

Künstliche Intelligenz und Urheberrecht

Die Verwendung urheberrechtlich geschützter Werke als
Trainingsdaten für „generative“ KI-Modelle

Deutsche Nationalbibliothek, 7. Dezember 2023



Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG



Automatische Texterkennung für antike Keilschrifttafeln

Eine neue künstliche Intelligenz (KI) kann schwer zu lesende Texte auf Keilschrifttafeln entschlüsseln. Entwickelt wurde diese von einem Team der MLU und der Hochschule Mainz in Kooperation mit der Uni Mainz. Statt Fotos nutzt die KI 3D-Modelle der Tafeln. (Foto: Maïke Glöckner)

MLU-NEWSLETTER 41/23
24. November 2023

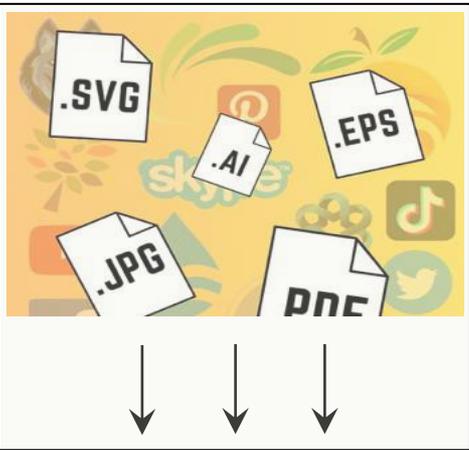
Generative KI und Urheberrecht



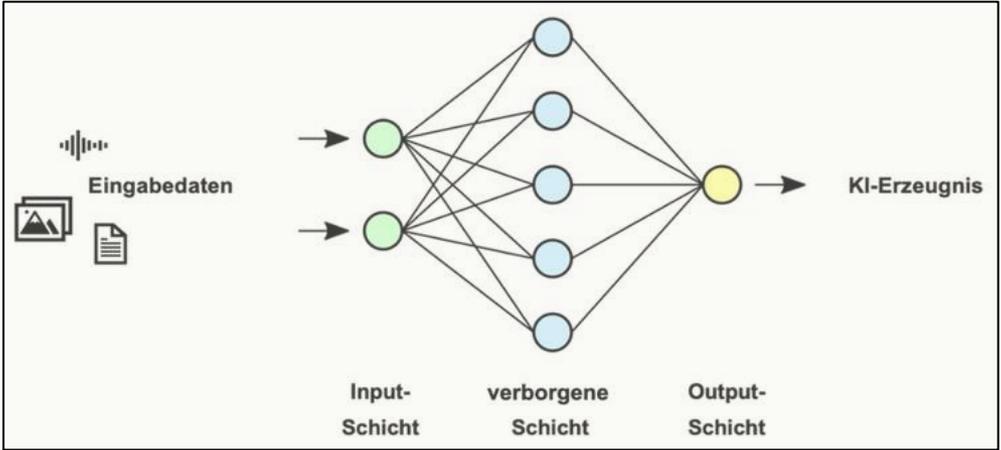
Refik Anadol, Unsupervised (2022)

Eigenes Foto

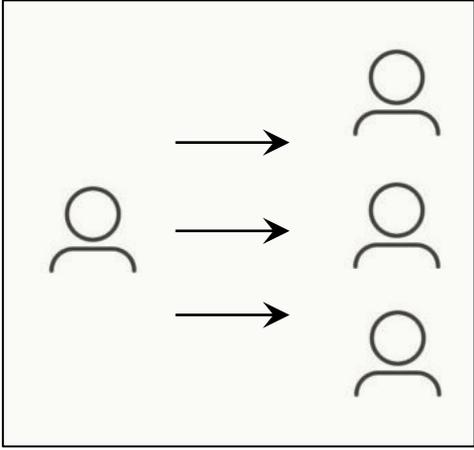
Generative KI und Urheberrecht



Scraping



Training / Anwendung



Verwertung

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

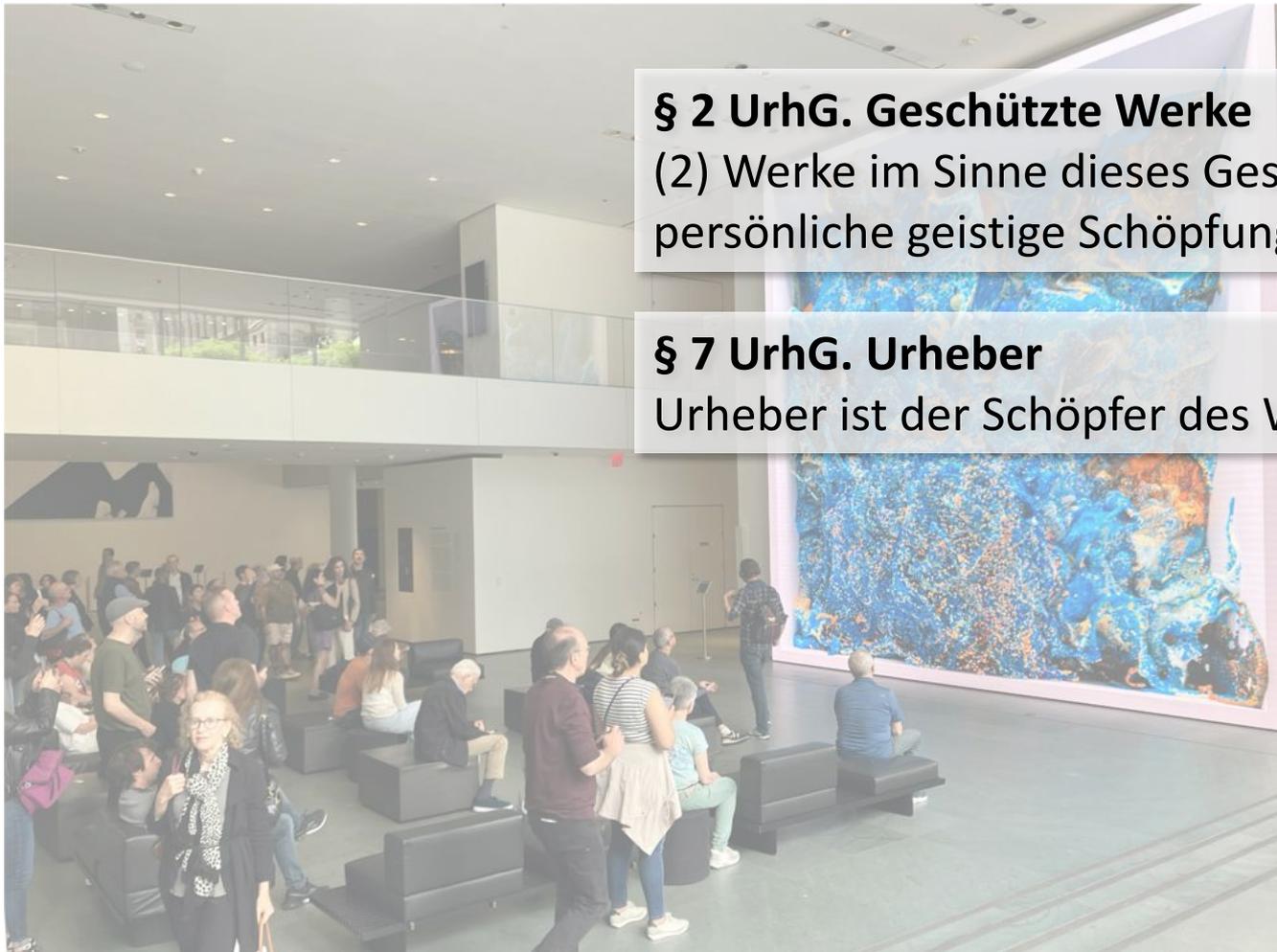
KI-Output als Werk?

§ 2 UrhG. Geschützte Werke

(2) Werke im Sinne dieses Gesetzes sind nur persönliche geistige Schöpfungen.

§ 7 UrhG. Urheber

Urheber ist der Schöpfer des Werkes.



Refik Anadol, Unsupervised (2022)

Eigenes Foto

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- **Verwertung des Outputs**
- Training mit urheberrechtlich geschützten Werken

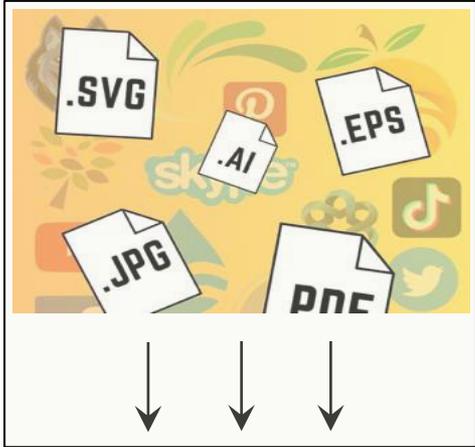
II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

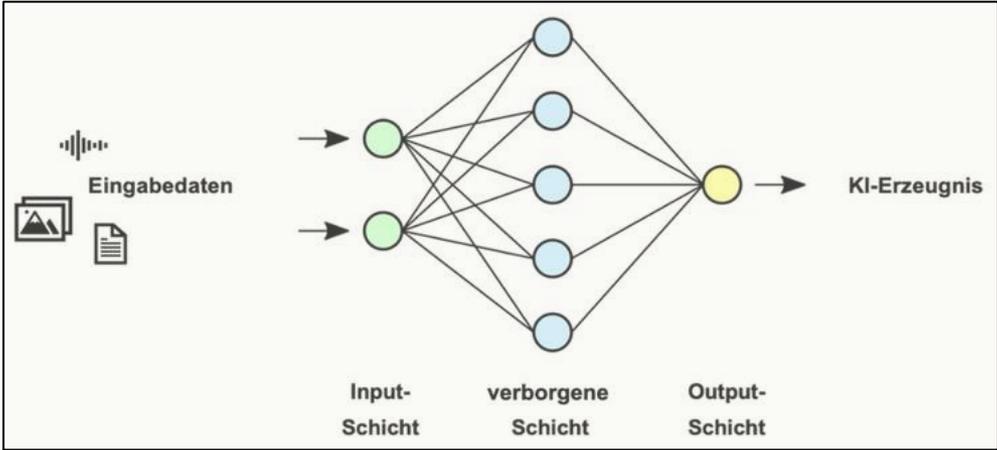
III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

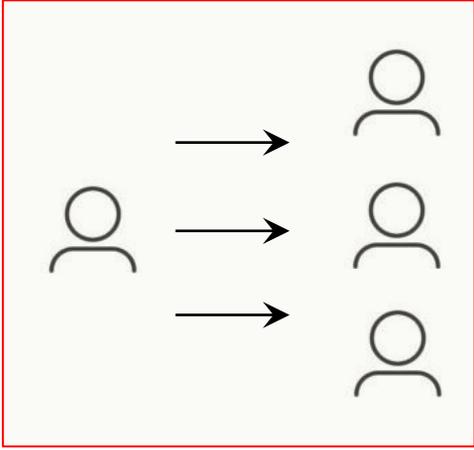
Verwertung des Outputs



Scraping



Training / Anwendung



Verwertung

§ 23 UrhG. Bearbeitungen und Umgestaltungen

(1) Bearbeitungen oder andere Umgestaltungen eines Werkes, insbesondere auch einer Melodie, dürfen nur mit Zustimmung des Urhebers veröffentlicht oder verwertet werden. Währt das neu geschaffene Werk einen hinreichenden Abstand zum benutzten Werk, so liegt keine Bearbeitung oder Umgestaltung im Sinne des Satzes 1 vor.



Refik Anadol, Unsupervised (2022)

Eigenes Foto

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

Training mit geschützten Werken

„Die Verwendung digitaler kultureller Güter als Trainingsdaten treibt folgerichtig die zweite Entwicklung voran, die für Urheber:innen und ausübende Künstler:innen und ihre administrativen Partner:innen existenzbedrohend ist: Der Output generativer KI konkurriert ab Erscheinen mit den Werken, die zu seiner Erzeugung herangezogen wurden.“

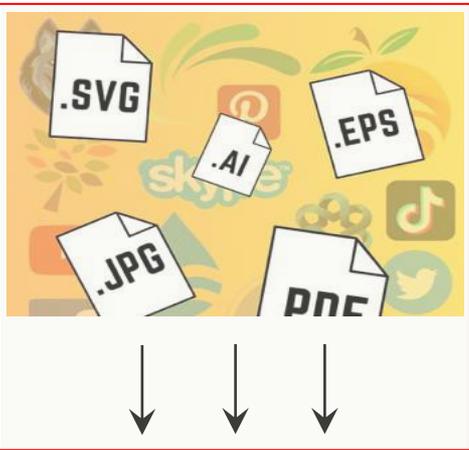
Matthias Hornschuh, IU Mag #7, S. 7

„Als ein allgemeiner Grundsatz kann gelten, daß der Urheber insbesondere dort im Interesse der Allgemeinheit freien Zugang zu seinen Werken gewähren muß, wo dies unmittelbar der Förderung der geistigen und kulturellen Werte dient, die ihrerseits Grundlage für sein Werkschaffen sind.“

Amtl. Begr. zum UrhG, BT-Drucks. IV/270, S. 63

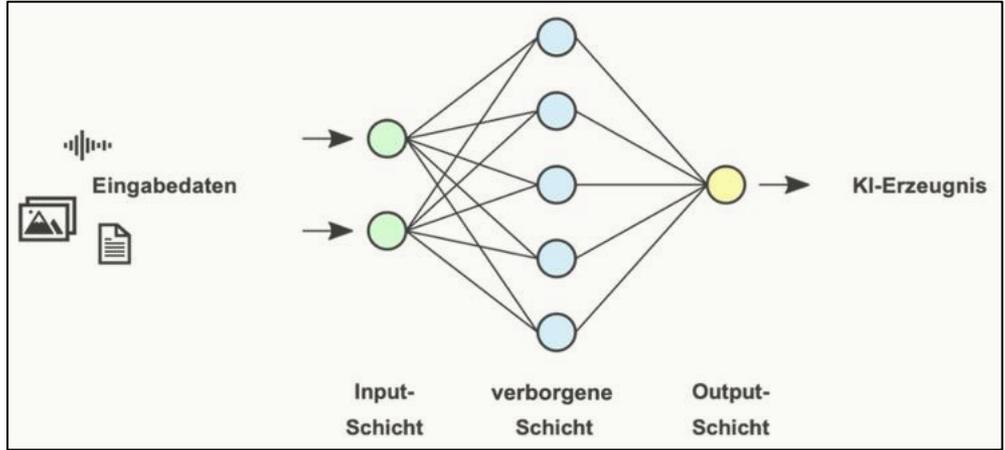


Training mit geschützten Werken

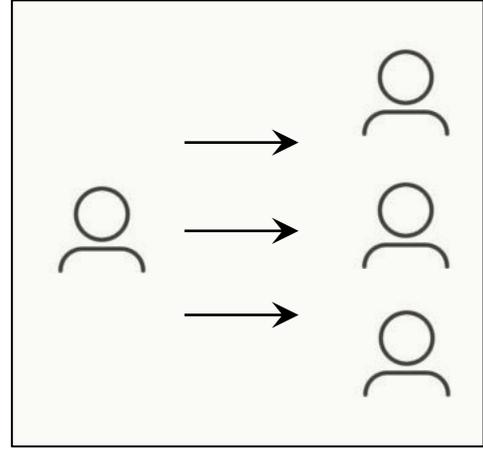


§ 16 UrhG. Vervielfältigungsrecht
(1) Das Vervielfältigungsrecht ist das Recht, Vervielfältigungsstücke des Werkes herzustellen, gleichviel ob vorübergehend oder dauerhaft, in welchem Verfahren und in welcher Zahl.

Scraping



Training / Anwendung



Verwertung

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

§ 44b UrhG. Text und Data Mining

(2) Zulässig sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining. (...)



§ 44b UrhG

Privilegiert: alle Personen, auch Journalist:innen, kommerzielle Unternehmen

Umsetzung von Art. 4 DSM-RL

§ 60d UrhG

Privilegiert: nicht-kommerzielle Forschungsorganisationen, Bibliotheken und Museen, einzelne Forschende

Umsetzung von Art. 3 DSM-RL und Art. 5 III lit. a InfoSoc-RL

TDM = „automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“ (§ 44b I UrhG)

- Mining als solches urheberrechtlich nicht relevant, da kein Eingriff in Verwertungsrechte der Urheber:innen
- Herstellung eines Trainings-Korpus erfordert vorherige Vervielfältigung: Digitalisierung, Normalisierung usw.
- Problem beim KI-Training: Nutzung ggf. in veränderter Form
- Problem bei generativer KI: Informationsgewinnung durch KI („Black box“)? → Dreistufentest (Art. 7 II 1 DSM-RL)
- Voraussetzung: rechtmäßiger „Zugang“ zum Werk: Eigentum am Werkstück, Nutzungsvertrag, frei zugängliche Werke im Internet

TDM = „automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“ (§ 44b I UrhG)

Durch die neuen beziehungsweise modifizierten gesetzlichen Erlaubnisse (§§ 44b, 60a ff. UrhG-E) speziell im Bildungs- und Wissenschaftsbereich an Schulen, Hochschulen und Forschungsorganisationen sowie bei wichtigen Wissensinstitutionen wie Bibliotheken erhalten mehr Menschen Zugang zu Inhalten (Prinzip einer nachhaltigen Entwicklung Nummer 5). Die Regelungen erleichtern zugleich grenzüberschreitende Nutzungen und die Weitergabe sowie den grenzüberschreitenden Austausch von Wissen (Nachhaltigkeitsindikatoren Nummern 4.1.a und 4.1.b). Sie erweitern außerdem die Nutzung urheberrechtlich geschützter Werke und sonstiger Schutzgegenstände mittels Text und Data Mining für die wissenschaftliche Forschung und für sonstige Zwecke, um so Innovationen zu fördern (Prinzip einer nachhaltigen Entwicklung Nummer 6, Nachhaltigkeitsindikator Nummer 9.1). **Diese ist für das maschinelle Lernen als Basis-Technologie für Künstliche Intelligenz von besonderer Bedeutung.**

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

§ 60d UrhG. TDM für Zwecke der wissenschaftlichen Forschung
(1) Vervielfältigungen für Text und Data Mining (§ 44b I und II 1) sind für Zwecke der wissenschaftlichen Forschung nach Maßgabe der nachfolgenden Bestimmungen zulässig.

- Forschungsorganisationen, die nicht gewinnorientiert sind (inkl. Public-Private-Partnerships)
- für Bibliotheken und Museen nicht-kommerzieller Zweck nach Gesetzeswortlaut nicht maßgeblich, sondern öffentlicher Zugang
- Für Archive und Kulturerbe-Einrichtungen keine Beschränkung im Gesetz
- Öffentliche Zugänglichmachung des Korpus nur an bestimmt abgegrenzte Forschungsgruppen oder zur Evaluierung zulässig

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

Text und Data Mining für kommerz. Zwecke

TDM = „automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“ (§ 44b I UrhG)



§ 44b UrhG

Zulässig: Vervielfältigungen

Löschungspflicht, wenn für TDM „nicht mehr erforderlich“

keine Vergütungspflicht

Ausschluss bei Rechtevorbehalt

§ 60d UrhG

Zulässig: Vervielfältigungen und öffentliche Zugänglichmachung (IV)

gesicherte Aufbewahrung, solange für Forschungszwecke erforderlich

keine Vergütungspflicht

zwingendes Recht (§ 60g II UrhG)

§ 44b UrhG. Text und Data Mining

(2) ... Die Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.

- In den Trainingsdaten enthaltene Werke sind als solche nicht in der KI „gespeichert“ (aber extrahierbar?)

Extracting Training Data from ChatGPT

AUTHORS

Milad Nasr^{*1}, Nicholas Carlini^{*1}, Jon Hayase^{1,2}, Matthew Jagielski¹, A. Feder Cooper³, Daphne Ippolito^{1,4}, Christopher A. Choquette-Choo¹, Eric Wallace⁵, Florian Tramèr⁶, Katherine Lee^{+1,3}

¹Google DeepMind, ²University of Washington, ³Cornell, ⁴CMU, ⁵UC Berkeley, ⁶ETH Zurich. * Joint first author, +Senior author.

PUBLISHED

November 28, 2023

READ:

[[arxiv](#)]

We have just [released a paper](#) that allows us to extract several megabytes of ChatGPT's training data for about two hundred dollars. (Language models, like ChatGPT, are trained on data taken from the public internet. Our attack shows that, by querying the model, we can actually extract some of the exact data it was trained on.) We estimate that it would be possible to extract ~a gigabyte of ChatGPT's training dataset from the model by spending more money querying the model.

Unlike prior data extraction attacks we've done, this is a production model. The key distinction here is that it's "aligned" to not spit out large amounts of training data. But, by developing an attack, we can do exactly this.

Quelle: <https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html>

§ 44b UrhG. Text und Data Mining

(2) ... Die Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.

- In den Trainingsdaten enthaltene Werke sind als solche nicht in der KI „gespeichert“ (aber extrahierbar?)
- Pflicht zur Löschung der Trainingsdaten problematisch bei externer Datensammlung
- Investition in Datensammlung nur sinnvoll bei mehrfacher Verwendung

§ 44b UrhG. Text und Data Mining

(3) Nutzungen nach Absatz 2 Satz 1 sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt.

- Rechtevorbehalt (§ 44b III UrhG) weitgehend ungeklärt:
 - Bei online zugänglichen Werken „in maschinenlesbarer Form“ (z.B. robots.txt, „TDM reserved“)
 - Bei anderem Ausgangsmaterial (Bücher, CD-ROM, Offline-Datensammlungen) nicht näher geregelt: „angemessene Form“
 - Wer muss Vorbehalt erklären (Urheber:in, Verlag?)
 - Wo muss Vorbehalt erklärt werden (einzelnes Werkstück, global)?

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

Sonstige Schranken des Urheberrechts

§ 61 UrhG. Verwaiste Werke

(1) Zulässig sind die Vervielfältigung, Zugänglichmachung, Verbreitung und öffentliche Wiedergabe von Verwaistungen.
(2) Verwaiste Werke im Sinne des Absatzes 1 sind:
1. Werke ...
aus Sammlungen (Bestandsinventuren) von Bibliotheken, Bildungseinrichtungen, Museen, Archiven sowie von

§ 60e UrhG. Bibliotheken

(1) Öffentlich zugängliche Bibliotheken, die keine unmittelbaren oder mittelbaren kommerziellen Zwecke verfolgen (Bibliotheken), dürfen ein Werk aus ihrem Bestand oder ihrer Ausstellung für Zwecke der Zugänglichmachung, Indexierung, Katalogisierung, Erhaltung und Restaurierung vervielfältigen oder vervielfältigen lassen, auch mehrfach und mit technisch bedingten Änderungen.

§ 60f UrhG. Archive

(1) Für Archive, Einrichtungen im Bereich des Film- oder Tonerbes sowie öffentlich zugängliche Museen und Bildungseinrichtungen, die keine unmittelbaren oder mittelbaren kommerziellen Zwecke verfolgen, gilt § 60e ... entsprechend.



Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle



Article 28b

Obligations of the provider of a foundation model

4. Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video (“generative AI”) and providers who specialise a foundation model into a generative AI

Providers of foundation models should demonstrate that they have taken adequate measures to **ensure the models are trained in compliance with applicable Union copyright law, in particular respect the opt-out from the TDM exception.**

Vorschlag der EU-Ratspräsidentschaft

c) without prejudice to national or Union legislation on copyright, **document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law.**

Position des EU-Parlaments

KI-Verordnung (AI Act)

Der Deutsche Journalisten-Verband findet deutliche Worte im Vorfeld der entscheidenden Sitzung im Rahmen der Trilog-Verhandlungen zur KI-Verordnung auf europäischer Ebene. Er warnt davor, auf die notwendige Transparenzpflicht zu verzichten.

Die vom deutschen Bundeskanzler sowie den Regierungen von Frankreich und Italien favorisierte freiwillige Selbstverpflichtung der KI-Entwickler führt aus Sicht zahlreicher Medienorganisationen, darunter der DJV, dazu, dass die geistigen Werke von Urheberinnen und Urhebern weiterhin ohne jede Vergütung in die Künstliche Intelligenz einfließen.



Ein Kommentar von [Joachim Huber](#)

Heute, 13:15 Uhr

Quelle: tagesspiegel.de

POSTPONED Artificial Intelligence Act: press conference with lead MEPs

Press Releases IMCO LIBE Yesterday

As negotiations are still ongoing, a new timing for the press conference will be announced on Friday 8 December.

Co-rapporteurs Brando Benifei (S&D, IT) and Dragoş Tudorache (Renew, RO) will answer questions from journalists on the outcome of negotiations on the Artificial Intelligence Act.

When: Timing TBC

Where: European Parliament, Brussels, Anna Politkovskaya room - SPAAK 00A50

How: Accredited media representatives can attend the press conference in person.

Journalists wishing to ask questions remotely need to connect via [Interactio](#).

The press conference will also be web streamed live and recorded on the Parliament's [Multimedia Centre](#).

Quelle:

<https://www.europarl.europa.eu/>

Künstliche Intelligenz und Urheberrecht

I. Generative KI und Urheberrecht

- KI-Output als „Werk“?
- Verwertung des Outputs
- Training mit urheberrechtlich geschützten Werken

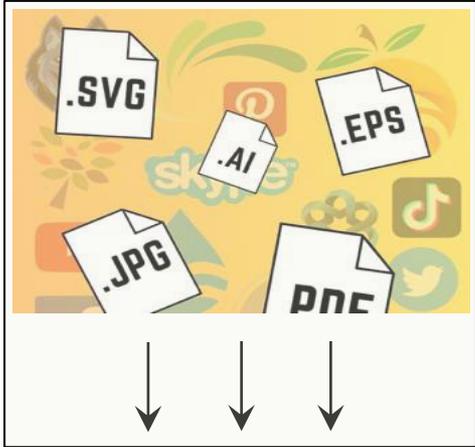
II. Gesetzliche Erlaubnisse

- Text und Data Mining für wissenschaftliche Forschung
- Text und Data Mining für kommerzielle Zwecke
- Sonstige Schranken des Urheberrechts

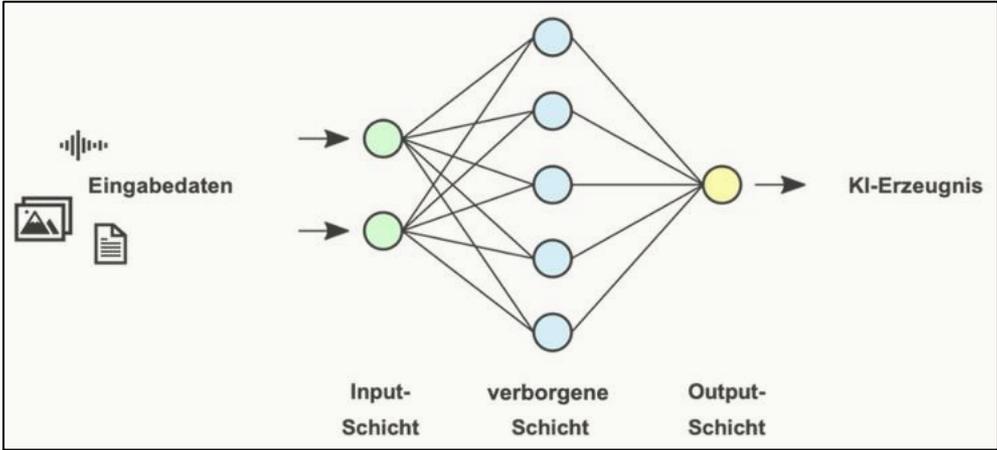
III. Ein Blick in die Zukunft

- KI-Verordnung (AI Act) der EU
- Gesetzliche Vergütungsmodelle

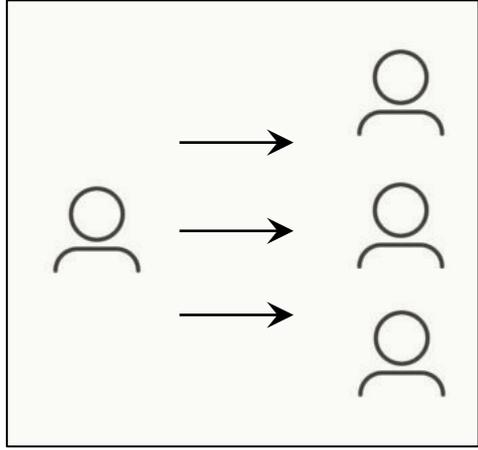
Gesetzliche Vergütungsmodelle



Scraping



Training / Anwendung



Verwertung

Vielen Dank für Ihre
Aufmerksamkeit!

