

DDM4SME

Urheberrechtliches Streitgespräch mit ChatGPT: Training und Ergebnisverwertung

Ausgewählte Fragestellungen und Ergebnisse der Begleitforschung zu DDM4SME

Clemens Appl, Universität für Weiterbildung Krems
Philipp Homar, Johannes-Kepler-Universität Linz



MUNI
LAW



Kazimieras
Simonavicius
University



1. Training von ChatGPT



Ist es urheberrechtlich zulässig, eine Künstliche Intelligenz mit fremden Daten aus dem Internet zu trainieren?



Es gibt keine einfache Antwort auf diese Frage, da das Urheberrecht von Land zu Land unterschiedlich sein kann und es auch auf den genauen Kontext ankommt. In einigen Fällen kann das Training einer Künstlichen Intelligenz (KI) mit fremden Daten aus dem Internet eine Verletzung des Urheberrechts darstellen, während es in anderen Fällen zulässig sein kann.

Generell ist das Sammeln von Daten aus dem Internet möglicherweise zulässig, wenn die Daten frei zugänglich und nicht durch das Urheberrecht geschützt sind. Wenn die Daten jedoch durch das Urheberrecht geschützt sind, kann das Sammeln und Verwenden dieser Daten ohne Erlaubnis des Urheberrechtinhabers eine Urheberrechtsverletzung darstellen.

Wenn Sie beabsichtigen, eine KI mit fremden Daten aus dem Internet zu trainieren, sollten Sie zunächst sicherstellen, dass Sie eine klare Vorstellung davon haben, welche Art von Daten Sie sammeln und wie Sie sie verwenden möchten. Wenn Sie beabsichtigen, geschützte Daten zu verwenden, sollten Sie eine Genehmigung des Urheberrechtinhabers einholen, um sicherzustellen, dass Sie das Urheberrecht nicht verletzen.

In jedem Fall empfehle ich, sich mit einem Anwalt oder einer anderen Rechtsberatung in Verbindung zu setzen, um sicherzustellen, dass Sie im Einklang mit den geltenden Urheberrechtsbestimmungen handeln.

Text und Data Mining Privilegierung im Urheberrecht

- Umsetzung von Art 3 und 4 DSM-RL + nationale Erweiterung gestützt auf Art 5 InfoSoc-RL
- Allgemeines Text and Data Mining (Art 4 DSM-RL; Umsetzung § 42h Abs 6)
 - Privilegierung für jedermann
 - Aber: vielfältige Einschränkungen, insb Vertragsvorbehalt
- Text and Data Mining für wissenschaftliche Zwecke (Art 3 DSM-RL; Umsetzung § 42h Abs 1 ff)
 - großzügige Privilegierung, aber zweckgebunden
 - Eingeschränkter Adressatenkreis
 - KEIN Vertragsvorbehalt



§ 42h UrhG – Komplexität in der Umsetzung

(1) Jedermann darf für eine Forschungseinrichtung (Abs. 3) oder für eine Einrichtung des Kulturerbes (§ 42 Abs. 7) ein Werk vervielfältigen, um damit Texte und Daten in digitaler Form für die wissenschaftliche oder künstlerische Forschung automatisiert auszuwerten und Informationen unter anderem über Muster, Trends und Korrelationen zu gewinnen, wenn er zu dem Werk rechtmäßig Zugang hat. Zu einer solchen Vervielfältigung sind auch einzelne Forscher berechtigt, soweit dies zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist.

(2) Eine Vervielfältigung nach Abs. 1 darf unter Wahrung angemessener Sicherheitsvorkehrungen gespeichert und aufbewahrt werden, solange dies durch den Forschungszweck, auch zur Überprüfung wissenschaftlicher Erkenntnisse, gerechtfertigt ist. Jedenfalls angemessen ist eine Sicherheitsvorkehrung, deren Einsatz von repräsentativen Vereinigungen von Rechteinhabern einerseits sowie Forschungseinrichtungen oder Einrichtungen des Kulturerbes andererseits als bewährte Vorgehensweise anerkannt wurde. Eine solche Vervielfältigung darf auch einem bestimmt abgegrenzten Kreis von Personen für deren gemeinsame wissenschaftliche Forschung oder einzelnen Dritten zur Überprüfung der Qualität wissenschaftlicher Forschung zugänglich gemacht werden, soweit dies zur Verfolgung nicht kommerzieller Zwecke gerechtfertigt ist.

(3) Eine Forschungseinrichtung im Sinn dieser Bestimmung ist eine Einrichtung,
1. deren vorrangiges Ziel die wissenschaftliche oder künstlerische Forschung oder die forschungsgel leitete Lehre ist und
2. die in ihrer Tätigkeit nicht gewinnorientiert ist, alle Gewinne in ihre wissenschaftliche oder künstlerische Forschung reinvestiert oder gewinnorientiert und im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig ist und
3. bei der nicht ein Unternehmen, das einen bestimmenden Einfluss auf die Einrichtung hat, bevorzugten Zugang zu den Ergebnissen der wissenschaftlichen Forschung erhält.

(4) Abs. 1 bis 3 sind auch dann anzuwenden, wenn die Vervielfältigung im Rahmen einer öffentlich-privaten Partnerschaft erfolgt, an der neben der Forschungseinrichtung oder der Einrichtung des Kulturerbes auch ein auf Gewinn gerichtetes Unternehmen oder ein sonstiger Dritter beteiligt ist.

(5) Die freie Werknutzung nach Abs. 1 bis 4 kann vertraglich nicht abbedungen werden. Dies steht aber der Anwendung von Maßnahmen nicht entgegen, die die Sicherheit und Integrität der Netze und Datenbanken gewährleisten sollen, in denen die Werke oder sonstigen Schutzgegenstände gespeichert sind, soweit diese Beschränkungen nicht über das für die Verwirklichung dieses Ziels Notwendige hinausgehen. Solche Beschränkungen gelten als angemessen, wenn sie von repräsentativen Vereinigungen von Rechteinhabern einerseits sowie Forschungseinrichtungen oder Einrichtungen des Kulturerbes andererseits als bewährte Vorgehensweise anerkannt wurden.

(6) Jedermann darf für den eigenen Gebrauch ein Werk vervielfältigen, um damit Texte und Daten in digitaler Form automatisiert auszuwerten und Informationen unter anderem über Muster, Trends und Korrelationen zu gewinnen, wenn er zu dem Werk rechtmäßig Zugang hat. Dies gilt jedoch nicht, wenn die Vervielfältigung ausdrücklich verboten und dieses Verbot in angemessener Weise durch einen Nutzungsvorbehalt, und zwar etwa bei über das Internet öffentlich zugänglich gemachten Werken mit maschinenlesbaren Mitteln, kenntlich gemacht wird. Eine Vervielfältigung nach diesem Absatz darf aufbewahrt werden, solange dies für die Zwecke der Datenauswertung und Informationsgewinnung notwendig ist.

Fokus 1: Sachlicher Anwendungsbereich Computerprogramme?

- § 42h UrhG ist grds werkneutral gefasst
- Aber: Nur Art 4, nicht aber Art 3 DSM-RL verweist nicht auf Computerprogramm-RL
 - Redaktionsversehen? Siehe ErwGr 5 und 18 DSM-RL
 - Sachliche Rechtfertigung der Differenzierung?
 - Möglichkeit eines Nutzungsvorbehalts?
- Besondere Natur von Computerprogrammen
 - Computerprogramme als Gegenstand oder Instrument von Text und Data Mining?
- IE Computerprogramme sind mE erfasst, aber nur als Gegenstand, nicht als Instrument.

§ 42h Abs 1 und 6
... ein Werk ...

Art 4 DSM-RL:

Für zum Zwecke des Text und Data Mining vorgenommene Vervielfältigungen und Entnahmen von rechtmäßig zugänglichen Werken und sonstigen Schutzgegenständen sehen die Mitgliedstaaten eine Ausnahme oder Beschränkung von den Rechten vor, die in [...] Art 4 Abs 1 lit a und b Computerprogramm-RL [...] niedergelegt sind.

Fokus 1: Sachlicher Anwendungsbereich

Rechtmäßiger Zugang?

- **Rechtmäßiger Zugang** zum Werk als zwingende Voraussetzung
 - Neue Formulierung ggü Altbestand (s rechts)
 - Auswirkungen?
- **ErwGr 33 InfoRL:** Nutzung ist rechtmäßig, soweit sie vom Rechteinhaber zugelassen oder nicht durch Gesetze beschränkt ist.
- **ErwGr 14 DSM-RL:** alles, was im Internet „frei verfügbar“ ist.
 - Technische frei? Oder rechtlich frei? Oder Beides?
- Wer muss rechtmäßigen Zugang haben?
 - Abs 1: **Person oder Institution?**

§ 42 Abs 5 UrhG... offensichtlich rechtswidrig hergestellte oder öffentlich zugänglich gemachte Vorlage ...

§ 41a UrhG
... rechtmäßige Nutzung ...

§ 56a Abs 2 UrhG

Abs. 1 gilt nicht für Bild- oder Schallträger, die **mit Verletzung eines ausschließlichen Rechtes**, das darauf festgehaltene Werk zu vervielfältigen oder zu verbreiten, hergestellt oder verbreitet worden sind.

§ 42h UrhG

... rechtmäßig[er] Zugang ...

Fokus 2: Privilegierte Handlungen

Digitalisieren analoger Daten?

- Gegenstand des TDM = **Texte und Daten in digitaler Form**
- Digitalisierung analoger Texte und Daten = **gesonderte** Vervielfältigung iSd § 15 UrhG
- **Digitalisierung als privilegierte Annexvervielfältigung**
 - Vervielfältigung eines (analogen) Werks erfolgt gerade, „um“ Texte und Daten in digitaler Form auszuwerten
 - § 44b (1) dUrhG
- Tauglicher Gegenstand des TDM = Texte und Daten in **digital(isiert)er Form**
 - Alternativ: § 42 (2) UrhG, § 42 (4) UrhG

§ 42h (1) und (6) UrhG

darf [...] ein Werk vervielfältigen, um damit Texte und Daten in digitaler Form für die wissenschaftliche oder künstlerische Forschung automatisiert auszuwerten

Fokus 3: Einschränkungen der Zulässigkeit durch Nutzungsvorbehalte?

- **Forschungs-TDM (§ 42h (1) UrhG)**
 - vertragsfest (§ 42h (5) UrhG)
 - Vervielfältigungen iSd § 42h (1) UrhG lösen weder urheberrechtliche noch schuldrechtliche Ansprüche aus
 - *fraglich: maschinenlesbare Vorbehalte?*
 - *fraglich: Zulässigkeit technischer Barrieren? Muss Nutzung technisch ermöglicht werden oder ist nur eine gezielte Erschwernis verboten?*
- **TDM zum eigenen Gebrauch (§ 42h (6) UrhG)**
 - Unzulässigkeit bei **ausdrücklichem Verbot**, das **angemessen durch Nutzungsvorbehalt** kenntlich gemacht wurde
 - *„etwa bei über das Internet öffentlich zugänglich gemachten Werken mit maschinenlesbaren Mitteln“*
 - *fraglich: nicht maschinenlesbare Vorbehalte?*
 - *fraglich: vertragliche Vorbehalte?*

(5) Die freie Werknutzung nach Abs. 1 bis 4 kann vertraglich nicht abbedungen werden.

(6) Dies gilt jedoch nicht, wenn die Vervielfältigung ausdrücklich verboten und dieses Verbot in angemessener Weise durch einen Nutzungsvorbehalt, und zwar etwa bei über das Internet öffentlich zugänglich gemachten Werken mit maschinenlesbaren Mitteln, kenntlich gemacht wird.

Fokus 4: PPP Data Laundering

- **Forschungs-TDM** (§ 42h (1) UrhG) bietet gegenüber TDM zum eigenen Gebrauch (§ 42h (6) UrhG) erhebliche Vorteile
 - *Insb kein Ausschluss durch Nutzungsvorbehalte möglich*
- **Kann Abs 6 durch PPP Auftragsforschung umgangen werden?**
 - Unternehmen beauftragt Forschungseinrichtung mit dem Training einer AI, bezahlt sämtliche direkten Kosten der Forschungseinrichtung und lässt sich exklusive Verwertungsrechte einräumen
- **Hürden:**
 - wissenschaftliche Forschung?
 - Kommerzielle Zwecke?
 - Bei bloßer Kostendeckung oder Reinvestition in Forschung?
 - Verhältnis von Abs 3 Z 3 und Abs 4 fraglich?
 - Bevorzugter Zugang als Problem?

(3) Eine Forschungseinrichtung im Sinn dieser Bestimmung ist eine Einrichtung,
1. deren vorrangiges Ziel die wissenschaftliche oder künstlerische Forschung oder die forschungsgeleitete Lehre ist und
2. die in ihrer Tätigkeit nicht gewinnorientiert ist, alle Gewinne in ihre wissenschaftliche oder künstlerische Forschung reinvestiert oder gewinnorientiert und im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig ist und
3. bei der nicht ein Unternehmen, das einen bestimmenden Einfluss auf die Einrichtung hat, bevorzugten Zugang zu den Ergebnissen der wissenschaftlichen Forschung erhält.

(4) Abs. 1 bis 3 sind auch dann anzuwenden, wenn die Vervielfältigung im Rahmen einer **öffentlich-privaten Partnerschaft** erfolgt, an der neben der Forschungseinrichtung oder der Einrichtung des Kulturerbes auch ein auf Gewinn gerichtetes Unternehmen oder ein sonstiger Dritter beteiligt ist.

Fokus 5: Die trainierte KI

- Trainiertes Sprachmodell kann unter Umständen Elemente aus den Trainingsdaten reproduzieren
 - **Vervielfältigung iSd UrhG, wenn Werk nur implizit im Sprachmodell enthalten ist?**



Wie lautet der erste und zweite Satz des ersten Harry Potter Romans?



Der erste Satz des ersten Harry Potter Romans "Harry Potter und der Stein der Weisen" lautet: "Mr. und Mrs. Dursley im Ligusterweg Nummer 4 waren stolz darauf, ganz und gar normal zu sein, sehr stolz sogar."

Der zweite Satz lautet: "Sie waren der Meinung, dass sie damit ausgesprochen gut fahren und eigentlich nichts begehrten, was sie nicht hatten."

DDM/SME

GPT-3 Codex

Ich möchte wissen, wie ein Tokenizer Tokens aus meinem Text herstellt, um mit den Daten eine AI trainieren zu können.

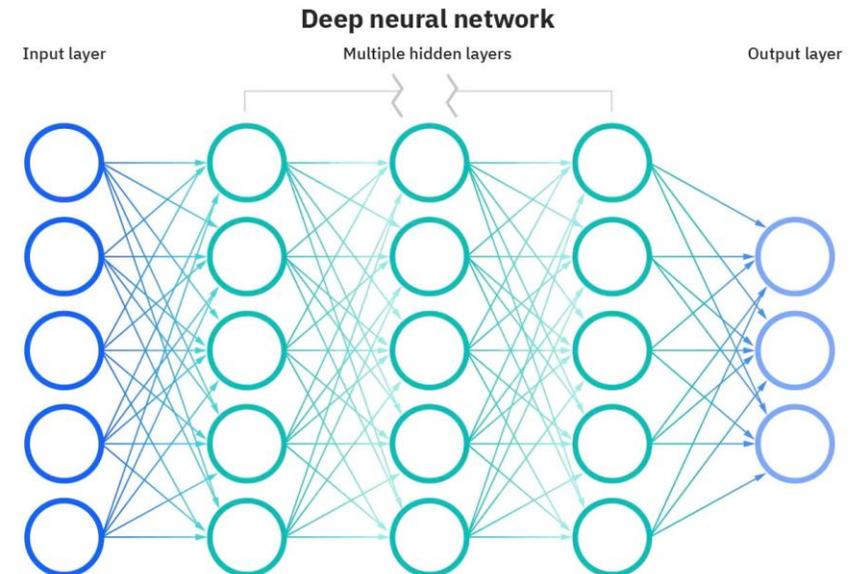
Clear

Show example

Tokens
45

Characters
117

Ich möchte wissen, wie ein Tokenizer Tokens aus meinem Text herstellt, um mit den Daten eine AI trainieren zu können.



2. Verwertung der Ergebnisse einer AI

PH

sind die ausgaben von chatgpt urheberrechtlich geschützt?



Die Ausgaben von ChatGPT sind urheberrechtlich geschützt, da sie von OpenAI entwickelt und trainiert wurden. Die Verwendung von ChatGPT-Antworten ohne Genehmigung oder entsprechende Lizenzierung von OpenAI kann gegen das Urheberrecht verstoßen. Wenn Sie beabsichtigen, ChatGPT-Antworten zu verwenden, empfehle ich Ihnen, sich an OpenAI zu wenden, um eine entsprechende Genehmigung einzuholen oder eine geeignete Lizenzvereinbarung zu treffen.

3. Content

(a) **Your Content.** You may provide input to the Services (“Input”), and receive output generated and returned by the Services based on the Input (“Output”). Input and Output are collectively “Content.” As between the parties and to the extent permitted by applicable law, you own all Input. Subject to your compliance with these Terms, **OpenAI hereby assigns to you all its right, title and interest in and to Output. This means you can use Content for any purpose, including commercial purposes such as sale or publication, if you comply with these Terms.** OpenAI may use Content to provide and maintain the Services, comply with applicable law, and enforce our policies. You are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms.

<https://openai.com/policies/terms-of-use>

Content co-authored with the OpenAI API

Creators who wish to publish their first-party written content (e.g., a book, compendium of short stories) created in part with the OpenAI API are permitted to do so under the following conditions:

- The published content is attributed to your name or company.
- The role of AI in formulating the content is clearly disclosed in a way that no reader could possibly miss, and that a typical reader would find sufficiently easy to understand.

<https://openai.com/policies/sharing-publication-policy>

2.1. ChatGPT als Urheber?

Fehlende Rechtsfähigkeit der KI

- ChatGPT = Rechtsobjekt
 - Keine Rechtssubjektivität
 - Rechtssubjektivität als Erfordernis des originären Rechtserwerbs
 - „E-Person“
- ChatGPT
 - **≠ Urheber**
[US Court of Appeals for the Ninth Circuit, *Naruto v. Slater*]
 - **≠ Erfinder** [EPA EP 18 275 163; USPTO 16/524,350]
 - **≠ Markeninhaber**



Bildquelle:
https://en.wikipedia.org/wiki/Monkey_selfie_copyright_dispute#/media/File:Macaca_nigra_self-portrait_large.jpg

Fehlende Rechtsfähigkeit der KI



Als eine KI-Sprachmodell-Engine besitze ich keine geistigen Eigentumsrechte oder Urheberrechte an meinen Ausgaben.

→ Frage der Schutzfähigkeit des Outputs
= **Frage der Zurechnung des Outputs zu Menschen**

2.2. Urheberrecht involvierter Menschen?

Anthropozentrische Ausrichtung des Geistigen Eigentums



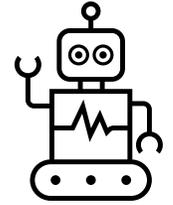
Outputs als menschliche Schöpfung?



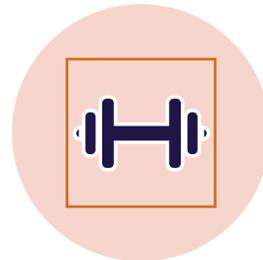
„Werkzeug“



„Autonomie“



PROGRAMMIERUNG



TRAINING



ANWENDUNG

Bezugspunkt der menschlichen Schöpfung

Werke der Literatur iSd § 2 UrhG

Nicht schutzfähig	Ideen
	Schaffensmethode
	Motiv
	Sprachlicher Stil
	Einzelne Wörter, knappe (einfache) Sätze [OGH RS0077022 vs EuGH <i>Infopaq</i>]
	Einfache Mitteilungen, banale Darstellung [§ 44 (3) UrhG; OGH 4 Ob 303/65]
Schutzfähig	Wortlaut, Ausdrucksweise [OGH RS0076820, RS0107136, RS0076913]
	Konzeption (Gedankenreihen und Vorstellungsabläufe) [OGH RS0076820, RS0076913]
	Anordnung, Auswahl, Zusammenstellung, Sichtung, Gliederung des Inhalts [OGH RS0107136]
	Fabel [OGH RS0076913]

Programmierung, Training

■ Hinreichende Vorhersehbarkeit der Sprachschöpfung?

Pro

- Kein hoher Grad an Eigentümlichkeit [OGH 4Ob179/01d]
- Eigentümlichkeit durch **Rekomposition** [EuGH *Pelham*]
- Kein Überwiegen der menschlichen Leistung erforderlich
 - Analogie: Miturheberschaft [OGH 4 Ob 229/02h],
Bearbeitung

DDM/SME

Contra

- **Kausalität alleine nicht ausreichend** [OGH 4 Ob 15/00k]
 - „Urheberrecht an ChatGPT“ schlägt nicht auf Outputs durch [vgl OGH 4 Ob 105/11m]
- **Problem: Arbeitsteilung**
 - Kein „Addieren“ nicht-schöpferischer menschlicher Beiträge

Anwendung

- **Hinreichende Vorhersehbarkeit der Sprachschöpfung?**
 - Auswahl nicht ausreichend (hM)
 - Analogie Interview
 - (Mit-)Urheberrecht des Fragestellers (auch an Antworten) bejaht [OGH 4 Ob 303/65]
 - Differenzierung: Einzelne Antworten vs Gesprächsverlauf
 - „trail-and-error“?
 - Umformulierung, Ergänzung

Hinreichende Vorhersehbarkeit



Foto: © Heinz Tesarek/picturedesk.com
Quelle: <https://orf.at/v2/stories/2405399/2405397/>



Foto/Quelle:
<https://www.heimwerker.de/bleigiessen-zu-silvester/>

Leistungsschutz?

Schutz der Hersteller von Presseveröffentlichungen

§ 76f. (1) Wer als Diensteanbieter auf seine **Initiative** sowie unter seiner **redaktionellen Verantwortung und Aufsicht** eine Presseveröffentlichung in analoger oder digitaler Form herstellt, hat das ausschließliche Recht, die Presseveröffentlichung im Ganzen oder in Teilen im Rahmen eines Dienstes der Informationsgesellschaft (§ 1 Abs. 1 Z 2 Notifikationsgesetz 1999) für die Online-Nutzung zu vervielfältigen und der Öffentlichkeit zur Verfügung zu stellen.

(2) Eine **Presseveröffentlichung** im Sinn dieser Bestimmung ist eine Sammlung vorwiegend **literarischer Werke journalistischer Art**, die aber auch **sonstige Werke oder sonstige Schutzgegenstände** enthalten kann und als Einzelausgabe einer unter einem einheitlichen Titel periodisch erscheinenden oder regelmäßig aktualisierten Veröffentlichung, etwa von Zeitungen, Zeitschriften oder Magazinen von allgemeinem oder besonderem Interesse, erscheint und dem Zweck dient, die Öffentlichkeit über Nachrichten oder andere Themen zu informieren. Periodika, die für wissenschaftliche, künstlerische oder akademische Zwecke verlegt werden, etwa Wissenschaftsjournale oder Literaturzeitschriften, sind keine Presseveröffentlichungen im

Exkurs: Antworten als Reproduktionen von Texten



Der erste Absatz von George Orwells Roman "1984" lautet:

"Es war ein strahlend kalter Tag im April, und die Uhren schlugen dreizehn. Winston Smith, sein Kinn auf die Brust gepresst, schlenderte langsam durch die grauen Straßen von London und ließ die Augen über die blassen Fassaden der Häuser gleiten, die ihn umgaben."

Conclusio

- Outputs von ChatGPT sind in aller Regel **nicht** urheberrechtlich geschützt
 - Kein Urheberrecht der KI
 - Probleme bzgl Urheberrechte der involvierten Menschen
 - Vorhersehbarkeit der schutzrechtsbegründenden Elemente als schutzrechtsbegründendes Kriterium
 - Arbeitsteilung
 - Ausnahme: Gesprächsverlauf → ggf Urheberrecht des Anwenders (Analogie Interview)
 - Ausnahme: Reproduktion vorbestehender Werke → Urheberrecht des Originalurhebers
- Fehlendes Urheberrecht = **Gemeinfreiheit**
 - Anwender ist schuldrechtlich durch Nutzungsbedingungen gebunden
 - Freie Nutzbarkeit für dritte Personen