

Fünfter Teil: Zusammenfassung und Schlussfolgerungen aus der Untersuchung

Urheberrechtlich geschützte Werke werden in erheblichem Umfang für das Training von KI, also von auf künstlichen neuronalen Netzen basierenden Systemen maschinellen Lernens, benötigt. Außerdem können sie als Inputdaten und damit als Berechnungsgrundlage für die Anwendung von KI-Systemen im konkreten Einzelfall dienen. Entwicklung sowie Einsatz von KI und das urheberrechtliche Regelungsregime stehen daher in einer engen Beziehung zueinander. Vor diesem Hintergrund muss untersucht werden, welche Rahmenbedingungen das Urheberrecht für die Verwendung von Werken als Trainings- und Inputdaten zur Verfügung stellt. Fraglich ist außerdem, ob Perspektiven für den urheberrechtlichen Rechtsrahmen aufgezeigt werden können. Konkret sind die aufgeworfenen Fragestellungen dabei in Bezug auf die praktisch relevante Speicherung urheberrechtlich geschützter Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern, die eigentliche Durchführung des Trainings auf Basis von Werken, die Speicherung urheberrechtlich geschützter Inputdaten für die Anwendung von KI-Systemen in konkreten Anwendungsszenarien sowie die notwendige Vorverarbeitung dieser gewonnenen Rohdaten in den Blick zu nehmen. Um Perspektiven für den urheberrechtlichen Rechtsrahmen aufzuzeigen, müssen für die aufgezeigten Fallgruppen außerdem regulatorische Freistellungsbedürfnisse identifiziert werden. Aus einem Vergleich dieser Freistellungsbedürfnisse, also des Soll-Zustands, mit der urheberrechtlichen *Lex lata*, dem Ist-Zustand des Regelungsregimes, lassen sich Änderungsbedarfe im Urheberrecht ermitteln. Auf Basis dieser Änderungsbedarfe können anschließend konkrete Anpassungsvorschläge für den urheberrechtlichen Regelungskomplex erarbeitet werden.

Die Speicherung urheberrechtlich geschützter Trainingsdaten mit Hilfe von Webcrawlern sollte urheberrechtlich freigestellt werden. Grund hierfür ist, dass die Innovationsinteressen der Trainingsdatensammler sowie bestehende praktische Lizenzierungsschwierigkeiten das nur geringfügig beeinträchtigte Verwertungsinteresse betroffener Urheber überwiegen. Das Freistellungsinteresse wird dem Grunde nach durch die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfüllt. Dies ergibt sich aus einer richtlinienkonformen Auslegung des Schrankentatbestands. Die Re-

gelung privilegiert auch Vervielfältigungshandlungen kommerziell tätiger Trainingsdatensammler. Dennoch können nicht sämtliche der identifizierten Regelungsbedürfnisse mit der Schrankenbestimmung aus § 44b UrhG erfüllt werden. Denn zum einen ist eine Kompensation der freigestellten Speicherung urheberrechtlich geschützter Trainingsdaten zugunsten betroffener Urheber mittels eines gesetzlichen Vergütungsanspruchs notwendig. Einen solchen Vergütungsanspruch sieht § 44b UrhG aber nicht vor. Darüber hinaus sind der in § 44b Abs. 3 UrhG angelegte Opt-out-Mechanismus, der durch Vorbehaltserklärungen in natürlicher Sprache ausgelöst werden kann, sowie die in § 44b Abs. 2 S. 2 UrhG vorgesehene Löschungspflicht für Trainingsdatensammler nicht interessengerecht. Im Einzelfall kann eine Freistungswirkung zugunsten von Trainingsdatensammlern zwar auch über die Schrankenbestimmung für die Nutzung von Werken als unwesentliches Beiwerk aus § 57 UrhG erreicht werden. Das strukturelle Frestellungsinteresse kann hierdurch *de lege lata* allerdings nicht erfüllt werden. Die Schrankenbestimmung für Text und Data Mining ist *de lege ferenda* deswegen durch einen § 44b Abs. 4 UrhG-E zu ergänzen. Er soll für das Sammeln urheberrechtlich geschützter Trainingsdaten eine Bereichsausnahme von der Löschungspflicht und dem Opt-out-Mechanismus vorsehen. Da § 44b UrhG den Art. 4 DSM-RL in nationales Recht umsetzt, kann das allerdings nur im Zusammenspiel mit einer vorangehenden europäischen Rechtssetzungsinitiative erfolgen. Der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG ist mit §§ 44c bis 44e UrhG-E darüber hinaus eine gesetzliche Vergütungspflicht für die Vervielfältigung von Werken zum Training von KI anzufügen. Sie ist nach dem Vorbild der Hersteller- und Betreibervergütung (§§ 54 ff. UrhG) auszugestalten. Die Einführung dieses Vergütungsmechanismus ist unionsrechtlich bereits jetzt zulässig.

Auch für die eigentliche Durchführung des KI-Trainings, also die Anpassung der numerischen Parameter des KNN auf Basis der zum Training gesammelten Werke, konnte ein regulatorisches Frestellungsinteresse ermittelt werden. Sie ist tatsächlich und rechtlich von der Aggregation der Trainingsdaten abzugrenzen. Zweckmäßig ist dabei eine vergütungsfreie Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte. Damit das Frestellungsinteresse im geltenden Urheberrecht abgebildet wird, müssen alle vier Teilaspekte des Trainings diese regulatorischen Bedürfnisse erfüllen. Das ist bei der Umwandlung der Trainingsdaten in numerische Tensoren, dem sogenannten Feature Encoding, der eigentlichen Analyse der Trainingswerke im Rechen-

prozess, hierbei angefertigter, vorübergehender Kopien der in Tensoren umgewandelten Trainingsdaten sowie insbesondere auch der Anpassung der Gewichtungswerte und anderen Parameter innerhalb des KNN der Fall. Denn im Rahmen dieser Teilschritte werden keine Vervielfältigungsstücke der Werke im Sinne des § 16 UrhG angefertigt. Das gilt selbst dann, wenn einzelne Werke in den Strukturen des KNN subsymbolisch gespeichert und im Einzelfall wieder rekonstruiert werden können. Diese Teilaspekte des Trainingsvorgangs berühren demzufolge nicht den Schutzbereich des urheberrechtlichen Ausschließlichkeitsrechts. Etwas anderes gilt nur für die Normalisierung von urheberrechtlich geschützten Trainingsdaten, also das Transponieren der Werke auf eine vergleichbare Skala. Je nach Normalisierungsverfahren und Werkart können hierbei Vervielfältigungshandlungen im Sinne des § 16 UrhG vorgenommen werden. Auch wenn die Normalisierung der Trainingsdaten von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfasst und damit freigestellt ist, wird der geltende urheberrechtliche Rechtsrahmen damit nicht dem *Wie* des identifizierten Regelungsbedürfnisses gerecht. Als Folge dessen ist § 16 UrhG durch einen § 16 Abs. 3 UrhG-E zu ergänzen. Er nimmt unter anderem die Normalisierung von als Trainingsdaten erfassten Werken vom Vervielfältigungsbegriff aus. Da das Vervielfältigungsrecht unionsrechtlich harmonisiert ist, muss dem ebenfalls eine europäische Rechtssetzungsinitiative vorangehen. Eine isoliert nationale Anpassung des UrhG ist im konkreten Fall nicht möglich.

Ebenfalls von § 16 Abs.3 UrhG-E umfasst werden sollte die Normalisierung von als Inputdaten für die Anwendung von KI-Systemen gespeicherten Werken. Sie muss de lege lata ebenfalls unter den Vervielfältigungsbegriff aus § 16 UrhG gefasst werden. Auch wenn die Normalisierungsvervielfältigungen der als Inputdaten gespeicherten Werke ebenfalls so lange von § 44b UrhG privilegiert und damit freigestellt werden, wie kein wirksamer Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist, löst dies vor dem Hintergrund des regulatorischen Freistellungsinteresses ein Regelungsdefizit im geltenden urheberrechtlichen Rechtsrahmen aus. Solange die Inputdaten nicht zur Verwirklichung von Sekundärzwecken abgesondert und nur vorübergehend vervielfältigt werden, ergibt sich eine Erlaubnis der Normalisierung urheberrechtlich geschützter Inputdaten weiterhin aus der Schrankenbestimmung für vorübergehende Vervielfältigungen, § 44a UrhG. Damit kann das regulatorische Interesse nach einer Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertrrechte allerdings gleichsam nicht erfüllt werden.

Im Übrigen können die für die Vorverarbeitung von Trainingsdaten herausgearbeiteten Grundsätze auf die Vorverarbeitung von Werken übertragen werden, die als Input für KI-Systeme verwendet werden sollen. Das führt dazu, dass bei der Umwandeln als Input verwendeter Werke in numerische Repräsentationen, also Tensoren, keine Vervielfältigungsstücke im Sinne des § 16 UrhG hergestellt werden. Das Feature Encoding der als Inputdaten verwendeten Werke greift demzufolge nicht in das urheberrechtliche Ausschließlichkeitsrecht ein. Diesbezüglich besteht infolgedessen kein Konflikt zwischen den identifizierten regulatorischen Freistellungsinteressen und dem geltenden Urheberrecht. Für das Feature Encoding der urheberrechtlich geschützten Inputdaten ist daher keine Anpassung des urheberrechtlichen Rechtsrahmens erforderlich.

Auch die vorgelagerte Speicherung urheberrechtlich geschützter Inputdaten beispielsweise durch Sensoren wie Kameras oder mit Hilfe von Webcrawlern sollte urheberrechtlich freigestellt sein. Infolge einer Zusammenschau verschiedener Begründungsansätze erweist sich die Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte als interessengerecht. Diesem Regelungsbedürfnis wird der geltende Rechtsrahmen jedoch nicht gerecht. Bei der Speicherung urheberrechtlich geschützter Inputdaten werden vorübergehende Vervielfältigungen hergestellt, die in das ausschließliche Vervielfältigungsrecht der Urheber nach § 16 UrhG eingreifen. Die Anwendung einer KI kann allerdings sowohl unter den Begriff des Text und Data Mining (§ 44b Abs. 1 UrhG) subsumiert als auch als rechtmäßige Nutzung (§ 44a Nr. 2 UrhG) eingeordnet werden. Solange mit den Werkstücken keine Sekundärverwertung bezweckt wird, die eine dauerhafte Speicherung der Inhalte notwendig macht, sind die Vervielfältigungshandlungen daher von den Schrankenbestimmungen für vorübergehende Vervielfältigungen aus § 44a UrhG sowie für Text und Data Mining aus § 44b UrhG erlaubnisfrei gestellt. Letzteres aber ist aber nur dann der Fall, wenn kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist. Dennoch wird mit der Anwendung der Schrankenbestimmungen nur dem *Ob*, nicht aber dem *Wie* der identifizierten Freistellungsbedürfnisse Rechnung getragen. Infolgedessen ist in dem neu einzufügenden § 16 Abs. 3 UrhG-E festzulegen, dass bei der vorübergehenden Zwischenspeicherung von Werken zu dem ausschließlichen Zweck, sie als Inputdaten für die Anwendung eines KI-Systems zu verwenden, keine Vervielfältigungsstücke im Sinne des § 16 Abs. 1 UrhG hergestellt werden. Aufgrund der unionsrechtlichen Vollharmonisierung des Vervielfältigungsrechts ist die Freistellung der Speicherung von Inputdaten für die Anwendung eines KI-Systems zu verwenden, die keine Vervielfältigungsstücke im Sinne des § 16 Abs. 1 UrhG herstellt, als eine Ausgestaltung der Schrankenbestimmungen für vorübergehende Vervielfältigungen zu verstehen.

fältigungsrechts kann auch das allerdings nur im Zusammenspiel mit einer unionsrechtlichen Rechtssetzungsinitiative erfolgen.

Alles in allem stellt das Urheberrecht einen Rechtsrahmen zur Verfügung, der sowohl das Training als auch den Einsatz von KI auf Basis urheberrechtlich geschützter Trainings- beziehungsweise Inputdaten grundsätzlich ermöglicht. Von besonderer Bedeutung hierfür ist die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG. Maßgebend für die urheberrechtliche Erlaubnisfreiheit des Trainings ist auch, dass keine Vervielfältigungsstücke der zum Training verwendeten Werke dauerhaft im KNN im Sinne des § 16 Abs.1 UrhG hinterlegt werden. Ein Eingriff in das urheberrechtliche Vervielfältigungsrecht findet hierdurch demzufolge nicht statt. Hierfür kommt es nicht entscheidend auf die technische Funktionsweise des KI-Trainings an. Bedeutung erlangt vielmehr ein nicht rein-technisches, sondern normatives Verständnis des urheberrechtlichen Vervielfältigungsbegriffs.

Insbesondere der auch in natürlicher Sprache erklärbarer Nutzungsvorbehalt im Sinne des § 44b Abs.3 UrhG sowie das Fehlen einer Vergütungspflicht für die Speicherung von Werken als Trainingsdaten zugunsten betroffener Urheber, aber auch andere Bestandteile des urheberrechtlichen Regelungsregimes führen allerdings dazu, dass sich die aus den Sachverhaltskonstellationen extrahierbaren Interessenslagen nicht vollständig im geltenden Urheberrecht abbilden. Nationaler und europäischer Gesetzgeber sind deswegen angehalten, den urheberrechtlichen Rechtsrahmen für das Training und die Anwendung von KI einer Revision zu unterziehen. Aufgrund des unionsrechtlichen Mehrebenensystems kommt es hierbei insbesondere auf den europäischen Richtliniengeber an. Der nationale Gesetzgeber kann jedoch bereits jetzt ein Vergütungsmodell für die Nutzung von Werken als Trainingsdaten für KI-Entwicklung einführen, ohne dass dies unionsrechtlich vorgezeichnet sein müsste. Bei dieser Gelegenheit sollte er in § 44b Abs.1 UrhG auch eine Klarstellung dahingehend einfügen, dass das KI-Training unter den Text und Data Mining-Begriff der Regelung fällt. Dabei kann er auch die sich aus Art. 4 Abs.3 DSM-RL ergebenden Voraussetzungen der Angemessenheit und Ausdrücklichkeit von Nutzungsvorbehaltserklärungen im nationalen Regelungstext umsetzen. Sie können aktuell nur im Wege einer richtlinienkonformen Auslegung des Tatbestands berücksichtigt werden.

