

DIE LICENSE.TXT-FALLE

DER HEIMLICHE KILLER DES KI-FORTSCHRITTS ?!

Kristian Borkert, Anastasia Nomerowskaja
JURIBO Legal & Consulting

Herbstakademie 2025



DIE LICENSE.TXT-FALLE

DER HEIMLICHE KILLER DES KI-FORTSCHRITTS ?!

LICENSE.TXT

KI im Fokus

Meta-CEO Zuckerberg: "KI wird besseren Code schreiben als Entwickler"

Laut CEO Nadella schreibt KI rund 30 Prozent des Microsoft-Codes. Zuckerberg schätzt den Anteil bei Meta höher und glaubt an das Potenzial von KI-Code.

<https://www.heise.de/news/Meta-CEO-Zuckerberg-KI-wird-besseren-Code-schreiben-als-Entwickler-10371134.html>

X sammelt ungefragt Daten fürs KI-Training

Wer die eigenen Beiträge und Kommentare nicht zum Training von Grok beitrugen möchte, muss die Einstellungen im Browser ändern.

<https://www.heise.de/news/X-sammelt-ungefragt-Daten-fuers-KI-Training-9815225.html>

Chatbot-Streit

Elon Musk klagt gegen Apple und OpenAI

26.08.2025 | 04:23

← | ☆

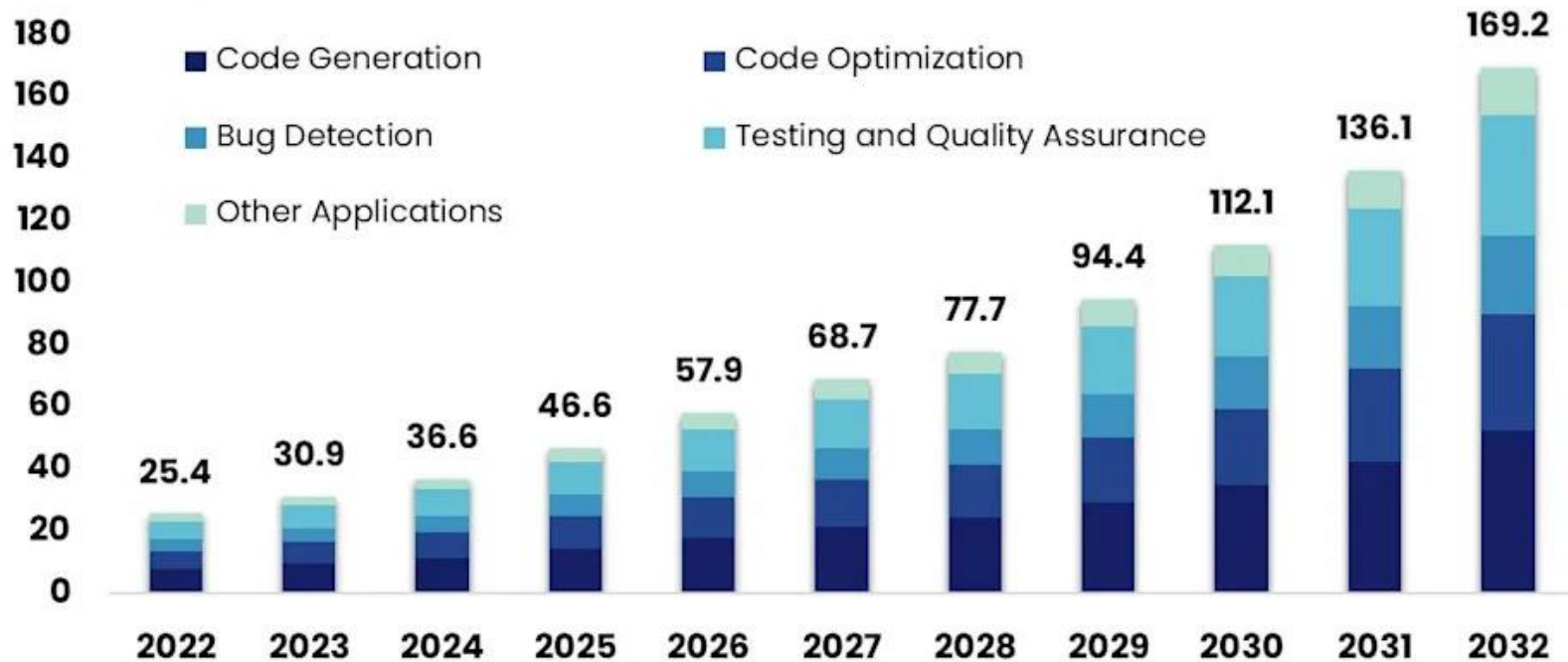
Tech-Milliardär Elon Musk zieht gegen Apple und OpenAI vor Gericht. Der iPhone-Hersteller und das Unternehmen hinter ChatGPT hätten sich verbündet, um ein KI-Monopol zu schaffen.

<https://www.zdfheute.de/wirtschaft/unternehmen/usa-internet-ki-xai-apple-openai-musk-klage-100.html>

Generative KI in der Softwareentwicklung bis 2032

Generative AI in Software Development Market

Size, by application, 2022-2032 (USD Billion)



The Market will Grow
At the CAGR of:

21.4%

The forecasted market
size for 2032 in USD:

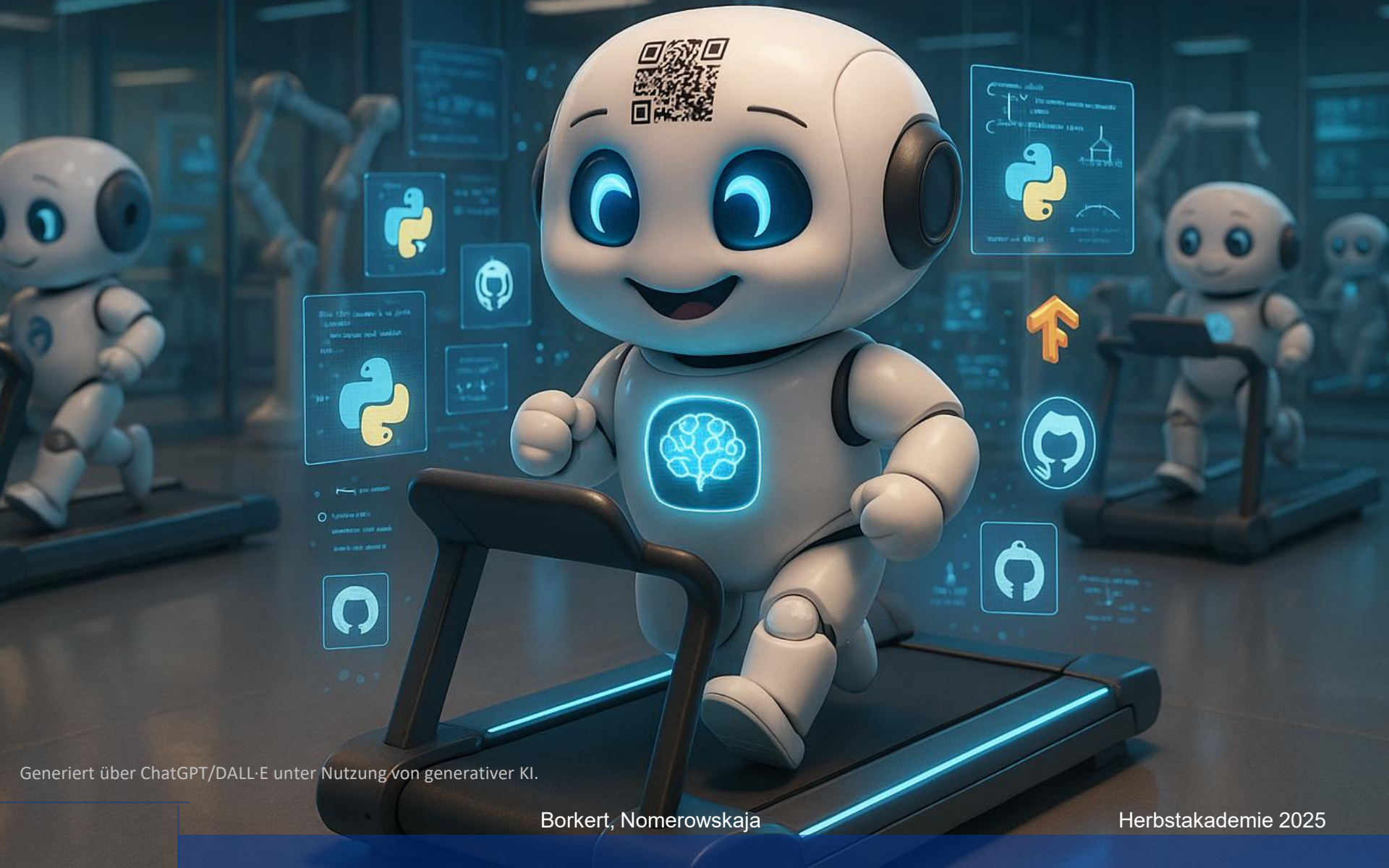
\$169.2M



MarketResearch **BIZ**
WIDE RANGE OF GLOBAL MARKET REPORTS

Quelle: <https://finance.yahoo.com/news/generative-ai-software-development-market-064400309.html>

Coding Agenten trainieren mit Open Source



Generiert über ChatGPT/DALL·E unter Nutzung von generativer KI.

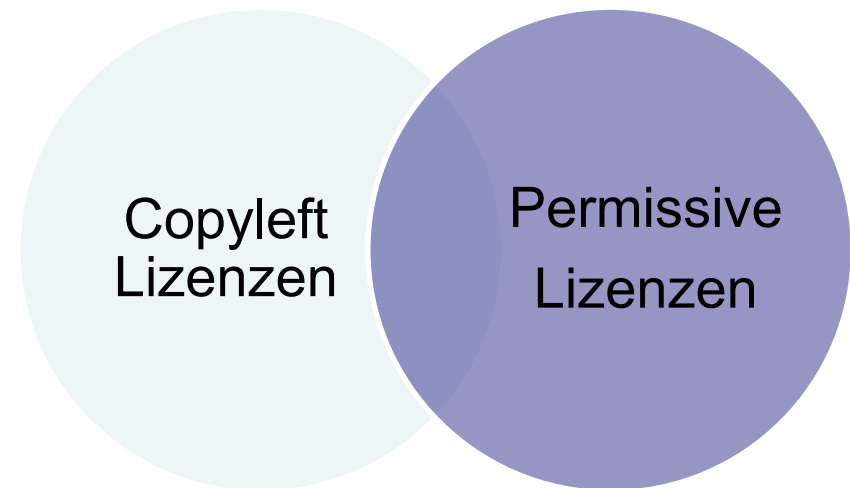
Die license.txt – ...wenn KI mit Open Source trainiert!

Open Source Lizenz gewährt Nutzern das Recht, die Software frei zu nutzen, zu vervielfältigen, zu verändern und weiterzuverarbeiten

FOSS-Lizenzen räumen ein
nichtausschließliches Nutzungsrecht an der
FOSS ein.



Verpflichtung die Lizenzbedingungen
einzuhalten und Ableitungen ggfs. unter der
gleichen Lizenz zu veröffentlichen / verbreiten.



Bereitstellung Lizenzbedingungen in
der Datei license.txt zum Download
bzw. im Installationspaket

**Ausgangsfrage:
Schränkt die license.txt die Nutzung von Open
Source für das KI-Training ein?**

Open Source Quellcode als Trainingsdaten für KI

Das perfekte Zusammenspiel von KI und Open Source

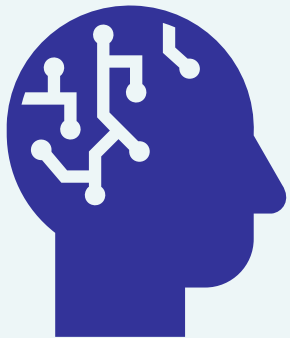
Heutzutage werden die meisten KI-Modelle mit Hilfe von Open-Source-Projekten trainiert. Diese bieten eine ideale Trainingsumgebung mit Zugang zu einer großen Anzahl von Codebasen, die weltweit in verschiedenen Infrastrukturen eingesetzt werden. Gleichzeitig wird die Software durch den Einsatz von KI effizienter und leistungsfähiger, was zu schnelleren Prozessen und besseren Funktionalitäten führt.

<https://www.it-daily.net/it-management/ki/ki-meets-open-source>

KI wirbelt die die Communities von Open-Source-Projekten durcheinander und stellt Entwicklerinnen und Entwickler vor wichtige Fragen. Darf Open-Source-Code zum Trainieren von KI-Modellen uneingeschränkt verwendet werden? Wo hilft KI und wo schadet sie Open-Source-Projekten – beim Beantworten von Fragen zum Code, beim Programmieren und beim Projektmanagement? Was ist Open Washing und welche Rolle spielt Open Source im EU AI Act? Und schließlich die entscheidende Frage: Wird KI Open Source verdrängen?

<https://www.heise.de/hintergrund/Wie-KI-die-Open-Source-Entwicklung-veraendert-10444253.html>

► Konforme Nutzung von FOSS unter Copyleft-Bedingungen?



- Training von KI-Systemen mit FOSS ist Eingriff in Urheberrecht.
- Bearbeitung und Weiterverbreitung lediglich unter der Bedingung zulässig, dass Veröffentlichung erneut unter den Bedingungen der ursprünglichen Lizenz erfolgt.
- Ohne Rückverfolgbarkeit, wie die Software trainiert wurde, ist Sicherstellung der Einhaltung der Lizenzanforderungen oder der Softwarelizenzierung problematisch.

Relevante urheberrechtliche Schranken



Beachte: Schrankenatalog enumerativ und abschließend
Im Kontext des KI-Trainings relevant:

- ▶ Text- und Data-Mining (TDM)-Schranken aus § 44b, § 60d UrhG
- ▶ die allgemeine Ausnahme für vorübergehende technische Vervielfältigung aus § 44a UrhG

§ 44a UrhG

Vorübergehende Vervielfältigungshandlungen

§ 44b UrhG

Text und Data Mining

§ 60d UrhG

Text und Data Mining für Zwecke der wissenschaftlichen Forschung

Vorbehalt des Text & Data Mining nach § 44b III UrhG

Werke von GEMA Mitgliedern dürfen nicht für kommerzielles Text- und Datamining (TDM) genutzt werden. Darunter fällt auch das Training von generativer künstlicher Intelligenz (KI). Den entsprechenden Nutzungsvorbehalt hat die GEMA öffentlich erklärt.

<https://www.gema.de/de/w/text-and-data-mining-vorbehalt-ki>



Automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insb. über Muster, Trends und Korrelationen zu gewinnen



Nutzungsvorbehalt des Rechteinhabers

Wirksamer Nutzungsvorbehalt bei online zugänglichen Werken nur in maschinenlesbarer Form.



Opt-out: Metadaten, license.txt, robots.txt, o.ä. technische Maßnahmen



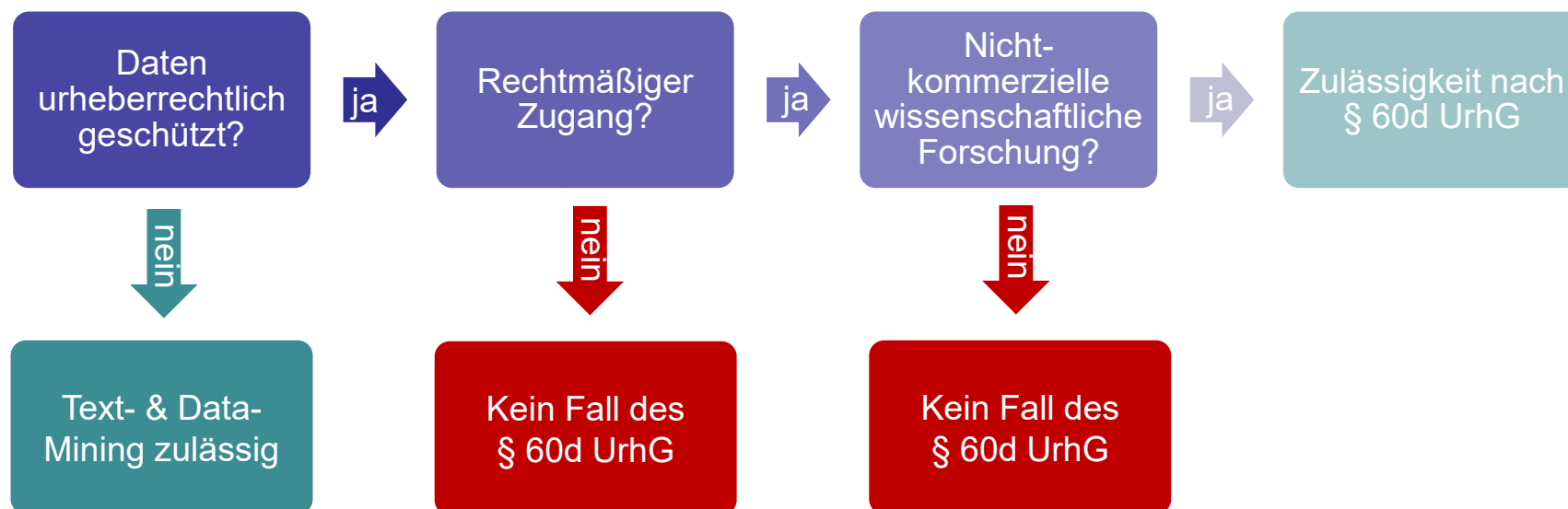
Nutzungsbedingungen?

LG Hamburg – Urteil „LAION“ (27.09.2024, Az. 310 O 227/23)

Rolle der wissenschaftlichen Forschung nach § 60d UrhG



Berechtigte, u.a. Forschungsorganisationen, wie Hochschulen & Forschungsinstitute, Gedächtnisorganisationen, wie Bibliotheken & Museen, sowie einzelne Forscher.

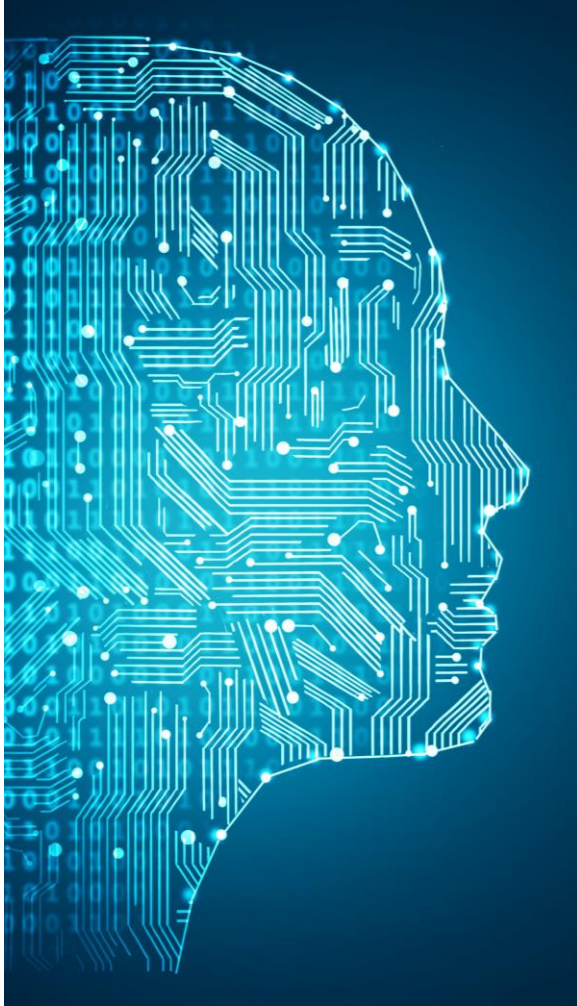


Erlaub: jede Vervielfältigung und Bearbeitung, die zum Zweck des TDM für die Forschung erforderlich ist.



Beachte: Rechteinhabern steht **kein Opt-out-Mechanismus** zu.

Bedeutung des Art. 53 I KI-VO und ErwG 105 KI-VO für das Text und Data Mining



Art. 53 I lit. c KI-VO

Anbieter müssen sicherstellen, dass beim Training und der Entwicklung ihrer KI-Modelle alle einschlägigen Urheberrechtsvorgaben eingehalten werden.

ErwG 105 KI-VO

Wurde (...) Opt-out ausdrücklich und in geeigneter Weise vorbehalten, müssen Anbieter (...) eine Genehmigung von den Rechteinhabern einholen, wenn sie Text- und Data-Mining mit solchen Werken durchführen wollen.

Folgen

- Fairer Wettbewerb durch **gleiche Urheberrechtsstandards** in der EU
- Anbieter müssen **proaktive Maßnahmen** ergreifen, um u.a. TDM-Opt-out aus § 44b III UrhG gerecht zu werden.
- **Marktübliche technische Mittel** einzusetzen, um in Trainingsdatensätzen mögliche Nutzungsvorbehalte maschinell zu erkennen und zu berücksichtigen.

Auswirkungen aktueller Rechtsprechungen



LG Hamburg – Urteil „LAION“ (27.09.2024, Az. 310 O 227/23)

- ▶ **Sachverhalt:** Der Verein LAION nutzte große Mengen urheberrechtlich geschützter Werke durch Webscraping zum Training von KI-Modellen.
 - ▶ **Streitpunkt:** Zulässigkeit nach § 44b, § 60d UrhG und **Vorliegen eines maschinenlesbaren Nutzungsvorbehalts.**
- ▶ **Ergebnis:** LG Hamburg bejahte die Zulässigkeit aufgrund von **§ 60d UrhG** (nicht-kommerzielle Forschung). Abschließende Klärung des wirksamen Vorbehaltes nicht entschieden.
- ▶ **Ausblick:** Rechtssicherheit fehlt weiterhin; eine höchstrichterliche Klärung steht noch aus.



Exkurs: US-Recht und das Fair Use Prinzip

The more transformative the new work, the less will be the significance of other factors, like commercialism (...)“ Campbell v. Acuff-Rose Music, Inc., 1994, 510 U.S. 569 (569)

- ▶ Vier nicht abschließende Faktoren, die bei der Bestimmung von Fair Use berücksichtigt werden müssen:

1. **Zweck und Charakter** der Nutzung, einschließlich der Frage, ob sie kommerziell oder nicht erwerbsorientiert erfolgt;
2. **Art** des urheberrechtlich geschützten Werks;
3. **Umfang und Wesentlichkeit** des genutzten Teils im Verhältnis zum Gesamtwerk; sowie
4. **Auswirkung** der Nutzung auf den potentiellen Markt oder Wert des geschützten Werks.

„undoubtedly the single most important element of fair use“
- Thomson Reuters vs. Ross Intelligence

Vergleichende Betrachtung

Aspekt	US-Recht (Fair Use)	EU-/ deutsches Recht
Rechtsgrundlage	Generalklauselartiges Fair-Use-Prinzip (§ 107 U.S. Copyright Act)	abgeschlossener Katalog an Schranken (§§ 44a ff. UrhG)
Flexibilität	<ul style="list-style-type: none"> • Hohe Flexibilität • technologieoffen • Gerichte können neue Nutzungsarten leichter einordnen 	<ul style="list-style-type: none"> • weniger anpassungsfähig • nur explizit geregelte Schranken anwendbar
Risiken/ Kritik	<ul style="list-style-type: none"> • Vorteil: Anpassungsfähigkeit; • Nachteil: Geringere Vorhersehbarkeit von Entscheidungen 	<ul style="list-style-type: none"> • Vorteil: Hohe Rechtssicherheit • Nachteil: Weniger Flexibilität
Aktuelle Entwicklung	<ul style="list-style-type: none"> • Weiterhin Einzelfallentscheidungen ohne spezielle KI-Gesetzgebung 	<ul style="list-style-type: none"> • KI-VO bringt erste Klarheit zu TDM (§§ 44b, 60d UrhG) • LG Hamburg („LAION“) schafft zusätzliche Orientierung

license.txt als maschinenlesbarer Vorbehalt

Maschinenlesbarkeit

AGB

Impressum

license.txt

Metadaten

robots.txt

Umkehrargument aus 2 ERVV:
„Dem elektronischen Dokument
soll ein **strukturierter maschinen-
lesbarer Datensatz im
Dateiformat XML** beigelegt
werden...“

natürliche Sprache



**technisch codierte
Informationen**



.csv, .xml,
.json, .txt,...

In ihrem Inhalt finden sich die Lizenzbedingungen, bei denen es sich rechtlich um AGB im Sinne der § 305 ff. BGB handelt.

ERVV = Elektronischer Rechtsverkehr-Verordnung, Regelt das besondere elektronische Behördenpostfach

license.txt als inhaltlicher Vorbehalt im Sinne des § 44 III UrhG am Beispiel der GPL 2

1

► Einordnung Ziffer 4 GPL 2 umstritten:

- auflösend bedingte Einräumung der Nutzungsrechte i.S.d. § 158 II BGB
- eigene Nutzungsart gem. § 31 UrhG
- lediglich bedingte Einräumung von Nutzungsrechten

You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. [...]

2

- Ausschließlichkeitsrecht des Urhebers: **nur bestimmte Nutzung gestattet** → andere Nutzungen ausgeschlossen
- Nach Ziffer 4 **entfallen** explizit alle Nutzungsrechte, wenn Lizenzbedingungen nicht eingehalten
- **Kombination dieser Bedingungen mit enger Auslegung der Schranken spricht für hinreichende Bestimmung der license.txt der der GPL 2 als Nutzungsvorbehalt nach § 44 III UrhG**

license.txt als inhaltlicher Vorbehalt im Sinne des § 44 III UrhG - Zusammenfassung



Generiert über ChatGPT/DALL-E unter Nutzung von generativer KI.

Abschluss & Ausblick

- ▶ **Open-Source-Software**, speziell unter sog. **Copy-Left-Lizenzen**, kann beim KI-Training erhebliche **lizenzrechtliche Risiken** auslösen, sowohl für die **Nutzung der Trainingsdaten** als auch für den **Output**
 - ➔ **Einzelfallprüfung je Lizenz**, neues Feld, noch ohne gerichtliche Klärung
- ▶ Unternehmen sollten **proaktives Risikomanagement** betreiben u. a. durch
 1. **sorgfältige Lizenzprüfungen** ➔ Code-Scans mit Abgleich zu den Trainingsdaten und Lizenzmodellen,
 2. **vertragliche Absicherungen** ➔ Haftung für Code Scans, verlässlicher Ausschluss von erzeugtem Code, der ähnlich zu Open Source Code (permissive oder copy left) ist, bzw. Nennung entsprechend der Lizenzbedingungen
 3. **transparente Dokumentation** ➔ der Prompts für die Code-Generation sowie der der Trainingsdaten zum Abgleich
 4. **eine klare KI-Governance** ➔ z.B. Informations- und Auditrechte gegenüber dem Hersteller des Code Generators, zusätzlich Punkte 1-3.
- ▶ Der Trend zu **KI-Agenten** und **Agentic Automation** verstärkt die Problematik, da Systeme zunehmend **eigenständig Code und Inhalte generieren**.



Vielen Dank für Ihre Aufmerksamkeit!



Kristian Borkert



Anastasia Nomerowskaja

📍 JURIBO Legal & Consulting
Am Rötelpark 82
71332 Waiblingen

✉ info@juribo.de

☎ +49 (0) 7151 966 9380

