

Künstliche Intelligenz und das Urheberrecht:

Wo stehen wir?

Erläuterungen zur aktuellen Rechtslage

geschrieben von Jonathan Pukas

Kaum ein Rechtsgebiet steht durch die Entwicklungen im Bereich künstlicher Intelligenz so sehr im Rampenlicht wie das Urheberrecht. Das hängt nicht zuletzt damit zusammen, dass weitreichende technische Veränderungen diese Rechtsmaterie seit jeher besonders prägen. Der Zeitstrahl solcher Umwälzungen ließe sich bis zur Erfindung des mechanischen Buchdrucks im 15. Jahrhundert zurückverfolgen. Er legte den Grundstein für das urheberrechtliche Ausschließlichkeitsrecht, kraft dessen beispielsweise Autorinnen und Autoren eine Vergütung für die Nutzung ihrer Werke verlangen oder konkurrierende Nutzungen verbieten können.

Dieser Tradition folgend kommt es nicht von ungefähr, dass sich das Urheberrecht mit dem Aufkommen von Large Language Models und KI-gestützten Chatbots wie OpenAI's »ChatGPT« zu einem der Taktgeber der Debatte rund um künstliche Intelligenz entwickelt hat. Im Zentrum steht dabei die hochbrisante Frage, ob künstliche Intelligenz ohne direkte Zustimmung betroffener Rechteinhaber (also zum Beispiel Urheber oder Verwerter wie Verlage) mit deren Werken trainiert werden darf. An diese Kernproblematik knüpfen auch eine Vielzahl weiterer praxisrelevanter Fragen an.

Die Bedürfnisse der verschiedenen Interessengruppen werden seit einiger Zeit politisch diskutiert. Gleichzeitig arbeitet auch die juristische Fachwelt daran, belastbare Antworten auf die drängenden Fragen zu finden.

Die Hauptfrage: KI-Training als TDM von § 44b UrhG erfasst?

Ob KI mit Werken trainiert werden darf, hängt davon ab, ob eine gesetzliche Ausnahme – eine sogenannte Schrankenbestimmung – eingreift. Diese erlaubt die Nutzung von Werken trotz des urheberrechtlichen Ausschließlichkeitsrechts. Ist das der Fall, könnten insbesondere im Internet aufgefundene Werke ohne ausdrückliche Zustimmung der Rechteinhaber zum Training von KI genutzt werden.

Trotz diverser Unklarheiten im Gesetzeswortlaut geht der überwiegende Teil der Rechtswissenschaft mittlerweile davon aus, dass die notwendige Erlaubnis – jedenfalls grundsätzlich – von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG gewährt wird. Auch Ausführungen des Hamburger Landgerichts zum ersten KI-Trainings-Prozess in Deutschland (»Laion-Entscheidung«) deuten in diese Richtung. Entscheidend ist, dass das KI-Training unter den Begriff des »Text und Data Mining« gefasst werden kann. Die Speicherung von Werken zum KI-Training ist daher grundsätzlich zulässig, ohne dass von betroffenen Rechteinhabern individuell eine Erlaubnis eingeholt werden muss. Nach dem seit Juni 2021 geltenden Urheberrecht ist das KI-Training mit Werken daher grundsätzlich zustimmungsfrei möglich.

Zum »maschinenlesbaren« Nutzungsvorbehalt

Die vorgenannte Regelung kommt allerdings nur dann zur Anwendung, wenn eine Reihe anderer Voraussetzungen ebenfalls erfüllt sind. Insbesondere darf kein Vorbehalt gegen die Nutzung der Werke abgegeben worden sein. Die Erklärung des Nutzungsvorbehalts muss für online veröffentlichte Inhalte in maschinenlesbarer Form erfolgen. Das umfasst nach Auffassung des Verfassers auch Vorbehaltserklärungen in natürlicher Sprache. Eine Erklärung »in Code« ist zwar möglich, aber rechtlich nicht notwendig. Einige Einschränkungen hierzu sind allerdings deswegen zu machen, weil das Gesetz auch vorschreibt, dass die abgegebenen Nutzungsvorbehaltserklärungen »angemessen« sein müssen. Nicht jede Erklärung kann daher die gesetzliche Erlaubniswirkung beenden. Voraussetzung ist beispielsweise, dass die gewählte

Formulierung gewisse Klarheitsanforderungen erfüllt. Es sollten stets übliche Phrasen wie beispielsweise »Text und Data Mining vorbehalten« genutzt werden. Wer sichergehen will, sollte die Vorbehaltserklärung zudem auch in englischer Sprache abgeben.

Wirksame Vorbehaltserklärungen sind jedoch im Ausgangspunkt nur von sogenannten ausschließlichen Rechtsinhabern möglich. Sie zeichnen sich dadurch aus, dass sie als einzige Person berechtigt sind, ein Werk auf eine bestimmte Art und Weise zu nutzen.

Mit Spannung erwartet wurde außerdem der am 10. Juli 2025 von der EU-Kommission veröffentlichte »Code of Practice for General-Purpose AI Models«. Dieser sollte unter anderem konkretisieren, welche Arten von Vorbehaltserklärungen von KI-Anbietern bei der Datensammlung beachtet werden müssen. Eine entscheidende Klärung der Fragestellung ist durch den auf europäischer Ebene verhandelten Verhaltenskodex allerdings nicht gelungen. Interessant ist jedoch, dass auch mit Hilfe des Robots Exclusion Protocols, also in robots.txt-Dateien abgegebene Vorbehaltserklärungen, gemäß dem Verhaltenskodex als wirksam angesehen und von Datensammlern beachtet werden sollen. Diese Sichtweise war jedenfalls in Deutschland bereits vor Veröffentlichung des Dokuments verbreitet. Der große Wurf ist hier daher – jedenfalls betreffend der Vorbehaltsregelung – ausgeblieben. Bemerkenswert ist aber, dass das Training auf Grundlage von Inhalten, die von Webseiten mit illegalem Geschäftsmodell extrahiert worden sind, im Verhaltenskodex als nicht rechtskonform angesehen wird.

Zur Inputverarbeitung

Auch wenn sich die rechtswissenschaftliche Forschung der Fragestellung bisher nur am Rande zugewandt hat, umfasst die gesetzliche Erlaubnis für Text und Data Mining außerdem die Verwendung von urheberrechtlich geschützten Inhalten als Inputdaten für KI-Systeme.

Als Inputdaten werden solche Daten bezeichnet, die nicht bei der Entwicklung, sondern bei der Anwendung einer KI als Berechnungsgrundlage genutzt werden. Sie bilden also die Basis für die Erzeugung konkreten Outputs durch KI-Nutzer wie beispielsweise Anwender eines KI-gestützten Chatbots. Im Kontext dieser Chatbots werden die Inputdaten oftmals auch als »Prompts« bezeichnet. Da die gesetzliche Erlaubnis die Nutzung der Werke als Inputdaten unabhängig vom konkreten Typ des KI-Systems erfasst, gilt die Freistellung auch für die Verwendung von Werken als Berechnungsgrundlage generativer KI-Systeme wie beispielsweise KI-gestützter Bildgeneratoren. Wichtig ist in diesem Zusammenhang aber, dass hiervon die Frage zu trennen ist, ob ein generierter KI-Output Teile eines urheberrechtlich geschützten Werks wiedergibt und seine Nutzung daher möglicherweise urheberrechtsverletzend ist. Die Input- und die Output-Frage müssen stets unabhängig voneinander beantwortet werden. Dies steht in einem gewissen Widerspruch zu ersten US-amerikanischen Gerichtsentscheidungen, die die Rechtmäßigkeit des Inputs unter anderem direkt davon abhängig machen, welche Art von Erzeugnis das betroffene KI-System ausgegeben hat. Hier kommen also die unterschiedlichen Grundkonzeptionen der verschiedenen (Urheber-)Rechtssysteme zum Vorschein.

Ungeachtet der grundsätzlichen Anwendbarkeit der gesetzlichen Erlaubnis für Text und Data Mining aus dem Urheberrechtsgesetz müssen auch bei der Nutzung von urheberrechtlich geschützten Werken als Inputdaten die übrigen Voraussetzungen der Regelung aus § 44b UrhG beachtet werden. Erst dann ist die Nutzung tatsächlich urheberrechtskonform möglich, ohne individuelle Erlaubnisse von den betroffenen Rechteinhabern einholen zu müssen.

Vorausgesetzt wird unter anderem, dass die genutzten Werke auch »rechtmäßig zugänglich« gewesen sein müssen. Obwohl diese Voraussetzung auf den ersten Blick einfach erscheint, ergeben sich aus ihr in der Praxis oft komplexe Fragestellungen – besonders, wenn mehrere Personen an der Nutzung beteiligt sind. Auch der bereits erwähnte, auf europäischer Ebene entwickelte KI-Verhaltenskodex präzisiert die Voraussetzung weiter. Darüber hinaus sind bei der Verwendung urheberrechtlich geschützter Werke als Inputdaten nach geltendem Recht ebenfalls erklärte Nutzungsvorbehalte zu beachten.

Wohin bewegt sich das Urheberrecht?

Wendet man den Blick ab von der kontrovers diskutierten und hier nur skizzierten Rechtslage hin zu Entwicklungsperspektiven für das Urheberrecht, muss vor allem eine Schlussfolgerung gezogen werden: Der aktuell anwendbare Rechtsrahmen weist erhebliche Defizite auf. Diese machen ein Tätigwerden sowohl des nationalen als auch des europäischen Gesetzgebers dringend notwendig. Eine gute Gelegenheit hierfür wäre die Evaluation der letzten Urheberrechtsrichtlinie, die im Jahr 2019 in Kraft getreten ist und bereits damals diverse Diskussionen, insbesondere über die Plattformverantwortlichkeit und den berühmten »Artikel 17«, auslöste. Sie ist auch Grundlage des für KI maßgeblichen Urheberrechts.

Hiervon unabhängig könnte der nationale Gesetzgeber aber bereits heute tätig werden und eine gesetzliche Vergütungspflicht für vom KI-Training betroffene Urheber einführen. Der aktuelle Koalitionsvertrag enthält bereits eine Forderung, die Rechteinhaber für KI-Training zu vergüten. Insofern nach geltendem Recht kein Nutzungsvorbehalt erklärt worden ist, könnte die Vergütung nach dem Vorbild der für den Ausgleich der Privatkopiefreiheit eingeführten Abgabe auf Geräte und Speichermedien ausgestaltet werden. Trainingsdatensammler sollten danach verpflichtet werden, eine pauschale Abgabe an Verwertungsgesellschaften zu entrichten, wenn sie Daten sammeln, bei denen davon ausgegangen werden kann, dass sie zumindest teilweise urheberrechtlich geschützt sind. Das wäre beispielsweise bei der Sammlung von Bildern aus frei verfügbaren Internetquellen der Fall. Die Abgabe könnte sodann in Form einer Pauschalzahlung an diejenigen Urheber ausgekehrt werden, die ihre Rechte in das Repertoire einer Verwertungsgesellschaft wie der VG Bild-Kunst oder der VG Wort eingebracht haben. Ein solches System würde nicht zuletzt die im kreativen Ökosystem dringend benötigten Verwertungsgesellschaften stärken. Soweit die Beteiligten dies als zweckmäßig und interessengerecht ansehen, lassen die nach geltendem Recht möglichen Nutzungsvorbehalte zudem schließlich auch Raum für den Abschluss individueller Lizenzverträge. Die Geschichte um das Urheberrecht im Wandel durch künstliche Intelligenz ist daher »to be continued«.

Mehr dazu

Jonathan Pukas, »Werknutzung und künstliche Intelligenz« (Diss.), Nomos 2025. Das Werk ist auch online im Open Access abrufbar.

Dieser Text ist zuerst erschienen in [Politik & Kultur 11/2025](#).