Text und Data Mining und von der Art der Organisation, die dieses vornimmt (Art. 4 Abs. 1 DSM-RL). Anders als bei der Schranke für die wissenschaftliche Forschung können die Rechteinhaber hier aber verhindern, dass ihre Werke oder anderen Schutzobjekte für Text und Data Mining genutzt werden, indem sie diese mit einem Nutzungsvorbehalt versehen. Dieser muss ausdrücklich und in angemessener Weise angebracht werden, bei online veröffentlichten Inhalten etwa mit maschinenlesbaren Mitteln (Art. 4 Abs. 3 DSM-RL).

Beim Erlass der DSM-RL war für den Unionsgesetzgeber nicht erkennbar, welche Tragweite die Schranken für das Text und Data Mining aufgrund der technischen Entwicklungen im Bereich von KI erlangen könnten. Klar ist jedenfalls, dass der Gesetzgeber sich beim Erlass der Schranken nicht bewusst war, dass er damit möglicherweise die *Nutzung von Werken für das Training von KI-Modellen* freistellt. Entsprechend war bzw. ist diese zentrale Frage bis heute umstritten – für das EU-Recht ebenso wie für die Umsetzung der unionsrechtlichen Vorgaben im Recht der Mitgliedstaaten, bspw. in Deutschland und Frankreich.[171]

Mit dem Erlass der KI-Verordnung[172] hat der Unionsgesetzgeber die Frage aber wohl nun geklärt: Nach Art. 53 Abs. 1 Bst. c KI-VO müssen die Anbieter von KI-Modellen mit generellem Verwendungszweck *(General Purpose AI, GPAI)* «eine Strategie zur Einhaltung des Urheberrechts der Union und damit zusammenhängender Rechte und insb. zur Ermittlung und Einhaltung eines gemäss Artikel 4 Absatz 3 der Richtlinie (EU) 2019/790 geltend gemachten Rechtsvorbehalts, auch durch modernste Technologien, auf den Weg» bringen. Auch wenn die (urheberrechtliche) Bedeutung dieser Bestimmung weitgehend unklar ist,[173] scheint der Unionsgesetzgeber damit doch unzweideutig zum Ausdruck zu bringen, dass er das Training von KI-Modellen als Text und Data Mining im Sinn der DSM-RL versteht.[174] Bestätigt wird dies durch die Ausführungen in den Erwägungsgründen 105 f. der KI-VO. Dieses Verständnis von Text und Data Mining wird in der Literatur teilweise kri-

169 Richtlinie (EU) 2019/790 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG.

170 Mitteilung der europäischen Kommission über Inhalte im digitalen Binnenmarkt, COM (2012) 789 final, 18. Dezember 2012, ‹https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52012DC0789› (abgerufen am 5. Januar 2026).

171 Siehe dazu für das europäische Recht: Dornis/Stober (Fn. 6), 98 ff.; T.W. Dornis, The Training of Generative AI is not Text and Data Mining, European Intellectual Property Review (E.I.P.R.), 2025, 66 ff.; T. Dreier, § 44b Abs. 1, in: T. Dreier/G. Schulze (Hg.), Urheberrechtsgesetz, 7. Aufl., München 2022; N. Maamar, Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen ZUM 2023, 483; K. de la Durantaye, «Garbage in, garbage out» – Die Regulierung generativer KI durch Urheberrecht, ZUM 2023, 651; F. Hofmann, Zehn Thesen zu Künstlicher Intelligenz (KI) und Urheberrecht, WRP 2024, 13 f.; M. Sanftleben, Compliance of National TDM Rules with International Copyright Law – An Overrated Nonissue?, International Review of Intellectual Property and Competition Law (IIC) 2022, 1478 ff.; M. Leistner/L. Antoine, TDM and AI Training in the European Union – From 'LAION' to Possible Ways Ahead?, GRUR Int. 2025, 1031 ff.; für Deutschland siehe hinten, III.3.a); für Frankreich siehe hinten, III.4.a).

172 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) (Text von Bedeutung für den EWR).

173 Statt Vieler: A. Buick, Copyright and AI training data – transparency to the rescue?, Journal of Intellectual Property Law & Practice, 2025, 190; J.P. Quintais, Generative AI, copyright and the AI Act, Computer Law & Security Review Nr. 56 2025, 6 ff.

174 Ebenso M. Leistner/L. Antoine, TDM and AI Training in the European Union – From 'LAION' to Possible Ways Ahead?, GRUR Int. 2025, 1032; A. Guadamuz, The EU's Artificial Intelligence Act and copyright, The Journal of World Intellectual Property 2025, 216.

tisiert, unter anderem weil die Anwendung der Schranken für Text und Data Mining auf das Training von KI-Modellen nicht mit dem Dreistufentest vereinbar sei.[175] In dieser zentralen Frage besteht damit weiterhin Rechtsunsicherheit. Der EuGH wird die Frage aber voraussichtlich schon bald in der Rechtssache C-250/25, *«Like Company v. Google Ireland»*, klären können.

b) Nutzungsvorbehalt (Opt-out)

Dem in Art. 4 Abs. 3 DSM-Richtlinie vorgesehenen Nutzungsvorbehalt (Opt-out) kommt bei den Schranken für das Text und Data Mining eine zentrale Rolle zu, im EU-Recht ebenso wie im harmonisierten Recht der Mitgliedstaaten.[176] Die praktische Umsetzung dieser Vorgabe ist allerdings (zumindest) bei online zugänglichen Werken mit erheblichen Herausforderungen verbunden. Eine *Best Practice* hat sich noch nicht herausgebildet.[177] Seit einiger Zeit werden verschiedene Ansätze diskutiert und entwickelt, die sich hinsichtlich des technischen Aufwands, der Reichweite und der Durchsetzbarkeit unterscheiden.[178]

Zum einen können Nutzungsvorbehalte in natürlicher Sprache angebracht werden, etwa durch öffentliche Erklärungen der Rechteinhaber.[179] Die deutsche Verwertungsgesellschaft GEMA hat bspw. einen Vorbehalt auf ihrer Website angebracht[180], und der Verlag Penguin Random House sieht einen solchen im Impressum seiner Bücher vor.[181] Möglich ist auch das Anbringen eines Nutzungsvorbehalts in den AGB einer Website.[182] Diese Ansätze können allerdings nicht sicherstellen, dass der Vorbehalt von allen Akteuren zur Kenntnis genommen wird, welche die Werke für Text und Data Mining nutzen wollen. In der Literatur wird zudem teilweise vertreten, dass Erklärungen in natürlicher Sprache den Anforderungen von Art. 4 Abs. 3 DSM-RL nicht genügen, weil die Maschinenlesbarkeit fehle.[183] Das französische *Syndicat national de l'édition* empfiehlt deshalb die Verwendung von sog. «booleschen Funktionen», die in die AGB von Websites aufgenommen werden können.[184] Diese Funktionen haben zwei Zustände: 0 und 1. Die Zahl 0 bedeutet «nein/aus/falsch», die Zahl 1 «ja/ein/wahr». Mit der Klausel «TDM-RESERVATION: 1» soll kommuniziert werden, dass der Rechteinhaber vom Nutzungsvorbehalt Gebrauch macht. Diese Art von Vorbehalt ist allerdings domainspezifisch, er bezieht sich also nur auf Werke, die auf der jeweiligen Website gehostet werden.[185] Werden die Werke von Dritten gespeichert und auf anderen Websites zugänglich gemacht, ist der Nutzungsvorbehalt nicht mehr ersichtlich.

Eine viel diskutierte technische Massnahme ist die Implementierung des *Robot Exclusion Protocol* («REP» oder «robots.txt»), das Webcrawlern kommuniziert, in welchem Mass der Zugriff auf eine Website erwünscht ist.[186] Ein Nachteil dieser Variante ist, dass die Webcrawler richtig benannt, den Rechteinhabern also bekannt sein müssen, damit das REP den Vorbehalt kommunizieren kann. Dies führt dazu, dass sich Rechteinhaber oder Websitebetreiber nicht immer wirksam gegen die Verwendung ihrer Werke wehren können und sich laufend über die neusten Webcrawler informieren müssen.[187] Sinnvoll wäre zudem ein Wechsel von einer identitätsbasierten zu einer verwendungsbasierten Kommunikation des Nutzungsvorbehalts,[188] bei welchem das REP den Webcrawlern nicht nur angibt, dass sie ein bestimmtes Werk nicht verwenden dürfen, sondern differenziert kommunizieren kann, zu welchem Zweck ein Werk genutzt werden darf. Mit diesem Ansatz könnte bspw. den Webcrawlern von Suchmaschinen kommuniziert werden, dass die Listung eines Werks in der Trefferliste der Suchmaschine zulässig, die Nutzung für Text und Data Mining aber ausgeschlossen ist. Da das REP den Zugriff auf Werke nicht verhindert, müssen Webcrawler so programmiert sein, dass sie sich an Nutzungsvorbehalte halten.[189] Um Webcrawler und Bots, die das REP ignorieren, dennoch von der Verwendung von Werken abzuhalten, wurden Dienste entwickelt, die Webcrawler blockieren, bspw. der Dienst von *Cloudflare*.[190] Trotz der skizzierten Einschränkungen wird die Verwendung des REP von der *Internet Engineering Task Force (IETF)* als möglicher Standard für das Anbringen eines Nutzungsvorbehalts vorgeschlagen.[191] Ähnlich funktioniert auch das Protokoll

175 N. Lucchi, Generative AI and Copyright, Training, Creation, Regulation, Brüssel 2025, 43; Dornis/Stober (Fn. 6), 94 ff.; T. W. Dornis/N. Lucchi, Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of the Referral in *Like Company v. Google*, International Review of Intellectual Property and Competition Law 2025, 23; E. Alonso/N. Lucchi, AI and Copyright 'Hallucinations': Does the Text and Data Mining Exception Really Support Generative AI Training?, European Intellectual Property Review (E.I.P.R.) 2025, 521 f.

176 Siehe dazu hinten, für Deutschland: III.3.a); für Frankreich: III.4.a).

177 Lucchi (Fn. 175), 23 f., 55; Hamann, Nutzungsvorbehalte für KI-Training in der Rechtsgeschäftslehre der Maschinenkommunikation, Dogmatische und praktische Schwächen von Art. 4 Abs. 3 DSM-RL und § 44 Abs. 3 UrhG, Zeitschrift für Geistiges Eigentum 2024, 145 ff.; P. Keller/Z. Warso, Defining Best Practices for Opting Out of ML Training, Open Future Policy Brief 2023, 8.

178 Ausführlich zu den wichtigsten Ansätzen EUIPO, The Development of Generative Artificial Intelligence from a Copyright Perspective, Alicante 2025, 15 ff., 164 ff.

179 EUIPO (Fn. 178), 164 ff.

180 ‹www.gema.de/de/w/text-and-data-mining-vorbehalt-ki?%E2%80%BA› (abgerufen am 5. Januar 2026).

181 ‹www.thebookseller.com/news/penguin-random-house-underscores-copyright-protection-in-ai-rebuff› (abgerufen am 5. Januar 2026).

182 EUIPO (Fn. 178), 170 ff.

183 Maamar (Fn. 171), 484; EUIPO (Fn. 178), 181.

184 ‹www.sne.fr/actu/une-clause-type-pour-sopposer-a-la-fouille-de-textes-et-de-donnees-par-les-intelligences-artificielles/› (abgerufen am 5. Januar 2026).

185 EUIPO (Fn. 178), 170; weitergehend, Hamann (Fn. 177), 148, wonach der Vorbehalt nicht einmal für die gesamte Menge der Dokumente einer Website gelten könne.

186 EUIPO (Fn. 178), 173 f.; A. Peukert/C. Castets-Renard, The Copyright Chapter of the EU's Code of Practice for General-Purpose AI Models: A Commentary, International Review of Intellectual Property and Competition Law (IIC) 2025, 1971, 1974.

187 S. Longpre et al., Consent in Crisis: The Rapid Decline of the AI Data Commons, Proceedings of the 38th International Conference on Neural Information Processing Systems, NIPS 2024, 7.

188 Longpre et al. (Fn. 187), 10.

189 Peukert/Castets-Renard (Fn. 186), 1974; ausführlich zu verschiedenen Problemen EUIPO (Fn. 178), 174 ff.

190 ‹https://blog.cloudflare.com/declaring-your-aindependence-block-ai-bots-scrapers-and-crawlers-with-a-single-click/› (abgerufen am 5. Januar 2026).

191 ‹https://datatracker.ietf.org/doc/html/rfc9309› (abgerufen am 5. Januar 2026).

«ai.txt» von *Spawning.ai*, das (anders als das REP) erst unmittelbar vor dem Herunterladen eines Werks interveniert. Dieser Ansatz ist vorteilhaft, wenn ein Datensatz (wie bspw. LAION-5B) Direktlinks auf Werke enthält und der Webcrawler deshalb den Zeitpunkt des Abrufs der Website, in dem das REP aktiv würde, überspringt. *Spawning.ai* bietet noch weitere Tools an, namentlich ein Register, eine Programmierbibliothek und einen Webcrawler-Blocker, die gemeinsam sinnvolle Lösungen für Rechteinhaber und KI-Entwickler bilden sollen.[192]

Weitere technische Massnahmen werden von der *W3C Community Group* vorgeschlagen: Nutzungsvorbehalte können bspw. in einem Dokument auf dem Ursprungsserver der Website oder als Metadaten im HTML-Code angebracht werden. Möglich ist auch ein Einbinden ins *http-Headerfeld*, was dazu führt, dass der Nutzungsvorbehalt als Antwort auf eine http-Anfrage an den Client gesendet wird. Teilweise kann der Vorbehalt auch im Dokument selbst festgehalten werden, bspw. bei Texten im Format *EPUB*.[193] Möglich sind auch sog. «asset-basierte» Massnahmen, die bspw. von *Liccium* angeboten werden. Bei diesem Ansatz wird der Nutzungsvorbehalt in den Metadaten der Werke angebracht und gleichzeitig in einem öffentlich zugänglichen Register vermerkt.[194] Vergleichbare Register könnten auch von Verwertungsgesellschaften oder IP-Behörden erstellt und betrieben werden.[195]

Vor dem Hintergrund der technischen Herausforderungen werden Nutzungsvorbehalte in der Literatur kritisch bewertet. Bisweilen wird nicht nur die praktische Umsetzung, sondern auch ihre grundsätzliche Eignung in Frage gestellt, die Nutzung von Werken für das Training von KI-Modellen zu verhindern,[196] weil die Nutzungsvorbehalte strukturell nicht für die massenhafte und automatisierte Nutzung von Daten beim Training von KI-Modellen konzipiert seien.[197] Namentlich fehle es an skalierbaren, harmonisierten und durchsetzbaren technischen Standards, um Nutzungsvorbehalte entlang komplexer Datenverarbeitungsketten wirksam zu erhalten, was sie im Bereich des Trainings von KI-Modellen weitgehend wirkungslos mache.[198] Ähnliche Befunde haben empirische Arbeiten ergeben, die aufzeigen, dass Nutzungsvorbehalte beim Aufbau grosser Trainingsdatensätze regelmässig verloren gehen, oder nicht zuverlässig berücksichtigt werden können.[199] In diesem Zusammenhang wird teils von einer *«consent crisis»* gesprochen, weil die bestehenden Ansätze den Rechteinhabern keine effektive Kontrolle und den KI-Entwicklern keine Rechtssicherheit bieten.[200]

Als Lösung wird allgemein eine stärkere Standardisierung der technischen Ausgestaltung von Nutzungsvorbehalten gefordert.[201] Nach Einschätzung des EUIPO wird ein kombiniertes Vorgehen erforderlich sein, bei dem verschiedene technische Mittel parallel eingesetzt werden, etwa «robots.txt», Metadaten-Protokolle und öffentlich zugängliche Register.[202]

c) Transparenz

Ob ein bestimmtes Werk für das Training eines KI-Modells verwendet wurde, ist für die Rechteinhaber in der Regel nicht erkennbar. Die Durchsetzung ihrer Urheberrechte und die Prüfung, ob ein allfälliger Nutzungsvorbehalt beachtet wurde, droht deshalb an der fehlenden Transparenz zu scheitern. Der Unionsgesetzgeber hat diese Problematik erkannt und die Anbieter von GPAI-Modellen verpflichtet, eine hinreichend detaillierte Zusammenfassung der für das Training eines KI-Modells verwendeten Inhalte zu erstellen und zu veröffentlichen (Art. 53 Abs. 1 Bst. d KI-VO). Diese Pflicht gilt für alle Anbieter von GPAI-Modellen, unabhängig davon, ob die Modelle proprietär oder unter einer freien oder einer Open-Source-Lizenz veröffentlicht werden.[203] Die Zusammenfassung ist nach einer Vorlage des Büros für KI zu erstellen. Das sog. *summary template* ist in drei Hauptabschnitte gegliedert: Der erste Abschnitt enthält eine allgemeine Beschreibung der Modellkategorie und des Trainingsansatzes. Im zweiten Abschnitt werden zentrale Datensammlungen oder Datenquellen aufgeführt, die für das Training herangezogen wurden, bspw. grosse öffentliche oder private Datenbanken oder Archive. Im dritten Abschnitt sind Angaben zu wesentlichen Aspekten der Datenverarbeitung zu machen, die der Wahrnehmung berechtigter Interessen, insb. der Entfernung illegaler Inhalte, dienen sollen.[204] Aus Sicht des Urheberrechts sind die Angaben im zweiten Abschnitt zentral. Diese sollen es Rechteinhabern ermöglichen, zumindest auf abstrakter Ebene einschätzen zu können, ob ihre Werke Teil der Trainingsdaten gewesen sein könnten.[205] Zugleich trägt die Ausgestaltung des *summary template* dem Schutz von Geschäftsgeheimnissen und vertraulichen Informationen Rechnung, indem keine Offenlegung einzelner Werke, konkreter Trainingsdaten oder technischer Details verlangt wird.[206]

192 EUIPO (Fn. 178), 191 ff.
193 ‹www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/#sec-protocol› (abgerufen am 5. Januar 2026).
194 ‹https://docs.liccium.com/documentation/› (abgerufen am 5. Januar 2026); ausführlich dazu EUIPO (Fn 178), 199 ff.
195 EUIPO (Fn. 178), 259 ff.
196 Lucchi (Fn. 175), 8; Buick (Fn. 173), 187.
197 Lucchi (Fn. 175), 12.
198 Lucchi (Fn. 175), 23 f., 54, 56 f., 61; zur Wichtigkeit von klaren Marktstandards bei der Umsetzung von Nutzungsvorbehalten siehe S. Havlikova, Technical Challenges of Rightsholders' Opt-out From Gen AI Training after Robert Kneschke v. LAION, Journal of Intellectual Property, Information Technology and E-Commerce Law Nr. 16 2025, Rz. 50.
199 Longpre et al. (Fn. 187), 1 f.
200 Longpre et al. (Fn. 187), 1, 9 f.
201 Havlikova (Fn. 198), Rz. 50; P. Keller, Considerations for Opt-Out Compliance Policies By AI Model Developers, Open Future Policy Brief 2024, 11.
202 EUIPO (Fn. 178), 230 ff.
203 Europäische Kommission, Explanatory Notice and Template for the Public Summary of Training Content for General-Purpose AI Models required by Article 53(1)(d) of Regulation (EU) 2024/1689 (AI Act), Communication to the Commission, C(2025) 5235 final, Annex, Brüssel, 24. Juli 2025, Rz. 3; zum sachlichen Anwendungsbereich siehe Art. 2 KI-VO.
204 Europäische Kommission (Fn. 203), Explanatory Notice, Rz. 15 f.
205 Erwägungsgrund 107 KI-VO; Europäische Kommission (Fn. 203), Explanatory Notice, Rz. 8.
206 Erwägungsgrund 107 KI-VO; Europäische Kommission (Fn. 203), Explanatory Notice, Rz. 17 ff.

3. Deutschland

a) Gesetzliche Regelung

Deutschland hat die Vorgaben von Art. 3 und Art. 4 der DSM-RL in den §§ 44b und 60d UrhG im nationalen Recht umgesetzt. Die Bestimmung von § 44b UrhG wurde 2021 durch das Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31. Mai 2021[207] ins UrhG eingefügt, während die Vorschrift des § 60d UrhG schon 2017 – auf Grundlage der InfoSoc-RL[208] – durch das Urheberrechts-Wissensgesellschafts-Gesetz vom 1. September 2017[209] eingeführt wurde. In der Folge wurde § 60d UrhG durch das Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31. Mai 2021 an die Vorgaben von Art. 3 DSM-RL[210] angepasst.

Den Vorgaben der DSM-RL entsprechend sieht das deutsche Recht eine allgemeine Schranke für das Text und Data Mining (§ 44b UrhG) und eine Schranke für das Text und Data Mining für Zwecke der wissenschaftlichen Forschung (§ 60d UrhG) vor. Die für beide Normen geltende Definition von Text und Data Mining weicht in der Formulierung leicht von der Definition in Art. 2 Nr. 2 DSM-RL ab, entspricht ihr aber inhaltlich. Nach deutschem Recht ist Text und Data Mining die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insb. über Muster, Trends und Korrelationen zu gewinnen (§ 44b Abs. 1 UrhG).

Die *allgemeine Schranke für Text und Data Mining* stellt Vervielfältigungen von rechtmässig zugänglichen Werken für das Text und Data Mining frei (§ 44b Abs. 2 Satz 1 UrhG). Diese Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr gebraucht werden (§ 44b Abs. 2 Satz 2 UrhG). Die freigestellten Nutzungen sind allerdings nur zulässig, wenn der Rechteinhaber sich diese nicht vorbehalten hat (§ 44b Abs. 3 Satz 1 UrhG). Anders als in Art. 4 Abs. 3 DSM-RL, wonach die Rechteinhaber die Werke «ausdrücklich und in angemessener Weise» mit einem Nutzungsvorbehalt versehen müssen,[211] enthält das deutsche Recht keine allgemeinen Vorgaben für das Anbringen eines Nutzungsvorbehalts. Vorgaben bestehen nur für online zugängliche Werke, bei denen der Nutzungsvorbehalt nur wirksam ist, wenn er in maschinenlesbarer Form erfolgt (§ 44b Abs. 3 Satz 2 UrhG).

Die *Schranke für Text und Data Mining für Zwecke der wissenschaftlichen Forschung* übernimmt die Vorgaben von Art. 3 DSM-RL und präzisiert sie in verschiedener Hinsicht. So wird der Begriff der Forschungsorganisation näher definiert. Dieser umfasst nach deutschem Recht Hochschulen, Forschungsinstitute und sonstige Einrichtungen, die wissenschaftliche Forschung betreiben, wenn sie (i) nicht kommerzielle Zwecke verfolgen, (ii) sämtliche Gewinne in die Forschung reinvestieren und (iii) im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sind (§ 60d Abs. 2 Satz 1 UrhG). Neben diesen Forschungsorganisationen können sich auch einzelne Forschende auf die Schranke berufen, wenn sie nicht kommerzielle Zwecke verfolgen (§ 60d Abs. 3 Nr. 2 UrhG). Nicht auf die Schranke berufen können sich hingegen Forschungsorganisationen, die mit einem privaten Unternehmen zusammenarbeiten, das einen bestimmenden Einfluss auf die Forschungsorganisation und einen bevorzugten Zugang zu den Ergebnissen der wissenschaftlichen Forschung hat (sog. *Public Privat Partnerships*, siehe: § 60d Abs. 2 Satz 3 UrhG). Mit dieser Bestimmung soll verhindert werden, dass die im Vergleich zu § 44b UrhG grosszügigere Schranke des § 60d UrhG rein privatwirtschaftlich tätige Unternehmen mitbegünstigt.

Wie der Begriff der Forschungsorganisation wird auch der in der DSM-RL verwendete Begriff der Einrichtung des Kulturerbes näher definiert. Er umfasst nach deutschem Recht öffentlich zugängliche Bibliotheken und Museen sowie Archive und Einrichtungen im Bereich des Film- oder Tonerbes (§ 60d Abs. 3 Nr. 1 UrhG). Wie Art. 3 DSM-RL erlaubt auch das deutsche Recht die Aufbewahrung der Vervielfältigungen nach Abschluss des Text und Data Mining, sofern dies ausschliesslich für Zwecke der wissenschaftlichen Forschung erfolgt, einschliesslich für die Überprüfbarkeit von Forschungsergebnissen (§ 60d Abs. 5 UrhG). Darüber hinaus erlaubt es das deutsche Recht, die im Rahmen der Schranke erstellten Vervielfältigungen Dritten zugänglich zu machen. Der Kreis der Dritten ist allerdings beschränkt; zulässig ist nur das Zugänglichmachen in einem bestimmt abgegrenzten Kreis von Personen für die gemeinsame wissenschaftliche Forschung und an einzelne Dritte für die Überprüfung der Qualität der wissenschaftlichen Forschung (§ 60d Abs. 4 UrhG).

Ob die Nutzung von Werken für das Training von KI-Modellen als Text und Data Mining qualifiziert werden kann, ist im deutschen Recht umstritten. Die Frage wird in der rechtswissenschaftlichen Literatur nunmehr überwiegend bejaht,[212] teils aber auch verneint.[213] Für die zweite Rechtsauffassung spricht, dass die Schranken für das Text und Data Mining vom europäischen Gesetzgeber zweifellos nicht eingeführt wurde, um das Training von KI-Modellen zu erfassen. Allerdings hat der deutsche Gesetzgeber in der

207 BGBl. I 1204.

208 RL 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft (ABl. L 167/10).

209 BGBl. 2017 I 3346.

210 RL 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG (ABl. L 130/92).

211 Siehe dazu vorne, III.2.b).

212 So etwa M. Becker, Generative KI und Deepfakes in der KI-VO, Für eine Positivkennzeichnung authentischer Inhalte, CR 2024, 357 f.; Maamar (Fn. 171), 483 ff.; J. Siglmüller/D. Gassner, Softwareentwicklung durch Open-Source-trainierte KI – Schutz und Haftung, RDi 2023, 127; so wohl auch S. Grimm/L. Münster, Training von KI in der EU: Fragen zum Nutzungsvorbehalt beim Text und Data Mining sowie zur Beweislast, GRUR-Prax 2025, 443 ff.

213 So insb. (noch) H. Schack, Auslesen von Webseiten zu KI-Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda, NJW 2024, 114; Dornis/Stober (Fn. 6), 72 f.; T. W. Dornis, Generatives KI-Training und der TDM-Trugschluss, GRUR 2024, 1676 ff.

Begründung zum Regierungsentwurf des Umsetzungsgesetzes[214] ausdrücklich festgehalten, dass Schranken für das Text und Data Mining «das maschinelle Lernen als Basis-Technologie für Künstliche Intelligenz» erleichtern und fördern solle. Aus dieser Aussage wird überwiegend gefolgert, dass die Nutzung von Werken für das Training von KI-Modellen als Text und Data Mining zu qualifizieren sei.[215] Mittlerweile hat der europäische Gesetzgeber die Frage, ob das KI-Training Text und Data Mining ist, mit Erlass der KI-Verordnung geklärt, indem er in Art. 53 Abs. 1 Bst. c KI-VO ausdrücklich auf Art. 4 Abs. 3 DSM-RL verweist.[216] Damit ist klar, dass die Vorgänge beim Training von KI-Modellen nach europäischem und damit auch nach deutschem Recht[217] als Text und Data Mining zu verstehen sind.

b) Gerichtsverfahren

In Deutschland sind zwei erstinstanzliche Entscheide zur urheberrechtlichen Zulässigkeit des Trainings generativer KI-Modelle mit urheberrechtlich geschützten Werken ergangen, eine weitere Klage ist vor erster Instanz hängig.[218] Der erste Entscheid wurde bereits von der zweiten Instanz bestätigt, eine höchstrichterliche Klärung steht aber noch aus.

Die erste europäische Gerichtsentscheidung zur Frage, ob die Nutzung von Werken beim Training von KI-Modellen als Text und Data Mining zu qualifizieren ist, erging beim *Landgericht Hamburg* mit Urteil vom 27. September 2024.[219] Der Kläger ist Urheber eines Fotos, das zumindest nach § 72 Abs. 1 UrhG Schutz als Lichtbild geniesst. Dieses Foto wurde auf der Website eines Stockfoto-Anbieters öffentlich zugänglich gemacht. Der Beklagte, der Verein «Laion e.V.», lud das Bild von dort herunter und verwendete es, um einen KI-Trainingsdatensatz zu erstellen. Der Kläger machte geltend, dass die damit verbundene Vervielfältigung des Fotos seine Urheberrechte verletze. Dabei brachte er unter anderem vor, dass auf einer Unterseite der Website des Stockfoto-Anbieters ein Nutzungsvorbehalt gegen das *Webscraping* der veröffentlichten Inhalte in natürlicher (englischer) Sprache erklärt worden sei. Der Beklagte wendete dagegen ein, das Foto befinde sich gar nicht im Trainingsdatensatz; dieser enthalte nur die URL zur Website des Stockfoto-Anbieters und eine Bildbeschreibung sowie andere Metadaten. Zweck der Vervielfältigung sei lediglich der Abgleich des Fotos mit einer von einem Drittanbieter angefertigten Bildbeschreibung gewesen.

Das *Landgericht Hamburg* hat die Klage des Fotografen abgewiesen. Zur Begründung führte es an, dass die Nutzung durch den Beklagten zwar nicht als vorübergehende Vervielfältigung (§ 44a UrhG) zulässig sei, aber eine Freistellung über § 60d UrhG in Betracht komme, weil der technische Vorgang, bei dem ein Bild mit einer Bildbeschreibung automatisiert abgeglichen werde, als Text und Data Mining i.S.v. § 44b Abs. 1 UrhG zu qualifizieren sei.[220] Dieses Text und Data Mining betreibe der Beklagte zum Zweck der wissenschaftlichen Forschung, weil er die Trainingsdatensätze unentgeltlich zur Verfügung stelle.[221] Der Beklagte könne sich damit auf die Schranke von § 60d UrhG berufen.[222] Da diese die Möglichkeit eines Nutzungsvorbehalts nicht vorsehe, komme es nicht darauf an, ob der erklärte Nutzungsvorbehalt die Voraussetzungen nach § 44b Abs. 3 UrhG erfülle, insb. ob er «maschinenlesbar» sei. Das Gericht liess allerdings erkennen, dass es die Maschinenlesbarkeit wohl auch dann als gegeben erachten würde, wenn ein Nutzungsvorbehalt in natürlicher Sprache erklärt wird, weil sonst ein «gewisser Wertungswiderspruch» entstehe.[223] Diesen sieht es darin, dass KI-Anbietern einerseits durch die Schranke des § 44b UrhG die Entwicklung immer leistungsfähigerer KI-Modelle ermöglicht werde, von ihnen aber andererseits die Anwendung dieser KI-Modelle nicht verlangt würde, um einen Nutzungsvorbehalt in natürlicher Sprache auszulesen.[224]

Eine Berufung gegen das landgerichtliche Urteil wurde vom *Hanseatischen Oberlandesgericht* mit Urteil vom 10. Dezember 2025 als unbegründet zurückgewiesen.[225] Das Oberlandesgericht hat die Argumentation des Landgerichts im Wesentlichen gestützt, aber angemerkt, dass der Nutzungsvorbehalt in natürlicher Sprache nicht die in § 44b Abs. 3 S. 2 UrhG vorgesehene Form erfülle.[226] Ein Nutzungsvorbehalt sei bei online zugänglichen Werken nur wirksam, wenn er in maschinenlesbarer Form erfolge.[227] Der *Senat* hat die Revision zum Bundesgerichtshof zugelassen, weshalb die Entscheidung noch nicht rechtskräftig ist.

214 Entwurf der Bundesregierung eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 9. März 2021, RegE BT-Drs. 19/27426, 60.

215 Siehe dazu LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», ZUM 2025, 67 ff.; M. Baumann, Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld, NJW 2023, 3675 f.; D. Bomhard, KI-Training mit fremden Daten – IP-rechtliche Herausforderungen rund um § 44 b UrhG, DSRITB 2023, 259 f.; de la Durantaye (Fn. 171), 651; L. Käde, Training generativer KI-Modelle ist (auch) Text- und Data-Mining – Anwendbarkeit der TDM-Schranke des § 44b UrhG, KIR 2024, 162 ff.; Maamar (Fn. 171), 481 ff., 483; a.A. Dornis (Fn. 213), 1678 ff.; Schack (Fn. 213), 114 f.; offenlassend, aber zweifelnd J. Pukas, KI-Trainingsdaten und erweiterte kollektive Lizenzen – Generierung von Werken als KI-Trainingsdaten auf Basis erweiterter kollektiver Lizenzen, GRUR 2023, 615.

216 Siehe dazu vorne III.2.a.

217 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», ZUM 2025, 67 ff.; Baumann (Fn. 215), 3674; Becker (Fn. 212), 357 f.; Käde (Fn. 215), 162 ff.; Maamar (Fn. 171), 483 ff.; Siglmüller/Gassner (Fn. 212), 127.

218 Klage der GEMA vor dem Landgericht München I gegen die Suno Inc., Aktenzeichen unbekannt.

219 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», GRUR 2024, 1710 ff.

220 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», GRUR 2024, 1712 ff., N 39 ff.

221 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», GRUR 2024, 1715, N 77.

222 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», GRUR 2024, 1715, N 75.

223 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», GRUR 2024, 1715, N 71.

224 LG Hamburg vom 29. Oktober 2024, 310 O 227/23, «LAION», GRUR 2024, 1714, N 65 ff.

225 HansOLG vom 10. Dezember 2025, 310 O 227/23, «LAION», GRUR-RS 2025, 33887.

226 HansOLG vom 10. Dezember 2025, 310 O 227/23, «LAION», GRUR-RS 2025, 33887, N 80 ff.

227 HansOLG vom 10. Dezember 2025, 310 O 227/23, «LAION», GRUR-RS 2025, 33887, N 80.

Den zweiten Entscheid hat das *Landgericht München* I am 11. November 2025 gefällt.[228] Die Klägerin, die Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte (GEMA), klagte gegen OpenAI und machte geltend, dass das Unternehmen die Urheberrechte an neun Liedtexten verletzt habe, indem es diese für das Training seiner Sprachmodelle GPT-4 und GPT-4o verwendete. Dies zeige sich darin, dass bereits die Eingabe einfacher Prompts dazu führe, dass der auf diesen Sprachmodellen basierende Chatbot «ChatGPT» die fraglichen Liedtexte nahezu wortgleich reproduziere. OpenAI hat dem entgegengehalten, die Ausgaben beruhten auf einer «sequenziell-analytischen, iterativ-probabilistischen Synthese» und stellten eine eigenständige, originäre maschinelle Generierung eigener Art dar.

Das *Landgericht München I* hiess die Klage gut und untersagte OpenAI, die streitgegenständlichen Liedtexte ganz oder teilweise in grossen Sprachmodellen zu vervielfältigen und diese ganz oder teilweise in veränderter oder unveränderter Form in den Ausgaben (Output) des Chatbots öffentlich zugänglich zu machen und/oder zu vervielfältigen.[229] Das Gericht hielt fest, dass die Memorisierung der Liedtexte in den Sprachmodellen als Vervielfältigung im Sinn von § 16 UrhG zu qualifizieren sei.[230] Diese sei nicht durch die Schranken für Text und Data Mining (§§ 44b, 60d UrhG) freigestellt, weil die Memorisierungen im Sprachmodell über ein Text und Data Mining hinausgingen und daher kein solches darstellten.[231] In seiner Begründung bezog sich das *Landgericht München I* auf den Entscheid des *Landgerichts Hamburg* und führte aus, dass sein Entscheid nicht im Widerspruch zu jenem Entscheid stehe, weil dort ein anderer Sachverhalt zu beurteilen gewesen sei, nämlich die Erstellung von Trainingsdatensätzen als Vorstufe des Trainings.[232] Dieser Vorgang sei (auch) aus Sicht des *Landgerichts München* als Text und Data Mining anzusehen. Die Vervielfältigung von Werken (hier: Liedtexte) durch die Generierung eines entsprechenden Outputs sei nicht Gegenstand des Hamburger Verfahrens gewesen. Im vorliegenden Fall würden die Liedtexte nicht nur als Trainingsdaten ausgewertet, sondern vollständig in die Parameter des Modells übernommen, was erheblich in die Verwertungsinteressen der Urheber eingreife. Die Bestimmung von Art. 4 Abs. 1 DSM-RL, auf der die Schranke von § 44b UrhG beruht, erfordere «zum Zwecke des Text und Data Mining vorgenommene Vervielfältigungen».[233] Die hier infrage stehenden Vervielfältigungen der Texte im Modell dienen nach Auffassung des Gerichts nicht der weiteren Datenanalyse, weshalb eine Freistellung über die Schranke für Text und Data Mining ausgeschlossen sei.[234]

4. Frankreich

a) Gesetzliche Regelung

Frankreich hatte bereits 2016 mit dem Gesetz zur «*République Numérique*»[235] zwei Schranken für Text und Data Mining ins französische Urheberrecht eingeführt, das im *Code de la propriété intellectuelle CPI* geregelt ist.[236] Diese orientierten sich stark an der damaligen Regelung des britischen Rechts in s29A CDPA.[237] Diese Schranken stellten das Vervielfältigen von Werken oder Datenbanken zum Zweck von Text und Data Mining zu wissenschaftlichen Zwecken frei.[238] Vorausgesetzt wurde, dass ein rechtmässiger Zugang zur Quelle bestand und keine kommerziellen Zwecke verfolgt wurden.[239] Die Ausnahme war auf Texte und wissenschaftliche Schriften beschränkt, weshalb unklar war, ob beim Text und Data Mining bspw. auch Bilder oder Lieder verwendet werden durften.[240]

Frankreich hat die Vorgaben von Art. 3 und Art. 4 der DSM-RL mit der Verordnung Nr. 2021-1518 vom 24. November 2021[241] in Art. L122-5-3 CPI im nationalen Recht umgesetzt. Dabei wurde in Art. L122-5-3 I CPI die Definition des Text und Data Mining aus Art. 2 Nr. 2 DSM-RL übernommen.[242] Von dieser Definition werden im französischen Recht alle Verfahren erfasst, die dazu dienen, Informationen aus Daten zu gewinnen.[243] In Art. L122-5-3 III CPI wird die in Art. 4 DSM-RL vorgesehene allgemeine Schranke für Text und Data Mining umgesetzt, in Art. L122-5-3 II CPI die in Art. 3 DSM-RL vorgesehene Schranke für die wissenschaftliche Forschung.

228 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1917 ff.
229 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1920 ff., N 124 ff.
230 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1923, N 165 ff.
231 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1925 ff., N 192 ff., 201 ff.
232 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1923, N 166.
233 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1926 f., N 206.
234 LG München I vom 11. November 2025, «GEMA ./. OpenAI», GRUR 2025, 1926 f., N 206.
235 LOI nº 2016–1321 du 7 octobre 2016 pour une République numérique.
236 AltCPI Art. L122-5 nº 10 lautete: «Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites; ces fichiers constituent des données de la recherche.»; für Datenbanken siehe altCPI Art. L342-3 nº 5.
237 Siehe dazu M. Vivant/J.-M. Bruguière., Droit d'auter et droits voisins, 5. Aufl., Paris 2024, 805.
238 AltCPI Art. L122-5 nº 10 verlangte «pour les besoins de la recherche publique», altCPI Art. L342-3 nº 5 sprach von «dans un cadre de recherche».
239 AltCPI Art. L122-5 nº 10; AltCPIArt. L342-3 nº 5.
240 Vivant/Bruguière (Fn. 237), 809; C. Bernault, Les nouvelles exceptions au droit d'auteur, Juris art etc. 2017, nº 47, 22 ff., pdf print, 2.
241 Ordonnance nº 2021–1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.
242 Siehe dazu vorne, III.2.a).
243 N. Binctin, Droit de la propriété intellectuelle, 8. Aufl., Paris 2024, Rz. 180.

Die Schranke für *Text und Data Mining zum Zweck der wissenschaftlichen Forschung* übernimmt die Vorgaben der DSM-RL und präzisiert sie in verschiedener Hinsicht. Wie in Art. 3 DSM-RL werden Forschungsorganisationen allgemein als privilegierte Organisationen genannt, die Einrichtungen des Kulturerbes werden aber spezifisch aufgeführt; auf die Schranke berufen können sich öffentlich zugängliche Bibliotheken, Museen, Archivdienste und Einrichtungen, die das filmische, audiovisuelle oder akustische Erbe wahren (Art. L122-5-3 II al. 1 CPI). Nach französischem Recht können die privilegierten Organisationen Private mit dem Durchführen von Text und Data Mining beauftragen, wenn die Zusammenarbeit keinen Gewinnzweck verfolgt (Art. L122-5-3 II al. 1 CPI).[244] Diese Möglichkeit entfällt allerdings, wenn ein Unternehmen, das Aktionär oder Gesellschafter der Organisation ist, die das Text und Data Mining durchführt, einen privilegierten Zugang zu den Ergebnissen hat (Art. L122-5-3 II al. 2 CPI). Mit dieser Einschränkung soll den Vereinbarungen zwischen Google und grossen Bibliotheken in Frankreich und den zahlreichen Forschungskooperationen zwischen Big Tech-Unternehmen und öffentlichen Forschungseinrichtungen Rechnung getragen werden.[245] Der französische Gesetzgeber wollte eine Freistellung von Text und Data Mining im Rahmen solcher Kooperationen ausdrücklich ausschliessen, weil ihm das Gleichgewicht zwischen den Forschungsinteressen und den Interessen der Inhaber von Urheberrechten gefährdet schien.[246] Wie Art. 3 DSM-RL erlaubt auch die französische Schranke die Aufbewahrung der Vervielfältigungen nach Abschluss des Text und Data Mining, sofern dies ausschliesslich für Zwecke der wissenschaftlichen Forschung erfolgt, einschliesslich für die Überprüfbarkeit von Forschungsergebnissen (Art. L122-5-3 II al. 3 CPI).

Die *allgemeine Schranke für Text und Data Mining* übernimmt im Wesentlichen die Vorgaben von Art. 4 DSM-RL. Die Regelung stellt Vervielfältigungen von Werken für das Text und Data Mining durch beliebige Organisationen zu beliebigen Zwecken frei, sofern der Urheber keinen Nutzungsvorbehalt angebracht hat (Art. L122-5-3 III CPI). Der Nutzungsvorbehalt muss nicht begründet werden und kann in jeder Form erfolgen.[247] Bei online zugänglichen Inhalten kann er durch maschinenlesbare Verfahren kommuniziert werden, bspw. in den Metadaten oder in den Nutzungsbedingungen.[248] Der Nutzungsvorbehalt soll dem Schutz der berechtigten Interessen der Rechteinhaber im Sinn des Dreistufentests dienen.[249] In der Lehre wird allerdings die grosse Rechtsunsicherheit bei der Umsetzung und Durchsetzung kritisiert.[250] Nach Abschluss des Text und Data Mining müssen die Vervielfältigungen vernichtet werden.[251]

Ob die allgemeine Schranke für das Text und Data Mining (Art. L122-5-3 III CPI) auf das *Training von KI-Modellen* anwendbar ist, ist umstritten.[252] Zwar stellt die Bestimmung die Nutzung von Werken für das Text und Data Mining unabhängig von dessen Zweck frei. Nach einem Teil der französischen Lehre ist aber fraglich, ob damit auch die Entwicklung von KI-Modellen gemeint sein kann, zumal bei der Einführung der Bestimmung nicht absehbar gewesen sei, dass diese Modelle genutzt werden könnten, um Inhalte zu erstellen, welche die für das Training verwendeten Werke direkt konkurrenzieren.[253] Andere Autoren machen hingegen geltend, dass der EU-Gesetzgeber mit den Schranken für das Text und Data Mining die Entwicklung von KI-Modellen in Europa gegenüber der Konkurrenz in den USA und in China fördern wollte.[254] Der Begriff des Text und Data Mining erfasst nach dieser Auffassung alle Verfahren, die dazu dienen, Informationen aus Daten zu gewinnen, auch das Training von KI-Systemen.[255] Die französischen Verwertungsgesellschaften Société des auteurs, compositeurs et éditeurs de musique (SACEM) und Société des auteurs et compositeurs dramatiques (SACD)scheinen jedenfalls davon auszugehen, dass die Schranken für das Text und Data Mining auch die Verwendung von Werken für das Training von KI-Modellen erfasst, zumal sie nach einem Treffen ankündigten, den Nutzungsvorbehalt für die Werke in ihrem Repertoire geltend zu machen, um so eine unentgeltliche Nutzung der Werke für das KI-Training zu verhindern.[256]

Beide Schranken für Text und Data Mining setzen einen rechtmässigen Zugang zu den Werken voraus. Diese Voraussetzung ist nach der Lehre erfüllt, wenn die Zustimmung des Rechteinhabers vorliegt. Diese wird angenommen, wenn das Werk erworben, ein Abonnement für eine Datenbank abgeschlossen oder das Werk Open Access zu-

244 Siehe dazu Vivant/Bruguière (Fn. 237), 808; N. Binctin, TDM a challenge for artificial intelligence, RIDA 2019, nº 262, 5 ff., 21; Ouvrir la science, La fouille de textes et de données à des fins de recherche: une pratique confirmée et désormais opérationnelle en droit français, ‹www.ouvrirlascience.fr/la-fouille-de-textes-et-de-donnees-a-des-fins-de-recherche-une-pratique-confirmee-et-desormais-operationnelle-en-droit-francais/› (abgerufen am 5. Januar 2026).

245 Vivant/Bruguière (Fn. 237), 808; Binctin (Fn. 244), RIDA, 22.

246 Binctin (Fn. 244), RIDA, 22.

247 Décret nº 2022-928 du 23 juin 2022 portant modification du code de la propriété intellectuelle et complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

248 Décret nº 2022-928 (Fn. 247); Vivant/Bruguière (Fn. 237), 811; Binctin (Fn. 243), Rz. 180.

249 A. Deroudille/F. Fatah, Intelligence artificielle: environnement réglementaire en droit français et en droit de l'Union européenne – Pratiques et prospectives, Revue de l'Union européenne 2025, 326 ff., pdf print, 17.

250 Deroudille/Fatah (Fn. 249), pdf print, 17; S. Le Cam/F. Maupomé, IA génératives de contenus: pour une obligation de transparence des bases de données!, Dalloz Actualité, pdf print, 5 f., ‹www.dalloz-actualite.fr/node/ia-generatives-de-contenus-pour-une-obligation-de-transparence-des-bases-de-donnees› (abgerufen am 5. Januar 2026).

251 Art. L122-5-3 III Satz 2 CPI «Les copies et reproductions sont stockées avec un niveau de sécurité approprié puis détruites à l'issue de la fouille de textes et de données».

252 Dafür Binctin (Fn. 243), Rz. 180; Ders. (Fn. 244), RIDA, 10 f.; Deroudille/Fatah (Fn. 249), pdf print, 17; zweifelnd: Le Cam/Maupomé (Fn. 250), pdf print, 2.

253 Siehe dazu Le Cam/Maupomé (Fn. 250), pdf print, 2.

254 Vivant/Bruguière (Fn. 237), 810 f.

255 Binctin (Fn. 243), Rz. 180.

256 Pour une intelligence artificielle vertueuse, transparente et équitable, la Sacem exerce son droit d'opt-out, ‹https://societe.sacem.fr/actualites/notre-societe/pour-une-intelligence-artificielle-vertueuse-transparente-et-equitable-la-sacem-exerce-son-droit› (abgerufen am 5. Januar 2026).

gänglich gemacht wurde.[257] Dasselbe soll gelten, wenn Daten von sozialen Netzwerken genutzt und die Nutzungsbedingungen eingehalten werden.[258] Nicht rechtmässig ist der Zugang, wenn er unter Verstoss gegen Bestimmungen des Straf- oder Wettbewerbsrechts erfolgt, etwa beim Eindringen in eine fremde Datenbank.[259]

b) Gerichtsverfahren

Der *Syndicat national de l'édition (SNE)*, die *Société des gens de lettres (SGDL)* und der *Syndicat national des auteurs et des compositeurs (SNAC)* haben im März 2025 vor dem Pariser Gerichtshof gegen Meta geklagt. Sie werfen dem Konzern vor, urheberrechtlich geschützte Werke, darunter viele Bücher, ohne Genehmigung zum Training generativer KI-Modelle wie Llama genutzt zu haben.[260] Die Klageschrift ist zwar nicht öffentlich zugänglich, die Verwertungsgesellschaften haben aber in einer Pressemitteilung festgehalten, dass sie Urheberrechtsverletzungen sowie «*parasitisme*», also einen Verstoss gegen das französische Lauterkeitsrecht, geltend machen.[261] Mit der Klage soll der Verbreitung von KI-generierten Büchern *(«faux livres»)* entgegengewirkt werden.[262] Der Fall ist derzeit hängig; bisher wurden keine Verhandlungstermine bekannt gegeben.

c) Mögliche Weiterentwicklung

Am 12. September 2023 haben acht Abgeordnete der *Assemblée nationale* einen Gesetzesentwurf eingereicht, der verschiedene Änderungen der bestehenden Regelung zur Nutzung von Werken durch KI-Systeme vorsieht.[263] Begründet wurde der Vorstoss damit, dass die exponentielle Entwicklung im Bereich der KI eine Bedrohung für das kreative Schaffen sei.[264] Die Änderungen sollten den Urheberinnen und Urhebern bei Werknutzungen im Zusammenhang mit KI einen besseren Schutz gewähren und eine Vergütung sicherstellen.

Kern des Vorstosses ist eine Bestimmung, die ausdrücklich festhält, dass die Nutzung von urheberrechtlich geschützten Werken für das Training von KI-Modellen nur mit Zustimmung der Rechteinhaber zulässig ist.[265] Zudem soll vorgesehen werden, dass Werke, die mit einem KI-System erstellt wurden, gekennzeichnet und die Urheber der dafür verwendeten Werke genannt werden müssen.[266] Darüber hinaus soll eine Vergütungspflicht eingeführt werden, die greift, wenn Werke durch ein KI-System auf der Grundlage von Werken erzeugt werden, deren Ursprung ungewiss ist. Diese Vergütung soll von der zuständigen Verwertungsgesellschaft erhoben und vom Unternehmen bezahlt werden, welches das entsprechende KI-System betreibt.[267]

Der Vorschlag wurde allerdings als unrealistisch und als aus technischen Gründen nicht umsetzbar kritisiert.[268] Namentlich sei die Rückverfolgbarkeit der beim Training von KI-Modellen verwendeten Werke mangels Transparenz schwierig.[269] Zudem sei es wenig sinnvoll, in einzelnen Mitgliedstaaten der EU isolierte Regelungen zu erlassen, weil dies zu einer Zersplitterung der Rechtsordnungen und zu Rechtsunsicherheit führe.[270] Der Vorschlag wurde bisher nicht weiterverfolgt, und es haben, soweit ersichtlich, keine parlamentarischen Beratungen stattgefunden.

5. Vereinigtes Königreich

a) Gesetzliche Regelung

Das Vereinigte Königreich (UK) hat bereits 2014 mit Section 29A des *Copyright, Designs and Patents Act 1988 (CDPA)* eine Schranke für Vervielfältigungen für nichtkommerzielle computergestützte Analysen *(computational analysis)* von Werken eingeführt. Diese Bestimmung lehnt sich an die allgemeine Schranke für nichtkommerzielle Forschung an (s29 CDPA). Nach dem Austritt aus der EU wurden im UK keine Bemühungen mehr unternommen, die Vorgaben der DSM-RL ins nationale Urheberrecht umzusetzen.[271]

Die Schranke von s29A CDPA stellt Vervielfältigungen von Werken für die computergestützte Analyse der in diesen Werken enthaltenen Informationen *(«of anything recorded in the work»)* frei, wenn dies allein *für den Zweck der nichtkommerziellen Forschung* erfolgt. Voraussetzung ist ein rechtmässiger Zugang zu den Werken (s29A (1)(a) CDPA). Zudem muss die Quelle angegeben werden, wenn dies nicht aus praktischen Gründen unmöglich ist (s29A (1)(b) CDPA). Die im Rahmen dieser Schranke erstellten Vervielfältigungen eines Werks dürfen nicht zu anderen Zwecken verwendet (s29A (2)(b) CDPA) und nicht an andere weitergegeben

257 Vivant/Bruguière (Fn. 237), 809; Ouvrir la science, La fouille de textes et de données à des fins de recherche: une pratique confirmée et désormais opérationnelle en droit français, ‹www.ouvrirlascience.fr/la-fouille-de-textes-et-de-donnees-a-des-fins-de-recherche-une-pratique-confirmee-et-desormais-operationnelle-en-droit-francais/› (abgerufen am 5. Januar 2026).

258 Vivant/Bruguière (Fn. 237), 809.

259 Vivant/Bruguière (Fn. 237), 809.

260 Unis, auteurs et éditeurs assignent Meta pour imposer le respect du droit d'auteur aux développeurs d'outils d'intelligence artificielle générative, 22.5.2025, ‹www.sne.fr/actu/unis-auteurs-et-editeurs-assignent-meta-pour-imposer-le-respect-du-droit-dauteur-aux-developpeurs-doutils-dintelligence-artificielle-generative/› (abgerufen am 5. Januar 2026).

261 Unis, auteurs et éditeurs assignent Meta pour imposer le respect du droit d'auteur aux développeurs d'outils d'intelligence artificielle générative (Fn. 260).

262 Unis, auteurs et éditeurs assignent Meta pour imposer le respect du droit d'auteur aux développeurs d'outils d'intelligence artificielle générative (Fn. 260).

263 Assemblé Nationale, Proposition de loi, visant à encadrer l'intelligence artificielle par le droit d'auteur, 12. septembre 2023, nº 1630.

264 Exposé de motifs, Proposition de loi, nº 1630.

265 Art. 1 Proposition de loi, nº 1630.

266 Art. 3 Proposition de loi, nº 1630.

267 Art. 4 Proposition de loi, nº 1630.

268 E. Migliore, Proposition de loi visant à encadrer l'intelligence artificielle par le droit d'auteur: une initiative louable mais perfectible, Dalloz actualité, 4 octobre 2023, pdf print, 2.

269 Migliore (Fn. 268), pdf print, 3.

270 Migliore (Fn. 268), pdf print, 3.

271 «[…] United Kingdom will not be required to implement the Directive, and the Government has no plans to do so», Copyright: EU Action, ‹https://questions-statements.parliament.uk/written-questions/detail/2020-01-16/4371› (abgerufen am 5. Januar 2026).

(s29A (2)(a) CDPA), namentlich nicht zum Kauf oder zur Miete angeboten, werden (s29A (3) i.V.m. (4) CDPA). Vertragsbestimmungen, welche das Erstellen einer Kopie für die freigestellten Zwecke verhindern oder einschränken, sind ungültig (s29A (5) CDPA).

Der genaue Umfang der Schranke von s29A CDPA ist nicht geklärt. Soweit ersichtlich, wurde sie bisher noch nie von einem Gericht angewendet. In der jüngeren Literatur wird geltend gemacht, dass die Schranke den praktischen Anforderungen moderner datengetriebener Forschung nur begrenzt Rechnung trage. Kritisiert werden namentlich die Beschränkung auf die nichtkommerzielle Forschung, das Verbot der Weitergabe von im Rahmen der Schranke erstellten Kopien und die Unklarheiten im Zusammenhang mit dem Erfordernis des rechtmässigen Zugangs zu den genutzten Werken.[272]

b) Gesetzgeberische Entwicklungen

In der «*National AI Strategy*»[273] vom September 2021 wurde das Ziel formuliert, das UK innerhalb der nächsten zehn Jahre als führende Nation im Bereich der KI-Technologie zu etablieren. 2022 wurde im Rahmen einer Konsultation des *UK Intellectual Property Office* (UKIPO) die Einführung einer *neuen Schranke für Text und Data Mining* vorgeschlagen, welche die Nutzung von Werken zu allen Zwecken von Text und Data Mining erlauben sollte.[274] Diese Lösung würde nach Auffassung UKIPO die Entwicklung von KI und Innovation am stärksten fördern, weil das Text und Data Mining von urheberrechtlich geschützten Werken insgesamt erleichtert würde. Das UKIPO verwies dabei auf die Schranken für Text und Data Mining in Singapur, Japan und der EU sowie auf die *Fair Use*-Doktrin des US-amerikanischen Rechts, die sich seiner Ansicht nach positiv auf die Entwicklung von KI ausgewirkt haben.[275]

Die vom UKIPO vorgeschlagene Schranke für Text und Data Mining sah keine Möglichkeit zur Erklärung eines Rechtevorbehalts vor und nahm den Rechteinhabern auch die Möglichkeit, Lizenzen für Text und Data Mining zu erteilen.[276] Dem Schutz der Rechteinhaber diente nur die Voraussetzung des rechtmässigen Zugangs zu den Werken.[277] Die Rechteinhaber hätten damit wählen können, auf welcher Plattform sie ihre Werke zur Verfügung stellen und ob sie diese bspw. nur gegen Entgelt zugänglich machen oder den Zugriff beschränken wollen.[278] Mit dem Verzicht auf einen Rechtevorbehalt ging der Vorschlag bewusst über die Vorgaben von Art. 4 DSM-RL[279] hinaus. Die durch den Brexit gewonnenen Spielräume sollten genutzt werden, um bei Text und Data Mining einen Standortvorteil zu schaffen.[280] Neben dem vom UKIPO bevorzugten Vorschlag für eine neue Schranke stellte das Amt weitere Optionen zur Diskussion, namentlich den Verzicht auf eine Änderung der bestehenden Regelung, einen Ausbau der Lizenzierung für Text und Data Mining und eine Anpassung der bestehenden Schranke.[281] Für diese dritte Option wurden unter anderem eine Ausdehnung der Regelung in s29A CDPA auf kommerzielle Forschung, die Einführung einer neuen Schranke für alle Zwecke von Text und Data Mining mit einem Rechtevorbehalt und alternativ eine entsprechende Regelung ohne Rechtevorbehalt vorgeschlagen.[282]

Der Vorschlag für eine neue Schranke für Text und Data Mining wurde in der öffentlichen Debatte stark kritisiert.[283] Namentlich wurde vorgebracht, dass die potenziellen Auswirkungen auf die Kreativwirtschaft unzureichend berücksichtigt worden seien, weshalb das UKIPO aufgefordert wurde, den Vorschlag nicht weiterzuverfolgen und stattdessen eine Regulierungsfolgenabschätzung durchzuführen.[284] In der rechtswissenschaftlichen Literatur wurde zudem geltend gemacht, dass die Gegenüberstellung eines Opt-out- und eines Opt-in-Modells die praktischen Auswirkungen der unterschiedlichen Regelungsansätze nur unvollständig abbilde, insb. mit Blick auf bestehende Marktkonzentrationen und die technische Umsetzbarkeit von Rechtevorbehalten.[285] Wegen der Kritik der Kreativwirtschaft wurde der Vorschlag für eine neue Schranke für Text und Data Mining nicht weiterverfolgt.[286] Im Dezember 2024 publizierte das UKIPO eine weitere Konsultation zu *Copyright and Artificial Intelligence*, die drei Ziele für eine künftige Regelung formulierte: (i) die Unterstützung der Rechteinhaber bei der Durchsetzung ihrer Rechte; (ii) die Förderung der Entwicklung führender KI-Modelle im UK; und (iii) die Stärkung von Vertrauen und Transparenz.[287] In der Konsultation sprach sich das UKIPO für die Einführung einer neuen Schranke aus, die für alle Zwecke von Text und Data Mining – insb. auch für kommerzielle Nutzungen – gelten soll, den Rechteinhabern aber die Möglichkeit belässt, einen

272 Zum Ganzen M. Kretschmer/B. Meletti/L. Bently et al., Copyright and AI in the UK: Opting-In or Opting-Out?, GRUR Int. 2025, 1056 ff.
273 National AI Strategy, ‹https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf› (abgerufen am 5. Januar 2026).
274 Consultation outcome, Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation, ‹www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation#text-and-data-mining› (abgerufen am 5. Januar 2026), Rz. 58.
275 Consultation outcome (Fn. 274), Rz. 34.
276 Consultation outcome (Fn. 274), Rz. 61.
277 Consultation outcome (Fn. 274), Rz. 61.
278 Consultation outcome (Fn. 274), Rz. 61.
279 Siehe dazu vorne, III.2.a).
280 Consultation outcome (Fn. 274), Rz. 60.
281 Consultation outcome (Fn. 274), Rz. 39 ff.
282 Consultation outcome (Fn. 274), Rz. 48 ff.
283 Siehe dazu House of Lords, Communications and Digital Committee, At risk: our creative future, January 2023, ‹https://publications.parliament.uk/pa/ld5803/ldselect/ldcomm/125/125.pdf› (abgerufen am 5. Januar 2026), Rz. 30 ff.
284 Communications and Digital Committee (Fn. 283), Rz. 35.
285 Kretschmer/Meletti/Bently et al. (Fn. 272), 1055 f., 1062 f.
286 Siehe dazu UK Parliament, Artificial Intelligence: Intellectual Property Rights, Volume 727: debated on Wednesday 1 February 2023, ‹https://hansard.parliament.uk/commons/2023-02-01/debates/7CD1D4F9-7805-4CF0-9698-E28ECEFB7177/ArtificialIntelligenceIntellectualPropertyRights› (abgerufen am 5. Januar 2026).
287 Copyright and AI: Consultation, Dezember 2024, ‹https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf› (abgerufen am 5. Januar 2026).

Rechtevorbehalt zu erklären.[288] Nach Auffassung des UKIPO trägt diese Option den Interessen der Rechteinhaber und der Entwickler von KI-Modellen am besten Rechnung. Zudem sollen Transparenzpflichten für KI-Entwickler eingeführt werden, insb. in Bezug auf die Offenlegung der für das Training von KI-Modellen verwendeten Inhalte.[289]

Die vorgeschlagene Schranke entspricht im Wesentlichen der Regelung in Art. 4 DSM-RL. Wie sich in der EU gezeigt hat, ist die praktische Umsetzung von Rechtevorbehalten aber mit Herausforderungen verbunden.[290] Das UKIPO verweist auf die heute bestehenden Möglichkeiten zur technischen Ausgestaltung und nennt namentlich den bereits weit verbreiteten robots.txt-Standard, die Möglichkeit der Verknüpfung von Werken mit Metadaten und das «Do Not Train»-Register von *Spawing.AI*.[291] Als problematisch erscheint dem UKIPO, dass der robots.txt-Standard keine detaillierte Kontrolle der Werknutzung erlaubt.[292] Denn der Standard ermöglicht zwar den Hinweis, dass das *Webcrawling* auf einer ganzen Website unerwünscht sei, erlaubt aber weder eine granulare Information auf Ebene einzelner Werke noch eine Differenzierung nach Nutzungsarten, etwa wenn ein Rechteinhaber das *Webcrawling* für die Indexierung in Suchmaschinen zulassen möchte, nicht aber für Anwendungen seiner Werke für generative KI.[293] Vor diesem Hintergrund betont das UKIPO die Bedeutung der Entwicklung einheitlicher technischer Standards, die es Rechteinhabern ermöglichen, ihre Rechte wirksam vorzubehalten, und KI-Entwicklern, die Vorbehalte zuverlässig zu berücksichtigen.[294]

Die Konsultation zu *Copyright and Artificial Intelligence* lief bis zum 25. Februar 2025,[295] das Feedback wird derzeit analysiert.[296] Der Bericht und das Ergebnis der Regulierungsfolgenabschätzung sollen bis zum 18. März 2026 publiziert werden.[297] Einstweilen hat die Regierung angekündigt, dass Expertengruppen mit Vertreterinnen und Vertretern der Kreativindustrie und der KI-Branche gebildet werden sollen, um gemeinsam praktikable Lösungen zu entwickeln, die Innovation fördern und die Rechte von Urheberinnen und Urhebern schützen.[298] Ein konkreter Gesetzesvorschlag liegt aber noch nicht vor.

Zeitlich parallel zu dieser Konsultation wurde auch das Parlament aktiv. Am 6. Februar 2025 wurde die *Data (Use and Access) Bill* im *House of Commons* eingebracht. Im Verlauf der Beratungen wurden insb. im *House of Lords* verschiedene Vorschläge für Transparenz- und Informationspflichten bei der Nutzung urheberrechtlich geschützter Werke beim Training von KI-Modellen diskutiert.[299] Damit griffen die Vorschläge zentrale Fragestellungen der laufenden Konsultation auf, ohne deren Ergebnisse vorwegzunehmen. Der *Data (Use and Access) Act* wurde am 19. Juni 2025 angenommen.[300] In der endgültigen Fassung sind keine weitergehenden Transparenzpflichten vorgesehen. Stattdessen verpflichtet das Gesetz die Regierung, innerhalb von neun Monaten nach Inkrafttreten einen Bericht über die Nutzung urheberrechtlich geschützter Werke bei der Entwicklung von KI-Modellen und eine Abschätzung der wirtschaftlichen Folgen der gesetzgeberischen Optionen vorzulegen.[301]

c) Gerichtsverfahren

Am 9. Juni 2025 eröffnete der *High Court of Justice (EWHC)* das Verfahren zwischen Getty Images und Stability AI.[302] Getty Images machte geltend, dass das Trainieren und Entwickeln des KI-Modells *Stable Diffusion* massenhaft ihre Urheberrechte verletze.[303] Einerseits behauptete Getty, dass geschützte Werke für das Training und die Entwicklung von Stable Diffusion auf Server und Computer im UK heruntergeladen und verarbeitet worden seien *(Training and Development Claim)*, was eine direkte Urheberrechtsverletzung *(primary infringement)* sei. Zudem liege eine sekundäre Urheberrechtsverletzung *(secondary infringement)* vor, weil die trainierte Software als Gegenstand *(«article»)* im Sinn von s22 CDPA einzustufen sei, und ihr Import in das UK eine Rechtsverletzung darstelle *(Secondary Infringement Claim)*. Als dritten Klagepunkt machte die Klägerin geltend, dass von Nutzern im UK generierte Ausgabe-Bilder *(Outputs)* ihr Urheberrecht verletzten *(Outputs Claim)*.[304] Neben den urheberrechtlichen Ansprüchen erhob Getty Images auch Klage wegen einer Verletzung von Datenbankrechten *(Database Rights Claim)* und Markenrechten *(Trade Mark Infringement Claim)*.[305]

In diesem Verfahren war unbestritten, dass Werke für das Training und die Entwicklung von *Stable Diffusion* vervielfältigt worden waren.[306] Umstritten war aber, ob die

288 Consultation 2024 (Fn. 287), Rz. 74.
289 Consultation 2024 (Fn. 287), Rz. 103 ff.
290 Siehe dazu vorne, III.2.a).
291 Consultation 2024 (Fn. 287), Rz. 82 ff.
292 Consultation 2024 (Fn. 287), Rz. 86.
293 Consultation 2024 (Fn. 287), Rz. 86, wobei fälschlicherweise angenommen wird, dass robots.txt zu einer Blockierung von Webcrawlern führe.
294 Consultation 2024 (Fn. 287), Rz. 87.
295 Consultation 2024 (Fn. 287), Rz. 21.
296 Closed Consultation: Copyright and Artificial Intelligence, ‹www.gov.uk/government/consultations/copyright-and-artificial-intelligence› (abgerufen am 5. Januar 2026).
297 Copyright and artificial intelligence statement of progress under Section 137 Data (Use and Access) Act, ‹www.gov.uk/government/publications/copyright-and-artificial-intelligence-progress-report/copyright-and-artificial-intelligence-statement-of-progress-under-section-137-data-use-and-access-act› (abgerufen am 5. Januar 2026).
298 Press release: Creative and AI sectors kick-off next steps in finding solutions to AI and copyright, July 2025, ‹www.gov.uk/government/news/creative-and-ai-sectors-kick-off-next-steps-in-finding-solutions-to-ai-and-copyright› (abgerufen am 5. Januar 2026).
299 Zum Ganzen House of Commons Library, Data (Use and Access) Bill [HL]: progress of the bill, April 2025, ‹https://researchbriefings.files.parliament.uk/documents/CBP-10212/CBP-10212.pdf› (abgerufen am 5. Januar 2026), 17 ff.
300 Parliamentary Bills, Data (Use and Access) Act 2025, News, 2025, ‹https://bills.parliament.uk/bills/3825/news› (abgerufen am 5. Januar 2026).
301 Data (Use and Access) Act 2025 (UK) c 18, ss 135–136.
302 Die Klage wurde im September 2023 am EWHC eingereicht.
303 Siehe dazu die prozessuale Zwischenentscheidung *Getty Images (US) Inc v Stability AI Ltd* [2025] EWHC 38 (Ch) [9] und die Entscheidung zum Antrag auf Eilverfahren (summary judgement) *Getty Images (US) Inc v Stability AI Ltd* [2023] EWHC 3090 (Ch) [11].
304 *Getty Images v Stability AI* [2025] 38 (Fn. 303) [9].
305 *Getty Images v Stability AI* [2023] 3090 (Fn. 303) [11].
306 *Getty Images v Stability AI* [2023] 3090 (Fn. 303) [59 ff.].

rechtlich relevanten Vervielfältigungen im UK erfolgt waren.[307] Zwar ist Stability AI im UK registriert, das Training soll nach den Angaben von Stability AI aber ausschliesslich auf Servern in den USA stattgefunden haben.[308]

Am 4. November 2025 veröffentlichte der *High Court* sein Urteil. Nachdem Getty Images anerkannt hatte, dass keine Anhaltspunkte dafür bestanden, dass Trainingstätigkeiten von Stability AI im UK stattgefunden haben, liess die Klägerin den *Training and Development Claim* fallen.[309] Mit dem *Outputs Claim* wollte Getty Images erreichen, dass bestimmte, als urheberrechtsverletzend beanstandete Ausgaben des Modells von Stability AI – konkret jene Beispielbilder, die im Prozess als Beweismittel vorgelegt worden waren – nicht mehr erzeugt werden können. Diese Bilder liessen sich jedoch nur durch die Eingabe bestimmter Prompts generieren, die inzwischen durch Stability AI blockiert worden waren, sodass das Modell nicht mehr in der Lage war, die beanstandeten Ausgaben zu erzeugen. Der *Outputs Claim* war damit gegenstandslos geworden und wurde nicht weiterverfolgt. Aufgrund des engen Zusammenhangs mit dem *Training and Development Claim* und dem *Outputs Claim* konnte auch der Anspruch einer Verletzung von Datenbankrechten *(Database Rights Claim)* nicht weiterverfolgt werden.[310] Bei der geltend gemachten Verletzung von Markenrechten *(Trade Mark Infringement Claim)* war die Klägerin teilweise erfolgreich.[311]

Der einzig verbleibende urheberrechtlich relevante Klagepunkt war damit der Vorwurf einer sekundären Urheberrechtsverletzung *(Secondary Infringement Claim)*. Der *High Court* stellte fest, dass Stable Diffusion keine Kopien der geschützten Werke enthalte und daher nicht als «*infringing copy*» qualifiziert werden könne.[312] Ebenso lehnte das Gericht ein weites Verständnis des Begriffs «*article*» ab, nach dem auch ein immaterielles KI-Modell aufgrund seines Trainings als rechtsverletzender Gegenstand einzustufen wäre.[313] Da dieses zentrale Element nicht erfüllt war, scheiterte der *Secondary Infringement Claim*.[314]

Der Entscheid des *High Court* ist erstinstanzlich und berufungsfähig. Auf Antrag von Getty Images hat der High Court die Zulassung zur Berufung erteilt und es ist anzunehmen, dass Getty Berufung einlegen wird.

6. USA

a) Gesetzliche Regelung

Der US-amerikanische Gesetzgeber hat bisher keine spezifische Bestimmung zu KI und Urheberrecht erlassen. Das *U.S. Copyright Office* hat aber 2024 und 2025 zwei viel beachtete Berichte zu AI und Copyright veröffentlicht. Der erste Bericht[315] befasst sich mit *digital replicas* von Personen,[316] der zweite mit dem urheberrechtlichen Schutz des Outputs von KI-Systemen.[317] Ein dritter Bericht, der noch nicht publiziert wurde, konzentriert sich auf die Rechtsfragen beim Training von KI-Modellen. Eine *Pre-Publication-Version* des Berichts ist auf der Website des *U.S. Copyright Office* abrufbar.[318] Aus dieser ergibt sich, dass mehrere Vorgänge bei der Entwicklung und Nutzung von KI-Modellen – etwa das Sammeln und Aufbereiten der Trainingsdaten, das eigentliche Training des Modells und die Generierung von Output – aus Sicht des *Copyright Office* als *prima facie*-Urheberrechtsverletzung qualifiziert werden können, insb. als Eingriff in das Vervielfältigungsrecht.[319]

In den USA wurden in den letzten Jahren mehrere Vorschläge für gesetzliche Regelungen zu KI und Urheberrecht vorgelegt.[320] Nach der Aufhebung der von Präsident Biden erlassenen *Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*[321] am 20. Januar 2025 durch Präsident Trump[322] ist allerdings nicht zu erwarten, dass diese Vorschläge in nächster Zeit weiterverfolgt werden.

b) Gerichtsverfahren

In den USA sind derzeit 65 Klagen zu Fragen an der Schnittstelle von Urheberrecht und KI hängig. Mehrere betreffen

307 *Getty Images v Stability AI* [2023] 3090 (Fn. 303) [44].
308 *Getty Images v Stability AI* [2023] 3090 (Fn. 303) [44].
309 *Getty Images v Stability AI* [2025] EWHC 2863 (Ch) [9].
310 Zum Ganzen siehe *Getty Images v Stability AI* [2025] 2863 (Fn. 309) [9].
311 *Getty Images v Stability AI* [2025] 2863 (Fn. 309) [453], [456], [538].
312 *Getty Images v Stability AI* [2025] 2863 (Fn. 309) [548 ff.], [571 ff.].
313 *Getty Images v Stability AI* [2025] 2863 (Fn. 309) [568 ff.].
314 *Getty Images v Stability AI* [2025] 2863 (Fn. 309) [595 ff.], [610].
315 U.S. Copyright Office, Copyright and Artificial Intelligence, Part 1: Digital Replicas, ‹www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf› (abgerufen am 5. Januar 2026).
316 Das U.S. Copyright Office verwendet in diesem Bericht die Bezeichnungen «digital replica» und «deepfake» synonym; siehe dazu U.S. Copyright Office (Fn. 315), 2.
317 U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability, ‹www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf› (abgerufen am 5. Januar 2026).
318 U.S. Copyright Office, Copyright and Artificial Intelligence, Part 3: Generative AI Training, Pre Publication Version, ‹www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf› (abgerufen am 5. Januar 2026).
319 U.S. Copyright Office (Fn. 318), 26 ff.
320 Siehe dazu House, AI Foundation Model Transparency Act of 2023, ‹www.congress.gov/bill/118th-congress/house-bill/6881/text› (abgerufen am 5. Januar 2026); House, Generative AI Copyright Disclosure Act of 2024, ‹www.congress.gov/118/bills/hr7913/BILLS-118hr7913ih.pdf› (abgerufen am 5. Januar 2026); Senate, Artificial Intelligence Research, Innovation, and Accountability Act of 2024, ‹www.congress.gov/118/bills/s3312/BILLS-118s3312rs.pdf› (abgerufen am 5. Januar 2026).
321 Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (the «Biden Order»), ‹www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence› (abgerufen am 5. Januar 2026); siehe auch The White House, FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, ‹https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/› (abgerufen am 5. Januar 2026).
322 Presidential Actions, Initial Rescissions of Harmful Executive Orders and Actions, ‹www.whitehouse.gov/presidential-actions/2025/01/initial-rescissions-of-harmful-executive-orders-and-actions/› (abgerufen am 5. Januar 2026).

das Training von KI-Modellen.[323] Kernpunkt ist dabei die Frage, ob und unter welchen Voraussetzungen die Nutzung von urheberrechtlich geschützten Werken beim Training von KI-Modellen als *Fair Use* im Sinn von *U.S. Copyright Act*, 17 U.S.C. § 107 qualifiziert werden kann.

Die *Fair Use*-Schranke ist eine sog. *affirmative defense*, also eine rechtliche Einwendung, bei welcher der Beklagte Tatsachen behauptet, die zur Abweisung der Klage führen. Kann der Rechteinhaber eine *prima facie*-Urheberrechtsverletzung nachweisen, muss der Beklagte die Voraussetzungen für das Vorliegen von *Fair Use* beweisen.[324] Der Kläger muss lediglich nachweisen, dass er Inhaber der Urheberrechte ist und dass wesentliche Teile des Werkes übernommen wurden;[325] das ist der Fall, wenn das Verletzungsobjekt im Wesentlichen gleichartige Züge aufweist *(substantial similarity)* wie das geschützte Werk.[326] Verlangt ist zudem eine tatsächliche Übernahme des geschützten Werkes *(actual copying)*.[327] Eine solche liegt vor, wenn der Beklagte das fremde Werk tatsächlich genutzt hat, um sein eigenes Werk zu schaffen.[328] Die tatsächliche Übernahme kann direkt oder indirekt bewiesen werden. Der direkte Beweis erfordert den Nachweis der tatsächlichen Nutzung; beim indirekten Beweis reicht der Nachweis, dass der Beklagte Zugang zum geschützten Werk hatte und etwas Ähnliches geschaffen hat *(probative similarity)*.[329] Bei den Anforderungen an die Ähnlichkeit wird differenziert: Weist ein Werk nur wenige schutzfähige Elemente auf und hat es damit einen eher kleinen Schutzbereich *(thin copyright)*, so muss die Ähnlichkeit sehr hoch, das Verletzungsobjekt also praktisch identisch sein *(virtually identical)*.[330] Hat das Werk hingegen einen grossen Schutzbereich *(thick copyright)*, so sind die Anforderungen an die Ähnlichkeit tiefer und es genügt, dass das Verletzungsobjekt im Wesentlichen gleichartige Züge *(substantial similarity)* aufweist.[331]

Liegt eine *prima facie*-Urheberrechtsverletzung vor, so ist zu prüfen, ob die Voraussetzungen für *Fair Use* erfüllt sind. Das ist nach vier Faktoren zu beurteilen *(four factor test)*: (1) Zweck und Art der Nutzung; (2) Art des urheberrechtlich geschützten Werkes; (3) Menge und Wesentlichkeit des verwendeten Teils im Verhältnis zum gesamten geschützten Werk; und (4) Auswirkungen der Nutzung auf den potenziellen Markt oder den Wert des geschützten Werkes.[332] Dabei kommt dem ersten Faktor besonderes Gewicht zu.[333] Bei dessen Bewertung sind verschiedene Aspekte *(subfactors)* zu berücksichtigen,[334] namentlich (i) der kommerzielle Charakter der Handlung, (ii) der transformative oder produktive Charakter der Handlung, (iii) die Motive des Verletzers und (iv) das Vorliegen einer Zwecksetzung, die in der Präambel von *Section* 107 des *Copyright Act* erwähnt wird.[335] Die US-amerikanische Lehre ist sich weitgehend einig, dass es bei der Beurteilung der Nutzung von Werken beim Training von KI-Modellen entscheidend ist, ob diese Nutzung als *transformative* qualifiziert wird.[336] Nach der Rechtsprechung des U.S. Supreme Court kann diese Voraussetzung erfüllt sein, wenn das verwendete Werk umgestaltet wird oder die Nutzung einem völlig anderen, gesellschaftlich wertvollen Zweck dient.[337] Dienen die ursprüngliche Nutzung und die Zweitnutzung dem gleichen oder einem ähnlichen Zweck und hat die Zweitnutzung kommerziellen Charakter, so spricht der erste Faktor eher gegen das Vorliegen von *Fair Use*.[338]

Das erste Urteil zum Training von KI-Modellen hat der *U.S. District Court for the District of Delaware* gefällt. In seinem Entscheid vom 11. Februar 2025 «*Thomson Reuters v. Ross Intelligence*»[339] hatte das Gericht die Frage zu entscheiden, ob die Nutzung urheberrechtlich geschützter Werke aus der Datenbank von *Westlaw* für das Training eines KI-gestützten juristischen Recherchetools als *Fair Use* zu qualifizieren ist. Um seine KI zu trainieren, erwarb Ross rund 25'000 *Bulk Memos* von *LegalEase* und nutzte diese als Trainingsdaten. Bei den *Bulk Memos* handelte es sich um von Anwälten erstellte Rechtsfragen mit dazugehörigen guten und schlechten Antworten. *LegalEase* hatte die Anwälte angewiesen, die *Bulk Memos* mithilfe der *Headnotes* von *Westlaw* zu erstellen, in welchen die Kernpunkte von Gerichtsentscheiden zusammengefasst werden. Das Gericht bejahte eine *prima facie*-Urheberrechtsverletzung, weil die von Ross erworbenen *Bulk Memos* den *Headnotes* von Westlaw im Wesentlichen ähnlich *(substantial similarity)* waren. Zudem hatte *LegalEase* Zugang zu den Werken von Reuters, was das Gericht als Indizienbeweis *(circumstantial evidence)* für die Vervielfältigung der Werke von Reuters genügen liess.[340] Bei der Prüfung des *Fair Use* verneinte das Gericht eine transformative Nutzung, weil der Dienst von Ross denselben Zweck wie der von der Klägerin betriebene Dienst von *Westlaw* habe, nämlich das Zugänglichmachen juristischer Informationen. Die Nutzung der Werke durch Ross diene weder einem weitergehenden Zweck noch habe sie einen anderen Charakter *(«further purpose or different character»)*. Vielmehr

323 Übersicht über die laufenden Verfahren auf ‹https://chatgptiseatingtheworld.com/category/map-of-ai-copyright-lawsuits/› (abgerufen am 5. Januar 2026).
324 L. Pallas Loren, Fair Use: An Affirmative Defense?, in: 90 Washington Law Review 685, 688 (2015); *Campbell v. Acuff-Rose Music Inc.*, 510 U.S. 569 (1994).
325 *Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc.*, 1:20-cv-00613, 5 (D. Del. Feb. 11, 2025) (Memorandum Opinion).
326 *Reuters v. Ross* (Fn. 325), 15.
327 *Reuters v. Ross* (Fn. 325), 11; *Clayton Prince Tanksley v. Lee Daniels*, 902 F.3d 165 (3rd Cir. 2018).
328 *Tanksley v. Daniels* (Fn. 327).
329 *Reuters v. Ross* (Fn. 325), 11; *Tanksley v. Daniels* (Fn. 327), 173.
330 *Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435, 1439 ff. (9th Cir. 1994); *Reuters v. Ross* (Fn. 325), 13.
331 *Apple v. Microsoft* (Fn. 330), 1439 ff.; *Reuters v. Ross* (Fn. 325), 13 f.
332 § 107: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.
333 *Campbell v. Acuff-Rose Music Inc.* (Fn. 324); B. Beebe, An Empirical Study of U.S. Copyright Fair Use Opinions, 1978–2005, UPLR 2008, 597.
334 Beebe (Fn. 333), 597.
335 Beebe (Fn. 333), 597.
336 *Reuters v. Ross* (Fn. 325).
337 *Campbell v. Acuff-Rose Music Inc.* (Fn. 324).
338 *Campbell v. Acuff-Rose Music Inc.* (Fn. 324).
339 *Reuters v. Ross* (Fn. 325).
340 *Reuters v. Ross* (Fn. 325), 14.

entwickle die Beklagte mithilfe der Werke der Klägerin ein konkurrierendes Produkt, das kommerziellen Zwecken diene, womit keine transformative Nutzung vorliege.[341] Ross argumentierte, dass die kopierten *Headnotes* nicht im Endprodukt enthalten seien. Die Nutzung erfolge lediglich in einem Zwischenschritt, bei dem die *Headnotes* in numerische Daten umgewandelt würden, um das KI-Modell zu trainieren.[342] Das Gericht räumte zwar ein, dass ein solches «Zwischenkopieren» *(intermediate copying)* in früheren Fällen[343] als *transformative* qualifiziert und als *Fair Use* zugelassen wurde. Diese Fälle betrafen allerdings das Kopieren von *Source Code*.[344] Entscheidend war dabei, dass die Vervielfältigungen des *Source Code* für die innovative Tätigkeit der Beklagten notwendig war. Das war hier nicht der Fall.[345]

Der zweite und der dritte Entscheid zum Training von KI-Modellen stammen vom *U.S. District Court for the Northern District of California*. In «*Bartz v. Anthropic*»[346] und in «*Kadrey v. Meta*»[347] klagten verschiedene Rechteinhaber wegen der Nutzung ihrer Werke beim Training von generativen KI-Modellen. In «*Bartz v. Anthropic*» war zum einen die Verwendung von Büchern zu beurteilen, welche die Beklagte rechtmässig erworben und nach Entfernung des Einbandes eingescannt hatte. Zum anderen machte die Klägerin geltend, die Beklagte habe digitale Versionen von Werken aus illegalen Quellen heruntergeladen und gespeichert. Alle diese Daten nutzte die Beklagte für den Aufbau einer «zentralen Bibliothek», die sie für das Training ihrer LLMs verwendete.[348] Mit *Summary Judgement* vom 23. Juni 2025 fällte das Gericht einen Teilentscheid zum Vorliegen von *Fair Use*.[349] Es qualifizierte die Verwendung von Werken für das Training von KI-Modellen als «*exceedingly transformative*».[350] Bei den verwendeten Daten differenzierte das Gericht: Das Einscannen und Speichern der rechtmässig erworbenen Bücher und die Verwendung der Texte für das Training der KI-Modelle qualifizierte es als *Fair Use*. Anders entschied es hinsichtlich der rechtswidrig beschafften digitalen Kopien,[351] weil das Beschaffen von Inhalten zum Aufbau einer «zentralen Bibliothek» und das Speichern dieser Kopien für ein späteres Training eines KI-Modells eine eigenständige Nutzungshandlung sei, die nicht als *transformativ* gelten könne.[352]

In «*Kadrey v. Meta*» hatte das Gericht die Nutzung von Werken zu beurteilen, die *Meta* aus sog. «Schattenbibliotheken» *(shadow libraries)* bezogen hatte. Dabei handelt es sich um Online-Archive, die Bücher, Zeitschriftenartikel, Filme und weitere Werke ungeachtet des urheberrechtlichen Schutzes und ohne Zustimmung der Rechteinhaber kostenlos zum *Download* anbieten.[353] Am 25. Juni 2025 entschied das Gericht wiederum in Form eines *Partial Summary Judgement* und bejahte das Vorliegen von *Fair Use*. Wie in «*Bartz v. Anthropic*» bewertete es die Nutzung von Werken beim Training von KI-Systemen als «*highly transformative*».[354] Im Vordergrund standen die Erwägungen zum vierten Faktor, der Auswirkung der Nutzung auf den potenziellen Markt oder Wert des urheberrechtlich geschützten Werkes. Das Gericht hielt namentlich fest, dass die technischen Massnahmen im LLM der Beklagten die Reproduktion von Trainingsdaten verhindere und dass ein potenzieller, schutzwürdiger Lizenzmarkt für KI-Training nicht erkennbar sei.[355] Damit fehle es an Belegen für indirekte wirtschaftliche Schäden der Kläger.

In beiden Entscheiden konzentrierte sich der *U.S. District Court for the Northern District of California* auf die Anwendbarkeit der *Fair-Use*-Schranke. Ob die Nutzung von Werken beim Training von KI-Modellen zu einer Memorisierung der Werke im trainierten Modell[356] führe, liess das Gericht jeweils mit der Begründung offen, dass dies an der rechtlichen Bewertung nichts ändere, weil die Nutzung der Werke auch in diesem Fall *transformative* sei.[357]

7. Erkenntnisse

Der Blick auf das europäische, deutsche, französische, britische und US-amerikanische Recht hat gezeigt, dass die urheberrechtliche Beurteilung der Nutzung von Werken beim Training von KI-Modellen in allen diesen Rechtsordnungen intensiv und kontrovers diskutiert wird.

Die gesetzliche Regelung ist in Deutschland und Frankreich aufgrund der Harmonisierung durch die DSM-RL weitgehend deckungsgleich. Das britische Recht sieht für computergestützte Analysen von Werken für Zwecke der nichtkommerziellen Forschung eine Schranke vor, die derjenigen für Text und Data Mining für Zwecke der wissenschaftlichem Forschung in Deutschland und Frankreich entspricht, es kennt aber keine allgemeine Schranke für Text und Data Mining, auf die sich kommerzielle Unternehmen stützen könnten. Ob diese Regelungen in den nächsten Jah-

341 *Reuters v. Ross* (Fn. 325), 17 f.
342 *Reuters v. Ross* (Fn. 325), 18.
343 *Google LLC v. Oracle America, Inc.*, 593 U.S. 1, 30–32 (2021); *Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 599, 606–07 (9th Cir. 2000); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1514–1515, 1522–23 (9th Cir. 1992).
344 *Reuters v. Ross* (Fn. 325), 18.
345 *Reuters v. Ross* (Fn. 325).
346 Das Gerichtsverfahren ist noch nicht abgeschlossen. Die Parteien haben sich auf einen Vergleich geeinigt, der vom Gericht vorläufig genehmigt wurde *(preliminary approval)*. Die Anhörung zur endgültigen Genehmigung *(final approval hearing)* ist für April 2026 angesetzt; siehe auch D. Hansen, The Bartz v. Anthropic Settlement: Understanding America's Largest Copyright Settlement, Kluwer Copyright Blog, ‹https://legalblogs.wolterskluwer.com/copyright-blog/the-bartz-v-anthropic-settlement-understanding-americas-largest-copyright-settlement/› (abgerufen am 5. Januar 2026).
347 Das Gerichtsverfahren ist noch nicht abgeschlossen. Die neusten Entwicklungen können unter ‹www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/› (abgerufen am 5. Januar 2026) eingesehen werden.
348 *Andrea Bartz, et al. v. Anthropic PBC*, Civil Action No. 24-05417 WHA, 1 (N.D. Cal. Jun. 23, 2025).
349 *Bartz v. Anthropic* (Fn. 348).
350 *Bartz v. Anthropic* (Fn. 348), 9 ¶13, 30 ¶22.
351 *Bartz v. Anthropic* (Fn. 348), 19.
352 *Bartz v. Anthropic* (Fn. 348), 19 ¶13 f.
353 *Richard Kadrey et al. v. Meta Platforms, Inc.*, Civil Action No. 23-cv-03417-VC, 11.
354 *Kadrey v. Meta* (Fn. 353), 16, 40.
355 *Kadrey v. Meta* (Fn. 353), 36.
356 Siehe dazu vorne, II.3.
357 *Bartz v. Anthropic* (Fn. 348), 11 ¶21 ff.; *Kadrey v. Meta* (Fn. 353), 12 f.

ren revidiert werden, ist derzeit offen. Zumindest im UK ist wohl mit einer Revision zu rechnen, die möglicherweise eine weitere Annäherung an die Vorgaben des EU-Rechts bringen wird. Einen anderen Ansatz verfolgt das US-amerikanische Urheberrecht, das keine besondere Regelung für Text und Data Mining kennt und die Herausforderungen beim Training von KI-Modellen durch Anwendung der *Fair-Use*-Schranke bewältigen muss.

Die ersten Gerichtsentscheide in Deutschland, im UK und in den USA gelangen zu teilweise grundlegend anderen Erkenntnissen. Bemerkenswert ist namentlich, dass der britische *High Court of Justice* in *Getty vs. Stability AI* feststellte, dass das KI-Modell *Stable Diffusion* keine Kopien der geschützten Werke enthalte und daher nicht als *«infringing copy»* qualifiziert werden könne,[358] während das Landgericht München I zum Schluss kam, dass die beim Training verwendeten Werke in den Sprachmodellen von OpenAI vervielfältigt werden.[359] Die Gerichte in den USA gehen in ersten Entscheiden zwar davon aus, dass die Nutzung von Werken für das Training von KI-Modellen *«transformative»* sei und als *Fair Use* qualifiziert werden könne, verneinen diesen aber, wenn das Training dazu dient, ein konkurrierendes Produkt zu entwickeln.[360] Angesichts der zahlreichen hängigen Klagen wird sich im US-amerikanischen Recht wohl bald ein deutlicheres Bild ergeben. Das gilt auch für das europäische Recht, zumal der EuGH in der Rechtssache C-250/25, *«Like Company v. Google Ireland»*, voraussichtlich klären wird, ob die Nutzung von Werken für das Training von KI-Modellen als Text und Data Mining qualifiziert und von den entsprechenden Schranken freigestellt werden kann.

Während sich in Europa bei der gesetzlichen Regelung ein recht einheitliches Bild zeigt, lässt sich in den ersten Gerichtsentscheiden in Europa und den USA noch kein klarer Trend erkennen. Aus der Rechtsprechung ergeben sich für den Schweizer Gesetzgeber damit keine klaren Orientierungspunkte. Sinnvoll erscheint aber eine Orientierung an der gesetzlichen Regelung in Deutschland, Frankreich und im UK.

IV. Stand der Diskussion in der Schweiz

1. Vorbemerkungen

Das schweizerische Urheberrecht kennt weder eine Schranke für Text und Data Mining noch eine Schranke für die Nutzung von Werken beim Training von KI-Modellen. Wie das deutsche, französische und britische sieht aber auch das schweizerische Recht eine Schranke für die Verwendung von Werken zum Zweck der wissenschaftlichen Forschung vor, welche die Vervielfältigung von Werken freistellt, wenn diese durch die Anwendung eines technischen Verfahrens bedingt sind und zu den zu vervielfältigenden Werken ein rechtmässiger Zugang besteht (Art. 24*d* URG). Diese Schranke orientiert sich an der Regelung in Art. 4 DSM-RL, geht aber bewusst weiter und ist nicht auf Text und Data Mining beschränkt.[361]

Vor diesem Hintergrund wird auch in der Schweiz intensiv diskutiert, wie die Nutzung von urheberrechtlich geschützten Werken beim Training von KI-Modellen zu beurteilen ist.[362] Im Zentrum stehen zwei Fragen: Weitgehend einig ist sich die Lehre, dass die Nutzung von Werken beim Training von KI-Modellen eine urheberrechtlich relevante Handlung ist, weil die Werke zumindest beim Erstellen der Trainingsdatensätze in einer Datenbank gespeichert werden. Umstritten ist hingegen, ob die Nutzung beim Training zu einer Vervielfältigung der Werke in den trainierten KI-Modellen führt.[363] Vereinzelt wird zudem vertreten, dass die Kuratierung der Daten vor dem Training eine urheberrechtlich relevante Änderung sei[364] und dass die Werke in den trainierten Modellen zugänglich gemacht werden.[365] Da beim Training von KI-Modellen urheberrechtlich relevante Handlungen vorgenommen werden, stellt sich die Frage, ob diese Nutzungen von einer (oder mehreren) Schranke(n), insb. von der Schranke des internen Gebrauchs

358 Siehe dazu vorne, III.5.c).
359 Siehe dazu vorne, III.3.b).
360 Siehe dazu vorne, III.6.b).
361 Botschaft zur Änderung des Urheberrechtsgesetzes sowie zur Genehmigung zweier Abkommen der Weltorganisation für geistiges Eigentum und zu deren Umsetzung vom 22. November 2017, BBl 2018, 628 f.
362 Y. Benhamou/A. Andrijevic, The protection of AI-generated pictures (photograph and painting) under copyright law, in: R. Abott/D. Geffen (Hg.), Research Handbook on Intellectual Property and Artificial Intelligence, Cheltenham 2022, 198 ff.; M. Berger, Künstliche Intelligenz und Immaterialgüterrecht, in: Jusletter IT vom 4. Juli 2024; E. W. Brem/E. J. Brem, Künstliche Intelligenz und künstlerische Darbietung, sic! 2024, 407 ff.; Cherpillod (Fn. 2), 445 ff.; J. de Werra, Artificial Intelligence & Intellectual Property, ZSR 2025, 355 ff.; P. Gilliéron, Intelligence artificielle: la titularité des données, in: A. Richa/D. Canapa (Hg.), Aspects juridiques de l'intelligence artificielle, Bern 2024, 13 ff.; P. Gilliéron, Réflexions autour de la contractualisation des projets d'intelligence artificielle, sic! 2024, 423 ff.; D. Hartmann, Text and Data Mining and Copyright in Switzerland and the European Union, sic! 2023, 157 ff.; M. Isler, Wissenschaftsschranke, in: P. Mosimann (Hg.), Das revidierte Urheberrecht, Basel 2020, 87 ff.; P. Kübler, Wie generative KI-Systeme Rechte nutzen, medialex 05/23, 1 ff.; C. Marti, Öffentliches Zurverfügungstellen eines Datensatzes zum Training von KI – Wie würden Schweizer Gerichte entscheiden?, sic! 2025, 145 ff.; S. Marmy-Brändli/I. Oehri, Das Training künstlicher Intelligenz, sic! 2023, 655; D. Rosenthal/L. Veraldi, Training von KI-Sprachmodellen: Was das geltende Urheberrecht & Co. erlauben, ABI Technik 2025, 391 ff.; Rosenthal/Veraldi, (Fn. 93), Jusletter vom 3. Februar 2025; V. Salvadé, Comment licencier l'utilisation d'œuvres préexistantes pour entraîner l'intelligence artificielle?, sic! 2024, 87 ff.; Ders., Intelligence artificielle et droit d'auteur: de l'input à l'output, en passant par le droit international privé et la motion Gössi, sic! 1/2026, 1 ff., D. Schönberger, Deep Copyright: Up- and Downstream – Questions Related to Artificial Intelligence (AI) and Machine Learning (ML), in: J. De Werra (Hg.), Droit d'auteur 4.0/Copyright 4.0, Geneva 2018, 1 ff.; Semmelmann (Fn. 2), 637 ff.; M. Städeli/L. Mary, Künstliche Intelligenz und Urheberrecht, SJZ 2024, 244 ff.; Thouvenin (Fn. 4), 381 ff.; Thouvenin/Picht, (Fn. 2), 507 ff.; M. von Welser, KI-Verordnung und Urheberrecht, sic! 2025, 429 ff.; ferner EJPD/BJ, Rechtliche Basisanalyse im Rahmen der Auslegeordnung zu den Regulierungsansätzen im Bereich künstliche Intelligenz, 31.08.2024; AIPPI Swiss Group, Study Question Q295 – Copyright and Artificial Intelligence, sic! 2025, 629 ff.
363 Siehe dazu hinten, IV.2.a).
364 Siehe dazu hinten, IV.2.a) aa).
365 Siehe dazu hinten, IV.2.a) cc).

(Art. 19 Abs. 1 lic. C URG) oder von der Wissenschaftsschranke (Art. 24*d* URG), freigestellt werden.

In der Folge wird der Stand der Diskussion in der Schweiz zusammengefasst. Der Fokus liegt dabei auf den beiden zentralen Fragen, dem Vorliegen einer Vervielfältigung und der Anwendung der Wissenschaftsschranke.

2. Urheberrechtlich relevante Handlung

a) Vervielfältigung von Werken

Bei der Prüfung, ob Werke beim Training von KI-Modellen vervielfältigt werden, ist zwischen den verschiedenen Schritten des Trainings zu differenzieren, namentlich zwischen dem Erstellen von Datenbanken, dem eigentlichen Training und dem trainierten Modell.

aa) Beim Erstellen von Datenbanken

Bevor Werke für das Training von KI-Modellen verwendet werden können, müssen sie gesammelt und in einer Datenbank gespeichert werden. In der Regel werden die Daten vor dem Speichern zudem kuratiert und normalisiert.[366] Die Speicherung erfolgt dabei nicht nur vorübergehend, sondern während längerer Zeit oder gar dauerhaft.[367] Es ist unbestritten, dass diese Speicherung als Vervielfältigung der Werke zu qualifizieren ist.[368]

bb) Im Trainingsprozess

Ob im Trainingsprozess Werke vervielfältigt werden, ist umstritten. Allgemein anerkannt ist, dass nicht jeder Umgang mit urheberrechtlich geschützten Werken eine Verwendung im Sinn von Art. 10 URG ist.[369] Nicht erfasst ist namentlich der Werkgenuss, also die blosse Wahrnehmung von Werken mit den menschlichen Sinnen durch das Lesen, Hören oder Ansehen von Werken.[370] Vor diesem Hintergrund werden in der Lehre im Wesentlichen zwei Meinungen vertreten:

Ein Teil der Lehre geht davon aus, dass es beim Training zu technischen Vervielfältigungen der Trainingsdaten komme, bei denen jeweils unkörperliche Kopien entstehen.[371] Diese zielten zwar nicht unmittelbar auf den Werkkonsum ab, ein solcher wäre aber an sich möglich, weshalb die Kopien als Vervielfältigungen im Sinn des Urheberrechts einzustufen seien.[372] Zum gleichen Schluss kommen andere Autorinnen und Autoren aufgrund einer weiten Auslegung des Vervielfältigungsrechts.[373] Sie argumentieren mit Sinn und Zweck des Urheberrechts: Die Speicherungen und Kopien beim Trainingsvorgang seien mittelbar auf die Wahrnehmung von Werken, insb. auf die Vermittlung der Inhalte, ausgerichtet. Durch einen geeigneten Prompt könne bspw. der Stil einer Künstlerin massenweise auf neue Motive angewendet werden.[374] Dadurch würden Künstlerinnen und Künstler langfristig die Anreize für das Werkschaffen genommen, ihr Ruf genutzt, die wirtschaftlichen Verwertungsmöglichkeiten beschränkt und Originalinhalte verdrängt, was den Grundsätzen des Urheberrechts widerspreche.[375]

Ein anderer Teil der Lehre vertritt die Auffassung, dass im Trainingsprozess keine urheberrechtlich relevanten Handlungen vorgenommen werden, weil kein Werkgenuss ermöglicht werde.[376] Nicht jeder Vorgang, bei dem es sich rein technisch betrachtet um eine Vervielfältigung handle, sei auch als Vervielfältigung im Sinn des Urheberrechts zu qualifizieren.[377] Die Vervielfältigungen, die bei der Verwendung eines Trainingsdatensatzes anfielen, dienten weder unmittelbar noch mittelbar der menschlichen Wahrnehmung und eigneten sich auch nicht dazu, eine solche zu ermöglichen.[378] So seien die Werke während des Trainingsprozesses nicht in einer Form vorhanden, die für die menschliche Wahrnehmung geeignet sei; vielmehr werden die Werke (bzw. Teile davon) zu Datenpunkten einer statistischen Erhebung gemacht. Dabei wird betont, dass es gerade nicht Ziel des Trainings eines KI-Modells sei, den Konsum von Werken durch Dritte zu ermöglichen.[379]

Darüber hinaus wird vertreten, dass die Nutzung von Werken beim Training als freier Werkgenuss zu verstehen sei. Dieser Standpunkt beruht auf einem Analogieschluss zum menschlichen Werkkonsum. Das Training von Sprachmodellen sei mit dem Lernen des menschlichen Gehirns vergleichbar, das ebenfalls zuerst Texte lesen müsse, bevor es selbst Texte verfassen könne. Allerdings sei beim maschinellen Lernen anders als bei Menschen eine technische Vervielfältigung der Trainingsdaten erforderlich.[380]

Teilweise wird darauf hingewiesen, dass die Qualifikation der technischen Vervielfältigung als Vervielfältigung im urheberrechtlichen Sinn weitreichende Konsequenzen hätte, weil das Trainieren von KI-Modellen nur noch möglich wäre, wenn eine unüberschaubare Vielzahl von Rech-

366 Siehe dazu vorne, II.2.a).
367 Dornis/Stober (Fn. 6), 54.
368 Städeli/Mary (Fn. 362), 248; Cherpillod (Fn. 2), 446 f.; Hartmann (Fn. 362), 157 ff., 160; Kübler (Fn. 362), Rz. 5; Marmy-Brändli/Oehri (Fn. 362), 658 f.; Benhamou/Andrijevic (Fn. 362), 204; Brem/Brem (Fn. 362), 416; Rosenthal/Veraldi (Fn. 93), Rz. 20.
369 D. Barrelet/W. Egloff, Das neue Urheberrecht, Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4. Aufl., Bern 2020, URG 10 N 8; M. Rehbinder/L. Haas/K. Uhlig, Orell Füssli Kommentar, URG, Urheberrechtsgesetz mit weiteren Erlassen und internationalen Abkommen, 4. Aufl., Zürich 2022, URG 10 N 3.
370 Rehbinder/Haas/Uhlig (Fn. 369), URG 10 N 3; Barrelet/Egloff (Fn. 369), URG 10 N 8; R. M. Hilty, Urheberrecht, 2. Aufl., Bern 2020, Rz. 292 ff.; I. Cherpillod, in: J. de Werra/P. Gilliéron (Hg.), Commentaire Romand, Propriété intellectuelle, Basel 2013, URG 10 N 3; F. Thouvenin, Immaterialgüterrecht, Zürich 2025, Rz. 613.
371 Marmy-Brändli/Oehri (Fn. 362), 658; Hartmann (Fn. 362), 158; Berger (Fn. 362), Rz. 7.
372 Marmy-Brändli/Oehri (Fn. 362), 659, die zugleich die Auffassung zurückweisen, die urheberrechtliche Relevanz technischer Vervielfältigungen allein vom Vorliegen einer wirtschaftlichen Verwertung abhängig zu machen.
373 Semmelmann (Fn. 2), 639; Benhamou/Andrijevic (Fn. 362), 206.
374 Semmelmann (Fn. 2), 639.
375 Semmelmann (Fn. 2), 639.
376 Rosenthal/Veraldi (Fn. 93), Rz. 22; Thouvenin/Picht (Fn. 2), 517.
377 Rosenthal/Veraldi (Fn. 93), Rz. 23; Thouvenin/Picht (Fn. 2), 517 f.
378 Rosenthal/Veraldi (Fn. 93), Rz. 22 f.; Thouvenin/Picht (Fn. 2), 517.
379 Zum Ganzen Rosenthal/Veraldi (Fn. 93), Rz. 22.
380 Zum Ganzen Rosenthal/Veraldi (Fn. 93), Rz. 24; Semmelmann (Fn. 2), 638; Marmy-Brändli/Oehri (Fn. 362), 658 f.

teinhabern identifiziert und kontaktiert werden könnte und diese die Zustimmung zur Nutzung ihrer Werke erteilen würden. Das wäre mit sehr hohen Transaktionskosten verbunden und wegen der sehr grossen Menge an Werken, die für das Trainieren von KI-Modellen benötigt wird, in der Praxis nahezu unmöglich.[381]

cc) Im trainierten Modell

Bisher haben sich nur wenige Autoren mit der Frage befasst, ob die beim Training verwendeten Werke im trainierten Modell repräsentiert oder gar enthalten sind und ob das trainierte Modell damit Vervielfältigungen der Werke im urheberrechtlichen Sinn enthält. Die Autoren kommen dabei alle zum Schluss, dass die Werke im trainierten Modell nicht mehr repräsentiert seien und damit keine Vervielfältigungen vorliegen.[382] Jedenfalls seien die Werke im trainierten Modell nicht mehr in einer Form enthalten, die in den Schutzbereich der Originalwerke falle.[383] Sobald die Werke «tokenisiert» und nur noch die *Tokens* und die Wahrscheinlichkeiten ihrer Verknüpfung gespeichert seien, liege keine Vervielfältigung mehr vor.[384] Die Werke seien damit nicht mehr «als solche» im trainierten Modell vorhanden.[385] Eine gegenteilige Auffassung wurde für das Schweizer Recht – soweit ersichtlich – bisher nicht vertreten.[386]

dd) Freistellung als vorübergehende Vervielfältigung

Geht man davon aus, dass Werke im Trainingsprozess und allenfalls auch im trainierten Modell vervielfältigt werden, stellt sich die Frage, ob diese Vervielfältigungen als vorübergehende und damit nach Art. 24*a* URG zulässige Vervielfältigungen zu qualifizieren sind.[387]

Ebenso klar wie unbestritten ist, dass die Vervielfältigungen, die beim Erstellen der Trainingsdatensätze anfallen, nicht als vorübergehende Vervielfältigungen angesehen werden können.[388] Unklar ist nach dem gegenwärtigen Stand der Diskussion dagegen, ob die beim Trainingsprozess und/oder im trainierten Modell erstellten Vervielfältigungen als vorübergehende Vervielfältigungen zu qualifizieren sind. Denn die Lehre hat es bisher unterlassen, die beiden Fragestellungen hinreichend klar auseinanderzuhalten, sodass oft unklar bleibt, welcher Vorgang beurteilt wird und wie das Ergebnis zu verstehen ist. Die nachfolgende Darstellung der Lehrmeinungen ist deshalb mit einer gewissen Vorsicht zur Kenntnis zu nehmen.

Als vorübergehend und damit zulässig gelten Vervielfältigungen, wenn sie (i) flüchtig oder begleitend sind, (ii) einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen, (iii) ausschliesslich der Übertragung in einem Netz zwischen Dritten durch einen Vermittler oder einer rechtmässigen Nutzung dienen und (iv) keine eigenständige wirtschaftliche Bedeutung haben. Eine Vervielfältigung ist flüchtig, wenn sie besonders kurzlebig ist, und begleitend, wenn sie lediglich der Erleichterung eines anderen Nutzungsvorgangs dient.[389] Bei der Anwendung dieser Tatbestandsmerkmale auf das Training von KI-Modellen sind sich die Autorinnen und Autoren uneinig. Ein Teil der Lehre verneint den flüchtigen und begleitenden Charakter generell,[390] ein anderer Teil ist der Ansicht, dass diese Vervielfältigungen durchaus als flüchtig oder begleitend gelten können.[391] Einigkeit besteht bei der zweiten Voraussetzung. Dass die Kopien einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen ist unbestritten, zumal die technischen Abläufe beim Training ohne Kopien der Werke nicht möglich wären.[392] Bei der dritten Voraussetzung ist fraglich, ob die Vervielfältigung einer rechtmässigen Nutzung dient. Eine solche liegt vor, wenn der Rechteinhaber der Nutzung zugestimmt hat oder die Nutzung von einer Schranke freigestellt wird.[393] Die Beurteilung dieser Voraussetzung wird allerdings meist offengelassen.[394] Die vierte Voraussetzung sieht vor, dass den Kopien keine eigenständige wirtschaftliche Bedeutung zukommen darf. Diese Voraussetzung ist nach einem Teil der Lehre nicht erfüllt, weil den Vervielfältigungen regelmässig eine wesentliche wirtschaftliche Bedeutung zukomme.[395] Gegen die Qualifikation als vorübergehende Vervielfältigung spreche auch eine historische Auslegung:[396] Denn bei der Einführung der Wissenschaftsschranke sei angenommen worden, dass die Vorgänge beim Text und Data Mining nicht

381 Thouvenin/Picht (Fn. 2), 517.
382 Rosenthal/Veraldi (Fn. 93), Rz. 27; Thouvenin/Picht (Fn. 2), 517; so auch Städeli/Mary (Fn. 362), 249, sofern sich die urheberrechtlich geschützten Werke auf einem externen Cloud-Server befinden.
383 Rosenthal/Veraldi (Fn. 93), Rz. 27, die auch dann eine Vervielfältigung verneinen, wenn in den Trainingsdaten enthaltene Werke bei einer Memorisierung im Output wieder auftauchen.
384 Cherpillod (Fn. 2), 446, der die Situation mit einem Text vergleicht, dessen Wörter zerschnitten, mit Werten zur Wahrscheinlichkeit ihrer Verknüpfung versehen und in einen «Beutel» gelegt werden. In diesem Fall sei das Risiko, dass der ursprüngliche Text aus den isolierten Wörtern und deren Verknüpfungswahrscheinlichkeiten rekonstruiert werden könne, sehr gering.
385 Rosenthal/Veraldi (Fn. 93), Rz. 27; Thouvenin/Picht (Fn. 2), 517.
386 Anders aber Dornis/Stober (Fn. 6), 80 ff.; J. Quang, Does Training AI Violate Copyright Law?, Berkeley Technology Law Journal 2021; 1413 ff.; de la Durantaye (Fn. 171), 5 f.; P. J. Pesch/R. Böhme, Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern, GRUR 2003,1004 f.; ebenso von Welser (Fn. 362), 433.
387 Zur umstrittenen Frage, ob Art. 24a URG als Schranke oder als Konkretisierung des Begriffs der Vervielfältigung zu verstehen ist, siehe Hilty (Fn. 370), Rz. 473, der die Bestimmung als Konkretisierung des Vervielfältigungsbegriffs versteht; ebenso Thouvenin (Fn. 370), Rz. 621; als Schranke qualifizierend dagegen Botschaft URG 2006, BBl 2006, 3389 ff., 3403 und 3430 f.; R. Oertli, in: B. K. Müller/R. Oertli (Hg.), Urheberrechtsgesetz (URG) 2. Aufl., Bern 2012, URG 24a N 3.
388 Marmy-Brändli/Oehri (Fn. 362), 660; Hartmann (Fn. 362), 160 f.; Gilliéron (Fn. 362), Intelligence artificielle, 22.
389 Rehbinder/Haas/Uhlig (Fn. 369), URG 24a N 4.
390 Gilliéron (Fn. 362), Intelligence artificielle, 22; Kübler (Fn. 362), Rz. 8.
391 Marmy-Brändli/Oehri (Fn. 362), 660; Salvadé (Fn. 362), 89.
392 Marmy-Brändli/Oehri (Fn. 362), 661.
393 Rehbinder/Haas/Uhlig (Fn. 369), URG 24a N 8.
394 Eher dafür Marmy-Brändli/Oehri (Fn. 362), 661; eher dagegen Cherpillod (Fn. 2), 447; Gilliéron (Fn. 362), Intelligence artificielle, 22 f.
395 Marmy-Brändli/Oehri (Fn. 362), 661, mit Verweis auf die hohen Investitionen; Cherpillod (Fn. 2), 447, stellt darauf ab, dass die Modelle der Herstellung einer kommerziellen Anwendung dienen, die Bilder, Text oder andere Inhalte generieren; Salvadé (Fn. 362), 89.
396 Siehe dazu Marmy-Brändli/Oehri (Fn. 362), 661; Isler (Fn. 362), Rz. 219 ff., Rz. 228.

unter Art. 24*a* URG fielen, weshalb die Einführung der neuen Schranke von Art. 24*d* URG erforderlich erschien.[397] Analoges gelte auch für das Training von KI.[398]

b) Zugänglichmachen von Werken

Geht man – entgegen der heute herrschenden Lehre – davon aus, dass die trainierten Modelle Vervielfältigungen von Werken enthalten,[399] dann stellt sich die Frage, ob mit dem Zugänglichmachen der Modelle auch die darin enthaltenen Werke im Sinn von Art. 10 Abs. 2 lit. c URG zugänglich gemacht werden. Soweit ersichtlich, hat sich in der Schweiz bisher nur ein Autor zu dieser Frage geäussert. Nach seiner – allerdings nicht näher begründeten – Auffassung führt das Training von Modellen mit urheberrechtlich geschützten Werken zu einem Zugänglichmachen dieser Werke in den trainierten Modellen.[400]

c) Änderung von Werken

Ein Teil der Lehre versteht die Umgestaltung der Werke bei der Vorbereitung des Trainingsmaterials, bspw. die Umwandlung der Texte in *Tokens*, als Bearbeitung im Sinn von Art. 11 Abs. 1 lit. a URG.[401] Ein anderer Teil der Lehre verneint das Vorliegen von Änderungen oder Bearbeitungen im urheberrechtlichen Sinn, weil vergleichbare Umwandlungen bei jeder digitalen Vervielfältigung auftreten, bspw. auch wenn ein Text gescannt und mit einer Texterkennung bearbeitet werde.[402] Solche technisch bedingten Umwandlungen seien keine Bearbeitungen im Sinn des Urheberrechts.[403]

3. Schranken

Die Vervielfältigungen, die beim Training von KI-Modellen entstehen, können möglicherweise durch eine Schranke freigestellt werden. Kontrovers diskutiert wird namentlich, ob die Schranke für den internen Gebrauch (Art. 19 Abs. 1 lit. c URG) oder die Wissenschaftsschranke (Art. 24*d* URG) greift.

a) Interner Gebrauch

Die Schranke zugunsten des internen Gebrauchs stellt das Vervielfältigen von Werken in Betrieben, öffentlichen Verwaltungen, Instituten, Kommissionen und ähnlichen Einrichtungen für die interne Information und Dokumentation frei (Art. 19 Abs. 1 lit. c URG). Nicht freigestellt ist allerdings die vollständige oder weitgehend vollständige Vervielfältigung im Handel erhältlicher Werkexemplare (Art. 19 Abs. 3 lit. a URG).

In der Literatur besteht Einigkeit darüber, dass die Nutzung von Werken beim Training von KI-Modellen nicht als interner Gebrauch freigestellt werden kann. Der Anwendung der Schranke stehe entgegen, dass Werke beim Training von KI-Modellen in aller Regel nicht für die interne Information oder Dokumentation vervielfältigt werden, sondern für das Generieren von Output, der in der Regel auch an Dritte ausserhalb der Organisation gerichtet sei.[404] Diesem Standpunkt wird zwar entgegengehalten, dass jede betriebsinterne Nutzung einem gewissen kommerziellen Zweck diene.[405] Auch aus Vervielfältigungen, die zunächst der Information dienten, würden die daraus gewonnenen Informationen weiterverarbeitet, um entsprechenden Output zu erzeugen.[406] Die Literatur ist sich aber einig, dass die Schranke selbst dann nicht greifen würde, wenn man die Nutzung als interne Information oder Dokumentation qualifizieren sollte, weil zumindest in einem ersten Schritt vollständige Kopien von Werken erstellt werden, die in der Regel im Handel erhältlich seien, was aufgrund der Gegenausnahme von Art. 19 Abs. 3 lit. a URG ausdrücklich nicht freigestellt sei.[407]

b) Wissenschaftsschranke

Der primäre Anknüpfungspunkt für eine mögliche Freistellung der Nutzung von Werken für das Training von KI-Modellen ist die Schranke für die Verwendung von Werken zum Zweck der wissenschaftlichen Forschung (Art. 24*d* URG). Diese stellt Vervielfältigungen von Werken frei, wenn die Vervielfältigungen durch die Anwendung eines technischen Verfahrens bedingt sind und zu den zu vervielfältigenden Werken ein rechtmässiger Zugang besteht (Art. 24*d* Abs. 1 URG). Die Schranke greift allerdings nicht für Computerprogramme (Art. 24*d* Abs. 3 URG). Die Wissenschaftsschranke sieht eine volle Freistellung vor, die Rechteinhaber erhalten für die Nutzung ihrer Werke also keine Vergütung. Unbestritten ist, dass die Schranke grundsätzlich alle Arten von Forschung erfasst, auch die angewandte Forschung und die Forschung zu kommerziellen Zwecken.[408] Kontrovers diskutiert wird dagegen, ob die weiteren Voraussetzungen der Schranke erfüllt sind.

Kern der Diskussion ist die Frage, ob und gegebenenfalls inwieweit das Training von KI-Modellen zum Zweck der Forschung erfolgt. Ein Teil der Lehre verneint das Vorliegen eines Forschungszwecks, weil das Training von KI-Modellen nicht der systematischen Suche nach neuen Erkenntnissen diene bzw. die Modelle für kommerzielle Anwen-

397 Siehe dazu BBl 2006, 3431; Marmy-Brändli/Oehri (Fn. 362), 661; Isler (Fn. 362), Rz. 228.
398 Marmy-Brändli/Oehri (Fn. 362), 661.
399 Siehe dazu vorne, IV.2.a).
400 Cherpillod, (Fn. 2), 449.
401 Kübler (Fn. 362), Rz. 6.
402 Rosenthal/Veraldi (Fn. 93), Rz. 37.
403 Rosenthal/Veraldi (Fn. 93), Rz. 37; Thouvenin (Fn. 4), 384.
404 Semmelmann (Fn. 2), 640; Kübler (Fn. 362), Rz. 12; Salvadé (Fn. 362), 89; differenzierend Marmy-Brändli/Oehri (Fn. 362), 660.
405 Rosenthal/Veraldi (Fn. 93), Rz. 36, mit Verweis auf HGer Zürich vom 6. September 2021, E. 4.3.5, wonach sich der Zweck der Information nicht in der Wissensvermittlung erschöpfen muss, sondern auch nur der Arbeitserleichterung dienen darf.
406 Rosenthal/Veraldi (Fn. 93), Rz. 36.
407 Marmy-Brändli/Oehri (Fn. 362), 660; Salvadé (Fn. 362), 89; Rosenthal/Veraldi (Fn. 93), Rz. 36.
408 Botschaft, BBl 2018, 628 f.

dungen genutzt werden.[409] Nach anderer Auffassung stellt die Schranke auch Projekte frei, die Forschung und kommerzielle Anwendungen miteinander verknüpfen, soweit das systematische Anlernen des Systems der Produktentwicklung diene und die gewonnenen Erkenntnisse für weitere Produktentwicklungen verallgemeinert würden.[410] Dann sei es auch unerheblich, wenn ein Hauptziel des Projekts in der späteren Kommerzialisierung der Anwendung liege.[411]

Unproblematisch ist dagegen das Tatbestandsmerkmal der technischen Bedingtheit, nach welchem die Vervielfältigungen im Rahmen eines technischen Verfahrens zu Forschungszwecken erfolgen oder für die Anwendung eines solchen Verfahrens notwendig sein müssen. Die Lehre ist sich einig, dass die beim Training von KI-Modellen entstehenden Vervielfältigungen diese Voraussetzung erfüllen.[412]

Nicht vollständig geklärt ist die Bedeutung des dritten Tatbestandsmerkmals, nach dem ein rechtmässiger Zugang zu den zu vervielfältigenden Werken bestehen muss. Diese Voraussetzung ist nach der Lehre erfüllt, wenn die Werke gekauft, lizenziert, ausgeliehen, rechtmässig heruntergeladen oder sonst wie erlaubterweise zur Verfügung gestellt worden sind.[413] Ein Teil der Lehre geht der Frage nach, ob auch frei im Internet zugängliche Inhalte im Rahmen der Wissenschaftsschranke verwertet werden dürfen.[414] Dafür spreche, dass der europäische Gesetzgeber in Erwägungsgrund 14 der DSM-RL für Text und Data Mining explizit festgehalten habe, dass als rechtmässiger Zugang auch der Zugang zu im Internet frei verfügbaren Inhalten gelte.[415] Dieser Auffassung sei auch für die Schweiz zu folgen, weil Rechteinhaber, die ihre Werke selbst im Internet frei zur Verfügung stellen, davon ausgehen müssen, dass diese nicht nur von Menschen konsumiert, sondern auch für das KI-Training genutzt werden.[416]

Auf der Grundlage einer weiten Auslegung des Forschungszwecks geht ein Teil der Lehre davon aus, dass die Nutzung von Werken beim Training von KI-Modellen von der Wissenschaftsschranke freigestellt werden kann.[417] Dieses Ergebnis wird aber teilweise kritisiert, weil es nicht mit dem Dreistufentest vereinbar sei, zumal eine vergütungsfreie Nutzung die normale Verwertung der Werke untergraben würde.[418]

4. Erweiterte Kollektivlizenz

In der Literatur wird teilweise vertreten, dass die Nutzung von Werken beim Training von KI-Modellen durch das Erteilen erweiterter Kollektivlizenzen (Art. 43*a* URG) gegen Bezahlung einer angemessenen Vergütung erlaubt werden könne.[419]

Die Voraussetzungen für die Erteilung einer erweiterten Kollektivlizenz dürften nach dieser Auffassung in der Regel erfüllt sein, zumal beim Training von KI-Modellen stets eine grössere Anzahl von Werken genutzt werde, die Nutzung die (bisherige) normale Verwertung der Werke nicht beeinträchtige und die Verwertungsgesellschaften im Anwendungsbereich der erweiterten Kollektivlizenz eine massgebende Anzahl von Rechteinhabern vertreten.[420] Voraussetzung sei allerdings, dass nur veröffentlichte Werke für das Training verwendet werden.

5. Erkenntnisse

Die Darstellung des Standes der Diskussion hat gezeigt, dass sich die Schweizer Lehre bei einigen Fragen weitgehend einig ist, während andere umstritten oder bisher kaum untersucht worden sind. Nachfolgend werden die wichtigsten Erkenntnisse festgehalten, die sich aus der Auseinandersetzung mit den technischen Grundlagen, dem Blick auf die ausländischen Rechtsordnungen und dem Stand der Schweizer Lehre ergeben. Diese bilden den Ausgangspunkt für die abschliessende Skizzierung der Lösungsansätze:

Klar und unbestritten ist, dass die Speicherung der für das Training erforderlichen Werke in einer Datenbank zu *Vervielfältigungen* der Werke im Sinn von Art. 10 Abs. 2 lit. a URG führt.[421] Dasselbe gilt für die Vorgänge im Trainingsprozess. Die beim Training anfallenden Vervielfältigungen sind aber als vorübergehende Vervielfältigungen im Sinn von Art. 24*a* URG zu qualifizieren.[422] Die Lehre ist sich zudem einig, dass das Training nicht zu einer Vervielfältigung der beim Training verwendeten Werke im trainierten Modell führt.[423] Auch der Erstautor dieses Beitrags hat bisher den Standpunkt vertreten, dass die Werke beim Training nicht

409 Rehbinder/Haas/Uhlig (Fn. 369), URG 24d N 6; nach Semmelmann (Fn. 2), 640, und Kübler (Fn. 362), Rz. 17, werden die Grenzen der Wissenschaftsschranke überschritten, wenn mit den Werken eine Internetanwendung «fabriziert» (Kübler) wird; Cherpillod (Fn. 2), 447, verneint einen Forschungszweck, wenn die Vervielfältigung für die *«réalisation d'un outil automatisé à vocation commerciale»* erfolgt; Marmy-Brändli/Oehri (Fn. 362), 662, äussern sich insgesamt kritisch zum Vorliegen eines Forschungszweckes; nach Städeli/Mary (Fn. 362), 249, soll die Wissenschaftsschranke Anwendung finden, wenn das Training von generativen KI-Modellen ausschliesslich für die wissenschaftliche Forschung erfolgt.

410 Rosenthal/Veraldi (Fn. 93), Rz. 36, die festhalten, dass ein KI-Modell per Definition eine Erkenntnis sei, weil es in Form seiner Parameter das Ergebnis einer Analyse der Trainingsinhalte verkörpere und anhand von generalisierten Regeln Auskunft darüber gebe, wie Sprache verwendet werde, und gestützt darauf neue Texte generiere; sinngemäss ebenso Marti (Fn. 362), 151 f.

411 Rosenthal/Veraldi (Fn. 93), Rz. 36.

412 Marmy-Brändli/Oehri (Fn. 362), 663; Marti (Fn. 362), 151.

413 Rehbinder/Haas/Uhlig (Fn. 369), URG 24d N 2; Barrelet/Egloff (Fn. 369), URG 24d N 7; Marmy-Brändli/Oehri (Fn. 362), 663.

414 Gilliéron (Fn. 362), 23; Marmy-Brändli/Oehri (Fn. 362), 663.

415 Erwägungsgrund 14 DSM-RL «Als rechtmässiger Zugang sollte auch der Zugang zu im Internet frei verfügbaren Inhalten gelten».

416 Marmy-Brändli/Oehri (Fn. 362), 663; im Ergebnis auch Gilliéron (Fn. 362), 23 f.

417 Marmy-Brändli/Oehri (Fn. 362), 663 f.; Marti (Fn. 362), 151 ff.; a.M. Semmelmann (Fn. 2), 640; Salvadé (Fn. 362), 88 f.

418 Semmelmann (Fn. 2), 640; Kübler (Fn. 362), Rz. 17.

419 Salvadé (Fn. 362), 90 ff.; Ders. (Fn. 362), 1 ff., 5; Semmelmann (Fn. 2), 642; siehe auch Thouvenin/Picht (Fn. 2), 515, die diese Möglichkeit aber sogleich als kaum geeignet verwerfen.

420 Ebenso Salvadé (Fn. 362), 90 ff.

421 Siehe dazu vorne, IV.2.a) aa).

422 Siehe dazu vorne, IV.2.a) bb).

423 Siehe dazu vorne, IV.2.a) cc).

«als solche» in KI-Modellen gespeichert werden.[424] Die vertiefte Auseinandersetzung mit den technischen Grundlagen hat nun aber gezeigt, dass sich dieser Standpunkt nicht aufrechterhalten lässt. Auch wenn das Ziel des Trainings von KI-Modellen (anders als bei der Speicherung in einer Datenbank) nicht darin besteht, die Werke aus dem Modell erneut abrufen zu können, sondern das Modell zu befähigen, neue Inhalte zu generieren, so werden die Werke im Modell doch in einer Form repräsentiert, die es erlaubt, die beim Training verwendeten Werke (oder Teile davon) dem trainierten Modell durch geeignete Prompts in identischer oder nahezu identischer Form zu entnehmen.[425] Dies belegt, dass die Werke in den trainierten Modellen in einer Weise repräsentiert sind, die urheberrechtlich als Vervielfältigung zu qualifizieren ist. Diese Repräsentationen sind allerdings nicht als Umgestaltungen oder Bearbeitungen im Sinn des Urheberrechts zu qualifizieren. Vielmehr handelt es sich um technische Umwandlungen,[426] die sich zwar graduell, aber nicht kategorisch von Umwandlungen unterscheiden, die bei allen Formen der digitalen Speicherung erfolgen.

Geht man davon aus, dass die beim Training verwendeten Werke in den trainierten Modellen vervielfältigt sind, dann führt das *Zugänglichmachen* der Modelle möglicherweise auch zu einem Zugänglichmachen der in den Modellen repräsentierten Werke. Das Recht des Zugänglichmachens (Art. 10 Abs. 2 lit. c URG) erfasst zwar in erster Linie die Nutzung von Werken über das Internet, namentlich das Bereitstellen von Werken zum Streaming oder zum Download. Mit welchen technischen Mitteln Werke zugänglich gemacht werden, ist aber irrelevant.[427] Zugänglich macht ein Werk, wer den Zugriff auf ein Werkexemplar und die Wiedergabe des Werks zum Zweck der Wahrnehmung (Werkgenuss) ermöglicht.[428] Die zentrale Handlung besteht dabei im Bereithalten von Inhalten zum Abruf durch Dritte.[429] Kein Zugänglichmachen liegt dagegen vor, wenn der Zugriff auf Werke mit geeigneten technischen Massnahmen unterbunden wird.[430] Denn in diesem Fall werden nicht Werke bereitgestellt, um Dritten den Zugriff auf und die Wahrnehmung der Werke zu ermöglichen, sondern vielmehr Massnahmen getroffen, um den Zugriff und die Wahrnehmung zu verhindern. Wendet man dieses Verständnis des Rechts auf Zugänglichmachen auf die vorliegende Fragestellung an, dann zeigt sich, dass das Zugänglichmachen von KI-Modellen nur dann als Zugänglichmachen der in diesen Modellen repräsentierten Werke zu verstehen ist, wenn die Anbieter keine geeigneten technischen Massnahmen getroffen haben, um die Ausgabe von Werken (oder Teilen davon) im Output der Modelle zu verhindern.

Da die Nutzung von Werken beim Training von KI-Modellen zu urheberrechtlich relevanten Vervielfältigungen in einer Datenbank und in den trainierten Modellen führt, stellt sich die Frage, ob diese Vervielfältigungen durch eine (oder mehrere) *Schranke(n)* freigestellt sind. Die Lehre ist sich einig, dass die Voraussetzungen der Schranke zugunsten des *internen Gebrauchs* (Art. 19 Abs. 1 lit. c URG) in der Regel nicht erfüllt sind.[431] Die Vervielfältigungen können aber unter Umständen durch die *Wissenschaftsschranke* (Art. 24*d* URG) freigestellt sein. Dies setzt allerdings voraus, dass der Begriff der wissenschaftlichen Forschung weit verstanden wird. Selbst wenn die Schranke auch die kommerzielle Forschung freistellen soll,[432] erscheint es fraglich, ob der Begriff der Forschung so weit ausgelegt werden kann, dass er auch die kommerzielle Entwicklung von KI-Modellen erfasst. Da die Wissenschaftsschranke eine volle Freistellung vorsieht, erhalten die Rechteinhaber für die Nutzung ihrer Werke im Rahmen dieser Schranke keine Vergütung. Das erscheint jedenfalls dann als problematisch, wenn die Werke für das Training kommerzieller KI-Modelle verwendet werden. Mit Blick auf die entstehenden Märkte, auf denen grosse Rechteinhaber Lizenzen für die Nutzung ihrer Werke für das Training kommerzieller KI-Modelle erteilen, dürfte eine volle Freistellung zudem die (zunehmend) normale Verwertung der Werke beeinträchtigen. Eine weite Auslegung der Wissenschaftsschranke, die auch die kommerzielle Entwicklung von KI-Modellen freistellt, lässt sich deshalb kaum mit dem Dreistufentest vereinbaren.[433] Damit zeigt sich, dass die Auslegung der Wissenschaftsschranke mit erheblicher Rechtsunsicherheit verbunden ist, und deren Anwendung auf das Training kommerzieller KI-Modelle nicht zu überzeugenden Ergebnissen führt.

Im Rahmen des geltenden Rechts bleibt damit die Frage, ob die Nutzung von Werken für das Training von KI-Modellen durch das Erteilen von *erweiterten Kollektivlizenzen* (Art. 43*a* URG) erlaubt werden könnte. Mit Blick auf die erwähnten Lizenzmärkte ist allerdings fraglich, ob die Voraussetzungen hierfür erfüllt sind, weil erweiterte Kollektivlizenzen die normale Verwertung der Werke nicht beeinträchtigen dürfen. Selbst wenn man annimmt, dass diese Voraussetzung allgemein (oder zumindest in bestimmten Fällen) erfüllt sein könnte, bleibt die Reichweite des Ansatzes beschränkt. Da erweiterte Kollektivlizenzen immer einer bestimmten Lizenznehmerin erteilt werden, kann dieses Instrument nur verwendet werden, um die Nutzung von Werken für das Training von KI-Modellen durch ein bestimmtes Unternehmen zu ermöglichen, es kann die vorliegende Pro-

424 Thouvenin/Picht (Fn. 2), 517; F. Thouvenin, White Paper Urheberrecht; abrufbar unter: ‹www.itsl.uzh.ch/de/Wissenstransfer-und-Veranstaltung/Publikationen.html› (abgerufen am 5. Januar 2026).
425 Siehe dazu vorne, II.5.
426 Siehe dazu: Barrelet/Egloff (Fn. 369), URG 11 N 8; Hilty (Fn. 370), Rz. 401; Rehbinder/Haas/Uhlig (Fn. 369), URG 11 N 6.
427 Thouvenin (Fn. 370), Rz. 628.
428 Rehbinder/Haas/Uhlig (Fn. 369), URG 10 N 20.
429 Hilty (Fn. 370), Rz. 349.
430 Siehe dazu EuGH vom 22. Juni 2021, C-682/18 und C-683/18, wonach der Betreiber einer *Video Sharing*-Plattform, auf der Nutzer geschützte Inhalte rechtswidrig öffentlich zugänglich machen können, keine «öffentliche Wiedergabe» dieser Inhalte im Sinn von Art. 3 Abs. 1 der Richtlinie 2001/29/EG (Fn. 208) vornimmt, sofern er «die geeigneten technischen Massnahmen ergreift, die von einem die übliche Sorgfalt beachtenden Wirtschaftsteilnehmer in seiner Situation erwartet werden können, um Urheberrechtsverletzungen auf dieser Plattform glaubwürdig und wirksam zu bekämpfen».
431 Siehe dazu vorne, IV.3.a).
432 Siehe dazu vorne, IV.3.b).
433 Zur entsprechenden Kritik im europäischen Recht siehe vorne, III.2.a).

blematik aber nicht allgemein lösen. Hinzu kommt, dass es keine Verwertungsgesellschaft gibt, welche die Urheberrechte an Computerprogrammen wahrnimmt. Ein in der Praxis besonders wichtiger Bereich von generativer KI kann damit über das Instrument der erweiterten Kollektivlizenz nicht erfasst werden.

V. Lösungsansätze

1. Ausgangslage

Die geltende Rechtslage ist für alle Beteiligten unbefriedigend und mit Rechtsunsicherheit verbunden. Mit Blick auf Nutzen und Potential von KI erscheint es zentral, die Nutzung von Werken für das Training von KI-Modellen ausdrücklich im URG zu regeln. Die Regelung sollte einen angemessenen Ausgleich der Interessen der Rechteinhaber sowie der Entwickler und Anbieter von KI-Modellen schaffen und dem gesellschaftlichen Interesse an der weiteren Entwicklung dieser Technologie Rechnung tragen. Ein solcher Interessenausgleich liesse sich durch Einführung einer neuen Schranke oder durch eine Revision der bestehenden Wissenschaftsschranke erreichen. In beiden Fällen sollte die Nutzung aller Werkarten freigestellt werden, insb. auch diejenige von Computerprogrammen.

Mit Blick auf die Rechtslage in der EU und im UK bietet sich die Einführung einer *neuen Schranke für das Training von KI-Modellen* an, bei unveränderter Beibehaltung der geltenden Wissenschaftsschranke. Damit liesse sich eine Regelung schaffen, welche die in Art. 3 und Art. 4 DSM-RL angelegte Differenzierung zwischen einer allgemeinen Schranke für Text und Data Mining und einer spezifischen Schranke für die wissenschaftliche Forschung (auch) im Schweizer Recht etablieren würde. Die Differenzierung auf Ebene der Tatbestände würde zudem erlauben, die Nutzungskonstellationen unterschiedlichen Regelungen zu unterstellen, namentlich die Nutzung von Werken für die wissenschaftliche Forschung weiterhin ohne Vergütung freizustellen, für andere Nutzungen für das Training von KI-Modellen aber eine Vergütung vorzusehen.

Bei der Ausgestaltung der Schranke stehen drei Ansätze im Vordergrund: (i) eine volle Freistellung mit Nutzungsvorbehalt (Opt-out), (ii) eine gesetzliche Lizenz und (iii) eine gesetzliche Lizenz mit Nutzungsvorbehalt (Opt-out). Diese Ansätze werden nachfolgend kurz skizziert. Zunächst ist aber zu klären, wie die Tatbestandsmerkmale einer Schranke ausgestaltet und welche Nutzungen freigestellt werden sollten.

2. Tatbestandsmerkmale und freigestellte Nutzungen

Von zentraler Bedeutung ist das Tatbestandsmerkmal, das den sachlichen Anwendungsbereich der Schranke festlegt. Dieser kann durch eine Wendung wie «für das Training von Modellen der Künstlichen Intelligenz (KI)» umschrieben werden. Der Begriff der KI sollte allerdings nicht im URG definiert werden. Mit der Ratifikation der KI-Konvention des Europarates[434] kennt das Schweizer Recht bereits eine Definition von KI, auf die sich das URG stützen kann. Der Verzicht auf eine Definition in der Schranke hat den Vorteil, dass deren Auslegung und Anwendung bei Bedarf an Weiterentwicklungen der Technologie angepasst werden kann, ohne dass der Gesetzgeber aktiv werden muss. Dieser Ansatz hat sich beim Begriff der Computerprogramme bewährt, der im URG bewusst nicht definiert worden ist. Alternativ könnte auch eine technologieneutrale Formulierung verwendet werden, bspw. die Wendung «für die Entwicklung von algorithmischen Systemen».

Wie bei der Wissenschaftsschranke des URG (Art. 24*d* URG) und bei der allgemeinen Schranke für Text und Data Mining in Art. 4 DSM-RL sollte die Freistellung auf Handlungen beschränkt sein, «die durch das technische Verfahren bedingt» sind oder «zum Zweck des Trainings von KI-Modellen» erfolgen. Wie in diesen Bestimmungen sollte die Schranke zudem nur greifen, wenn «ein rechtmässiger Zugang zu den Werken» besteht. Anders als bei der Wissenschaftsschranke sollte die Schranke die Nutzung aller Werkarten freistellen, insb. auch diejenige von Computerprogrammen.

Die Schranke sollte alle Vervielfältigungen von Werken freistellen, die für das Training von KI-Modellen erforderlich sind, die Speicherung in einer Datenbank ebenso wie die Vervielfältigungen der Werke im trainierten Modell.[435] Zudem sollte die Schranke auch nach dem Training erfolgende Vervielfältigungen der Modelle erfassen. Andernfalls würden Dritte, die ein trainiertes Modell erwerben und auf ihren eigenen Rechnern speichern (bspw. durch Download von einer *Open Source*-Plattform wie *Hugging Face*), eine Vielzahl von Urheberrechtsverletzungen begehen. Diese Konstellation ist vor allem bei kleineren Modellen relevant, während grössere Modelle den Nutzerinnen und Nutzern in der Regel nur über ein *User Interface* zugänglich gemacht werden, von diesen aber nicht vervielfältigt werden können.

Die Freistellung durch die Schranke ist auf das Training von KI-Modellen und die Vervielfältigung der trainierten Modelle zu beschränken. Nicht freigestellt werden sollte die Nutzung von Werken als *Input* bei der Verwendung von KI-Systemen durch deren Nutzer, bspw. die Eingabe eines Werkes in einem Prompt oder die Nutzung von Werken bei der sog. *Retrieval Augmented Generation (RAG)*. Diese Nutzungen sind nur zulässig, wenn sie von einer anderen Schranke freigestellt werden, bspw. von derjenigen des Privatgebrauchs (Art. 19 Abs. 1 lit. a URG). Ebenfalls nicht freizustellen ist der *von den trainierten Modellen erzeugte Output*. Generiert ein KI-System (bspw. ChatGPT) einen Output, der mit einem bestehenden Werk identisch oder einem solchen so ähnlich ist, dass er in dessen Schutzbereich fällt, liegt eine Urheberrechtsverletzung vor. Anderes gilt nur, wenn das Er-

434 Committee on Artificial Intelligence (CAI), Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law vom 5. September 2024, abrufbar unter: ‹https://rm.coe.int/1680afae3c› (abgerufen am 5. Januar 2026).

435 Siehe dazu vorne, IV.2.a).

stellen des Outputs von einer Schranke freigestellt wird, bspw. weil ein Privatgebrauch vorliegt (Art. 19 Abs. 1 lit. a URG) oder das Zitatrecht greift (Art. 25 URG).[436] Generiert ein KI-System aufgrund eines simplen Prompts einen urheberrechtsverletzenden Output, lässt sich dem Modell also auf einfache Weise ein geschütztes Werk (oder Teile davon) entnehmen, liegt ein Zugänglichmachen der Werke durch den Anbieter des KI-Systems vor.[437] In allen anderen Fällen wird der Nutzer des KI-Systems als primärer Verletzer anzusehen sein, wenn er mit dem System einen urheberrechtsverletzenden Output generiert. Der Anbieter des KI-Systems kann aber als indirekter Verletzer ebenfalls in Anspruch genommen werden. Die Gefahr einer direkten oder indirekten Haftung für Urheberrechtsverletzungen bei der Verwendung der Modelle sollte massgebliche Anreize setzen, damit die Anbieter hinreichende technische Massnahmen implementieren, um die Memorisierung von Werken in den Modellen und deren Wiedergabe im Output zu verhindern.

3. Volle Freistellung mit Nutzungsvorbehalt (Opt-out)

Die Nutzung von Werken für das Training von KI-Modellen könnte vollständig freigestellt, also vergütungsfrei erlaubt werden. Um den Interessen der Rechteinhaber angemessen Rechnung zu tragen, müssten diese die Nutzung ihrer Werke aber durch Anbringen eines Nutzungsvorbehalts (Opt-out) verhindern können. Dieser Ansatz entspricht der Regelung von Art. 4 DSM-RL und damit dem geltenden Recht in den Mitgliedstaaten der EU.

Dogmatisch würde damit eine neue Art von Schranke geschaffen, die den bestehenden Ansatz der vollen Freistellung mit einem Element der erweiterten Kollektivlizenz (Art. 43*a* URG) kombiniert. Im Ergebnis würde eine nicht zwingende Schranke geschaffen. Anders als das Bundesgericht vor einigen Jahren in einem *obiter dictum* festgehalten hat,[438] ist das allerdings nicht neu; vielmehr kennt das URG schon heute nicht zwingende Schranken.[439]

Eine volle Freistellung mit einem Nutzungsvorbehalt würde die Rechtssicherheit erhöhen, die Transaktionskosten verringern, Innovation ermöglichen und einen angemessenen Ausgleich der Interessen aller Beteiligten schaffen. Die Wirkung des Ansatzes würde allerdings wesentlich davon abhängen, wie viele und welche Rechteinhaber einen Nutzungsvorbehalt anbringen. Während ein Opt-out von Rechteinhabern an grossen Werkbeständen Anlass für das Verhandeln und Abschliessen von Lizenzen für die Nutzung beim Training von KI-Modellen sein könnte, würde es bei den meisten Rechteinhabern lediglich dazu führen, dass ihre Werke nicht genutzt werden dürften. Da die Transaktionskosten für das Verhandeln und Abschliessen von Lizenzverträgen mit einer Vielzahl kleinerer Rechteinhaber für die Entwickler von KI-Modellen in der Regel zu hoch sein werden, könnten diese Rechteinhaber die Nutzung ihrer Werke nur dulden oder mit einem Opt-out verbieten, aber nicht gegen Bezahlung einer angemessenen Vergütung erlauben.

Für eine volle Freistellung mit Opt-out spricht, dass dieser Ansatz dem Recht der EU-Mitgliedstaaten entsprechen würde. Die Übernahme der europäischen Regelung hätte den Vorteil, dass in der Schweiz nach Massgabe des URG trainierte KI-Modelle auch die Vorgaben der KI-VO einhalten würden, nach denen die Anbieter von KI-Modellen mit allgemeinem Verwendungszweck (GPAI) eine Strategie zur Einhaltung der Vorgaben des Urheberrechts der EU und insb. zur Einhaltung des Nutzungsvorbehalts (Opt-out) bei der Verwendung von Werken für das Text und Data Mining (Art. 4 Abs. 3 DSM-RL) «auf den Weg bringen» müssen (Art. 53 Abs. 1 Bst. c KI-VO).[440] Diese Vorgaben gelten allerdings nur für die Anbieter von GPAI und damit nur für wenige Anbieter und Modelle. Da die Anbieter von GPAI ihre Modelle in aller Regel auch auf dem europäischen Markt anbieten werden, müssen sie die Vorgaben der KI-VO ohnehin einhalten. Ob das Schweizer Recht der europäischen Regelung entspricht, ist deshalb irrelevant.

Gegen die Einführung einer vollen Freistellung mit Opt-out spricht, dass es bisher nicht gelungen ist, einen Standard für das Anbringen eines maschinenlesbaren Nutzungsvorbehalts zu entwickeln.[441] Das erstaunt insofern, als eine Lösung aus technischer Sicht nicht übermässig komplex erscheint. Da das EU-Recht eine solche Möglichkeit fordert und eine Lösung durchaus möglich scheint, kann angenommen werden, dass es in absehbarer Zeit gelingen wird, einen funktionierenden Standard zu etablieren. Gelingt dies nicht, sollte der Schweizer Gesetzgeber auf die Einführung eines Nutzungsvorbehalts verzichten.

Sollte sich der Gesetzgeber für eine volle Freistellung mit Opt-out entscheiden, müssten die Rechteinhaber nach dem Anbringen eines Nutzungsvorbehalts prüfen können, ob ihre Werke dennoch für das Training von KI-Modellen verwendet wurden. Diese Transparenz liesse sich durch eine Informationspflicht der Anbieter von KI-Modellen sicherstellen, wie sie in der EU nach Art. 53 Abs. 1 Bst. d KI-VO für GPAI gilt.[442] Sollte, wie in der EU, nur eine hinreichend detaillierte Zusammenfassung der für das Training genutzten Werke und keine Liste mit den konkret verwendeten Werken verlangt werden, könnte es sinnvoll sein, die Informationspflicht der Anbieter von KI-Modellen durch einen Auskunftsanspruch der Rechteinhaber zu ergänzen, der es diesen erlauben würde, von den Anbietern konkrete Angaben über die Nutzung ihrer Werke zu erhalten.

4. Gesetzliche Lizenz

Eine gesetzliche Lizenz würde die Nutzung von Werken für das Training von KI-Modellen erlauben und einer Vergütungspflicht unterstellen. Dieser Ansatz hätte den Vorteil,

436 Siehe dazu Thouvenin (Fn. 4), 390.
437 Siehe dazu vorne, IV.5.
438 BGE 127 III 26 ff. E. 4; ebenso Barrelet/Egloff (Fn. 369), Vor URG 19 N 6.
439 Siehe dazu Thouvenin (Fn. 370), Rz. 37, m.w.H.
440 Siehe dazu vorne, III.2.
441 Siehe dazu vorne, III.2.b).
442 Siehe dazu vorne, III.2.c).

dass alle Rechteinhaber eine Vergütung erhielten, sofern sie Mitglied einer Verwertungsgesellschaft wären.

Eine gesetzliche Lizenz liesse sich ohne Weiteres in den heutigen Schrankenkatalog des URG integrieren, der bereits mehrere gesetzliche Lizenzen kennt. Die Berechnung und die Höhe der Vergütung wären in einem gemeinsamen Tarif (GT) der Verwertungsgesellschaften zu regeln, der bei Bedarf an die technische Entwicklung und die Marktverhältnisse angepasst werden könnte. Ein Tarif würde erlauben, relevante Differenzierungen vorzunehmen, um bspw. die Entwicklung kleinerer KI-Modelle durch KMU und Start-ups tariflich zu begünstigen. Denkbar wären auch Differenzierungen zwischen Training und Finetuning und zwischen generativer und diskriminativer KI, die dem Umstand Rechnung tragen würden, dass der Output diskriminativer KI die Verwertung der beim Training genutzten Werke nicht konkurrenziert.

Diesen Vorteilen stehen gewisse Nachteile gegenüber. Da bei den grossen KI-Modellen (insb. LLM) äusserst grosse Mengen von Werken von einer unüberschaubaren Vielzahl von Rechteinhabern für das Training verwendet werden, wäre der Ertrag für die meisten Rechteinhaber verschwindend klein. Zudem würde der Erlass einer gesetzlichen Lizenz dazu führen, dass die Inhaber der Urheberrechte an grossen Werkmengen (bspw. Medienhäuser, Verlage oder Bildagenturen) die Nutzung ihrer qualitativ hochwertigen Datensätze für das Training von KI-Modellen nicht mehr vertraglich erlauben und monetarisieren könnten. Der Erlass einer gesetzlichen Lizenz könnte damit die Entwicklung eines funktionierenden Lizenzmarktes verhindern.

5. Gesetzliche Lizenz mit Nutzungsvorbehalt (Opt-out)

Angesichts der Vor- und Nachteile einer vollen Freistellung mit Opt-out und einer gesetzlichen Lizenz könnte sich eine Kombination der beiden Ansätze als Lösung anbieten, also eine gesetzliche Lizenz mit einem Nutzungsvorbehalt (Opt-out). Dieser Ansatz würde die Nutzung von Werken für das Training von KI-Modellen grundsätzlich erlauben und einer Vergütungspflicht unterstellen. Die Vergütung wäre auch hier in einem gemeinsamen Tarif (GT) zu regeln.[443] Die Rechteinhaber könnten die Nutzung ihrer Werke für das Training von KI-Modellen aber durch ein Opt-out verbieten. In diesem Fall würden sie zwar keine Vergütung über den GT erhalten, sie könnten aber mit den Entwicklern von KI-Modellen auf der Grundlage des nun geltenden Nutzungsverbots Vertragsverhandlungen führen und die Nutzung ihrer Werke im Rahmen einer vertraglich vereinbarten Lizenz erlauben. Dieser Ansatz wäre zwar dogmatisch neu, er würde aber lediglich Elemente kombinieren, die das schweizerische Urheberrecht bereits kennt.

Die Verbindung einer gesetzlichen Lizenz mit einem Opt-out könnte sinnvolle Anreize setzen, um unerwünschte Wirkungen zu verhindern, die bei den anderen Ansätzen eintreten: Bei der *vollen Freistellung mit Opt-out* besteht die Gefahr, dass viele Rechteinhaber einen Nutzungsvorbehalt anbringen werden. Aus ihrer Sicht gibt es wenige Gründe, die Nutzung ihrer Werke für das KI-Training zuzulassen, zumal sie dafür keine Vergütung erhalten und der Output der mit ihren Werken trainierten Modelle die Marktchancen ihrer eigenen Werke beeinträchtigen kann. Da die meisten Rechteinhaber nicht in der Lage sein werden, die Nutzung ihrer Werke nach dem Opt-out auf vertraglicher Grundlage und gegen Bezahlung einer Vergütung zuzulassen, wird die Nutzung dieser Werke unterbleiben. Eine *gesetzliche Lizenz ohne Opt-out* nimmt dagegen den Inhabern der Rechte an grossen Werkbeständen die Möglichkeit, die Nutzung ihrer Werke für das Training von KI-Modellen auf vertraglicher Grundlage zu monetarisieren. Der Ansatz einer *gesetzlichen Lizenz mit Opt-out* würde den grossen Rechteinhabern die Möglichkeit der vertraglichen Verwertung belassen und bei den Inhabern von kleineren Werkbeständen Anreize setzen, auf ein Opt-out zu verzichten, weil sie damit den Anspruch auf die tarifliche Vergütung verlieren würden. Bei diesem Ansatz müssten die Rechteinhaber also aktiv werden, um auf Einkommen zu verzichten, das ihnen sonst ohne Weiteres zufallen würde, sofern sie Mitglied einer Verwertungsgesellschaft sind. Gewisse Rechteinhaber mögen dies aus Prinzip und Überzeugung dennoch tun, weil sie verhindern möchten, dass (grosse) KI-Anbieter ihre Werke nutzen. Unterstellt man ein rationales und nutzenmaximierendes Verhalten, kann aber angenommen werden, dass die Kombination von gesetzlicher Lizenz und Opt-out dazu führen wird, dass die meisten Rechteinhaber auf ein Opt-out verzichten werden, wenn sie nicht in der Lage sind, auf vertraglicher Grundlage ein höheres Entgelt für die Nutzung ihrer Werke zu erzielen. Der Ansatz scheint deshalb geeignet, einen angemessenen Ausgleich der Interessen aller Beteiligter zu erreichen, indem die Menge der für das Training von KI-Modellen zur Verfügung stehenden Werke erhöht und den Rechteinhabern zugleich die Möglichkeit gewährt wird, ein angemessenes Entgelt zu erzielen – sei es auf dem Markt oder über eine tariflich festgesetzte Vergütung.

443 Siehe dazu vorne, E. 4.

Zusammenfassung

Die Nutzung von Werken der Literatur und Kunst für das Training von Modellen der Künstlichen Intelligenz (KI) kann nur auf der Grundlage eines hinreichenden Verständnisses der *technischen Vorgänge* beurteilt werden. Von zentraler Bedeutung sind drei Erkenntnisse: Erstens müssen die für das Training verwendeten Text-, Bild-, Audio- und Videodateien in einem ersten Schritt gesammelt und kuratiert sowie in einer Datenbank gespeichert werden. Dieser Vorgang führt zu Vervielfältigungen der Werke in den Datenbanken. Zweitens funktioniert das Training von generativen KI-Modellen für alle Werkarten (Text, Bild, Ton, Video und Source Code) im Grundsatz gleich. Es geht immer darum, dass die Modelle die statistische Verteilung in den Trainingsdaten lernen, um auf dieser Grundlage ähnliche neue Inhalte generieren zu können. Drittens hat sich gezeigt, dass die Modelle nicht nur neue und statistisch wahrscheinliche Inhalte generieren können, sondern auch in der Lage sind, exakte und nahezu exakte Kopien von in den Trainingsdaten enthaltenen Werken auszugeben. Dieser Memorisierung kann zwar durch verschiedene Massnahmen entgegengewirkt werden, sie lässt sich aber (bisher) nicht vollständig verhindern. Diese Problematik besteht bei allen bekannten Modellen und bei allen Werkarten. Die Memorisierung zeigt, dass die beim Training verwendeten Werke in den Modellen in gewisser Weise «gespeichert» sein können.

Wie die Nutzung von Werken beim Training von KI-Modellen urheberrechtlich zu beurteilen ist, wird in vielen Rechtsordnungen intensiv und kontrovers diskutiert. Besonders relevant sind aus Schweizer Sicht die gesetzgeberischen Entwicklungen und die laufenden Gerichtsverfahren in der EU sowie in Deutschland und Frankreich, im Vereinigten Königreich (UK) und in den USA. Die gesetzliche Regelung ist in Deutschland und Frankreich aufgrund der Harmonisierung durch die *Digital Single Market*-Richtlinie (DSM-RL) weitgehend deckungsgleich. Das britische Recht sieht für computergestützte Analysen von Werken für Zwecke der nichtkommerziellen Forschung eine Schranke vor, die derjenigen für Text und Data Mining für Zwecke der wissenschaftlichem Forschung in der DSM-RL entspricht, es kennt aber keine allgemeine Schranke für Text und Data Mining, auf die sich kommerzielle Unternehmen stützen können. Ob diese Regelungen in den nächsten Jahren revidiert werden, ist derzeit offen. Zumindest im UK ist wohl mit einer Revision zu rechnen, die möglicherweise eine weitere Annäherung an die Regelung der EU bringen wird. Einen anderen Ansatz verfolgt das US-amerikanische Urheberrecht, das keine besondere Regelung für Text und Data Mining kennt und die Herausforderungen bei der Nutzung von Werken für das Training von KI-Modellen durch Anwendung der *Fair Use*-Schranke bewältigen muss.

Die *Gerichte in Deutschland, im UK und in den USA* gelangen in ersten Entscheiden zu teilweise grundlegend anderen Erkenntnissen. Bemerkenswert ist namentlich, dass der britische *High Court of Justice* in *«Getty vs. Stability AI»* feststellte, dass das KI-Modell *Stable Diffusion* keine Kopien der geschützten Werke enthalte und daher nicht als *«infringing copy»* qualifiziert werden könne, während das Land-

Résumé

L'utilisation d'œuvres littéraires et artistiques pour l'entraînement de modèles d'intelligence artificielle (IA) ne peut être évaluée que sur la base d'une compréhension suffisante des processus techniques. Trois éléments sont essentiels: premièrement, les fichiers texte, image, audio et vidéo utilisés pour l'entraînement doivent d'abord être collectés, triés et stockés dans une base de données. Ce processus entraîne la reproduction des œuvres dans les bases de données. Deuxièmement, l'entraînement des modèles d'IA générative fonctionne en principe de la même manière pour tous les types d'œuvres (texte, image, son, vidéo et code source). Il s'agit toujours pour les modèles d'apprendre la distribution statistique des données d'entraînement afin de pouvoir générer de nouveaux contenus similaires sur cette base. Troisièmement, il a été démontré que les modèles peuvent non seulement générer des contenus nouveaux et statistiquement probables, mais aussi produire des copies exactes ou quasi exactes d'œuvres contenues dans les données d'entraînement. Si diverses mesures permettent de contrer cette mémorisation, il n'est toutefois pas possible (à ce jour) de l'empêcher complètement. Ce problème se pose pour tous les modèles connus et tous les types d'œuvres. La mémorisation montre que les œuvres utilisées lors de l'entraînement peuvent être «stockées» d'une certaine manière dans les modèles.

La question de l'évaluation, au regard du droit d'auteur, de l'utilisation d'œuvres dans l'entraînement des modèles d'IA fait l'objet de débats intenses et controversés dans de nombreux ordres juridiques. Du point de vue suisse, les développements législatifs et les procédures judiciaires en cours dans l'UE, en Allemagne, en France, au Royaume-Uni et aux États-Unis sont particulièrement pertinents. La réglementation légale en Allemagne et en France est largement similaire en raison de l'harmonisation prévue par la directive sur le marché unique numérique (directive DSM). Le droit britannique prévoit une restriction pour l'analyse assistée par ordinateur d'œuvres à des fins de recherche non commerciale, qui correspond à celle prévue par la directive DSM pour la fouille de textes et de données à des fins de recherche scientifique, mais il ne prévoit pas de restriction générale pour la fouille de textes et de données sur laquelle les entreprises commerciales pourraient s'appuyer. La question de savoir si ces réglementations seront révisées dans les prochaines années reste actuellement ouverte.

Au Royaume-Uni du moins, il faut s'attendre à une révision qui pourrait rapprocher davantage la réglementation de celle de l'UE. Le droit d'auteur aux États-Unis suit une autre approche, qui ne prévoit aucune réglementation particulière pour la fouille de textes et de données et doit relever les défis liés à l'utilisation d'œuvres pour la formation de modèles d'IA en appliquant la restriction du «fair use».

Les tribunaux en Allemagne, au Royaume-Uni et aux États-Unis sont parvenus à des conclusions parfois fondamentalement différentes dans leurs premières décisions. Il est notamment intéressant de noter que la Haute Cour de justice britannique a estimé, dans l'affaire Getty vs. Stability

gericht München I zum Schluss kam, dass die beim Training verwendeten Werke in den Sprachmodellen von OpenAI vervielfältigt werden. Die Gerichte in den USA gehen in ersten Entscheiden zwar davon aus, dass die Nutzung von Werken für das Training von KI-Modellen *«transformative»* sei und als *Fair Use* qualifiziert werden könne, verneinen diesen aber, wenn das Training dazu diene, ein konkurrierendes Produkt zu entwickeln. Angesichts der zahlreichen hängigen Klagen wird sich im US-amerikanischen Recht wohl bald ein genaueres Bild ergeben. Das gilt auch für das europäische Recht, zumal der EuGH in der Rechtssache C-250/25, *«Like Company v. Google Ireland»*, voraussichtlich klären wird, ob die Nutzung von Werken für das Training von KI-Modellen als Text und Data Mining qualifiziert und von den entsprechenden Schranken freigestellt werden kann.

Die Darstellung des *Standes der Diskussion in der Schweiz* hat gezeigt, dass sich die Lehre bei einigen Fragen weitgehend einig ist, während andere umstritten oder bisher kaum untersucht worden sind. Klar und unbestritten ist, dass die Speicherung der für das Training erforderlichen Werke in einer Datenbank zu Vervielfältigungen der Werke im Sinn von Art. 10 Abs. 2 lit. a URG führt. Dasselbe gilt für die Vorgänge im Trainingsprozess. Die beim Training anfallenden Vervielfältigungen werden aber zu Recht als vorübergehende Vervielfältigungen im Sinn von Art. 24*a* URG qualifiziert. Soweit die Frage untersucht wird, ist sich die Lehre bisher einig, dass die Nutzung von Werken beim Training nicht zu einer Vervielfältigung der Werke im trainierten Modell führt. Auch der Erstautor dieses Beitrags hat bisher den Standpunkt vertreten, dass die Werke beim Training nicht «als solche» in KI-Modellen gespeichert werden. Die vertiefte Auseinandersetzung mit den technischen Grundlagen hat nun aber gezeigt, dass sich dieser Standpunkt nicht aufrechterhalten lässt. Auch wenn das Ziel des Trainings von KI-Modellen (anders als bei der Speicherung in einer Datenbank) nicht darin besteht, die Werke aus dem Modell erneut abrufen zu können, sondern das Modell zu befähigen, neue Inhalte zu generieren, so werden die Werke im Modell doch in einer Form repräsentiert, die es erlaubt, die beim Training verwendeten Werke (oder Teile davon) dem trainierten Modell durch geeignete Prompts in identischer oder nahezu identischer Form wieder zu entnehmen. Dies belegt, dass die Werke in den trainierten Modellen in einer Weise repräsentiert sind, die urheberrechtlich als Vervielfältigung zu qualifizieren ist.

Geht man davon aus, dass die beim Training verwendeten Werke in den trainierten Modellen vervielfältigt sind, dann führt das *Zugänglichmachen* der Modelle möglicherweise auch zu einem Zugänglichmachen der in den Modellen repräsentierten Werke. Das Recht des Zugänglichmachens (Art. 10 Abs. 2 lit. c URG) erfasst zwar in erster Linie die Nutzung von Werken über das Internet, namentlich das Bereitstellen von Werken zum Streaming oder zum Download. Mit welchen technischen Mitteln Werke zugänglich gemacht werden, ist aber irrelevant. Zugänglich macht ein Werk, wer den Zugriff auf ein Werkexemplar und die Wie-

AI, que le modèle d'IA Stable Diffusion ne contenait pas de copies des œuvres protégées et ne pouvait donc pas être considéré comme «enfreignant le droit d'auteur», tandis que le tribunal régional de Munich I a conclu que les œuvres utilisées pour l'entraînement étaient reproduites dans les modèles linguistiques d'OpenAI. Dans leurs premières décisions, les tribunaux des États-Unis partent certes du principe que l'utilisation d'œuvres pour l'entraînement de modèles d'IA est «transformative» et peut être considérée comme respectant le «fair use», mais ils rejettent cette qualification lorsque l'entraînement sert à développer un produit concurrent. Compte tenu des nombreuses actions en justice en cours, leur position devrait bientôt se préciser. Il en va de même pour le droit européen, d'autant plus que la CJUE devrait clarifier, dans l'affaire C-250/25, «Like Company c. Google Ireland», si l'utilisation d'œuvres pour l'entraînement de modèles d'IA peut être qualifiée de fouille de textes et de données et être exemptée par des restrictions correspondantes.

La présentation de l'état actuel du débat en Suisse a montré que la doctrine s'accorde largement sur certaines questions, tandis que d'autres sont controversées ou n'ont guère été étudiées jusqu'à présent. Il est clair et incontesté que le stockage des œuvres nécessaires à l'entraînement dans une base de données conduit à des reproductions des œuvres au sens de l'art. 10 al. 2 let. a LDA. Il en va de même pour les opérations effectuées dans le cadre du processus d'entraînement. Toutefois, les reproductions effectuées lors de l'entraînement sont à juste titre qualifiées de reproductions temporaires au sens de l'art. 24*a* LDA. La question de savoir si l'entraînement conduit à une reproduction des œuvres utilisées lors de l'entraînement dans le modèle entraîné est controversée. L'auteur principal de cet article a jusqu'à présent défendu le point de vue selon lequel les œuvres ne sont pas stockées «en tant que telles» dans les modèles d'IA lors de l'entraînement. Une analyse approfondie des bases techniques a toutefois montré que ce point de vue ne pouvait être maintenu. Certes, l'objectif de l'entraînement des modèles d'IA, contrairement à ce qu'il en est lors du stockage dans une base de données, n'est pas de pouvoir récupérer les œuvres à partir du modèle, mais de permettre au modèle de générer de nouveaux contenus. Néanmoins, les œuvres utilisées lors de l'entraînement sont représentées dans le modèle sous une forme qui permet, en utilisant des prompts appropriés, de les récupérer à partir du modèle entraîné, ou d'en récupérer des parties, sous une forme identique ou presque identique. Cela prouve que les œuvres sont représentées dans les modèles entraînés d'une manière qui peut être qualifiée de reproduction en termes de droits d'auteur.

Si l'on part du principe que les œuvres utilisées lors de l'entraînement sont reproduites dans les modèles entraînés, la mise à disposition des modèles peut également entraîner la mise à disposition des œuvres représentées dans les modèles. Le droit de mise à disposition (art. 10 al. 2 let. c LDA) couvre certes principalement l'utilisation d'œuvres sur Internet, à savoir la mise à disposition d'œuvres en streaming

dergabe des Werks zum Zweck der Wahrnehmung (Werkgenuss) ermöglicht. Die zentrale Handlung besteht dabei im Bereithalten von Inhalten zum Abruf durch Dritte. Kein Zugänglichmachen liegt hingegen vor, wenn der Zugriff auf Werke mit geeigneten technischen Massnahmen unterbunden wird. Damit zeigt sich, dass das Zugänglichmachen von KI-Modellen nur dann als Zugänglichmachen der in diesen Modellen repräsentierten Werke zu verstehen ist, wenn die Anbieter keine geeigneten technischen Massnahmen getroffen haben, um die Ausgabe von Werken (oder Teilen davon) im Output der Modelle zu verhindern.

Fraglich ist, ob die Nutzung von Werken beim Training von KI-Modellen durch eine (oder mehrere) *Schranke(n)* freigestellt ist. Die Lehre ist sich einig, dass die Voraussetzungen der Schranke zugunsten des *internen Gebrauchs* (Art. 19 Abs. 1 lit. c URG) in der Regel nicht erfüllt sind. Die Vervielfältigungen könnten aber durch die *Wissenschaftsschranke* (Art. 24*d* URG) freigestellt sein. Das würde allerdings erfordern, dass der Begriff der wissenschaftlichen Forschung so weit verstanden würde, dass er auch die kommerzielle Entwicklung von KI-Modellen erfasst. Da die Wissenschaftsschranke eine volle Freistellung vorsieht, würden die Rechteinhaber für die Nutzung ihrer Werke im Rahmen dieser Schranke keine Vergütung erhalten. Das erscheint jedenfalls dann als problematisch, wenn die Werke für die Entwicklung kommerzieller KI-Modelle verwendet werden. Mit Blick auf die entstehenden Märkte, auf denen grosse Rechteinhaber Lizenzen für die Nutzung ihrer Werke für das Training von KI-Modellen erteilen, dürfte eine volle Freistellung zudem die (zunehmend) normale Verwertung der Werke beeinträchtigen. Eine weite Auslegung der Wissenschaftsschranke, die auch die kommerzielle Entwicklung von KI-Modellen freistellen würde, liesse sich deshalb kaum mit dem Dreistufentest vereinbaren.

Im Rahmen des geltenden Rechts bleibt damit die Frage, ob die Nutzung von Werken für das Training von KI-Modellen durch das Erteilen von *erweiterten Kollektivlizenzen* (Art. 43*a* URG) erlaubt werden könnte. Mit Blick auf die erwähnten Lizenzmärkte ist allerdings fraglich, ob die Voraussetzungen hierfür erfüllt sind, weil erweiterte Kollektivlizenzen die normale Verwertung der Werke nicht beeinträchtigen dürfen. Selbst wenn man annimmt, dass diese Voraussetzung allgemein (oder zumindest in bestimmten Fällen) erfüllt sein könnte, bleibt die Reichweite des Ansatzes beschränkt. Da erweiterte Kollektivlizenzen immer einer bestimmten Lizenznehmerin erteilt werden, kann dieses Instrument nur verwendet werden, um die Nutzung von Werken für das Training von KI-Modellen durch ein bestimmtes Unternehmen zu ermöglichen, es vermag die vorliegende Problematik damit nicht allgemein zu lösen. Hinzu kommt, dass es keine Verwertungsgesellschaft gibt, welche die Urheberrechte an Computerprogrammen wahrnimmt. Ein in der Praxis besonders wichtiger Bereich von generativer KI kann damit über das Instrument der erweiterten Kollektivlizenz nicht erfasst werden.

Die geltende Rechtslage ist unbefriedigend und mit Rechtsunsicherheit verbunden. Mit Blick auf Nutzen und

ou en téléchargement. Toutefois, les moyens techniques utilisés pour rendre les œuvres accessibles n'ont aucune importance. Rend accessible une œuvre quiconque permet l'accès à un exemplaire de l'œuvre et la reproduction de l'œuvre à des fins de perception (jouissance de l'œuvre). L'action centrale consiste ici à mettre des contenus à la disposition de tiers. Il n'y a en revanche pas de mise à disposition lorsque l'accès aux œuvres est empêché par des mesures techniques appropriées. Il apparaît ainsi que la mise à disposition de modèles d'IA ne peut être considérée comme une mise à disposition des œuvres représentées dans ces modèles que si les fournisseurs n'ont pas pris de mesures techniques appropriées pour empêcher la diffusion d'œuvres (ou de parties d'œuvres) dans les réponses fournies par les modèles.

La question se pose de savoir si l'utilisation d'œuvres lors de l'entraînement de modèles d'IA est exemptée par une (ou plusieurs) restriction(s). La doctrine s'accorde à dire que les conditions de la restriction en faveur de l'usage interne (art. 19 al. 1 let. c LDA) ne sont généralement pas remplies. Les reproductions pourraient toutefois être exemptées par la restriction scientifique (art. 24*d* LDA). Cela supposerait toutefois que la notion de recherche scientifique soit comprise de manière suffisamment large pour englober également le développement commercial de modèles d'IA. Étant donné que la restriction scientifique prévoit une exemption totale, les titulaires de droits ne recevraient aucune rémunération pour l'utilisation de leurs œuvres dans le cadre de cette restriction. Cela semble en tout cas problématique lorsque les œuvres sont utilisées pour le développement de modèles d'IA commerciaux. Compte tenu des marchés émergents sur lesquels les grands titulaires de droits accordent des licences pour l'utilisation de leurs œuvres pour la formation de modèles d'IA, une exemption totale risquerait en outre de nuire à l'exploitation (de plus en plus) normale des œuvres. Une interprétation large de la restriction scientifique, qui exempterait également le développement commercial de modèles d'IA, serait donc difficilement compatible avec le test en trois étapes.

Dans le cadre du droit en vigueur, la question reste donc de savoir si l'utilisation d'œuvres pour l'entraînement de modèles d'IA pourrait être autorisée par l'octroi de licences collectives étendues (art. 43*a* LDA). Au regard des marchés de licences mentionnés, on peut toutefois se demander si les conditions requises à cet effet sont remplies, car les licences collectives étendues ne doivent pas porter atteinte à l'exploitation normale des œuvres. Même en supposant que cette condition puisse être remplie de manière générale (ou du moins dans certains cas), la portée de cette approche reste limitée. Étant donné que les licences collectives étendues sont toujours accordées à un preneur de licence spécifique, cet instrument ne peut être utilisé que pour permettre l'utilisation d'œuvres pour l'entraînement de modèles d'IA par une entreprise spécifique et ne peut donc pas résoudre le problème de manière générale. À cela s'ajoute le fait qu'il n'existe aucune société de gestion collective qui gère les droits d'auteur sur les programmes infor-

Potential von KI erscheint es zentral, die Nutzung von Werken für das Training von KI-Modellen ausdrücklich im URG zu regeln. Die Regelung sollte einen angemessenen Ausgleich zwischen den Interessen der Rechteinhaber sowie der Entwickler und Anbieter von KI-Modellen schaffen und dem gesellschaftlichen Interesse an der weiteren Entwicklung dieser Technologie Rechnung tragen. Ein solcher Interessenausgleich liesse sich durch die Einführung einer neuen Schranke oder durch eine Revision der bestehenden Wissenschaftsschranke erreichen. In beiden Fällen sollte die Nutzung aller Werkarten freigestellt werden, insb. auch diejenige von Computerprogrammen.

Mit Blick auf die Rechtslage in der EU und im UK bietet sich die Einführung einer *neuen Schranke für das Training von KI-Modellen* an, bei unveränderter Beibehaltung der geltenden Wissenschaftsschranke. Damit liesse sich eine Regelung schaffen, welche die in Art. 3 und Art. 4 DSM-RL angelegte Differenzierung zwischen einer allgemeinen Schranke für Text und Data Mining und einer spezifischen Schranke für die wissenschaftliche Forschung (auch) im Schweizer Recht etablieren würde. Die Differenzierung auf Ebene der Tatbestände würde zudem erlauben, die verschiedenen Nutzungskonstellationen unterschiedlichen Regelungen zu unterstellen, namentlich die Nutzung von Werken für die wissenschaftliche Forschung weiterhin ohne Vergütung freizustellen, für andere Nutzungen für das Training von KI-Modellen aber eine Vergütung vorzusehen. Bei der Ausgestaltung der Schranke stehen drei Ansätze im Vordergrund: eine volle Freistellung mit Nutzungsvorbehalt (Opt-out), eine gesetzliche Lizenz und eine gesetzliche Lizenz mit Nutzungsvorbehalt (Opt-out).

Die Nutzung von Werken für das Training von KI-Modellen könnte vollständig freigestellt, also vergütungsfrei erlaubt werden. Um den Interessen der Rechteinhaber angemessen Rechnung zu tragen, müssten diese die Nutzung ihrer Werke aber durch Anbringen eines Nutzungsvorbehalts (Opt-out) verhindern können. Eine *volle Freistellung mit Nutzungsvorbehalt* würde die Rechtssicherheit erhöhen, Transaktionskosten verringern, Innovation ermöglichen und einen angemessenen Ausgleich der Interessen aller Beteiligten schaffen. Für eine volle Freistellung mit Opt-out spricht, dass dieser Ansatz der Regelung von Art. 4 DSM-RL und damit dem geltenden Recht in den Mitgliedstaaten der EU entspricht. Dagegen spricht, dass es bisher nicht gelungen ist, einen funktionierenden Standard für das Anbringen eines maschinenlesbaren Nutzungsvorbehalts zu entwickeln. Da das EU-Recht eine solche Möglichkeit fordert und eine Lösung technisch durchaus machbar scheint, kann allerdings angenommen werden, dass es in absehbarer Zeit gelingen wird, einen Standard zu etablieren. Gelingt dies nicht, sollte der Schweizer Gesetzgeber auf die Einführung eines Nutzungsvorbehalts verzichten.

Eine *gesetzliche Lizenz* würde die Nutzung von Werken für das Training von KI-Modellen erlauben und einer Vergütungspflicht unterstellen. Dieser Ansatz hätte den Vorteil, dass alle Rechteinhaber, die Mitglied einer Verwertungsgesellschaft sind, eine Vergütung erhalten würden. Die Bematiques. Un domaine particulièrement important de l'IA générative ne peut donc pas être couvert dans la pratique par l'instrument de la licence collective étendue.

La situation juridique actuelle est insatisfaisante et source d'insécurité juridique. Compte tenu de l'utilité et du potentiel de l'IA, il semble essentiel de réglementer expressément l'utilisation d'œuvres pour l'entraînement de modèles d'IA dans la LDA. Cette réglementation devrait établir un équilibre approprié entre les intérêts des titulaires de droits et ceux des développeurs et fournisseurs de modèles d'IA, tout en tenant compte de l'intérêt de la société pour le développement de cette technologie. Un tel équilibre pourrait être atteint par l'introduction d'une nouvelle restriction ou par une révision de la restriction scientifique existante. Dans les deux cas, l'utilisation de tous les types d'œuvres devrait être autorisée, en particulier celle des programmes informatiques.

Au vu de la situation juridique dans l'UE et au Royaume-Uni, une solution serait d'introduire une nouvelle restriction pour l'entraînement des modèles d'IA, tout en conservant la restriction scientifique en vigueur. Cela permettrait de créer une réglementation qui établirait dans le droit suisse la distinction prévue aux articles 3 et 4 de la directive DSM entre une restriction générale pour la fouille de textes et de données et une restriction spécifique pour la recherche scientifique. La distinction au niveau des faits permettrait en outre de soumettre les différentes configurations d'utilisation à des réglementations différentes, à savoir continuer à exempter l'utilisation d'œuvres à des fins de recherche scientifique sans rémunération, mais prévoir une rémunération pour d'autres utilisations à des fins d'entraînement de modèles d'IA. Trois approches sont envisagées pour la conception de la restriction: une exemption totale avec un mécanisme de droit d'opposition (opt-out), une licence légale, et une licence légale avec droit d'opposition.

L'utilisation d'œuvres pour l'entraînement de modèles d'IA pourrait être totalement exemptée, c'est-à-dire autorisée sans rémunération. Afin de tenir compte de manière appropriée des intérêts des titulaires de droits, ceux-ci devraient toutefois pouvoir empêcher l'utilisation de leurs œuvres en exerçant leur droit d'opposition. Une exemption totale avec droit d'opposition renforcerait la sécurité juridique, réduirait les coûts de transaction, favoriserait l'innovation et permettrait un équilibre approprié entre les intérêts de toutes les parties concernées. L'exemption totale avec droit d'opposition présente l'avantage de correspondre à la réglementation de l'article 4 de la directive DSM et donc au droit applicable dans les États membres de l'UE. L'argument contraire est qu'il n'a pas encore été possible de développer une norme fonctionnelle pour l'apposition d'une réserve d'utilisation correspondant au droit d'opposition qui soit lisible par machine. Cependant, comme le droit européen exige une telle possibilité et qu'une solution semble tout à fait réalisable sur le plan technique, on peut supposer qu'une norme sera établie dans un avenir proche. Si cela ne se produit pas, le législateur suisse devrait renoncer à l'introduction d'un droit d'opposition.

rechnung und die Höhe der Vergütung wären in einem gemeinsamen Tarif (GT) der Verwertungsgesellschaften zu regeln. Dieser könnte bei Bedarf an die technische Entwicklung und die Marktverhältnisse angepasst werden und relevante Differenzierungen vorsehen, um bspw. die Entwicklung kleinerer KI-Modelle durch KMU und Start-ups tariflich zu begünstigen. Diesen Vorteilen stehen gewisse Nachteile gegenüber. Da bei den grossen KI-Modellen (insb. *Large Language Models*, LLM) äusserst grosse Mengen von Werken von einer Vielzahl von Rechteinhabern für das Training verwendet werden, wäre der Ertrag der meisten Rechteinhaber verschwindend klein. Zudem könnten die Inhaber der Urheberrechte an grossen Werkmengen (bspw. Medienhäuser, Verlage oder Bildagenturen) die Nutzung ihrer Daten für das Training von KI-Modellen nicht mehr vertraglich erlauben und monetarisieren. Der Erlass einer gesetzlichen Lizenz könnte damit die Entwicklung eines funktionierenden Lizenzmarktes verhindern.

Angesichts der Vor- und Nachteile einer vollen Freistellung mit Opt-out und einer gesetzlichen Lizenz könnte sich eine Kombination der beiden Ansätze als Lösung anbieten, also eine *gesetzliche Lizenz mit Nutzungsvorbehalt (Opt-out)*. Dieser Ansatz würde die Nutzung von Werken für das Training von KI-Modellen grundsätzlich erlauben und einer Vergütungspflicht unterstellen. Die Vergütung wäre auch hier in einem gemeinsamen Tarif (GT) zu regeln, die Rechteinhaber könnten die Nutzung ihrer Werke für das Training von KI-Modellen aber durch ein Opt-out verbieten. In diesem Fall würden sie zwar keine Vergütung über den GT erhalten, sie könnten aber mit den Entwicklern von KI-Modellen auf der Grundlage des nun geltenden Nutzungsverbots Vertragsverhandlungen führen und die Nutzung ihrer Werke im Rahmen einer vertraglich vereinbarten Lizenz erlauben. Dieser Ansatz wäre zwar dogmatisch neu, würde aber lediglich Elemente kombinieren, die das schweizerische Urheberrecht bereits kennt. Die Verbindung einer gesetzlichen Lizenz mit einem Opt-out könnte sinnvolle Anreize setzen, um unerwünschte Wirkungen zu verhindern, die bei den anderen Ansätzen zu erwarten sind. Der Ansatz scheint geeignet, einen angemessenen Ausgleich der Interessen aller Beteiligter zu erreichen, indem die Menge der für das Training von KI-Modellen zur Verfügung stehenden Werke erhöht und den Rechteinhabern zugleich die Möglichkeit gewährt wird, ein angemessenes Entgelt zu erzielen – sei es auf dem Markt oder über eine tariflich festgesetzte Vergütung.

Une licence légale permettrait l'utilisation d'œuvres pour l'entraînement de modèles d'IA et serait soumise à une obligation de rémunération. Cette approche présenterait l'avantage que tous les titulaires de droits membres d'une société de gestion collective recevraient une rémunération. Le calcul et le montant de la rémunération seraient régis par un tarif commun (TC) au sein des sociétés de gestion collective. Celui-ci pourrait être adapté si nécessaire à l'évolution technique et aux conditions du marché et prévoir des différenciations pertinentes afin, par exemple, de fixer un tarif plus favorable au développement de petits modèles d'IA par les PME et les start-ups. Ces avantages s'accompagnent toutefois de certains inconvénients. Étant donné que les grands modèles d'IA (en particulier les grands modèles de langage, ou LLM) utilisent des quantités extrêmement importantes d'œuvres provenant d'un grand nombre de titulaires de droits pour leur entraînement, les revenus de la plupart des titulaires de droits seraient négligeables. En outre, les titulaires des droits d'auteur sur de grandes quantités d'œuvres (par exemple, les entreprises de médias, les maisons d'édition ou les agences photographiques) ne pourraient plus autoriser contractuellement l'utilisation de leurs données pour l'entraînement de modèles d'IA et les monétiser. L'octroi d'une licence légale pourrait donc empêcher le développement d'un marché des licences fonctionnel.

Compte tenu des avantages et des inconvénients d'une exemption totale avec droit d'opposition et d'une licence légale, une combinaison des deux approches pourrait constituer une solution, à savoir une licence légale avec droit d'opposition. Cette approche autoriserait en principe l'utilisation d'œuvres pour l'entraînement de modèles d'IA et soumettrait cette utilisation à une obligation de rémunération. La rémunération serait également régie par un tarif commun (TC), mais les titulaires de droits pourraient interdire l'utilisation de leurs œuvres pour l'entraînement de modèles d'IA par le biais d'un mécanisme de droit d'opposition. Dans ce cas, ils ne percevraient certes aucune rémunération via le TC, mais ils pourraient mener des négociations contractuelles avec les développeurs de modèles d'IA sur la base du droit d'opposition désormais en vigueur et autoriser l'utilisation de leurs œuvres dans le cadre d'une licence convenue contractuellement. Cette approche serait certes nouvelle sur le plan dogmatique, mais elle ne ferait que combiner des éléments déjà connus du droit d'auteur suisse. La combinaison d'une licence légale et d'un droit d'opposition pourrait créer des incitations utiles pour éviter les effets indésirables auxquels on peut s'attendre avec les autres approches. Cette approche semble appropriée pour parvenir à un équilibre raisonnable entre les intérêts de toutes les parties concernées, en augmentant la quantité d'œuvres disponibles pour l'entraînement des modèles d'IA tout en offrant aux titulaires de droits la possibilité d'obtenir une rémunération appropriée, que ce soit sur le marché ou par le biais d'une rémunération fixée par convention collective.