

## LEGAL UPDATE IP, IT, COMMERCIAL

Köln, Februar 2025

# KI: Urheberrechtsverletzung durch Verwendung von Daten für das Training von künstlicher Intelligenz in den USA.

Dr. Ricardo Vocke-Kerkhof

---

Am 11. Februar 2025 hat ein US-Bundesgericht ein entscheidendes Urteil zugunsten eines Urheberrechtinhabers getroffen, welcher sich gegen die Verwendung seiner urheberrechtlich geschützten Werke beim Training einer künstlichen Intelligenz zur Wehr gesetzt hat. Auch wenn Kern der Entscheidung die Reichweite des amerikanischen Rechtsbegriffs der freien Nutzung (fair use) war, lassen sich daraus auch Parallelen für Deutschland erkennen.

### Hintergrund der Entscheidung

Klägerin in dem Verfahren war der Medienkonzern Thomson Reuters, welcher unter anderem den streitgegenständlichen Online-Recherche-dienst „Westlaw“ betreibt. Dieser richtet sich an Juristen und enthält eine Vielzahl von Urteilen und Gesetzen. Zu den besonderen Merkmalen von „Westlaw“ gehört die Verwendung von sogenannten „headnotes“, bei welchen es sich um eine (von Menschen erstellte) kurze rechtliche Zusammenfassung eines Urteils handelt. Diese „headnotes“ werden zudem noch mit sogenannten „Key Numbers“ ergänzt, welche einzelne rechtliche Themenkomplexe bündeln und es den Nutzern so ermöglichen, anhand dieser

Nummern Entscheidungen und Gesetze zu dem gesuchten Themenkomplex zu finden.

Die Beklagte Ross Intelligence plante ihrerseits eine juristische Datenbank zu erstellen, welche allerdings durch die Verwendung eines KI-Modells die Nutzer in die Lage versetzen sollte, relevante Urteile mittels Anfragen in einfacher, nicht-juristischer Sprache zu finden. Um das KI-Modell in die Lage zu versetzen, diese Aufgabe zu erfüllen, musste das KI-Modell im Zuge des beaufsichtigten Lernens (supervised learning) mithilfe von markierten Daten trainiert werden. Hierfür werden dem KI-Modell Trainingsdaten mitsamt den erwarteten Ergebnissen mitgeteilt. Anhand dieser Daten kann das KI-Modell dann die Fähigkeit erwerben, eigenständige Ergebnisse zu ermitteln. Im Fall der Beklagten war es daher relevant, Daten zu generieren, anhand welcher dem KI-Modell der Zusammenhang zwischen einer Suchanfrage und dem dazu passenden Urteil aufgezeigt werden kann.

Für dieses Training waren die „headnotes“ der Klägerin besonders geeignet, so dass die Beklagte zunächst auf diese zuging und um eine Lizenzierung der Daten bat. Die Klägerin lehnte dies jedoch ab, sodass sich die Beklagte an einen anderen Anbieter wandte, um geeignete

Trainingsdaten zu generieren. Ohne Kenntnis der Beklagten und der Klägerin generierte dieser Anbieter die Trainingsdaten allerdings mithilfe der Datenbank von „Westlaw“. Die so generierten Trainingsdaten hatten weitreichende Überschneidungen mit den jeweiligen „headnotes“ von „Westlaw“.

Nachdem die Klägerin hiervon Kenntnis erlangte, reichte sie gegen die Beklagte Klage vor einem Bundesgericht in Delaware ein. Die Klägerin berief sich dabei auf eine Verletzung ihres Urheberrechts an den „headnotes“, wohingegen die Beklagte argumentierte, dass ein Fall der freien Verwendung (fair use) vorlag.

### Die Entscheidung des US-Bundesgerichts

Das Gericht gab der Klage statt. Es ging davon aus, dass die rechtliche Zusammenfassung in Form der „headnotes“ für sich genommen eine urheberrechtlich schutzfähige Leistung darstellt, da sie insoweit nicht bloß die Entscheidung des Gerichts wiedergibt, sondern eine eigenständige Leistung darstellt.

Darüber hinaus sah das Gericht in der Verwendung der Daten nach summarischer Prüfung keine freie Nutzung. Nach Ansicht des Gerichts folgte dies insbesondere daraus, dass die Daten für einen identischen Zweck (die Erstellung einer juristischen Datenbank) genutzt wurden. Hierbei verwies das Gericht explizit auf den Umstand, dass es sich im vorliegenden Fall um nicht generative KI handelt, so dass die verwendeten Daten nicht etwa dazu verwendet würden, neue Inhalte zu schaffen und so keine transformative Nutzung der Daten gegeben sei.

### Vergleich mit Deutschland

In den USA ist dies der erste Fall, in welchem ein Gericht in der Verwendung von Trainingsdaten eine Urheberrechtsverletzung angenommen hat. Ob dem Urteil eine Signalwirkung zukommen wird, bleibt indes abzuwarten.

Es lassen sich allerdings schon jetzt relevante Erkenntnisse aus dem Urteil gewinnen.

So ist festzuhalten, dass es für Anbieter von KI-Modellen und KI-Systemen von erheblicher Relevanz ist, sich bei der Generierung von Trainingsdaten bei vertrauenswürdigen Anbietern zu bedienen, welche bei der Erstellung der Daten das geistige Eigentum anderer beachten. Auch wenn die Beklagte in dem US-Verfahren keine Kenntnis von der Urheberrechtsverletzung hatte, hat das Gericht sie dennoch wegen der Urheberrechtsverletzung verurteilt.

Ebensolche Risiken drohen auch hier in Deutschland. Zwar existiert mit § 44b UrhG eine gesetzliche Regelung, welche die Vervielfältigung von rechtmäßig zugänglichen Werken das den Umgang mit Text- und Data Mining regelt. Zum jetzigen Zeitpunkt besteht aber noch erhebliche Unsicherheit, inwiefern sich diese Regelung auf die verschiedenen Ebenen des Trainings von KI-Modellen auswirkt. So ist im Besonderen ungeklärt, ob § 44b UrhG ausschließlich für die Erstellung der Trainingsdaten gelten soll oder auch für ein Training des KI-Modells anhand dieser Daten (wobei Art. 53 Abs. 1 lit. c) der KI-Verordnung darauf hindeutet, dass nach dem Willen des europäischen Gesetzgebers das Training von KI von § 44b UrhG erfasst sein soll). Eine erste Entscheidung des LG Hamburg bezog sich bisher nur auf das Erstellen von Trainingsdaten und nicht auf ein darauf aufbauendes KI-Training. Zudem enthält § 44b Abs. 3 UrhG auch die Möglichkeit für den Urheberrechtsinhaber, die zugänglichen Werke mithilfe eines maschinenlesbaren Hinweises von dem

Anwendungsbereich der Vorschrift auszuschließen, so dass nicht alle rechtmäßig zugänglichen Werke auch rechtmäßig für das Text- und Datamining genutzt werden dürfen.

## Folgen für Unternehmen

Das Urteil zeigt, dass Entwickler von KI-Modellen bedacht vorgehen müssen, um ihre wirtschaftlichen und rechtlichen Interessen zu wahren.

Für die Beklagte hat das vorliegende Verfahren bereits den wirtschaftlichen Ruin bedeutet.

Wenn erhebliche Ressourcen in die Entwicklung von KI-Modellen gesteckt werden, kann eine verhältnismäßig geringe Urheberrechtsverletzung bereits einen erheblichen Schaden verursachen. Entwickler sollten daher darauf achten, dass sie entweder direkt mit den Rechteinhabern robuste Lizenzvereinbarungen treffen oder aber sicherstellen, dass die Anbieter von vorgefertigten Trainingsdatensätzen ausreichende Sicherheiten dafür liefern, dass die angebotenen Trainingsdaten nicht rechtsverletzend sind.

### Hinweis

Dieser Überblick dient ausschließlich der allgemeinen Information und kann konkreten Rechtsrat im einzelnen Fall nicht ersetzen. Sprechen Sie bei Fragen bitte Ihren gewohnten Ansprechpartner bei GÖRG bzw. den Autor Dr. Ricardo Vocke-Kerkhof unter +49 221 33660 258 oder [rvocke-kerkhof@goerg.de](mailto:rvocke-kerkhof@goerg.de) an. Informationen zum Autor finden Sie auf unserer Homepage [www.goerg.de](http://www.goerg.de).

Wir verwenden das generische Maskulinum und sehen von einer Nennung aller Geschlechtsidentitäten ab, damit dieser Text besser lesbar ist und meinen damit ausdrücklich jeden in jeder Geschlechtsidentität.

## Unsere Standorte

GÖRG Partnerschaft von Rechtsanwälten mbB

### BERLIN

Kantstr. 164, 10623 Berlin  
Tel. +49 30 884503-0  
Fax +49 30 882715-0

### HAMBURG

Alter Wall 20 - 22, 20457 Hamburg  
Tel. +49 40 500360-0  
Fax +49 40 500360-99

### FRANKFURT AM MAIN

Ulmenstr. 30, 60325 Frankfurt am Main  
Tel. +49 69 170000-17  
Fax +49 69 170000-27

### KÖLN

Kennedyplatz 2, 50679 Köln  
Tel. +49 221 33660-0  
Fax +49 221 33660-80

### MÜNCHEN

Prinzregentenstr. 22, 80538 München  
Tel. +49 89 3090667-0  
Fax +49 89 3090667-90