

Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung

Retrieval-Augmented Generation (RAG) wird verwendet, um die Zuverlässigkeit und Aktualität des Outputs von Sprachmodellen zu erhöhen. RAG ist weit verbreitet, wurde aber bislang kaum urheberrechtlich untersucht. Dieser Beitrag vermittelt einen Überblick über die Rechtsfragen und nimmt eine erste Einordnung vor.

I. Einleitung

Die Entwicklung und Verwendung von Modellen und Systemen der generativen Künstlichen Intelligenz (KI) werfen komplexe rechtliche Fragen auf, auch im Urheberrecht. Seit einiger Zeit wird vor allem die Nutzung von urheberrechtlich geschützten Werken beim Training von

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

285

KI-Modellen intensiv diskutiert.¹ Andere Fragen wurden bisher kaum beleuchtet, so insbesondere die Nutzung von Werken im Rahmen der sogenannten Retrieval-Augmented Generation (RAG).² Dieser Beitrag nimmt eine urheberrechtliche Einordnung dieser weit verbreiteten Anwendung von KI vor. Er versteht sich ausdrücklich nicht als Beitrag zum rechtspolitischen Diskurs. Vielmehr wird – im Sinne einer klassischen Subsumtion – ein konkretes Phänomen urheberrechtlich eingeordnet und es werden erste Leitlinien für dessen urheberrechtliche Beurteilung skizziert. Der Beitrag erhebt nicht den Anspruch, alle Fragen zu klären. Nicht näher untersucht wird namentlich die Frage, welchem Akteur eine allfällige Verletzung von Urheberrechten als Täter, Teilnehmer oder Störer zuzurechnen ist und ob bestimmte Handlungen allenfalls durch eine Haftungsbeschränkung privilegiert werden sollten.

Retrieval-Augmented Generation (RAG) soll ein zentrales Problem von KI-Systemen, insbesondere von Chatbots, beheben: Diese neigen zu »Halluzinationen«, also dazu, falsche oder frei erfundene Aussagen zu produzieren, die zwar plausibel klingen, aber keinen Bezug zur Realität haben.³ Dies geschieht, weil große Sprachmodelle (Large Language Models, LLMs) Muster aus ihren Trainingsdaten extrapolieren und dabei statistische Zusammenhänge zwischen Wörtern und Wortteilen (Tokens) »erlernen«. ⁴ Auf dieser Grundlage sagen sie voraus, welcher Token angesichts der Abfolge der vorangegangenen Tokens mit einer hohen Wahrscheinlichkeit als nächstes folgt.⁵ Da Sprachmodelle auf statistischen Zusammenhängen und Wahrscheinlichkeitsrechnungen beruhen und für das Generieren von Text und nicht für die Wiedergabe richtiger Informationen optimiert werden,

ist es systemimmanent, dass ein gewisser Anteil der von ihnen generierten Aussagen nicht zutrifft.⁶

Das sogenannte »Halluzinieren« von KI-Modellen lässt sich mit verschiedenen technischen Ansätzen reduzieren, bislang aber nicht vollständig verhindern.⁷ So kann die Richtigkeit des Outputs von großen Sprachmodellen durch zusätzliches Training mit verlässlichen und kuratierten Daten (Fine-Tuning) verbessert werden.⁸ Auch spezialisierte Submodelle⁹ oder Fact-Checking-Algorithmen¹⁰ können zur Qualitätssicherung beitragen. Ein besonders vielversprechender und weit verbreiteter Ansatz zur Erhöhung der Zuverlässigkeit und Aktualität besteht darin, Sprachmodelle mit externen Datenquellen zu verknüpfen, etwa mit Datenbanken oder mit Inhalten von Websites. Dieser Ansatz wird als Retrieval-Augmented Generation (RAG) bezeichnet.¹¹ Dabei kombiniert das Modell sein allgemeines »Wissen« und seine Fähigkeit zur sprachlichen Formulierung mit spezifischen Informationen, die aus unternehmensinternen Datenbanken oder von Websites abgerufen werden.¹²

Einsatzmöglichkeiten von RAG finden sich in zahlreichen Branchen: Im juristischen Bereich sind KI-Systeme bspw. durch den Zugriff auf interne Vertragsdatenbanken in der Lage, qualitativ hochwertige Entwürfe für neue Dokumente zu generieren.¹³ Im Versicherungswesen können Chatbots relevante Informationen aus Richtlinien oder Versicherungsbedingungen abrufen, um Kundenanfragen präzise und kontextbezogen zu beantworten.¹⁴ Und im Mediensektor können Dienste wie Perplexity.ai mithilfe von RAG über das Internet auf die Inhalte von Medien

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

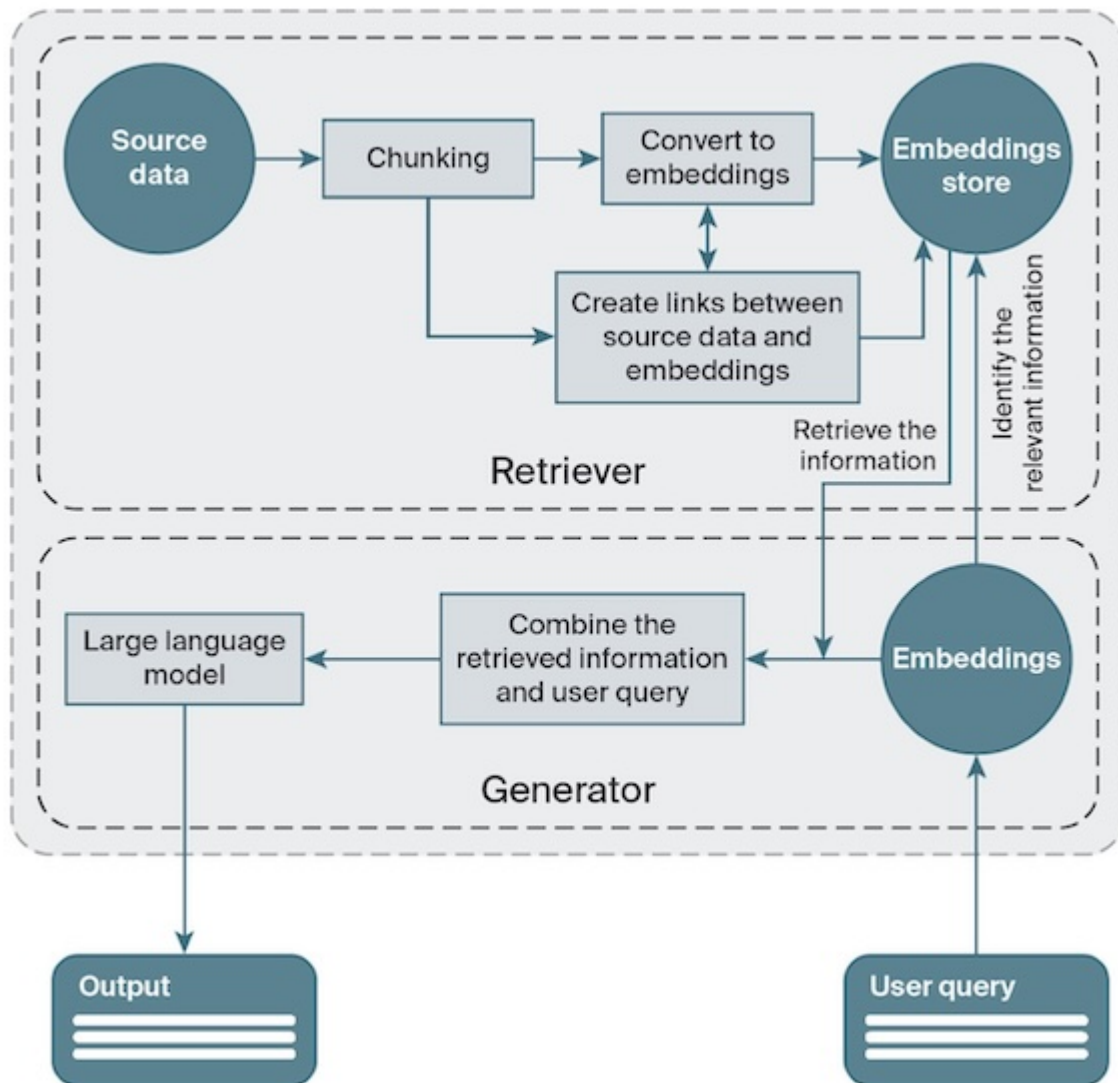
286

zugreifen und ihre Nutzer damit auf eine Anfrage hin zuverlässig über das tagesaktuelle Geschehen informieren.¹⁵

II. Retrieval-Augmented Generation (RAG)

Große Sprachmodelle wie GPT 5.1, Gemini 3, Mistral Medium 3.1 oder DeepSeek-V3.2-Exp werden mit riesigen Datenmengen trainiert.¹⁶ Da die Kosten für das Training sehr hoch sind,¹⁷ können diese Modelle nicht laufend mit aktuellen Daten nachtrainiert und aktualisiert werden. Die in Sprachmodellen enthaltenen Daten sind deshalb rasch veraltet.¹⁸ Hier setzt RAG an, indem es zwei Komponenten miteinander verbindet: Beim sogenannten Retrieval wird eine interne Datenbank oder das Internet (genauer: das World Wide Web) verwendet, um für das Beantworten einer Anfrage relevante Informationen zu identifizieren und abzurufen. Bei der sogenannten Generation verarbeitet das Sprachmodell die abgerufenen Informationen und generiert auf deren Grundlage eine (in der Regel) präzise und zutreffende Antwort.¹⁹

Die Funktionsweise von RAG lässt sich im Wesentlichen wie folgt grafisch darstellen:²⁰



(ZUM 2026, Seite 286, Abb. 1)

(1) Der Prozess beginnt mit der Eingabe eines Prompts durch einen Nutzer (User query). Diese Eingabe wird als Suchanfrage interpretiert und vom System verarbeitet. Dabei wird der eingegebene Text analysiert, um relevante Schlüsselbegriffe und semantische Beziehungen zu identifizieren.²¹ Das erfolgt mithilfe des sogenannten Embeddings. Dabei handelt es sich um numerische Vektordarstellungen von Texten oder anderen Daten, die es einem KI-Modell erlauben, Ähnlichkeiten mit den Inhalten der abzufragenden Datenbank zu identifizieren.²² Dies ermöglicht eine semantische Suche, bei der nicht nur nach exakten Schlüsselwörtern, sondern auch nach inhaltlich verwandten Informationen gesucht wird.²³

(2) Im Retriever werden die relevanten Informationen bereitgestellt, die für die Beantwortung des Prompts gebraucht werden. Der Retriever-Prozess beginnt mit der Vorbereitung der Source data (Quelldaten) und endet mit der Speicherung der generierten Embeddings in einem Vektorspeicher (Embeddings store).²⁴ Die Daten können aus verschiedenen Quellen stammen und in verschiedenen Formaten vorliegen, Texte bspw. als PDF oder als HTML-Code.²⁵ Die Daten werden verarbeitet, indem sie in kleinere Teile, bspw. Absätze eines Textes, aufgeteilt (Chunking) und anschließend in Vektordarstellungen umgewandelt (Embedding) und mit der Source data verlinkt werden. Die Embeddings

werden dann im Embeddings store gespeichert und mit den Embeddings der User query verglichen, um die relevanten Informationen in der Source data, bspw. ein (oder mehrere) Dokument(e), zu identifizieren. Weiterverarbeitet werden dabei meist nur die besten Treffer (häufig die besten fünf oder zehn); so können die relevantesten Dokumente für die Texterzeugung priorisiert werden. Diese werden dann für das Generieren der Antwort an den Generator weitergegeben.²⁶ Beim Retrieval mit Zugriff auf das

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung(ZUM 2026, 284)

287

Web werden die Quelldaten nicht vorgängig in einem Embeddings store dauerhaft gespeichert.²⁷ Stattdessen werden die relevanten Inhalte anlassbezogen über eine Application Programming Interface (API)²⁸ mithilfe einer Websuche ermittelt, von einer (oder mehreren) Website(s) abgerufen und in temporäre Repräsentationen (Embeddings) überführt, die ausschließlich für die jeweilige Anfrage verarbeitet und regelmäßig nicht dauerhaft gespeichert werden.²⁹

(3) Der Generator erstellt aufgrund der erhaltenen Dokumente eine Antwort auf die User query. Dabei werden die User query und die Dokumente miteinander verbunden und dem Sprachmodell als Prompt eingegeben. Dieses analysiert die bereitgestellten Informationen und generiert auf dieser Basis eine Antwort (Output).³⁰ Die Dokumente können dabei direkt und vollständig in den Prompt integriert werden;³¹ bei umfangreichen Dokumenten werden stattdessen relevante Abschnitte zusammengefasst und dem Sprachmodell in komprimierter Form zur Verfügung gestellt.³² Beim Erzeugen der Antwort (Output) nutzt das Sprachmodell linguistische Techniken, um klare und verständliche Formulierungen zu generieren.³³ Zusätzlich kann ein Validierungsmechanismus integriert werden, um die Richtigkeit und Konsistenz der Antworten zu überprüfen. Auch hier gibt es verschiedene Ansätze: Das Sprachmodell kann anhand interner Wahrscheinlichkeiten bewerten,³⁴ ob die Antwort plausibel ist; alternativ lassen sich externe Fact-Checking-Mechanismen verwenden, bspw. spezielle KI-Modelle³⁵ oder APIs³⁶, um die generierte Antwort mit vertrauenswürdigen Quellen abzugleichen. Zudem können unklare oder fehlerhafte Antworten aufgrund von Rückmeldungen der Nutzer erkannt und künftige Generationen des Sprachmodells verbessert werden.³⁷

Aus technischer Sicht sind zwei Arten von RAG-Systemen zu unterscheiden: Bei der »klassischen« Nutzung von RAG greift das Sprachmodell auf die Inhalte einer internen Datenbank zu, also auf eine Datenbank, die von einem Unternehmen oder einer anderen Organisation betrieben wird.³⁸ Daneben gibt es auch RAG-Systeme, die nicht auf eine interne Datenbank, sondern auf das Web zugreifen.³⁹

Aus Sicht des Urheberrechts lassen sich die Vorgänge eines RAG-Systems in drei Schritte unterteilen, wobei der erste Schritt nur für das »klassische« RAG relevant ist: erstens, das Speichern von Werken in einer Datenbank, zweitens, der Zugriff auf diese Werke, die

Identifikation der relevanten Inhalte im Retriever und die Integration dieser Inhalte in einen von einem Nutzer verfassten Prompt sowie, drittens, die allfällige Wiedergabe von Werken oder Teilen davon in der Ausgabe (Output). Diese drei Schritte werden nachfolgend urheberrechtlich eingeordnet.

III. Speichern von Werken in einer internen Datenbank

Beim »klassischen« RAG stellt sich als Erstes die Frage, ob das Speichern von Werken in einer internen Datenbank zulässig ist. Zu beurteilen sind dabei das Speichern der Werke als Source data und das Speichern als numerische Vektordarstellungen im Embeddings store.

Die digitale Speicherung ist als Vervielfältigung iSv § 16 UrhG zu qualifizieren.⁴⁰ Stets zulässig ist das Speichern

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

288

von Werken in einer internen Datenbank, wenn der Datenbankhersteller Inhaber des Vervielfältigungsrechts an den Werken ist, wenn er sich das Recht hat einräumen lassen (§ 31 UrhG) oder wenn ihm der Rechtsinhaber die Vervielfältigung schuldrechtlich gestattet hat (§ 29 Abs. 2 UrhG).⁴¹ Zulässig ist eine solche Speicherung aber auch, wenn die damit einhergehenden Vervielfältigungen von einer Schranke freigestellt werden (gesetzliche Nutzungserlaubnis).⁴²

Als Schranken für Vervielfältigungen beim Speichern von Werken in einer internen Datenbank kommen die Text und Data Mining-Schranken (§ 44b und § 60d UrhG) und die Schranke für Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch infrage.

1. Schranken für Text und Data Mining

Die Vervielfältigung von Werken für Text und Data Mining wird von zwei Schranken freigestellt: Nach § 44b Abs. 2 S. 1 UrhG sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining zulässig, sofern der Rechteinhaber keinen Nutzungsvorbehalt im Sinne von § 44b Abs. 3 UrhG angebracht hat. § 60d UrhG stellt Vervielfältigungen für Text und Data Mining für Zwecke der wissenschaftlichen Forschung frei; diese Vervielfältigungen können die Rechteinhaber nicht durch einen Nutzungsvorbehalt verhindern.⁴³

Text und Data Mining ist nach § 44b Abs. 1 UrhG die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen. Die Bedeutung der Text und Data Mining-Schranken für die Entwicklung und Verwendung von KI war bei deren Einführung in Art. 3 f. der DSM-Richtlinie⁴⁴ kaum vorhersehbar; entsprechend war lange Zeit umstritten, ob diese Schranken auf das Training von KI-Modellen angewendet werden können. Mittlerweile hat der europäische Gesetzgeber die Frage mit Erlass der KI-Verordnung geklärt, indem er in Art. 53 Abs. 1 lit. c KI-VO ausdrücklich auf Art. 4 Abs. 3 DSM-RL verweist.⁴⁵ Damit ist klar, dass die Vorgänge beim Training von KI-Modellen nach

europäischem und damit auch nach deutschem Recht⁴⁶ als Text und Data Mining zu verstehen sind.

Wie beim Training von KI-Modellen sind auch die Vorgänge bei der Verwendung eines RAG-Systems als Text und Data Mining zu qualifizieren. Auch hier werden digitale Werke automatisiert analysiert, um daraus Informationen zu gewinnen.⁴⁷ Die Schranken für Text und Data Mining stellen dabei nicht nur den eigentlichen Vorgang des Mining, also die automatisierte Analyse der Werke, frei,⁴⁸ sondern auch die für die Vorbereitung erforderlichen Vervielfältigungen, insbesondere die Bildung des Datenkorpus, der später beim Mining analysiert wird.⁴⁹ Damit stellen die Schranken auch die Speicherung von Werken in einer internen Datenbank frei, die später im Rahmen von RAG genutzt werden sollen. Voraussetzung ist allerdings, dass die Werke rechtmäßig zugänglich sind (§ 44b Abs. 1 S. 1 UrhG). Das ist der Fall, wenn die Werke frei im Web zugänglich sind oder der Nutzer im Rahmen einer Lizenz Zugang zu diesen hat.⁵⁰

Dient das Text und Data Mining der wissenschaftlichen Forschung und wird es von einer hierzu berechtigten Forschungsorganisation im Sinne von § 60d Abs. 2 und 3 UrhG durchgeführt, ist das vorgängige Speichern von Werken in einer Datenbank stets freigestellt.⁵¹ In allen anderen Fällen kann der Rechteinhaber die Speicherung allerdings durch Anbringen eines wirksamen Nutzungsvorbehalts verhindern (§ 44b Abs. 3 UrhG).⁵² Das Speichern von Werken in einer internen Datenbank ist damit nur zulässig, wenn der Rechteinhaber keinen Nutzungsvorbehalt angebracht hat oder die Werke für die wissenschaftliche Forschung durch eine berechnigte Forschungsinstitution genutzt werden.

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

289

2. Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch

Die beim Speichern von Werken in einer internen Datenbank entstehenden Vervielfältigungen könnten von der Schranke zugunsten des privaten und sonstigen eigenen Gebrauchs freigestellt werden. § 53 Abs. 1 S. 1 UrhG erlaubt, einzelne Vervielfältigungsstücke eines Werkes für den privaten Gebrauch herzustellen.⁵³ Privater Gebrauch ist die Benutzung innerhalb der privaten Sphäre durch eine natürliche Person, welche die Vervielfältigung selbst herstellt oder durch einen anderen herstellen lässt.⁵⁴ Da RAG-Systeme allerdings in aller Regel nicht dem privaten Gebrauch dienen, wird diese Schranke kaum je anwendbar sein.⁵⁵

Ausgeschlossen ist die Anwendung der Schranke zugunsten des sonstigen eigenen Gebrauchs. Nach § 53 Abs. 2 S. 1 Nr. 4 UrhG ist es zwar zulässig, einzelne Vervielfältigungsstücke eines Werkes zum sonstigen eigenen Gebrauch herzustellen oder herstellen zu lassen. Freigestellt wird aber nur die Vervielfältigung von kleinen Teilen eines Werkes oder einzelnen Beiträgen in Zeitungen oder Zeitschriften, wenn es sich dabei um ein seit mindestens zwei Jahren vergriffenes Werk handelt.⁵⁶ Zudem schränkt § 53 Abs. 2 S. 2

UrhG den Anwendungsbereich dieser Schranke auf fotomechanische Vervielfältigungen und ausschließlich analoge Nutzungen ein. Damit kommt die Schranke zugunsten des sonstigen eigenen Gebrauchs für die hier vorliegende digitale Speicherung nicht infrage.⁵⁷

3. Fazit

Die Speicherung von Werken Dritter in einer internen Datenbank zur späteren Nutzung im Rahmen von RAG wird grundsätzlich durch die Schranken des Text und Data Minings freigestellt. Das gilt allerdings nur, wenn der Rechteinhaber sich die Nutzung seiner Werke für Text und Data Mining nicht wirksam vorbehalten hat.⁵⁸ Hat er einen Nutzungsvorbehalt angebracht, dürfen die Werke nur gespeichert und genutzt werden, wenn ein RAG-System von einer natürlichen Person zum privaten Gebrauch oder von einer berechtigten Forschungsinstitution für die wissenschaftliche Forschung genutzt wird.⁵⁹

IV. Zugriff auf Werke, Identifikation relevanter Inhalte und Integration in Prompts

Beim Zugriff auf Werke, die in einer internen Datenbank gespeichert sind, bei der Identifikation der relevanten Inhalte und bei der Integration dieser Inhalte in einen von einem Nutzer verfassten Prompt werden systemintern Vervielfältigungen erstellt. Dasselbe gilt beim Zugriff auf Werke, die im Web verfügbar sind. Das Vorgehen kann dabei auf einer Identifikation der relevanten Inhalte beruhen, das demjenigen bei RAG mit Zugriff auf eine interne Datenbank weitgehend entspricht. Bisweilen werden die relevanten Inhalte aber gar nicht spezifisch identifiziert; stattdessen werden einfach Teile von Websites (bspw. die ersten Absätze), die aufgrund einer Websuche als relevant identifiziert wurden, direkt in den Prompt übernommen.

Werden Werke Dritter ohne deren Zustimmung genutzt, stellt sich die Frage, ob das Erstellen der damit verbundenen Vervielfältigungen durch eine Schranke freigestellt wird.⁶⁰ Im Vordergrund steht dabei die allgemeine Schranke für Text und Data Mining (§ 44b UrhG). Hat sich der Rechteinhaber allerdings die Nutzung der Werke für das Text und Data Mining wirksam vorbehalten⁶¹ und werden die Werke nicht von einer berechtigten Forschungsinstitution für die wissenschaftliche Forschung verwendet, stellt sich die Frage, ob die Nutzung im Rahmen von RAG von einer anderen Schranke freigestellt wird. Infrage kommen dabei die Schranke für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG) und die Schranke für den privaten und sonstigen eigenen Gebrauch (§ 53 UrhG).

1. Text und Data Mining-Schranken

Die Vervielfältigungen, die beim Zugriff auf Werke in einer internen Datenbank oder im Web, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt erstellt werden, dienen der automatisierten Analyse der vervielfältigten digitalen Werke, um aus diesen Informationen zu gewinnen. Diese Vervielfältigungen werden von der allgemeinen Schranke für Text und Data Mining (§ 44b UrhG) freigestellt, wenn der Rechte

inhaber keinen wirksamen Nutzungsvorbehalt iSv § 44b Abs. 3 UrhG angebracht hat.⁶² Andernfalls wird die Nutzung von Werken im Rahmen eines RAG-Systems nur freigestellt, wenn die dazu erforderlichen Vervielfältigungen von einer berechtigten Forschungsinstitution für Zwecke der wissenschaftlichen Forschung erstellt werden (§ 60d UrhG).⁶³

Voraussetzung für die Freistellung durch die Text und Data Mining-Schranken ist, dass die genutzten Werke rechtmäßig zugänglich sind (§ 44b Abs. 2 S. 1 UrhG; § 60d Abs. 1 iVm § 44b Abs. 2 S. 1 UrhG).⁶⁴ Dies ist beim Zugriff auf Werke in einer internen Datenbank stets der Fall. Beim Zugriff auf Werke im Web ist zu unterscheiden: Ohne Weiteres zulässig ist die Nutzung, wenn die Werke frei zugänglich sind.⁶⁵ Anderes gilt, wenn der Zugang durch technische Schutzmaßnahmen (§ 95a Abs. 2 UrhG) eingeschränkt ist.⁶⁶ In diesem Fall sieht § 95a Abs. 1 UrhG zwar ein Umgehungsverbot vor.⁶⁷ Gemäß § 95b Abs. 1 S. 1 Nr. 1, Abs. 3 Nr. 1 UrhG ist der Rechteinhaber aber verpflichtet, dem durch die Schranke von § 44b UrhG Begünstigten die notwendigen Mittel zur Verfügung zu stellen, um die freigestellte Nutzung vornehmen zu können.⁶⁸ Die eigenständige Umgehung der technischen Schutzmaßnahme durch den Nutzer ist zwar unzulässig.⁶⁹ Nutzt der durch die Schranke Begünstigte aber die ihm zur Verfügung gestellten Mittel, um Zugang zu den Werken zu erhalten, so liegt ein rechtmäßiger Zugang im Sinne von § 44b Abs. 2 S. 1 UrhG vor.⁷⁰

2. Vorübergehende Vervielfältigungshandlungen

Vorübergehende Vervielfältigungen von Werken sind gemäß § 44a UrhG zulässig, wenn vier Voraussetzungen erfüllt sind: Die Vervielfältigung muss flüchtig oder begleitend sein, sie muss einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen, sie muss ausschließlich der Übertragung in einem Netz zwischen Dritten durch einen Vermittler (§ 44a Nr. 1 UrhG) oder einer rechtmäßigen Nutzung (§ 44a Nr. 2 UrhG) dienen und sie darf keine eigenständige wirtschaftliche Bedeutung haben.

a) Flüchtig oder begleitend

Flüchtig sind Vervielfältigungen, die kurzlebig sind, also mit der Beendigung des technischen Vorgangs gleich wieder gelöscht werden.⁷¹ Begleitend sind Vervielfältigungen, die beiläufig im Rahmen eines technischen Verfahrens erstellt werden und damit der Erleichterung eines anderen Nutzungsvorgangs dienen.⁷² Die Vervielfältigungen, die beim Zugriff auf die Werke in einer Datenbank, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt erstellt werden, werden nach diesem Vorgang gleich wieder gelöscht. Sie sind damit flüchtig. Da sie zudem Teil eines technischen Vorgangs sind, der im Retriever abläuft, dürften sie auch als begleitend einzustufen sein.⁷³

b) Integraler und wesentlicher Teil eines technischen Verfahrens

Die Vervielfältigung muss einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen.⁷⁴ Das ist der Fall, wenn die Vervielfältigungshandlungen vollständig im Rahmen der Durchführung eines technischen Verfahrens vorgenommen werden und die

Vervielfältigungen in dem Sinn notwendig sind, dass das technische Verfahren ohne sie nicht einwandfrei und effizient funktionieren könnte.⁷⁵ Auch diese Voraussetzung ist erfüllt, weil RAG als technisches Verfahren nur funktioniert, wenn das System auf die Werke in einer Datenbank oder im Web zugreifen sowie die relevanten Inhalte identifizieren und in den Prompt integrieren kann. Ohne diese Vervielfältigungen könnte das technische Verfahren eines RAG-Systems nicht ordnungsgemäß ablaufen.

c) Übertragung in einem Netz oder rechtmäßige Nutzung

Zulässig sind vorübergehende Vervielfältigungen nur, wenn ihr alleiniger Zweck darin besteht, eine Übertragung in einem Netz zwischen Dritten durch einen Vermittler (§ 44a Nr. 1 UrhG) oder eine rechtmäßige Nut

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

291

zung (§ 44a Nr. 2 UrhG) zu ermöglichen. Die erste Konstellation zielt auf Vervielfältigungen beim Übermitteln von Werken über das Internet ab. Eine Übertragung in einem Netz zwischen Dritten durch einen Vermittler liegt beim »klassischen« RAG aber nicht vor; bei RAG mit Zugriff auf das Web findet zwar eine Übertragung in einem Netz statt, die Vervielfältigungen beschränken sich aber nicht auf diese. Damit ist für beide Arten von RAG zu prüfen, ob die Vervielfältigungen, die beim Zugriff auf die Werke, bei der Identifikation der relevanten Inhalte und bei deren Integration in den Prompt erstellt werden, im Sinne der zweiten Konstellation eine rechtmäßige Nutzung ermöglichen. Rechtmäßig sind Nutzungen, welche der Rechteinhaber erlaubt hat, und solche, die gesetzlich zulässig sind, die also keinem Ausschließlichkeitsrecht unterliegen oder von einer Schranke freigestellt werden.⁷⁶

aa) Werkgenuss als rechtmäßige Nutzung

Als rechtmäßige Nutzung iSv § 44a Nr. 2 UrhG gilt auch der Werkgenuss,⁷⁷ also die Wahrnehmung von Werken mit den menschlichen Sinnen. Der Werkgenuss ist urheberrechtlich frei; er wird von den Verwertungsrechten nicht erfasst.⁷⁸ Freigestellt sind damit auch vorübergehende Vervielfältigungen, die unmittelbar dem Werkgenuss dienen, bspw. das Anzeigen von Inhalten auf einem Bildschirm, das Zwischenspeichern beim Abspielen einer DVD oder das Speichern von Inhalten im Arbeitsspeicher beim Abrufen einer Website (Browsing).⁷⁹

Die Vervielfältigungen beim Zugriff auf Werke, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt dienen zwar letztlich dem Werkgenuss, also der Wahrnehmung der in den Werken enthaltenen Informationen mit den menschlichen Sinnen, dies aber nur mittelbar, indem den Werken die relevanten Informationen entnommen und beim Erzeugen des Outputs verwendet werden. Mit den menschlichen Sinnen wahrgenommen wird nur der mithilfe von RAG generierte Output eines KI-Systems, nicht aber die Werke, die beim Generieren des Outputs verwendet worden sind. Die hier zu

prüfenden Vervielfältigungen dienen damit nicht unmittelbar dem Werkgenuss. Das gilt für RAG mit Zugriff auf eine interne Datenbank ebenso wie für RAG mit Zugriff auf das Web.

bb) Freistellung durch eine Schranke

Eine rechtmäßige Nutzung liegt vor, wenn eine Schranke (gesetzliche Nutzungserlaubnis) greift. Infrage kommen die Text und Data Mining-Schranken (§ 44b und § 60d UrhG) und die Schranke für Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch (§ 53 UrhG).

Eine eigenständige Bedeutung hat die Freistellung vorübergehender Vervielfältigungen durch die Schranke von § 44a UrhG in der vorliegenden Konstellation allerdings nur, wenn § 44a UrhG die Handlungen von Personen freistellt, die nicht bereits von einer anderen Schranke (also der »Zielschranke«, welche die Nutzung freistellt, die durch die vorübergehende Vervielfältigung ermöglicht wird) erfasst sind. Damit ist zu unterscheiden:

Bei RAG mit Zugriff auf eine interne Datenbank werden alle urheberrechtlich relevanten Handlungen von derselben (in aller Regel juristischen) Person ausgeführt. Der Schranke von § 44a UrhG kommt damit keine eigenständige Bedeutung zu.⁸⁰

Bei RAG mit Zugriff auf das Web sind zwei Konstellationen zu unterscheiden: Ergibt sich die Rechtmäßigkeit einer Vervielfältigung aus der allgemeinen Schranke für Text und Data Mining (§ 44b UrhG), dann kommt der Schranke von § 44a UrhG keine eigenständige Bedeutung zu, weil § 44b UrhG die Vervielfältigungen für Text und Data Mining unabhängig von der Person freistellt, welche diese vornimmt. Ergibt sich die Rechtmäßigkeit aber aus der Schranke für Vervielfältigungen zum privaten Gebrauch (§ 53 UrhG) oder aus der Schranke für Text und Data Mining für wissenschaftliche Zwecke (§ 60d UrhG) und nutzt die natürliche Person oder die Forschungsinstitution dafür das Angebot eines Dritten (bspw. ChatGPT), dann können die von diesem Dritten vorgenommenen vorübergehenden Vervielfältigungen freigestellt werden, weil sie einer von diesen Schranken (§ 53 und § 60d UrhG) freigestellten und damit rechtmäßigen Nutzung dienen. Das gilt allerdings nur, wenn diese Nutzungen keine eigenständige wirtschaftliche Bedeutung haben.⁸¹

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

292

d) Keine eigenständige wirtschaftliche Bedeutung

Vorübergehende Vervielfältigungen sind nur von § 44a UrhG freigestellt, wenn sie keine eigenständige wirtschaftliche Bedeutung haben. Eine solche liegt nach der Rechtsprechung des EuGH vor, wenn die vorübergehenden Vervielfältigungen einen zusätzlichen wirtschaftlichen Vorteil schaffen.⁸² In der Lehre wird eine eigenständige wirtschaftliche Bedeutung bejaht, wenn die vorübergehenden Vervielfältigungen eine neue, eigenständige Nutzungsmöglichkeit eröffnen.⁸³

Das Tatbestandsmerkmal der eigenständigen wirtschaftlichen Bedeutung weist enge inhaltliche Bezüge zur zweiten und dritten Stufe des Drei-Stufen-Tests auf, nach denen

Schranken die normale Verwertung eines Werks nicht beeinträchtigen und die berechtigten Interessen der Rechteinhaber nicht unangemessen verletzen dürfen.⁸⁴ Die Auslegung dieses Tatbestandsmerkmals kann sich deshalb am Verständnis der zweiten und dritten Stufe des Drei-Stufen-Tests orientieren. Bei der Beurteilung der zweiten Stufe des Drei-Stufen-Tests ist anerkannt, dass auch künftige Verwertungsformen einzubeziehen sind, wenn sie nicht rein hypothetisch erscheinen, sondern aufgrund der technischen und wirtschaftlichen Entwicklung absehbar sind.⁸⁵ Diese Perspektive erscheint auch bei der Prüfung der eigenständigen wirtschaftlichen Bedeutung nach § 44a UrhG sachgerecht.⁸⁶

Angesichts der schon heute großen und wohl weiter zunehmenden Bedeutung von RAG ist ohne Weiteres denkbar, dass für das Recht zur Nutzung von Werken im Rahmen von RAG schon heute oder jedenfalls zukünftig Lizenzen erteilt werden, soweit die Nutzung nicht gesetzlich freigestellt ist. Eine weite Auslegung der Schranke von § 44a UrhG würde das Entstehen eines solchen Marktes und die damit verbundene Möglichkeit zur Verwertung von Urheberrechten verhindern.⁸⁷ Daher wird man davon ausgehen müssen, dass die vorübergehenden Vervielfältigungen, die im Rahmen von RAG erstellt werden, eine eigenständige wirtschaftliche Bedeutung haben oder zumindest künftig haben könnten.⁸⁸ Das gilt sowohl für das »klassische« RAG als auch für die Nutzung von RAG mit Zugriff auf das Web.

Zwar lässt sich durchaus argumentieren, dass die im Rahmen von RAG erstellten Vervielfältigungen keine Nutzung von Werken ermöglichen, die nicht auch ohne diese Vervielfältigungen möglich wäre, zumal der Unterschied allein in der Reduktion des Aufwandes liegt, der bei einer menschlichen Identifikation relevanter Dokumente und Inhalte ungleich größer ist als beim Einsatz eines Retrievers.⁸⁹ Andererseits erlaubt RAG gerade wegen der geringeren Kosten und der höheren Geschwindigkeit der Identifikation und Entnahme der relevanten Inhalte, Werke anders und intensiver zu nutzen, als dies bisher möglich war. Dies ist als wirtschaftlicher Vorteil und damit als eigenständige wirtschaftliche Bedeutung zu qualifizieren.⁹⁰ Besonders deutlich zeigt sich dies, wenn mittels RAG auf Inhalte von Medien zugegriffen wird, die im Web zugänglich sind: Der Zugriff auf online zugängliche Inhalte von Medien ermöglicht einem Chatbot, Anfragen von Nutzern tagesaktuell zu beantworten. Diese können sich so über das aktuelle Geschehen durch die Nutzung von Chatbots informieren und auf den direkten Konsum von Medien verzichten.⁹¹ Chatbots können damit das Angebot der Medien substituieren. Zugleich erhöht die Nutzung von Medieninhalten mittels RAG die Zuverlässigkeit und Aktualität der Antworten von Chatbots und damit die Attraktivität dieser Angebote. RAG leistet damit einen wesentlichen Beitrag zum wirtschaftlichen Erfolg dieser Dienste.

Auch wenn die Frage nicht abschließend beurteilt werden kann, wird man somit davon ausgehen müssen, dass die vorübergehenden Vervielfältigungen, die im Rahmen von RAG

erstellt werden, eine eigenständige wirtschaftliche Bedeutung haben oder zumindest künftig haben könnten.

e) Zwischenfazit

Die Analyse der vier Tatbestandsvoraussetzungen hat gezeigt, dass die Vervielfältigungen, die beim Zugriff auf Werke in einer internen Datenbank, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

293

Prompt erstellt werden, die Voraussetzungen für die Qualifikation als vorübergehende Vervielfältigungshandlungen im Sinne von § 44a UrhG wohl nicht erfüllen. Das gilt auch für die bei der Verwendung von RAG mit Zugriff auf das Web erstellten Vervielfältigungen.⁹²

3. Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch

Zu prüfen bleibt, ob die Vervielfältigungen, die beim Zugriff auf die Werke in einer Datenbank oder im Web, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt erstellt werden, von der Schranke des § 53 UrhG freigestellt werden. Wie beim Speichern von Werken in einer internen Datenbank sind die Voraussetzungen der Schranke zugunsten des sonstigen eigenen Gebrauchs (§ 53 Abs. 2 S. 1 Nr. 4 und § 53 Abs. 2 S. 2 UrhG) auch hier nicht erfüllt.⁹³ Dasselbe gilt für die Schranke des § 53 Abs. 2 Nr. 3 UrhG, die Vervielfältigungen zur eigenen Unterrichtung über Tagesfragen privilegiert, zumal diese Schranke nur die Vervielfältigung von durch Funk gesendeten Werken freistellt und nicht greift, wenn Werke im Web zugänglich sind.⁹⁴

Infrage kommt damit nur die Schranke zugunsten des privaten Gebrauchs (§ 53 Abs. 1 UrhG).⁹⁵ Diese setzt voraus, dass die Vervielfältigung eines Werkes durch eine natürliche Person zum privaten Gebrauch erfolgt.⁹⁶ Die Nutzung von Werken bei RAG mit Zugriff auf eine interne Datenbank kann damit kaum je über diese Schranke freigestellt werden, weil natürliche Personen in aller Regel keine solchen Datenbanken erstellen.

Möglich ist aber eine Anwendung bei RAG mit Zugriff auf das Web, namentlich wenn natürliche Personen die Dienste Dritter (bspw. von Chatbot-Anbietern) nutzen und diese aufgrund einer Anfrage eines Nutzers auf im Web zugängliche Werke zugreifen, die relevanten Inhalte identifizieren und in den Prompt integrieren. Bei der Anwendung dieser Schranke stellt sich allerdings die Frage, ob der Nutzer oder der Anbieter eines KI-Systems als Hersteller der Vervielfältigung zu qualifizieren ist und ob der Anbieter allenfalls als anderer im Sinne von § 53 Abs. 1 S. 2 UrhG verstanden werden kann.⁹⁷ Bei der maßgeblichen technischen Betrachtung wird man wohl zu dem Schluss kommen, dass der Anbieter als Hersteller der Vervielfältigungen zu qualifizieren ist, die beim Zugriff auf die Werke im Web, bei der Identifikation der relevanten Inhalte und bei deren Integration in den Prompt des Nutzers erstellt werden, zumal dieser »die körperliche Festlegung technisch bewerkstelligt«.⁹⁸ Fraglich ist allerdings, ob der Anbieter die Vervielfältigungen im Auftrag

des Nutzers herstellt, womit diese dem Nutzer zuzurechnen wären. Das erscheint aber zweifelhaft, zumal der Anbieter dem Nutzer damit eine urheberrechtlich relevante Nutzung von Werken in einem Ausmaß und in einer Intensität erschließen würde, die sich mit den Erwägungen, die eine Privilegierung des Privatgebrauchs rechtfertigen, wohl nicht vereinbaren ließe.⁹⁹ Diese Frage kann allerdings offenbleiben, zumal die Anwendung der Schranke von § 53 Abs. 1 S. 2 UrhG schon daran scheitert, dass der Output vom Anbieter nicht unentgeltlich generiert wird und die Vervielfältigung auch nicht auf Papier oder auf einem ähnlichen Träger mittels fotomechanischer oder ähnlicher Verfahren erstellt wird.¹⁰⁰ Vielmehr zahlen die natürlichen Personen für die Nutzung von Chatbots und ähnlichen Diensten und damit auch für die beim Einsatz von RAG im System erstellten Vervielfältigungen. Das Entgelt wird entweder im Rahmen eines Abonnements als wiederkehrende Vergütung geleistet oder bei frei zugänglichen Angeboten durch Einwilligung in die Nutzung von Personendaten. Möglich bleibt damit nur die Anwendung der Schranke von § 53 Abs. 1 S. 1 UrhG, was aber voraussetzt, dass der Nutzer selbst als Hersteller qualifiziert werden kann.

4. Fazit

Beim Zugriff auf Werke, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt werden systeminterne Vervielfältigungen von Werken erstellt. Das gilt für RAG mit Zugriff auf eine interne Datenbank ebenso wie bei RAG mit Zugriff auf das Web. Das Erstellen dieser Vervielfältigungen wird von der allgemeinen Text und Data Mining-Schranke (§ 44b UrhG) umfassend freigestellt. Diese Schranke greift aber nur, wenn der Rechteinhaber keinen wirksamen Nutzungsvorbehalt angebracht hat. Wurde ein Vorbehalt angebracht, so werden die beim Einsatz von RAG erstellten Vervielfältigun

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung(ZUM 2026, 284)

294

gen nur freigestellt, wenn RAG von einer berechtigten Forschungsinstitution für Zwecke der wissenschaftlichen Forschung (§ 60d UrhG) genutzt wird. Die Schranke zugunsten vorübergehender Vervielfältigungen (§ 44a UrhG) greift nicht, weil die Nutzung von Werken im Rahmen von RAG wohl eine eigenständige wirtschaftliche Bedeutung hat, während die Anwendung der Schranken zugunsten des privaten oder sonstigen eigenen Gebrauchs (§ 53 UrhG) wohl jeweils an einem (oder mehreren) Tatbestandsmerkmal(en) scheitern wird, beim privaten Gebrauch insbesondere an der Entgeltlichkeit der von den natürlichen Personen genutzten Dienste.

V. Werke im Output

Der Output von KI-Systemen kann Teile von Werken enthalten oder bestehenden Werken so ähnlich sein, dass er in deren Schutzbereich fällt.¹⁰¹ In der Regel wird das allerdings nicht der Fall sein, weil Sprachmodelle – auch bei der Nutzung von RAG – nicht Texte duplizieren, sondern neu generieren.¹⁰² Allerdings hat sich gezeigt, dass KI-Modelle nicht nur neue und statistisch wahrscheinliche Inhalte generieren, sondern aufgrund geeigneter Prompts auch

exakte oder nahezu exakte Kopien von in den Trainingsdaten enthaltenen Werken ausgeben können. Dieser Memorisierung kann zwar durch verschiedene Maßnahmen entgegengewirkt werden, sie lässt sich aber (bisher) nicht vollständig verhindern. Damit besteht die Gefahr, dass der Output von KI-Systemen in den Schutzbereich bestehender Werke fällt.

Eine Urheberrechtsverletzung liegt nach herrschender Auffassung selbst bei einem Eingreifen in den Schutzbereich nur vor, wenn ein bestehendes Werk bei der Erzeugung eines neuen Werks tatsächlich genutzt wurde.¹⁰³ Haben mehrere Personen unabhängig voneinander identische oder weitgehend identische Werke geschaffen, handelt es sich um eine Doppel- oder Parallelschöpfung;¹⁰⁴ in diesem Fall verletzt das jüngere Werk die Urheberrechte am älteren Werk nicht, weil Letzteres bei der Schöpfung des Ersteren nicht verwendet wurde.¹⁰⁵ Dieser Fall ist zwar äußerst selten – jedenfalls solange Werke (nur) von Menschen geschaffen wurden. Es ist aber denkbar, dass die Verwendung von generativer KI dazu führen wird, dass der Output von KI-Systemen einem bestehenden Werk häufiger so ähnlich sein wird, dass er in dessen Schutzbereich fällt. Als Urheberrechtsverletzungen sind aber auch diese Fälle nur zu qualifizieren, wenn das geschützte Werk bei der Erzeugung des Outputs tatsächlich verwendet wurde.¹⁰⁶

Eine Verwendung von Werken liegt bei KI-Systemen nur vor, wenn ein Werk beim Training des Modells oder bei dessen Einsatz genutzt wurde.¹⁰⁷ Bei RAG mit Zugriff auf eine interne Datenbank ist das der Fall, wenn das Werk in der internen Datenbank gespeichert und der Output unter Zugriff auf dieses Werk generiert worden ist, bei RAG mit Zugriff auf das Web, wenn der Output mithilfe eines im Web zugänglichen Werks generiert worden ist. Trifft das nicht zu und wurde das Werk auch nicht beim Training des KI-Modells verwendet, so verletzt der Output des KI-Systems selbst dann keine Urheberrechte, wenn er mit einem bestehenden Werk identisch oder diesem so ähnlich ist, dass er in dessen Schutzbereich fällt.¹⁰⁸

Enthält der Output für sich geschützte Teile von Werken oder ist er einem bestehenden Werk so ähnlich, dass er in dessen Schutzbereich fällt, und wurde das Werk (oder Teile davon) beim Training oder beim Einsatz eines KI-Modells tatsächlich verwendet, so liegt eine Urheberrechtsverletzung vor, sofern das Generieren eines solchen Outputs nicht von einer Schranke freigestellt wird. Im Vordergrund stehen dabei die Schranke zugunsten von Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch (§ 53 UrhG) und die Zitatschranke (§ 51 UrhG). Denkbar ist auch, dass der Output als Pastiche (§ 51a UrhG) freigestellt wird, wenn ein KI-System künstlerisch eingesetzt wird.¹⁰⁹ Diese spezifische Konstellation wird hier aber nicht untersucht. Ebenfalls nicht näher geprüft wird die Frage, wie Speicherungen des Outputs im Arbeitsspeicher des Anbieters oder Nutzers eines KI-Systems zu beurteilen sind. Wird das Erstellen des Outputs von einer Schranke freigestellt, werden Speicherungen im Arbeitsspeicher in der Regel als vorübergehende

Vervielfältigungen im Sinne von § 44a UrhG zu verstehen sein, zumal sie flüchtig oder begleitend sind, einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen und gegenüber dem Erstellen des Outputs keine eigenständige wirtschaftliche Bedeutung haben.

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

295

1. Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch

Wird der Output eines KI-Modells nur zu privaten oder sonstigen eigenen Zwecken genutzt, so können die damit verbundenen Vervielfältigungen durch eine der Schranken des § 53 UrhG freigestellt sein. Die Schranke zugunsten des sonstigen eigenen Gebrauchs greift allerdings nicht, weil der Output eines KI-Modells immer (zumindest in einem ersten Schritt) digital gespeichert wird und damit weder eine Vervielfältigung auf Papier oder auf einem ähnlichen Träger mittels fotomechanischer oder ähnlicher Verfahren noch eine ausschließlich analoge Nutzung vorliegt (§ 53 Abs. 2 S. 2 UrhG).¹¹⁰ Das Erstellen des Outputs kann aber als privater Gebrauch freigestellt sein (§ 53 Abs. 1 UrhG). Bei RAG mit Zugriff auf eine interne Datenbank wird die Anwendung der Schranke zwar in aller Regel (auch hier) schon daran scheitern, dass sie nur Nutzungen durch natürliche Personen erfasst, die in der Regel keine Datenbanken betreiben.¹¹¹ Bei RAG mit Zugriff auf das Web wird der Nutzer hingegen regelmäßig eine natürliche Person sein, die den Output eines Dienstes (bspw. eines Chatbots) für den privaten Gebrauch erstellt. Wie beim Zugriff auf die Werke, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt stellt sich auch hier die Frage, ob der Nutzer oder der Anbieter eines KI-Systems als Hersteller der Vervielfältigung zu qualifizieren ist und ob der Anbieter allenfalls als anderer im Sinne von § 53 Abs. 1 S. 2 UrhG verstanden werden kann. Hier wie dort sprechen wohl überwiegende Gründe dafür, den Anbieter des KI-Systems als Hersteller zu qualifizieren, womit die Anwendung der Schranke daran scheitert, dass dieser den Output nicht unentgeltlich generiert und die damit verbundenen Vervielfältigungen nicht auf Papier oder einem ähnlichen Träger erstellt.¹¹²

2. Output als Zitat

Die Nutzung eines Werks im Output eines KI-Systems kann allenfalls durch die Zitatschranke (§ 51 UrhG) freigestellt sein.¹¹³ Voraussetzung ist allerdings, dass die Wiedergabe eines Werks oder von Teilen davon einen Zitatzweck verfolgt.¹¹⁴ Das bedeutet namentlich, dass ein Werk nicht allein um seiner selbst willen wiedergegeben wird.¹¹⁵ Ein Zitatzweck liegt vielmehr nur vor, wenn der Zitierende eine innere Verbindung zwischen dem fremden Werk und den eigenen Gedanken herstellt und das Zitat als Belegstelle oder Grundlage für eigene Ausführungen dient.¹¹⁶ Eine solche innere Verbindung fehlt regelmäßig, wenn das zitierende Werk sich nicht näher mit dem fremden Werk auseinandersetzt, sondern es nur zur Illustration verwendet;¹¹⁷ dasselbe gilt, wenn das zitierte Werk in einer bloß äußerlichen, zusammenhanglosen Weise in einen eigenen Inhalt

eingefügt wird oder das Zitat ausschließlich eine informierende Berichterstattung bezweckt.¹¹⁸ Bei der Nutzung eines fremden Werks im Output eines KI-Systems wird es damit entscheidend sein, ob das Zitat in diesen Output eingebettet ist und zwischen Output und zitiertem Werk ein inhaltlicher Zusammenhang besteht.¹¹⁹ Der Umfang des Zitats darf zudem nur so weit reichen, wie es der Zitatzweck erfordert.¹²⁰ Damit darf ein Werk zwar im Output eines KI-Systems enthalten sein, der Output darf aber weder allein noch überwiegend aus dem zitierten Werk bestehen.¹²¹ Ob diese Voraussetzungen erfüllt sind, lässt sich nur im Einzelfall beurteilen.¹²²

Eine Berufung auf die Schranke der Zitierfreiheit setzt zudem voraus, dass die formellen Vorgaben von § 63 Abs. 1 und 2 UrhG eingehalten werden, dass also insbesondere die Quelle angegeben wird.¹²³ Das ist zwar technisch anspruchsvoll, neuere Sprachmodelle sind aber durchaus in der Lage, diese Vorgabe zu erfüllen.

VI. Erkenntnisse

Retrieval-Augmented Generation (RAG) eröffnet neue Möglichkeiten für die Verbesserung der Zuverlässigkeit und Aktualität des Outputs von KI-Systemen. Obwohl RAG bereits weit verbreitet ist, fehlt bislang eine vertiefte urheberrechtliche Analyse. Für die urheberrechtliche Beurteilung sind zwei Konstellationen zu unterscheiden: RAG mit Zugriff auf eine interne Datenbank und RAG mit Zugriff auf das Web.

Bei RAG mit Zugriff auf eine interne Datenbank stellen sich drei Fragen:

(1) Beim Speichern von Werken in einer internen Datenbank ergeben sich keine Besonderheiten. Die damit verbundenen Vervielfältigungen sind zulässig, wenn die Rechteinhaber zugestimmt haben oder die Vervielfälti

Thouvenin, Florent: Retrieval-Augmented Generation (RAG) – eine urheberrechtliche Einordnung (ZUM 2026, 284)

296

gungen von einer Schrankenbestimmung freigestellt werden. Dabei kommt regelmäßig nur die Schranke für den privaten Gebrauch (§ 53 Abs. 1 UrhG) infrage, die eine Nutzung von RAG in Unternehmen nicht privilegiert.

(2) Beim Zugriff auf die Werke, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt werden ebenfalls Vervielfältigungen erstellt. Diese können nicht allgemein als vorübergehende Vervielfältigungen im Sinne von § 44a UrhG qualifiziert werden, weil sie nur bisweilen einer rechtmäßigen Nutzung dienen und ihnen wohl in aller Regel eine eigenständige wirtschaftliche Bedeutung zukommt. Die Vervielfältigungen sind dennoch zulässig, sofern entweder die Zustimmung der Rechteinhaber vorliegt oder eine andere Schranke greift. Im Vordergrund steht dabei die allgemeine Schranke zugunsten des Text und Data Minings (§ 44b UrhG). Diese stellt die hier zu prüfenden Vervielfältigungen frei, sofern der Rechteinhaber keinen Nutzungsvorbehalt angebracht hat. Liegt ein solcher vor, kann die Schranke für Text und Data Mining für Zwecke der wissenschaftlichen Forschung (§ 60d UrhG) die Vervielfältigungen freistellen, wenn die entsprechenden

Voraussetzungen erfüllt sind. Die Schranke zugunsten des privaten Gebrauchs (§ 53 Abs. 1 UrhG) ist praktisch irrelevant, weil natürliche Personen in aller Regel keine Datenbanken erstellen und die Anwendung der Schranke zugunsten des sonstigen eigenen Gebrauchs (§ 53 UrhG) wird an einem (oder mehreren) Tatbestandsmerkmal(en) scheitern.

(3) Der Output von KI-Systemen kann Teile von Werken enthalten oder bestehenden Werken so ähnlich sein, dass er in deren Schutzbereich fällt. In der Regel wird dies zwar nicht der Fall sein, weil Sprachmodelle nicht Texte duplizieren, sondern neu generieren. Es hat sich aber gezeigt, dass KI-Modelle aufgrund geeigneter Prompts exakte oder nahezu exakte Kopien von in den Trainingsdaten enthaltenen Werken (oder Teilen davon) ausgeben können (sog. Memorisierung). Fällt der Output in den Schutzbereich eines bestehenden Werks und wurde dieses beim Generieren des Outputs tatsächlich genutzt, liegt grundsätzlich eine Urheberrechtsverletzung iSv § 97 UrhG vor. Die Nutzung kann allerdings durch die Zitatschranke (§ 51 UrhG) freigestellt sein, wenn die materiellen und formellen Voraussetzungen des Zitatrechts erfüllt sind.

Bei RAG mit Zugriff auf das Web werden keine Werke in internen Datenbanken gespeichert. Für die Nutzer und die Anbieter eines solchen Dienstes stellen sich daher nur zwei Fragen:

(1) Wie bei RAG mit Zugriff auf eine interne Datenbank werden auch bei RAG mit Zugriff auf das Web bei der Identifikation der relevanten Inhalte und bei deren Integration in einen vom Nutzer verfassten Prompt Vervielfältigungen von Werken erstellt. Diese lassen sich nicht allgemein als vorübergehende Vervielfältigungen im Sinne von § 44a UrhG qualifizieren, werden aber von der allgemeinen Schranke für Text und Data Mining (§ 44b UrhG) freigestellt, sofern der Rechteinhaber keinen Nutzungsvorbehalt angebracht hat. Nutzen Privatpersonen Angebote Dritter, etwa einen Chatbot wie ChatGPT, kommt es darauf an, ob der Nutzer oder der Anbieter als Hersteller der Vervielfältigungen verstanden wird. Wird der private Nutzer als Hersteller qualifiziert, können die im Rahmen von RAG erstellten Vervielfältigungen als Privatgebrauch nach § 53 Abs. 1 S. 1 UrhG freigestellt werden. Allerdings sprechen wohl überwiegende Gründe dafür, den Anbieter als Hersteller zu verstehen. Damit scheitert die Anwendung der Schranke von § 53 Abs. 1 S. 2 UrhG daran, dass die Vervielfältigungen vom Anbieter weder unentgeltlich noch auf Papier oder einem ähnlichen Träger erstellt werden.

(2) Für den Output gilt grundsätzlich dasselbe wie bei RAG mit Zugriff auf eine interne Datenbank. Fällt der Output in den Schutzbereich eines bestehenden Werks und wurde dieses Werk beim Generieren des Outputs tatsächlich genutzt, kann die Nutzung unter Umständen als Zitat (§ 51 UrhG) freigestellt sein. Die Anwendung der Schranke zugunsten des privaten Gebrauchs (§ 53 Abs. 1 UrhG) wird aus denselben Gründen scheitern, die deren Anwendung auf die Vervielfältigungen beim Zugriff auf Werke, bei der Identifikation der relevanten Inhalte und bei deren Integration in einen Prompt ausschließen.

* Der Verfasser Prof. Dr. Florent Thouvenin ist Rechtsanwalt und Ordinarius für Informations- und Kommunikationsrecht sowie Vorsitzender des Leitungsausschusses des Center for Information Technology, Society, and Law (ITS�) an der Universität Zürich. Beim Verfassen der technischen Ausführungen zu RAG (Kapitel II.) wurde ChatGPT genutzt. Dieser Beitrag beruht auf dem Beitrag des Verfassers, Retrieval-Augmented Generation (RAG) – ein urheberrechtliches 101, sic! 2025, 381, der die Rechtslage im schweizerischen Urheberrecht untersucht. Der Autor dankt Markus Danhel, Head AI Software – Austria & Switzerland, IBM Technology, für die Klärung technischer Fragen, Prof. Dr. Alexander Peukert für wertvolle Hinweise und Lennart Deutschmann, Mlaw, M.iur., für die wertvolle Unterstützung bei der Analyse des deutschen Rechts, bei der Recherche und beim Erstellen des Fußnotenapparates. Alle Internetquellen wurden am 9.2.2026 letztmals abgerufen.

1 S. dazu bspw. Dornis/Stober, Urheberrecht und Training generativer KI-Modelle: Technologische und juristische Grundlagen, 2024; Beurskens RDi 2025, 1; Leistner GRUR 2024, 1665; Ventroni MMR 2025, 163; LG Hamburg ZUM 2025, 64 (die Berufung wurde durch das OLG Hamburg mit Urteil vom 10.12.2025 zurückgewiesen, ZUM 2026, 210; die Revision beim BGH ist anhängig: BGH – I ZR 281/25); daneben ist zzt. eine Klage der GEMA beim LG München I gegen die Musik-KI-Anbieterin Suno Inc. anhängig; über die Klage der GEMA gegen »Open AI« wurde bereits entschieden (LG München I ZUM 2026, 38, s. auch die Anm. von Peukert ZUM 2026, 17).

2 S. aber Wietzke RDi 2024, 538; zum Schweizer Recht s. Thouvenin sic! 2025, 381.

3 Ji et al., Survey of Hallucination in Natural Language Generation, ACM Computing Surveys Nr. 12 2023, 1.

4 Harshvardhan/Gourisaria/Pandey/Rautaray, A comprehensive survey and analysis of generative models in machine learning, Computer Science Review Nr. 100285 2020, 1 (2).

5 Zhao et al., Survey of Large Language Models, arXiv 2025, 1 (26); Vaswani et al., Attention is All You Need, Advances in Neural Information Processing Systems 2017, 1.

6 Ji et al., Survey of Hallucination in Natural Language Generation, ACM Computing Surveys Nr. 12 2023, 1 (5 ff.).

7 Ji et al., Survey of Hallucination in Natural Language Generation, ACM Computing Surveys Nr. 12 2023, 1 (10 ff.).

8 Lin et al., FLAME: factuality-aware alignment for large language models, Proceedings of the 38th International Conference on Neural Information Processing Systems 2024, 1.

9 Manakul/Liusie/Gales, SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models, Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing 2023, 9004.

10 Dhuliawala et al., Chain-of-Verification Reduces Hallucination in Large Language Models, Findings of the Association for Computational Linguistics: ACL 2024, 3563; Xie et al., FIRE: Fact-checking with Iterative Retrieval and Verification, Findings of the Association for Computational Linguistics: NAACL 2025, 2901.

- 11 Ruparel/Patel, RAGCacheSim: A discrete-event simulator for evaluating caching strategies in Retrieval-Augmented Generation Systems, Software Impacts Nr. 26 2025, 1; Tan et al., HtmlRAG: HTML is Better Than Plain Text for Modeling Retrieved Knowledge in RAG Systems, Proceedings of the ACM on Web Conference 2025, 1733; Lewis et al., Retrieval-augmented generation for knowledge-intensive NLP tasks, Proceedings of the 34th International Conference on Neural Information Processing Systems 2020, 1.
- 12 Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1.
- 13 S. zB Buffa/Ferrara/Picascia/Riva/Castano, Enhancing legal document building with Retrieval-Augmented Generation, Computer Law & Security Review 2025, 1.
- 14 Koetter et al., Conversational Agents for Insurance Companies: From Theory to Practice, Agents and Artificial Intelligence, in: van den Herik/Rocha/Steels, Agents and Artificial Intelligence, Cham 2019, 338.
- 15 <https://www.perplexity.ai/>.
- 16 Brown et al., Language Models are Few-Shot Learners, arXiv 2020, 1 (8).
- 17 Strubell/Ganesh/McCallum, Energy and Policy Considerations for Deep Learning in NLP, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019, 3645; Cottier et al., The rising costs of training frontier AI models, arXiv 2025, 1.
- 18 Tan et al., HtmlRAG: HTML is Better Than Plain Text for Modeling Retrieved Knowledge in RAG Systems, Proceedings of the ACM on Web Conference 2025, 1733; Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (5).
- 19 Zum Ganzen s. Lewis et al., Retrieval-augmented generation for knowledge-intensive NLP tasks, Proceedings of the 34th International Conference on Neural Information Processing Systems 2020, 1; Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (4).
- 20 Lawrence, Understanding RAG: Retrieval-Augmented Generation Explained, <https://www.6clicks.com/resources/blog/understanding-rag-retrieval-augmented-generation-explained>.
- 21 Zum Ganzen s. Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (4).
- 22 Vaswani et al., Attention Is All You Need, Advances in Neural Information Processing Systems 2017, 1 (5); 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.
- 23 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.

24 Zum Ganzen s. Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (5).

25 Dabei werden die Textdaten meist als reiner Text ohne Formatierung (plain text) verarbeitet; s. dazu Tan et al., HtmlRAG: HTML is Better Than Plain Text for Modeling Retrieved Knowledge in RAG Systems, Proceedings of the ACM on Web Conference 2025, 1733.

26 Zum Ganzen s. Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (5 f.).

27 Lewis et al., Retrieval-augmented generation for knowledge-intensive NLP tasks, Proceedings of the 34th International Conference on Neural Information Processing Systems 2020, 1.

28 Eine API-Schnittstelle ist ein technischer Zugangspunkt, über den ein Softwaresystem automatisiert und regelgebunden Informationen von einem externen Dienst (zB Google-Suche) abrufen, ohne dessen Website unmittelbar aufzurufen.

29 Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (4).

30 Zum Ganzen s. Buffa/Ferrara/Picascia/Riva/Castano, Enhancing legal document building with Retrieval-Augmented Generation, Computer Law & Security Review 2025, 1 (2 f.); Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (5 f.).

31 Lewis et al., Retrieval-augmented generation for knowledge-intensive NLP tasks, Proceedings of the 34th International Conference on Neural Information Processing Systems 2020, 1 (2).

32 Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (5).

33 Peters et al., Deep Contextualized Word Representations, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2018, 2227; Petroni et al., 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.

34 Beigi et al., InternalInspector I2: Robust Confidence Estimation in LLMs through Internal States, Findings of the Association for Computational Linguistics: EMNLP 2024, 12847.

35 Xie et al., FIRE: Fact-checking with Iterative Retrieval and Verification, Findings of the Association for Computational Linguistics: NAACL 2025, 2901.

36 So etwa das Google Fact Check Tool: <https://developers.google.com/fact-check/tools/api?hl=de>.

37 Rafailov et al., Direct Preference Optimization: Your Language Model Is Secretly a Reward Model, Proceedings of the 37th International Conference on Neural Information Processing Systems 2023, 1.

38 Buffa/Ferrara/Picascia/Riva/Castano, Enhancing legal document building with Retrieval-Augmented Generation, Computer Law & Security Review 2025, 1 (4).

39 Oche/Folashade/Ghosal/Biswas, A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions, arXiv 2025, 1 (6).

40 BGH ZUM 2017, 236 Rn. 37 f.; Wandtke/Bullinger/Heerma, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 16 Rn. 16; Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG § 16 Rn. 17 f. mwN; Fromm/Nordemann/Dustmann, Urheberrecht, 13. Aufl. 2024, UrhG § 16 Rn. 12; Dreier/Schulze/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 16 Rn. 37; s. auch (zur Werkvervielfältigung beim Training von KI-Systemen) Pesch/Böhme GRUR 2023, 997 (1001 ff.); Hofmann WRP 2024, 11 Rn. 11 f., jeweils mwN.

41 S. dazu Peukert, Urheberrecht, 19. Aufl. 2023, § 36 Rn. 11 ff.

42 S. Lettl, Urheberrecht, 5. Aufl. 2024, § 6 Rn. 1 ff.; Peukert, Urheberrecht, 19. Aufl. 2023, § 24 Rn. 1 ff.

43 Raue ZUM 2020, 172 (173).

44 Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17.4.2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG.

45 S. ferner Erwgr. 105 f. zur KI-VO.

46 LG Hamburg ZUM 2025, 64 (67 ff.); BeckOK UrhR/Bomhard, 47. Ed. 1.9.2025, UrhG § 44b Rn. 11c; Baumann NJW 2023, 3673 (3674); Becker CR 2024, 353 (357 f.); Beurskens RDi 2025, 1; Bullinger/Vonthien MittdtPatA 2023, 446; de la Durantaye ZUM 2023, 645 (651); Käde KIR 2024, 162; Maamar ZUM 2023, 481 (483 ff.); Siglmüller/Gassner RDi 2023, 124 (127); Schippel ITRB 2023, 216 (217).

47 Ebenso Wietzke RDi 2024, 538 Rn. 31.

48 Teilweise wird zwar die Auffassung vertreten, dass die automatisierte Analyse von Informationen gar keine urheberrechtlich relevante Handlung sei (BT-Drs. 18/12329, 40; BeckOK UrhR/Bomhard, 47. Ed. 1.9.2025, UrhG § 44b Rn. 12; Maamar ZUM 2023, 481 (482); Kleinkopf/Jacke/Gärtner MMR 2021, 196 mwN). Diese Auffassung blendet allerdings aus, dass auch bei einer automatisierten Analyse Vervielfältigungen erstellt werden.

49 BeckOK UrhR/Bomhard, 47. Ed. 1.9.2025, UrhG § 44b Rn. 1; Spindler/Schuster/Kaesling/Pesch, Recht der elektronischen Medien, 5. Aufl. 2026, UrhG § 44b Rn. 18 f.; s. ferner Baumann NJW 2023, 3673 Rn. 13 f.

50 BT-Drs. 19/27426, 88; LG Hamburg ZUM 2025, 64 (69); Baumann NJW 2023, 3673 Rn. 18.

51 Die Schranke in § 60d Abs. 1 UrhG erstreckt sich nach § 60d Abs. 2 UrhG auch auf die Vervielfältigung von Datenbankwerken (§ 4 Abs. 2 UrhG) und Datenbanken (§ 87b Abs. 1 UrhG), in denen die auszuwertenden Quellen enthalten sind, s. dazu Schricker/Loewenheim/Stieper, Urheberrecht, 6. Aufl. 2020, UrhG § 60d Rn. 15.

52 Weitergehend Schöttle/Völker K&R 2025, 22; Käde CR 2024, 598.

53 Peukert, Urheberrecht, 19. Aufl. 2023, § 27 Rn. 29 ff.

54 BGH ZUM 2017, 236 Rn. 49 mwN.

55 Besonders gelagerte Einzelfälle, in denen ein rein privates RAG-System etwa bei der Ausübung eines Hobbys betrieben wird, sind zwar denkbar, dürften aber kaum je vorkommen.

56 Schricker/Loewenheim/Loewenheim/Stieper, Urheberrecht, UrhG § 53 Rn. 47.

57 BeckOK UrhR/Grübler, 47. Ed. 1.2.2025, UrhG § 53 Rn. 23; Wandtke/Bullinger/Lüft, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 53 Rn. 32, 22.

58 Weitergehend zum Nutzungsvorbehalt Heine GRUR-Prax 2024, 87; Wandtke/Bullinger/Bullinger, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 44b Rn. 10.

59 Wandtke/Bullinger/Bullinger, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 44b Rn. 10; Peukert, Urheberrecht, 19. Aufl. 2023, § 26 Rn. 10.

60 Hierzu Peukert, Urheberrecht, 19. Aufl. 2023, § 24 Rn. 1 ff.; Lettl, Urheberrecht, 5. Aufl. 2024, § 6 Rn. 1 ff.

61 S. dazu vorne, III. 1.

62 S. dazu vorne, III. 1.

63 Hierzu weitergehend Heesen/Jüngels RuZ 2021, 45; Kleinkopf/Jacke/Gärtner MMR 2021, 196; Raue ZUM 2021, 793; Steinhauer RuZ 2021, 5.

64 S. dazu vorne, III. 1.

65 BT-Drs. 19/27426, 88; LG Hamburg ZUM 2025, 64 (69).

66 Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 44b Rn. 8, 10; Dreyer/Kotthoff/Hentsch/Kotthoff, Urheberrecht, 5. Aufl. 2025, § 44b Rn. 22 ff.

67 Fromm/Nordemann/Czychowski, Urheberrecht, 13. Aufl. 2024, UrhG § 95a Rn. 37 ff.

68 Dreier/Schulze/Specht-Riemenschneider/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 95b Rn. 14 ff.; BeckOK UrhR/Lindhorst, 47. Ed. 1.9.2025, UrhG § 95a Rn. 13.

69 Schricker/Loewenheim/Götting, Urheberrecht, 6. Aufl. 2020, UrhG § 95b Rn. 1 ff.; Wandtke/Bullinger/Wandtke/Ohst, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 95b Rn. 10.

- 70 Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 44b Rn. 10; Dreier/Schulze/Specht-Riemenschneider/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 95b Rn. 9.
- 71 EuGH ZUM 2009, 945 Rn. 64; KG ZUM-RD 2004, 401 (406); BeckOK UrhR/Schulz, 46. Ed. 1.2.2025, UrhG § 44a Rn. 4.
- 72 OLG Dresden NJOZ 2008, 160; Spindler/Schuster/Kaesling/Wiebe, Recht der elektronischen Medien, 5. Aufl. 2026, § 44a Rn. 3; Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG, § 44a Rn. 8.
- 73 So wohl auch Wietzke RDi 2021, 538 Rn. 33 f., 46 f.
- 74 Spindler GRUR 2002, 105 (111); Wandtke/Bullinger/v. Welser, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 44a Rn. 7.
- 75 EuGH ZUM 2014, 681 Rn. 28; Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG § 44a Rn. 9.
- 76 Statt vieler Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG § 44a Rn. 12; Fromm/Nordemann/Dustmann, Urheberrecht, 13. Aufl. 2024, UrhG § 44a Rn. 17; s. auch Erwägungsgrund 33 der RL 2001/29/EG (InfoSoc-RL).
- 77 BeckOK UrhR/Schulz, 46. Ed. 1.2.2025, UrhG § 44a Rn. 13; Fromm/Nordemann/Dustmann, Urheberrecht, 13. Aufl. 2024, § 44a UrhG Rn. 17 mwN; weitergehend und kritisch zum »digitalen Werkgenuss« s. auch Kaiser FS Hilty, 2024.
- 78 BGH GRUR 1994, 363 (364 f.) – Holzhandelsprogramm; Dreier/Schulze/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 16 Rn. 39; Wandtke/Bullinger/Heerma, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 15 Rn. 10.
- 79 Wandtke/Bullinger/Heerma, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 15 Rn. 10; Hoeren/Sieber/Holznagel/Ernst, MMR-HdB, 62. EL Juni 2024, Teil 7.1 Rn. 57; Wandtke/Bullinger/v. Welser, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 44a Rn. 21; Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 44a Rn. 4; aA Peukert, Urheberrecht, 19. Aufl. 2023, § 25 Rn. 6.
- 80 Ebenso Wietzke RDi 2024, 538 Rn. 56.
- 81 Zur »eigenständigen wirtschaftlichen Bedeutung« s. Wandtke/Bullinger/Leenen, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, InfoSoc-RL Art. 5 Rn. 24 ff.; EuGH ZUM 2011, 803 Rn. 175.
- 82 EuGH ZUM 2011, 803 Rn. 177.
- 83 Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 44a Rn. 10; Pesch/Böhme GRUR 2023, 997 (1006 f.); Schack NJW 2024, 113 Rn. 6 mwN; Ventroni/Kaiser MMR 2025, 163 (164); s. weitergehend hierzu Hofmann ZUM 2024, 166 (169); Spindler GRUR 2016, 1112 (1115).
- 84 Art. 13 TRIPS; Art. 9 Abs. 2 RBÜ; Art. 10 Abs. 1 WCT (WIPO Copyright Treaty) und Art. 16 Abs. 2 WPPT (WIPO Performances and Phonograms Treaty); s. ferner Art. 5 Abs. 5 der Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22.5.2001 zur

Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft (InfoSoc-RL).

85 Ricketson/Ginsburg, International Copyrights and Neighbouring Rights, 3. Aufl. 2022, Rn. 13.16; WTO Panel Report vom 15.6.2000, WT/DS 160/R – United States Section 110(5) of the US Copyright Act, Rn. 6.166 ff.

86 S. auch Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG § 44a Rn. 13; Senftleben GRUR Int. 2004, 200.

87 Zum Schweizer Recht s. Thouvenin sic! 2025, 381 (386 f.).

88 Zum Schweizer Recht s. Thouvenin sic! 2025, 381 (386 f.).

89 So für das deutsche (und europäische) Recht Hofmann ZUM 2024, 166 (169); aA Wietzke RDi 2024, 538 Rn. 34.

90 S. ferner EuGH ZUM 2012, 398 Rn. 50 ff.; BeckOK IT-Recht/Paul, 20. Ed. 1.10.2025, UrhG § 44a Rn. 17.

91 S. auch Metzger JZ 2025, 328 (330).

92 Anhand des tatsächlichen Modelloutputs differenzierend Wietzke RDi 2024, 538 Rn. 53 ff.

93 S. dazu III. 2. sowie IV. 1. c) bb).

94 S. nur BeckOK UrhR/Grübler, 47. Ed. 1.2.2025, UrhG § 53 Rn. 26; Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 53 Rn. 30.

95 S. auch Wandtke/Bullinger/Lüft, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 53 Rn. 31.

96 Wandtke/Bullinger/Lüft, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, UrhG § 53 Rn. 18, 23 ff.

97 S. dazu etwa Fromm/Nordemann/Renner/Wirtz, Urheberrecht, 13. Aufl. 2024, § 53 UrhG Rn. 70 f.; Maamar ZUM 2023, 481 (488).

98 BGH ZUM 2025, 140 Rn. 68; ZUM-RD 2009, 369 Rn. 16 – Internet-Videorecorder I; BGH ZUM 2020, 626 Rn. 2 – Internet-Radiorecorder I; BGH ZUM 2024, 808 Rn. 15 – Internet-Radiorecorder II; BeckOK UrhR/Grübler, 47. Ed. 1.2.2025, UrhG § 53 Rn. 16; Dreier/Schulze/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 16 Rn. 46.

99 BGH ZUM 2020, 626 Rn. 22 – Internet-Radiorecorder I; BGH ZUM-RD 2009, 369 Rn. 17 und 52 – Internet-Videorecorder I; LG München I ZUM 2026, 38 Rn. 289; Fromm/Nordemann/Renner/Wirtz, Urheberrecht, 13. Aufl. 2024, § 53 UrhG Rn. 71.

100 Schricker/Loewenheim/Loewenheim/Stieper, Urheberrecht, 6. Aufl. 2020, UrhG § 53 Rn. 49 f.; Wandtke/Bullinger/Leenen, Praxiskommentar Urheberrecht: UrhR, 6. Aufl. 2022, InfoSoc-RL Art. 5 Rn. 32 ff.

101 Baumann/Nordemann/Pukas GRUR 2025, 955 (956); Kraetzig NJW 2024, 697 (700); s. ferner Scherer ZUM 2025, 437.

102 Zum Schweizer Recht s. Thouvenin sic! 2025, 381 (387).

103 BGH GRUR 1971, 266 (268) – Magdalenenarie; OLG Zweibrücken ZUM-RD 2016, 298 (301) – Piano-Lehrbuch; Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG § 23 Rn. 34, 36; Dreier/Schulze/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 23 Rn. 35.

104 Hierzu Schricker/Loewenheim/Loewenheim, Urheberrecht, 6. Aufl. 2020, UrhG § 23 Rn. 34 ff.; Fromm/Nordemann/Nordemann, Urheberrecht, 13. Aufl. 2024, UrhG § 23 Rn. 64 f.; Dreier/Schulze/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 23 Rn. 35 ff.

105 BGH GRUR 1988, 810 (811) – Fantasy; BGH GRUR 1971, 266 (268 f.) – Magdalenenarie; s. hierzu auch Haas, Die Verwendung von Bearbeitungen urheberrechtlich geschützter Werke, 2018, S. 105 ff.

106 Zum KI-Training etwa Ebers/Heinze/Krügel/Steinrötter, Künstliche Intelligenz und Robotik, 2020, § 9 Rn. 4 ff.; Beurskens RDi 2025, 1; Leistner GRUR 2024, 1665; Skupin GRUR-Prax 2024, 752.

107 Zur Beweislast Grimm/Münster GRUR-Prax 2025, 443.

108 S. in diesem Sinn Dreier/Schulze/Raue, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 23 Rn. 35 ff.

109 Aus aktuellem Anlass zum Begriff des Pastiches s. Lennartz/Kraetzig ZUM 2025, 668.

110 EuGH ZUM 2013, 786 Rn. 67 f.; BeckOK IT-Recht/Paul, 20. Ed. 1.10.2025, UrhG § 53 Rn. 8; Spindler/Schuster/Kaesling/Wiebe, Recht der elektronischen Medien, 5. Aufl. 2026, UrhG § 53 Rn. 16.

111 S. dazu vorne III. 2.; IV. 3.

112 S. dazu vorne IV. 3.

113 S. auch Wietzke RDi 2024, 538 Rn. 57 ff.

114 BGH ZUM 2020, 777 Rn. 82.

115 BGH ZUM-RD 2008, 337 Rn. 42 mwN.

116 BGH ZUM-RD 2017, 581 Rn. 55.

117 BeckOK UrhR/Schulz, 47. Ed. 1.9.2025, UrhG § 51 Rn. 13; OLG Köln ZUM-RD 2014, 98 (101).

118 BGH ZUM-RD 2017, 581 Rn. 55 mwN.

119 BGH ZUM-RD 2008, 337 Rn. 10, 31, 36 ff.

120 Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 51 Rn. 3.

121 S. auch BGH ZUM-RD 2016, 214 Rn. 31 mwN.

122 So wohl auch Wietzke RDi 2024, 538 Rn. 57 ff., 62.

123 Dreier/Schulze/Dreier, Urheberrechtsgesetz, 8. Aufl. 2025, UrhG § 51 Rn. 2; Schricker/Loewenheim/Spindler, Urheberrecht, 6. Aufl. 2020, UrhG § 51 Rn. 58.

Urheberrechtlich geschützter Inhalt. Verlag C.H.Beck GmbH & Co. KG ist Inhaber von Nutzungsrechten. Ohne gesonderte Berechtigung ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für

oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen. Verlag C.H.Beck GmbH & Co. KG behält sich alle Rechte vor, insbesondere die Nutzung zum Text-und-Data-Mining (TDM) nach § 44b Abs. 3 UrhG (Art. 4 DSM-RL).