Vierter Teil: Perspektiven für das Urheberrecht

Beim Vergleich des geltenden urheberrechtlichen Regelungsregimes für die Verwendung von Werken als Trainings- und Inputdaten mit den im zweiten Teil der Untersuchung identifizierten Regelungsbedürfnissen ergeben sich mehrere Regelungsdefizite. De lege ferenda sind deswegen Anpassungen des urheberrechtlichen Rechtsrahmens notwendig. Im Folgenden werden Änderungsvorschläge für das Urheberrechts- sowie das Verwertungsgesellschaftengesetz erarbeitet, mit deren Hilfe die Regelungsdefizite behoben werden können.[840]

A. Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern

Die Speicherung von urheberrechtlich geschützten Trainingsdaten mit Hilfe von Webcrawlern sollte aus einer regulatorischen Perspektive durch eine Schrankenbestimmung erlaubnisfrei gestellt sein. Interessengerecht ist dabei, dass die Regelung sowohl nicht-kommerziell als auch kommerziell tätige Trainingsdatensammler privilegiert. Ein Opt-out-Mechanismus ist nicht erforderlich. Urheber sollen für die Vervielfältigungen ihrer Werke zum Training von KI allerdings eine angemessene Vergütung erhalten.[841]

Das geltende Urheberrecht sieht mit der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG eine Regelung vor, deren Freistellungswirkung sich nach richtlinienkonformer Auslegung des Tatbestands auch auf das automatisierte Sammeln urheberrechtlich geschützter Trainingsdaten mit Hilfe von Webcrawlern erstreckt. Allerdings erfüllt die Schrankenbestimmung wegen der in § 44b Abs. 2 S. 2 UrhG vorgesehenen Löschungspflicht trotzdem nicht die herausgearbeiteten regulatorischen Anforderungen. Gleiches gilt für den in § 44b Abs. 3 UrhG vorgesehenen Opt-out-Me-

840 Änderungsvorschläge jeweils durch Fettdruck hervorgehoben.
841 Zu den identifizierten Regelungsbedürfnissen insgesamt siehe 2. Teil B.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

chanismus und die de lege lata ermöglichte vergütungsfreie Nutzung der Werke.[842]

I. Reichweite der Schrankenbestimmung

Da § 44b UrhG das Web Scraping von Werken als Trainingsdaten für KI dem Grunde nach freistellt, ist es ausreichend, die notwendigen Anpassungen durch punktuelle Änderungen des § 44b UrhG umzusetzen. Zusätzlich zum herausgearbeiteten Änderungsbedarf ist dabei zu berücksichtigen, dass sich die Freistellungswirkung des § 44b UrhG für das Sammeln von Werken als Trainingsdaten erst nach richtlinienkonformer Auslegung des Schrankentatbestands bestimmen lässt. Zur Erhöhung der Rechtssicherheit bei der Speicherung von Werken als Trainingsdaten für KI sollte daher außerdem eine Klarstellung in den Regelungstext aufgenommen werden. Unter Berücksichtigung des gesamten Änderungsbedarfs könnte die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG de lege ferenda wie folgt aussehen:

§ 44b UrhG-E: Text und Data Mining[843]

(1) Text und Data Mining ist die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, ~~um daraus~~ **mit deren Hilfe** Informationen insbesondere über Muster, Trends und Korrelationen ~~zu gewinnen~~ **gewonnen werden können. Hierzu gehört auch das Training von auf künstlichen neuronalen Netzen basierenden Systemen maschinellen Lernens (Systeme künstlicher Intelligenz).**

(2) Zulässig sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining. Die Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.

(3) Nutzungen nach Absatz 2 Satz 1 sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt.

(4) Absatz 2 Satz 2 und Absatz 3 gelten nicht, wenn die Vervielfältigungen für das Training von Systemen künstlicher Intelligenz angefertigt worden sind.

842 Zur geltenden Urheberrechtslage für die Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern siehe insgesamt 3. Teil A.

843 Hervorhebung steht für neu einzufügenden Normtext.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

II. Angemessene Vergütung für Urheber

Darüber hinaus ist es interessengerecht, den vom Web Scraping für das KI-Training betroffenen Urhebern für die stattfindenden Vervielfältigungshandlungen eine angemessene Vergütung zu gewähren.

1. Vergütungspflicht und Vergütungssystem

Problematisch ist dabei allerdings, dass weder betroffene Urheber noch Trainingsdatensammler aufgrund der Automatisierung der Datensammlung, oftmals fehlender Metadaten zu den erfassten Inhalten und eines Kontrolldefizits regelmäßig Kenntnis darüber haben, ob es überhaupt zu urheberrechtlich relevanten Vervielfältigungshandlungen gekommen ist und welche Werke genutzt worden sind. Es fehlt daher regelmäßig an einer Konkretisierung des Vergütungsgläubigers. Der einzuführende gesetzliche Vergütungsansprüche muss aus diesem Grund durch Verwertungsgesellschaften geltend gemacht werden.[844] Darüber hinaus ist die Abrechnung von Einzelnutzungen vor dem Hintergrund der skizzierten praktischen Schwierigkeiten und der Größe der erstellten Datensätze ebenfalls nur erschwert möglich. Der Vergütungsanspruch muss deswegen pauschaliert werden. Wichtig ist hierbei aber, keine Überkompensation der Urheber zulasten der Trainingsdatensammler zu verursachen.

Auch die Hersteller- und Betreibervergütung (§§ 54 ff. UrhG), mit Hilfe derer Urhebern eine gesetzliche Vergütung unter anderem für gemäß § 53 UrhG hergestellte Privatkopien ihrer Werke zuteil wird, bedient sich solcher Pauschalierungsverfahren. Denn auch bei der Herstellung von Privatkopien haben Urheber im Regelfall keine Kenntnis darüber, ob und wie viele Privatkopien ihrer Werke angefertigt wurden. Die Hersteller und Betreiber von Speichermedien oder Kopiergeräten, Vergütungsschuldner, können außerdem nicht wissen, inwieweit ihre Hilfsmittel für urheberrechtlich relevante Nutzungshandlungen verwendet worden sind. Aus diesen Gründen wird die Hersteller- und Betreibervergütung ebenfalls über Verwertungsgesellschaften abgewickelt. Im Ergebnis bestehen wesentliche Parallelen zwischen der Hersteller- und Betreibervergütung aus den §§ 54 ff. UrhG und der gesetzlichen Vergütungspflicht, die für die Vervielfältigung von Werken zum Training von KI eingeführt werden sollte. Letz-

844 Ebenso *Geiger*, IIC 2024, 1129 (1153).

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

tere ist daher nach dem Vorbild der Hersteller- und Betreibervergütung auszugestalten. Eine Regelung könnte danach wie folgt aussehen:

§ 44c UrhG-E: Vergütung für die automatisierte Vervielfältigung von Werken aus Internetquellen zum Training von Systemen künstlicher Intelligenz

(1) Lässt die Art eines im Internet öffentlich zugänglich gemachten Werkes eine nach § 44b Absatz 2 UrhG erlaubte automatisierte Vervielfältigung für das Training von Systemen künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E) erwarten, so hat der Urheber des Werkes gegen den Trainingsdatensammler Anspruch auf Zahlung einer angemessenen Vergütung.

(2) Trainingsdatensammler ist, wer Inhalte im Geltungsbereich des Urheberrechtsgesetzes für das Training von Systemen künstlicher Intelligenz vervielfältigt, bei denen erwartet werden kann, dass sie ganz oder teilweise nach diesem Gesetz schutzfähig sind. Das umfasst insbesondere denjenigen, der online zugängliche Inhalte erstmals auf Datenträger speichert, als auch denjenigen, der einen vorverarbeiteten Trainingsdatenkorpus von einem Dritten erlangt.

(3) Maßgebend für die Vergütungshöhe ist insbesondere, zum Training wie vieler Systeme künstlicher Intelligenz das Werk genutzt wird und welche Aufgabe mit dem jeweiligen System künstlicher Intelligenz erfüllt werden soll. Dabei ist zu berücksichtigen, inwieweit technische Schutzmaßnahmen nach § 95a auf die betreffenden Werke angewendet werden. Die Vergütung darf die Trainingsdatensammler nicht unzumutbar beeinträchtigen; sie muss in einem wirtschaftlich angemessenen Verhältnis zum Wert des angefertigten Trainingsdatenkorpus als solchem und dem System künstlicher Intelligenz stehen, das auf Basis des Trainingsdatenkorpus entwickelt worden ist.

(4) Die Vergütung ist so zu gestalten, dass sie auch mit Blick auf die Vergütungspflicht anderer Trainingsdatensammler insgesamt angemessen ist. Mehrere an einem Trainingsvorgang beteiligte Trainingsdatensammler haften als Gesamtschuldner.

§ 44d UrhG-E: Meldepflicht

(1) Wer Inhalte im Geltungsbereich des Urheberrechtsgesetzes als Trainingsdaten speichert, bei denen erwartet werden kann, dass sie ganz oder teilweise nach diesem Gesetz schutzfähig sind, ist dem Urheber gegenüber verpflichtet, monatlich bis zum zehnten Tag nach Ablauf jedes Kalendermonats schriftlich Mitteilung zu leisten über:

1. *Die Menge und Art der vervielfältigten Daten,*
2. *Die Anzahl der Systeme künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E), die mit den Daten trainiert werden sollen, sowie*
3. *soweit ersichtlich, die Aufgabe des jeweiligen Systems künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E).*

(2) Kommt der Meldepflichtige seiner Meldepflicht nicht, nur unvollständig oder sonst unrichtig nach, kann der doppelte Vergütungssatz verlangt werden.

(3) Die Meldung erfolgt an eine gemeinsame Empfangsstelle der Verwertungsgesellschaften. § 54h Absätze 3 bis 5 gelten entsprechend.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

§ 44e UrhG-E: Verwertungsgesellschaften; Handhabung der Mitteilungen

(1) Der Anspruch nach § 44c kann nur durch eine Verwertungsgesellschaft geltend gemacht werden.
(2) Jedem Berechtigten steht ein angemessener Anteil an den nach § 44b gezahlten Vergütungen zu. Soweit Werke mit technischen Maßnahmen gemäß § 95a geschützt sind, werden sie bei der Verteilung der Einnahmen nicht berücksichtigt.
(3) Das Deutsche Patent- und Markenamt kann Muster für die Mitteilungen nach § 44d im Bundesanzeiger bekannt machen. Werden Muster bekannt gemacht, sind diese zu verwenden.
(4) Die Verwertungsgesellschaften und die Empfangsstelle dürfen die gemäß § 44b erhaltenen Angaben nur zur Geltendmachung der Ansprüche nach § 44c verwenden.

2. Vergütungshöhe und ihre Konkretisierung durch Verwertungsgesellschaften

Für eine Pauschalisierung der angemessenen Vergütung müssen Annahmen beispielsweise darüber getroffen werden, bei wie vielen der für das Training gespeicherten Daten (Beispiel: Textinhalte) aus einer statistischen Perspektive davon ausgegangen werden kann, dass sie urheberrechtlich schutzfähig sind (Beispiel: Eine Auswertung mehrerer repräsentativer Trainingsdatensätzen könnte ergeben, dass durchschnittlich 83 % der gesammelten Textdaten urheberrechtlich schutzfähige Sprachwerke sind). Um Tarife für die Vervielfältigung von Werken zum Training von KI aufzustellen, sind demzufolge empirische Voruntersuchungen notwendig. Die hierfür erforderliche Regelung muss im Verwertungsgesellschaftengesetz umgesetzt werden. Sie könnten wie folgt aussehen:

§ 40a VGG-E: Gestaltung der Tarife für die Entwicklung von Systemen künstlicher Intelligenz (§ 44b Absatz 1 Satz 2 UrhG)

(1) Die Höhe der Vergütung für das Vervielfältigen von Werken zum Training von Systemen künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E) bestimmt sich nach § 44c UrhG-E. Die Verwertungsgesellschaften stellen hierfür Tarife auf Grundlage einer empirischen Untersuchung aus einem Verfahren gemäß § 93 auf. § 38 Satz 2 bleibt unberührt.
(2) Die Pflicht zur Tarifaufstellung entfällt, wenn zu erwarten ist, dass der dafür erforderliche wirtschaftliche Aufwand außer Verhältnis zu den zu erwartenden Einnahmen stehen würde.

Daneben sind noch weitere redaktionelle Folgeanpassungen im Verwertungsgesellschaftengesetz notwendig, die hier nicht näher dargestellt werden.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

Im Übrigen bestimmt sich die Höhe der angemessenen Vergütung nach dem wirtschaftlichen Wert der Nutzung.[845] Bei der Vervielfältigung eines Werks als Trainingsdatum für die Entwicklung von KI wird die Vergütung demzufolge danach bestimmt, welcher wirtschaftliche Vorteil von der Nutzung des Werkes als Trainingsdatum ausgeht. Für die Konkretisierung dieses Vorteils liefert § 44c Abs. 3 UrhG-E Anhaltspunkte. Danach kommt es insbesondere darauf an, zum Training wie vieler KI-Systeme der betreffende Trainingsdatenkorpus verwendet werden soll. Denn der wirtschaftliche Vorteil, der aus den Trainingsdaten folgt, steigert sich proportional hierzu. Darüber hinaus ist von Bedeutung, welche Aufgaben das entwickele KI-System erfüllen wird. Berücksichtigt werden kann beispielsweise aber auch, ob mit der KI direkte Konkurrenzprodukte zu den genutzten Trainingsdaten generiert werden sollen. Maßgebend ist auch, welche Qualität die genutzten Inhalte haben und welche Popularität ihnen zukommt.[846] Denn je hochqualitativer und populärer die zum Training genutzten Inhalte, desto wahrscheinlicher ist auch der wirtschaftlicher Erfolg der durch das KI-System erschaffenen Konkurrenzprodukte.[847] Die Informationen, die zur Bestimmung der von den Trainingsdatensammlern zu leistenden Vergütung benötigt werden, sind von der Meldepflicht im Sinne des § 44d UrhG-E erfasst.

Der Vergütungsanspruch darf allerdings nicht zu einer Überkompensation der betroffenen Urheber führen. Daher ist die Vergütung gemäß § 44c Abs. 4 UrhG-E so auszugestalten, dass sie auch mit Blick auf die Vergütungspflicht anderer Trainingsdatensammler insgesamt angemessen ist. Denkbar ist beispielsweise, dass ein intermediärer Trainingsdatensammler eine im Vergleich zum eigenhändig sammelnden KI-Entwickler geringere Vergütung zahlen muss, die sich vor allem nach der Anzahl der KI-Systeme bemisst, die mit den Daten trainiert werden sollen (vgl. § 44c Abs. 3 UrhG-E). Das bedeutet, dass er für jede Datenübertragung eine Vergütung zu leisten hätte. Die Vergütung der KI-Entwickler, welche die Trainingsdatensätze erworben haben, könnte danach bemessen werden, welche Aufgabe mit dem jeweiligen KI-System erfüllt werden soll (vgl. § 44c Abs. 3 UrhG-E). Auch sie müsste im Vergleich zur Ein-Personen-Konstellation

845 Zur Vergütungshöhe bei vergütungspflichtigen Schrankenbestimmungen: *Pflüger*, Gerechter Ausgleich und angemessene Vergütung, S. 145. Für die Tarife der Verwertungsgesellschaften ergibt sich das aus § 39 Abs. 1 S. 1 VGG.

846 *Geiger*, IIC 2024, 1129 (1154).

847 *Geiger*, IIC 2024, 1129 (1154); ganz nach dem Motto: „garbage in, garbage out", vgl. *de la Durantaye*, ZUM 2023, 645 (647).

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

reduziert werden. Im Ergebnis würde Urhebern dadurch eine angemessene Vergütung zuteilwerden, die interessengerecht zwischen intermediärem Trainingsdatensammler und KI-Entwickler aufgeteilt ist. Da gemäß § 44c Abs. 4 UrhG-E eine gesamtschuldnerische Haftung vorgesehen ist, könnte die Vergütung insgesamt vom intermediären Trainingsdatensammler geleistet werden, der bei der Veräußerung der Trainingsdatensätze einen Teil der wirtschaftlichen Belastung an die KI-Entwickler weiterreichen kann.

3. Erwartung der Vervielfältigung als Voraussetzung für die Vergütung (§ 44c Abs. 1 UrhG-E)

Voraussetzung des Vergütungsanspruchs ist gemäß § 44c Abs. 1 UrhG-E, dass die Art des im Internet veröffentlichten Werkes eine nach § 44b Absatz 2 UrhG erlaubte Vervielfältigung für das Training von KI erwarten lässt. Diese § 54 Abs. 1 UrhG nachempfundene Regelung trägt der Pauschalisierungsnotwendigkeit Rechnung, die der Vergütungsanspruch für die Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern mit sich bringt.[848]

Eine Vervielfältigung kann gemäß §§ 44c Abs. 1 UrhG-E erwartet werden, wenn die jeweilige Werkart üblicherweise zum Training von KI verwendet wird.[849] Ob sich die Vervielfältigung eines Werkes erwarten lässt, hängt damit entscheidend vom Trainingsdaten- und KI-Markt sowie dem Stand der Technik bei der KI-Entwicklung ab. Eine Beurteilung kann im Rahmen der empirischen Untersuchung gemäß § 40a VGG-E erfolgen. Sie sollte aufgrund des Fortschreitens der Entwicklung und Anwendung von KI regelmäßig überprüft werden.

Inwieweit die Vervielfältigung eines konkreten Werkes zum Training von KI wahrscheinlich ist, ist für die Begründung der Vergütungspflicht hingegen nicht von Bedeutung.[850] Sie kann allerdings bei Bestimmung der Vergütungshöhe berücksichtigt werden.[851] Die Vervielfältigungswahr-

848 Zur Pauschalisierungsnotwendigkeit siehe oben unter 4. Teil A. II. 1. Zu § 54 UrhG vgl. außerdem *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 17.

849 Vgl. zur entsprechenden Voraussetzung in § 54 UrhG *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 17.

850 Vgl. *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 18.

851 Vgl. *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 18.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

scheinlichkeit wird insbesondere dadurch beeinflusst, dass das Werk durch Zugangsbeschränkungen geschützt oder das Web Scraping beispielsweise mit Hilfe des Robots Exclusion Standards vorbehalten worden ist. Die Urheber sollten bei der Meldung der Werke demzufolge entsprechende Angaben tätigen müssen.

4. Trainingsdatensammler als Vergütungsschuldner (§ 44c Abs. 2 UrhG-E)

Vergütungsschuldner sind gemäß § 44c Abs. 1, 2 UrhG-E Trainingsdatensammler. Dies ist, wer Inhalte für das Training von KI speichert, bei denen erwartet werden kann, dass sie ganz oder teilweise urheberrechtlich schutzfähig sind.

Hervorzuheben ist dabei, im Sinne der Legaldefinition nicht vorausgesetzt wird, dass *tatsächlich* Werke für das Training von KI vervielfältigen werden. Konkrete Nutzungshandlungen bei der Erstellung und Verarbeitung der Trainingsdatensätze sollen nach dem der Vergütungspflicht zugrundeliegenden Pauschalisierungsgedanken gerade nicht dargelegt und bewiesen werden müssen. Der Nachweis könnte von einem Urheber auch kaum erbracht werden. Der Pauschalisierung der Vergütung entsprechend muss der Kreis der Vergütungsschuldner zunächst weit gefasst werden. Ob und inwieweit erwartet werden kann, dass bei der Speicherung von Inhalten für das Training von KI insgesamt oder teilweise Werke erfasst werden, bestimmt sich nach Art und Umfang der abgespeicherten Daten sowie danach, wie viele der erfassten Inhalte aus einer statistischen Perspektive typischerweise urheberrechtlich geschützt sind.[852] Die hierfür notwendigen empirischen Untersuchungen sind von Verwertungsgesellschaften vorzunehmen.

Der Kreis der Vergütungsschuldner umfasst insbesondere diejenigen, die online zugängliche Inhalte zur eigenen Verwendung erstmals auf Datenträger abspeichern. Hierunter fällt das Web Scraping von Trainingsdaten für die Entwicklung von KI. Darüber hinaus sind nach dem Wortlaut des Regelungsvorschlags aber auch Beteiligte erfasst, die vorverarbeitete Trainingsdatenkorpora von Dritten erlangen. Denn auch sie vervielfältigen die Inhalte im Sinne des § 44c Abs. 2 UrhG-E. Auch nachgelagerte KI-Entwickler, denen die Trainingsdatensätze von intermediären Trainingsdatensammlern zur Verfügung gestellt werden, sind damit Trainingsdatensammler im Sinne des § 44c

852 Dazu bereits siehe oben unter 4. Teil A. II. 2.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

Abs. 2 UrhG-E. Hierdurch wird auch dann eine Vergütung für Urheber sichergestellt, wenn die Trainingsdatensätze zwischen KI-Entwicklern weitergegeben werden. Von der Regelung außerdem erfasst werden solche KI-Entwickler, die ihre Trainingsdatenkorpora aus dem Ausland beziehen, wo gegebenenfalls keine Vergütung betroffener Urheber stattfindet.[853]

Hinterfragt werden kann allerdings, ob die Vergütungspflicht nicht direkt an die Tätigkeit der KI-Entwickler beziehungsweise -Anbieter oder gar die Nutzung von KI-Systemen angeknüpft werden könnte.[854] Dafür spricht, dass insbesondere Anbieter von KI-Systemen, die erkennbar auf dem Markt auftreten, von Urhebern und Verwertungsgesellschaften mit geringerem Aufwand identifiziert und adressiert werden können. Dies erleichtert die Durchsetzung des Vergütungsanspruchs. Allerdings werden Vervielfältigungen urheberrechtlich geschützter Trainingsdaten gemäß § 16 UrhG nur im Vorfeld des Trainings von KI vorgenommen. Vor allem sind, unabhängig von einer technischen Rekonstruierbarkeit, keine Vervielfältigungsstücke der zum Training genutzten Werke dauerhaft im KNN abgelegt.[855] Bei einer Anknüpfung der Vergütungspflicht an die KI-Nutzer, -Entwickler oder -Anbieter kann vor diesem Hintergrund nicht sichergestellt werden, dass urheberrechtlich relevante Nutzungen der zum Training verwendeten Werke in Deutschland stattgefunden haben. Hierauf hätte auch die Anerkennung eines domaine public payant, einer auf eine Vergütungspflicht beschränkten Ausformung eines kollektiven Urheberrechts sämtlicher Urheber,[856] keinen Einfluss.[857] Denn auch diese kollektivierte Rechtsposition wäre nicht berührt, wenn keine Nutzung der Werkgesamtheit im Inland stattfände. Ohne eine Beeinträchtigung des nationalen Urheberrechts als verfassungsrechtlich gewährleistetes Eigentumsrecht besteht jedoch kein legitimer Zweck

853 Im Rahmen der Privatkopievergütung erfüllt die Händler- und Importeurshaftung (§ 54b UrhG) diese Funktion, siehe *Wiebe*, in: Spindler/Schuster, Recht der elektronischen Medien, UrhG § 54b Rn. 2.

854 Ein solches Vergütungsmodell, welches an den Betrieb von KI-Systemen anknüpft, die geeignet sind, Werkschaffen zu substituieren, schlagen *Senftleben*, IIC 2023, 1535 (1549 ff.); *Jacques/Flynn*, GRUR Int 2024, 1137 (1141, 1148) vor. Die generierten Einnahmen sollten sodann beispielsweise in eine Fondsstruktur zur Unterstützung Kreativer fließen, vgl. instruktiv *Jacques/Flynn*, GRUR Int 2024, 1137 (1148).

855 Dazu im Detail siehe oben 3. Teil B. IV.

856 So insbesondere noch vorgesehen im Entwurf des Urheberrechtsgesetzes von 1965 BT-Drs. IV/270, S. 15 ff. wobei der Vorschlag auf Anregung des Rechtsausschusses allerdings gestrichen worden ist. Dazu siehe BT-Drs. IV/3401, S. 13; Vgl. zur Historie und Funktionsweise allgemein *Dillenz*, GRUR Int 1983, 920; *Marzetti*, GRUR Int. 2019, 343.

857 Mit einem solchen Ansatz *Senftleben*, IIC 2023, 1535 (1552 f.).

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

dafür, die Freiheitsrechte der KI-Nutzer, -Entwickler und -Anbieter durch eine Vergütungspflicht einzuschränken. Um eine Vergütungspflicht verfassungsrechtlich rechtfertigen zu können, ist damit ein Eingriff in das nationale urheberrechtliche Ausschließlichkeitsrecht notwendig. Er muss zumindest erwartbar sein. Eine Verlagerung der Vergütungspflicht auf andere Beteiligte als den Trainingsdatensammler wie insbesondere Anbieter oder Nutzer von KI-Systemen ist deswegen aus verfassungsrechtlichen Gründen nicht möglich.

Ohnehin wäre es dogmatisch inkonsistent, die Vergütungsbedürftigkeit einer Nutzungshandlung (Vervielfältigung zum Training) durch die Vergütung einer anderen, urheberrechtlich eigenständigen Nutzungshandlung (Generierung urheberrechtlich relevanten Outputs) zu realisieren. Es ist vielmehr stets klar zwischen den verschiedenen Nutzungshandlungen bei der Entwicklung und dem Einsatz von KI zu trennen. Auch die Passivlegitimation für Urheberrechtsverletzungen beim Training einerseits und bei der Anwendung einer KI andererseits könnte unterschiedliche Verantwortlichkeiten begründen.

Weiterhin ist davon auszugehen, dass intermediäre Trainingsdatensammler die anfallende Vergütung an nachgelagerte KI-Entwickler weiterreichen. Bringen diese das KI-System nicht selbst in Verkehr, haben sie ebenfalls die Möglichkeit, die durch die Vergütungspflicht entstehenden Kosten auf den KI-Anbieter umzulegen. Damit kommt es zu einer automatischen Zuordnung der wirtschaftlichen Belastung zu dem Beteiligten, der wirtschaftlich erheblich von der Nutzung der Werke zum Training von KI profitiert. Das ist ökonomisch interessengerecht. Aus diesen Gründen ist eine Anknüpfung der Vergütungspflicht an die Tätigkeit der Trainingsdatensammler notwendig.[858]

III. Unionsrechtliche Perspektive auf den Regelungsvorschlag

Fraglich ist jedoch, ob der entwickelte Regelungsvorschlag mit den unionsrechtlichen Grundlagen des Urheberrechts vereinbar ist. Hierfür muss er insbesondere mit Art. 4 DSM-RL vereinbar sein. Er wird in § 44b UrhG

858 Andere Ansicht aber *Schack*, NJW 2024, 113 Rn. 26, nach dem die Vergütungspflicht an die Generierung des KI-Outputs angeknüpft werden sollte.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

umgesetzt.[859] Die in § 44b Abs. 1 S. 2 UrhG-E eingefügte Klarstellung ist darüber hinaus an Art. 2 Nr. 2 DSM-RL zu messen.

1. Klarstellung zum KI-Training

Wie auch § 44b UrhG adressieren Art. 2 Nr. 2 und Art. 4 DSM-RL das Sammeln urheberrechtlich geschützter Trainingsdaten nicht ausdrücklich. Dennoch ist es von der in Art. 2 Nr. 2 DSM-RL normierten Legaldefinition des Text und Data Mining umfasst.[860] Die Rechtsänderung hat nur klarstellenden Charakter. Der nationale Gesetzgeber ist hinsichtlich der konkreten Ausgestaltung seiner Rechtsnormen jedoch weitestgehend frei, solange das in der Richtlinie vorgegebene Ziel erreicht wird (vgl. Art. 288 Abs. 3 AEUV). Durch die klarstellende Formulierung wird der Anwendungsbereich der Regelung jedoch nicht verändert. Ihre Einfügung ist demzufolge mit dem Unionsrecht vereinbar.

2. Verzicht auf Löschungspflicht und Opt-Out-Mechanismus

Der zur Umsetzung in § 44b Abs. 4 UrhG-E vorgeschlagene Verzicht auf das Löschungserfordernis und den Opt-out-Mechanismus bei der Vervielfältigung urheberrechtlich geschützter Trainingsdaten könnte jedoch mit dem Unionsrecht nicht vereinbar sein.

Die Rechtsänderung muss sich zum einen an Art. 4 DSM-RL messen lassen. Löschungspflicht und Opt-out-Mechanismus sind hier in Art. 4 Abs. 2, 3 DSM-RL geregelt. In den Bestimmungen der DSM-RL ist hingegen keine Ausnahme für das Sammeln von Trainingsdaten vorgesehen.

Der DSM-RL kommt in Bezug auf ihre Schrankenbestimmungen allerdings nur mindestharmonisierende Wirkung zu.[861] Gemäß Art. 25 DSM-RL können für Arten und Bereiche der Nutzung, für die die Ausnahmen und Beschränkungen der Richtlinie gelten, umfassendere Bestimmungen erlassen werden, insofern diese mit den in der Datenbank- und InfoSoc-RL vorgesehenen Ausnahmen und Beschränkungen vereinbar sind. Der nationale Gesetzgeber könnte zugunsten der Trainingsdatensammler daher auch ein Mehr an Freistellungswirkung realisieren, insofern die eingeführ-

859 BT-Drs. 19/27426, S. 50.
860 Dazu 3. Teil A. III. 2. a) bis e).
861 *Stieper*, GRUR 2020, 1 (3 f.).

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

ten Schrankenbestimmungen mit Art. 6 Abs. 2 Datenbank-RL sowie Art. 5 InfoSoc-RL in Einklang stehen.

Da § 44b Abs. 4 UrhG-E des Regelungsentwurfs für Vervielfältigungen von Werken sämtlicher Werkarten und nicht nur für Datenbankwerke gilt, kommt nur Art. 5 InfoSoc-RL als unionsrechtliche Grundlage in Betracht. Hier könnte auf den ersten Blick zwar auf Art. 5 Abs. 3 lit. a InfoSoc-RL zurückgegriffen werden. Die dort vorgesehene Schrankenwirkung kann jedoch nur zum Zwecke der wissenschaftlichen Forschung eingreifen. § 44b Abs. 4 UrhG-E, der auch kommerziell tätige Trainingsdatensammler freistellt, übersteigt den durch Art. 5 Abs. 3 lit. a InfoSoc-RL gewährten Freistellungsspielraum demzufolge. Im Ergebnis wird die Erweiterung der Schrankenbestimmung durch § 44b Abs. 4 UrhG-E daher auch nicht durch Bestimmungen der InfoSoc-RL abgedeckt.

Mangels Grundlage in den urheberrechtlichen Richtlinien des Unionsgesetzgebers wäre der vorgeschlagene Verzicht auf die Löschungspflicht und den Opt-out-Mechanismus daher nicht mit dem europäischen Urheberrecht vereinbar. Der Umsetzung des entwickelten Regelungsvorschlags müsste demzufolge eine europäische Normgebungsinitiative vorangehen.

3. Einführung einer Vergütungspflicht für das Sammeln von Trainingsdaten

Klärungsbedürftig ist darüber hinaus, ob jedenfalls die in § 44c UrhG-E vorgeschlagene Vergütungspflicht für das Web Scraping urheberrechtlich geschützter Trainingsdaten vor dem Hintergrund des Art. 4 DSM-RL umgesetzt werden könnte.

a) Vergütungspflicht als solche

Art. 4 DSM-RL sieht keinen gerechten Ausgleich der Urheber für die Vervielfältigungshandlungen vor. Entscheidend ist damit, ob Art. 4 DSM-RL es den Mitgliedstaaten offenlässt, eine Vergütungspflicht für das Mining von Werken als KI-Trainingsdaten einzuführen.

Hiergegen spricht, dass Art. 5 Abs. 4 DSM-RL (Schrankenbestimmung für die Nutzung von Werken und sonstigen Schutzgegenständen für digitale und grenzüberschreitende Unterrichts- und Lehrtätigkeiten) ausdrücklich eine Vergütungspflicht anordnet. Aus dem Fehlen einer entsprechenden Regelung in Art. 4 DSM-RL könnte abgeleitet werden, dass Urhebern

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

im Umkehrschluss kein gerechter Ausgleich für das Text und Data Mining gewährt werden darf.[862] Darüber hinaus ging der Richtliniengeber jedenfalls davon aus, dass das Text und Data Mining für wissenschaftliche Zwecke nur einen minimalen Schaden für Urheber verursache und daher vergütungsfrei vorzunehmen sei.[863] Dieser Gedanke ließe sich verallgemeinern und auf das allgemeine Text und Data Mining beziehungsweise das Sammeln urheberrechtlich geschützter Trainingsdaten übertragen.[864] Nicht zuletzt werden beim Text und Data Mining und auch beim eigentlichen KI-Training auf Basis von Werken nur die mit den Schöpfungen übermittelten semantischen Informationen genutzt. Sie sind nicht Teil des urheberrechtlichen Schutzbereich.[865] Dies macht es unwahrscheinlich, dass der Richtliniengeber eine Vergütung der Nutzung hat ermöglichen wollen.[866]

Allerdings geht ErwGr. 17 DSM-RL, der die Vergütungsfreiheit des Text und Data Mining zum Zwecke wissenschaftlicher Forschung begründet, auf den im Jahr 2016 veröffentlichten Richtlinienentwurf der EU-Kommission zurück.[867] Dieser sah eine Schrankenbestimmung für das Text und Data Mining allein zum Zwecke nicht-kommerziell tätiger Forschungsorganisationen vor.[868] Zwar wurde die Schrankenbestimmung auf Initiative des Europäischen Parlaments im Gesetzgebungsprozess auf allgemeines Text und Data Mining ausgeweitet. Dies brachte auch die Einfügung zusätzlicher Erwägungsgründe mit sich.[869] Die Formulierung des ErwGr. 17 DSM-RL (ex. ErwGr. 13 DSM-RL-E) wurde allerdings ausdrücklich auf das Text und Data Mining zu wissenschaftlichen Zwecken beschränkt. Diese Historie macht deutlich, dass der Richtliniengeber vor dem Hintergrund einer nur geringfügigen Beeinträchtigung der Urheberinteressen nur von einer Vergütungsfreiheit des Text und Data Mining zu wissenschaftlichen Zwecken (Art. 3 DSM-RL) ausgegangen ist. Ein allgemeiner Rechtsgedanke, der auch das allgemeine, in Art. 4 DSM-RL geregelte Text und Data Mining erfasst, lässt sich aus dem Erwägungsgrund daher nicht ableiten. Der Richtlinie kann demzufolge insgesamt kein Rechtsgedanke entnommen

862 BT-Drs. 19/27426, S. 88; *Raue*, ZUM 2020, 172 (173); *Raue*, ZUM 2021, 793 (795).
863 ErwGr. 17 DSM-RL.
864 *Kleinkopf*, Text- und Data-Mining, S. 234 f.; so auch *Stieper*, ZUM 2021, 776 (778), wobei sich die Aussage hier nur auf eine rechtspolitische Dimension erstreckt.
865 *Heesen/Jüngels*, RuZ 2021, 45 (53).
866 *Heesen/Jüngels*, RuZ 2021, 45 (53).
867 Vgl. ErwGr. 13 DSM-RL-E, siehe COM(2016) 593 final.
868 Art. 3 DSM-RL-E, siehe COM(2016) 593 final.
869 Siehe zum Parlamentsbeschluss in erster Lesung: P8_TA-PROV(2019)0231.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

werden, nach dem das allgemeine Text und Data Mining vergütungsfrei ausgestaltet sein sollte.[870] Ein Umkehrschluss aus ErwGr. 17 DSM-RL und die Entstehungsgeschichte des Erwägungsgrundes legen vielmehr einen gegenteiligen Schluss nahe.

Weiterhin bringt der Richtliniengeber in ErwGr. 36 InfoSoc-RL zum Ausdruck, dass die Mitgliedstaaten bei der spezifischen Umsetzung der Schrankentatbestände dort einen gerechten Ausgleich für Urheber vorsehen sollen, wo es im konkreten Fall insbesondere aufgrund des Dreistufentests notwendig ist; auch, wenn in Art. 5 InfoSoc-RL ein solcher Ausgleich nicht vorgesehen ist.[871] Hieraus kann ein allgemeiner Rechtsgrundsatz des Richtliniengebers beim Umgang mit urheberrechtlichen Schrankenbestimmungen abgeleitet werden, nach dem die Einführung einer Vergütungspflicht innerhalb der allgemeinen unionsrechtlichen Grenzen grundsätzlich den Mitgliedstaaten freisteht. Dort, wo er diesen durchbrechen wollte, hat er es ausdrücklich kenntlich gemacht. Das ist beispielsweise in ErwGr. 17 DSM-RL, nicht aber in Bezug auf Art. 4 DSM-RL der Fall.

Nicht zuletzt zielt das Text und Data Mining zwar darauf ab, die den vervielfältigten Werken zugrundeliegenden Informationen zu verarbeiten. Das ändert jedoch nichts daran, dass nicht nur die Informationen, sondern die gesamten Werke inklusive ihres eigenschöpferischen Gehalts vervielfältigt werden.

Insgesamt sprechen damit die besseren Gründe dafür, dass die DSM-RL den Mitgliedstaaten die Einführung einer Vergütungspflicht für das in Art. 4 DSM-RL geregelte allgemeine Text und Data Mining freistellt.[872]

b) Trainingsdatensammler als Vergütungsschuldner

Auch die Zuweisung der Kostenlast zu den Trainingsdatensammlern müsste mit den unionsrechtlichen Bestimmungen vereinbar sein.

---

870 So auch *Wandtke*, NJW 2019, 1841 (1842).

871 Zur Notwendigkeit einer angemessenen Vergütung der Urheber siehe oben unter 2. Teil B. II. 1.

872 *Wandtke*, MMR 2017, 367 (368); *Wandtke*, NJW 2019, 1841 (1842); *Würtenberger/Freischem*, GRUR 2019, 1140 (1142); halten es zwar für unionsrechtlich möglich, aber für rechtpolitisch nicht zielführend: *Stieper*, GRUR 2020, 1 (4); *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (29). Andere Ansicht *Hauk/Pflüger*, ZUM 2020, 383 (385); *Heesen/Jüngels*, RuZ 2021, 45 (53); *Raue*, ZUM 2021, 793 (795); *Kleinkopf*, Text- und Data-Mining, S. 234 f.; BT-Drs. 19/27426, S. 88.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

Weder die InfoSoc-RL noch die DSM-RL treffen selbst Aussagen darüber, welchem Beteiligten eine Vergütungspflicht für Vervielfältigungshandlungen beim Sammeln von urheberrechtlich geschützten Trainingsdaten auferlegt werden kann. Ziel der Vergütung ist aber der Ausgleich wirtschaftlicher Nachteile, die Urheber durch freigestellte Werknutzungen erleiden.[873] Im Umkehrschluss sollte die Belastung bei dem Beteiligten verortet werden, der die Nutzungshandlung im urheberrechtlichen Sinne vornimmt.[874] Denn er zieht die Vorteile aus der Werknutzung, dessen Kehrseite den Nachteil für betroffene Urheber begründet. Bestehen praktische Schwierigkeiten bei der Inanspruchnahme des Werknutzers, kann die Vergütungspflicht allerdings auch auf einen anderen Beteiligten verlagert werden. In diesen Fällen muss jedoch die Möglichkeit bestehen, die Vergütungsleistung an den eigentlichen Werknutzer weiterzureichen sowie einen Erstattungsmechanismus zu bedienen, wenn die Vergütung tatsächlich nicht geschuldet gewesen ist.[875]

Die vorgeschlagene Vergütungspflicht knüpft an den Trainingsdatensammler an. Dieser stellt die Vervielfältigungen für das Training von KI her. Er ist demzufolge Werknutzer. Die Anknüpfung an die Trainingsdatensammler ist infolgedessen mit den unionsrechtlichen Anforderungen vereinbar.

c) Zwischenergebnis

Alles in allem ist die Einführung der in §§ 44c ff. UrhG-E umzusetzenden Vergütungspflicht für das Web Scraping von urheberrechtlich geschützten Trainingsdaten auf Basis von § 44b UrhG unionsrechtlich zulässig.

IV. Zusammenfassung und kritische Würdigung des Regelungsvorschlags

Insgesamt können die identifizierten Regelungsbedürfnisse für das Web Scraping von Werken als Trainingsdaten insbesondere durch die Anfügung eines vierten Absatzes in der Schrankenbestimmung für Text und Data Mi-

873 ErwGr. 35 InfoSoc-RL.

874 Für die Privatkopievergütung EuGH, Urt. v. 24.03.2022 - C-433/20, GRUR 2022, 558 Rn. 43 – Austro-Mechana/Strato.

875 Für die Privatkopievergütung EuGH, Urt. v. 24.03.2022 - C-433/20, GRUR 2022, 558 Rn. 45, 47 – Austro-Mechana/Strato.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

ning aus § 44b UrhG verwirklich werden. § 44b Abs. 4 UrhG-E bestimmt, dass weder Löschungspflicht noch Opt-out-Mechanismus bei Vervielfältigungshandlungen für das Sammeln von Trainingsdaten eingreifen sollen. Außerdem ist eine Klarstellung zur Anwendbarkeit der Schrankenbestimmung auf das Sammeln urheberrechtlich geschützter Trainingsdaten für die Entwicklung von KI zweckmäßig. Sie sollte als Satz 2 in § 44b Abs. 1 UrhG eingefügt werden.

Die Bereichsausnahme von Löschungserfordernis und Opt-out-Mechanismus kann aufgrund der Harmonisierungswirkung der DSM-RL aber nicht ohne weiteres im nationalen Urheberrecht umgesetzt werden. Hierfür fehlt es an einer entsprechenden Regelung oder Öffnungsklausel auf unionsrechtlicher Ebene. Der Umsetzung des § 44b Abs. 4 UrhG-E muss demzufolge eine europäische Rechtssetzungsinitiative vorangehen.

Die gesetzliche Vergütung, die Urheber für die Verwendung ihrer Werke als Trainingsdaten erhalten sollen, kann durch Einfügung der §§ 44c bis 44e UrhG-E umgesetzt werden. Die Einfügung des Vergütungsmechanismus ist auch unionsrechtlich zulässig. Er ist nach dem Vorbild der Hersteller- und Betreibervergütung (§§ 54 ff. UrhG) ausgestaltet. Dies bietet sich an, weil auch die Hersteller- und Betreibervergütung Pauschalisierungen bei Vergütungshöhe und Anspruchsinhaberschaft vorsieht. Beides ist auch im vorliegenden Fall notwendig. Vorteilhaft an einer Umsetzung nach dem Vorbild der Hersteller- und Betreibervergütung ist außerdem, dass den Verwertungsgesellschaften, die die Vergütungsansprüche geltend machen sollen (§ 44e Abs. 1 UrhG-E), die grundlegenden Mechanismen bekannt sind.

B. Speicherung urheberrechtlich geschützter Inputdaten

Die Speicherung urheberrechtlich geschützter Inputdaten wird von der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a UrhG erfasst.[876] Dies gilt jedenfalls so lange, wie die erfassten Inhalte nicht zur Verwirklichung von Sekundärzwecken abgesondert und die Werke tatsächlich nur vorübergehend gespeichert werden. Außerdem sind die Vervielfälti-

876 Dazu 3. Teil D. II. 1.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

gungshandlungen in der Anwendungsphase des KI-Systems auch von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG gedeckt.[877] Damit sind sie im Grundsatz urheberrechtlich erlaubnisfrei möglich.

Regulatorisch zweckmäßig ist allerdings, die Speicherung von Werken als Inputdaten für KI durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte, also bereits in den §§ 15 ff. UrhG definitorisch aus dem Ausschließlichkeitsrecht betroffener Urheber auszuklammern.[878]

I. Unmittelbar begrenzende Ausgestaltung des urheberrechtlichen Vervielfältigungsrechts

Damit die Speicherung urheberrechtlich geschützter Inputdaten durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte freigestellt wird, muss eine Anpassung des Vervielfältigungsrechts aus § 16 UrhG erfolgen. Eine entsprechende Regelung könnte wie folgt aussehen:

§ 16 UrhG-E: Vervielfältigungsrecht[879]

(1) Das Vervielfältigungsrecht ist das Recht, Vervielfältigungsstücke des Werkes herzustellen, gleichviel ob vorübergehend oder dauerhaft, in welchem Verfahren und in welcher Zahl.

(2) Eine Vervielfältigung ist auch die Übertragung des Werkes auf Vorrichtungen zur wiederholbaren Wiedergabe von Bild- oder Tonfolgen (Bild- oder Tonträger), gleichviel, ob es sich um die Aufnahme einer Wiedergabe des Werkes auf einen Bild- oder Tonträger oder um die Übertragung des Werkes von einem Bild- oder Tonträger auf einen anderen handelt.

(3) Keine Vervielfältigung ist die vorübergehende Zwischenspeicherung von Werken zu dem ausschließlichen Zweck, sie als Eingabedaten (Art. 3 Nr. 33 VO [EU] 2024/1689) für die Anwendung eines Systems künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E) zu verwenden.

Ziel der Regelung ist dabei die Einschränkung des über § 95a UrhG gewährten Schutzes für technische Schutzmaßnahmen, die gegen die Speicherung von Inputdaten gerichtet sind.[880] Demzufolge kommt es nicht entscheidend auf die mit § 16 Abs. 3 UrhG-E hervorgerufene Darlegungs- und Beweislastverteilung an. Aufgrund der negativen Formulierung des

877 Dazu 3. Teil D. II. 2. a).
878 Dazu im Detail siehe oben unter 2. Teil D. II.
879 Hervorhebung steht für neu einzufügenden Normtext.
880 Siehe im Detail unter 2. Teil D. II. 2.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

Regelungsvorschlags wäre der Nutzer für die Freistellungsvoraussetzungen im Rahmen des § 16 Abs. 3 UrhG-E darlegungs- und ggf. beweisbelastet.

II. Unionsrechtliche Perspektive auf den Regelungsvorschlag und Schlussfolgerungen

Auch das Vervielfältigungsrecht ist durch Art. 2 InfoSoc-RL unionsrechtlich harmonisiert. Nationale Bestimmungen dürfen jedenfalls nicht hinter dem durch das Unionsrecht gewährten Schutzniveau zurückbleiben. Dies wäre bei einer Einschränkung des Vervielfältigungsrechts entsprechend § 16 Abs. 3 UrhG-E allerdings der Fall. Eine solche Rechtsänderung wäre daher unionsrechtswidrig. Auch die regulatorisch notwendigen Anpassungen am urheberrechtlichen Vervielfältigungsrecht könnten deswegen nur umgesetzt werden, wenn ihr eine europäische Rechtssetzungsinitiative vorangeht.

C. Normalisierung urheberrechtlich geschützter Trainings- und Inputdaten als Teil der Datenvorverarbeitung

Bei der Einordnung der Vorverarbeitung urheberrechtlich geschützter Trainings- sowie Inputdaten in den Rechtsrahmen des geltenden Urheberrechts muss zwischen der Normalisierung und dem Feature Encoding der Werke unterschieden werden. Bei der Normalisierung von Werken können – je nach eingesetztem Normalisierungsverfahren – Vervielfältigungshandlungen gemäß § 16 Abs. 1 UrhG stattfinden. Diese sind sowohl von der Schrankenbestimmung für vorübergehende Vervielfältigungen (§ 44a Nr. 2 UrhG) als auch von der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) gedeckt und daher im Grundsatz urheberrechtlich erlaubnisfrei möglich. Beim nicht-expressiven Feature Encoding der Werke fehlt es hingegen bereits an der Herstellung eines Vervielfältigungsstücks im Sinne des § 16 Abs. 1 UrhG. Auch diese unmittelbar aus der Reichweite der urheberrechtlichen Verwertungsrechte folgende Rechtslage macht das Feature Encoding im Ergebnis erlaubnisfrei möglich.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

Sowohl das Feature Encoding als auch die Normalisierung von als Input- sowie Trainingsdaten generierten Werken sollte aus einer regulatorischen Perspektive jedoch durch eine unmittelbar begrenzende, definitorische Ausgestaltung des Vervielfältigungsrechts aus dem urheberrechtlichen Schutzbereich ausgeklammert sein.[881] Für die Normalisierung der Werke bedarf es damit trotz der Anwendbarkeit mehrerer Schrankenbestimmungen de lege ferenda einer Anpassung des urheberrechtlichen Regelungsregimes.

I. Unmittelbar begrenzende Ausgestaltung des urheberrechtlichen Vervielfältigungsrechts

Damit die Normalisierung urheberrechtlich geschützter Werke durch eine unmittelbar begrenzende Ausgestaltung des Vervielfältigungsrechts freigestellt wird, muss eine Anpassung in § 16 UrhG erfolgen. Ziel der Regelung ist die mittelbare Einschränkung des über § 95a UrhG erfolgenden Schutzes von gegen die Normalisierungsschritte eingesetzten technischen Schutzmaßnahmen.[882] Auch hier kommt es demzufolge nicht auf die aus der Regelung folgende Darlegungs- und Beweislastverteilung an. Da sich aus dem Freistellungsbedürfnis für die Speicherung von Werken als Inputdaten bereits Änderungsbedarf in Bezug auf § 16 UrhG ergeben hat,[883] können beide Anpassungsvorschläge wie folgt zusammengeführt werden:

§ 16 UrhG-E: Vervielfältigungsrecht[884]

(1) Das Vervielfältigungsrecht ist das Recht, Vervielfältigungsstücke des Werkes herzustellen, gleichviel ob vorübergehend oder dauerhaft, in welchem Verfahren und in welcher Zahl.

(2) Eine Vervielfältigung ist auch die Übertragung des Werkes auf Vorrichtungen zur wiederholbaren Wiedergabe von Bild- oder Tonfolgen (Bild- oder Tonträger), gleichviel, ob es sich um die Aufnahme einer Wiedergabe des Werkes auf einen Bild- oder Tonträger oder um die Übertragung des Werkes von einem Bild- oder Tonträger auf einen anderen handelt.

(3) Keine Vervielfältigung sind folgende Werkverwendungen:

1. die vorübergehende Zwischenspeicherung von Werken zu dem ausschließlichen Zweck, sie als Eingabedaten (Art. 3 Nr. 33 VO [EU] 2024/1689) für die

881 Dazu im Detail siehe oben unter 2. Teil C. sowie E.
882 Siehe im Detail unter 2. Teil C. II. 2. sowie 2. Teil D. II. 2.
883 Dazu siehe oben unter 4. Teil B. I.
884 Hervorhebung steht für neu einzufügenden Normtext.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -

Anwendung eines Systems künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG) zu verwenden, und

2. die Transponierung von als Trainings- oder Eingabedaten (Art. 3 Nr. 33 VO [EU] 2024/1689) für ein System künstlicher Intelligenz erfassten Werken auf eine vergleichbare Skala (Normalisierung).

II. Unionsrechtliche Perspektive auf den Regelungsvorschlag und Schlussfolgerungen

Auch bei einer Freistellung der Normalisierung von urheberrechtlich geschützten Trainings- sowie Inputdaten muss allerdings berücksichtigt werden, dass das urheberrechtliche Vervielfältigungsrecht durch Art. 2 InfoSoc-RL harmonisiert ist. Eine dem § 16 Abs. 3 UrhG-E entsprechende Rechtsänderung wäre daher, wie bereits für die Speicherung von Werken als Inputdaten gezeigt werden konnte, unionsrechtswidrig. Auch der Anpassung des urheberrechtlichen Regelungsregimes für die Normalisierung von Werken zur Verwendung als Trainings- und Inputdaten müsste deswegen eine europäische Rechtssetzungsinitiative vorausgehen.

https://doi.org/10.5771/9783748963462-307 - am 12.01.2026, 11:20:47. https://www.inlibra.com/de/agb - Open Access -