

Gerechte Rahmenbedingungen für Text- und Data-Mining

Autor:innen und Übersetzer:innen fordern Nutzungsvorbehalt und Vergütungspflicht im Text- und Data-Mining.

Es ist nicht hinnehmbar, dass Urheber:innen, durch deren Leistungen „wettbewerbsfähige“ KI-Produkte überhaupt erst entwickelt werden können, rechtlich und wirtschaftlich deutlich schlechter gestellt werden als die Verreiber und Verwerter jener Produkte, die aus ihren Werken entstehen.

Berlin, 26. Februar 2021

Im Hinblick auf die geplanten Regelungsvorschläge zum **Text- und Data-Mining (TDM)** fordert das Netzwerk Autorenrechte eine dringend gebotene Korrektur der bisher nicht ausgestalteten Vergütungspflicht, außerdem die Erleichterung eines Widerspruchs gegen die Nutzung der eigenen Daten (opt-out). Der politische Wille stellt ansonsten menschliche Urheber:innen, die die Basisleistung für die imitativen und inhaltlich unreflektierten Programme liefern (so genannte Schwache Künstliche Intelligenz, KI), bewusst schlechter als jene, die von diesen Leistungen profitieren sollen.

In § 44b Abs. 1 UrhG (neu) wird TDM derzeit als automatisierte Analyse von Werken definiert, um „Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“.

TDM ist ein relativ junges Konzept und entspringt Marktforschungsstrategien von 1989/1990, um eine „Eins-zu-Eins“-Beziehung zu einem potenziellen Kunden herzustellen. Entsprechend erleichtert § 44b Abs 1 KI-basierte Marktforschung für Target-Werbung. Dies kann z.B. das automatische inhaltliche und damit rechtswidrige Auswerten öffentlich zugänglicher urheberrechtlich geschützter Texte umfassen, um auf Basis von deren semantischen und inhaltlichen Musterwiederholungen Wirtschafts- und Werbungsprodukte herzustellen (wie etwa Read-O)¹.

Die als Umsetzung der Richtlinie (EU) 2019/790 („DSM-RL“) vorgeschlagenen Nutzungsmöglichkeiten von urheberrechtlich geschütztem Material im Rahmen des TDM gehen über den reinen Informationsgewinn längst weit hinaus. Zusätzlich ist auch für die Bereiche, in denen eine Vergütungspflicht nach der DSM-RL möglich ist (Erwägungsgrund 17), im aktuellen Kabinettsentwurf keine Vergütung vorgesehen, trotz der Möglichkeit, eine solche zu schaffen.

§ 44 Abs2 UrhG (neu) besagt zudem: „Zulässig sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das TDM. Die Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.“ Um Textwerke auszuwerten, muss jedoch ein maschinenlesbarer Korpus hergestellt werden. Dies geschieht mittels Digitalisierung und Vervielfältigung der Werke. Die massenhafte Nutzung von Textwerken erhöht das Risiko einer illegalen Beschaffung. Die Aufbewahrung von Kopien dürfte kaum zu kontrollieren, das Löschen nicht zu gewährleisten sein. Im Bereich des wissenschaftlichen Verlegens stellte u.a. der Bundesverband der Digitalpublisher bereits eine deutliche Zunahme der Online-Piraterie fest, ebenso das vermehrte Aufkommen sogenannter Schattenbibliotheken².

¹ <https://read-o.com>, abgerufen am 24.2.2021

² https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Stellungnahmen/2020/Downloads/110620_Stellungnahme_BDZV-VDZ_RefE_Urheberrecht-ges.pdf?__blob=publicationFile&v=2, abgerufen am 24.2.2021

Auch für diesen Kopiervorgang braucht es demzufolge unbedingt (a) eine Vergütungspflicht, (b) ein Opt-Out/Nutzungsvorbehalt und (c) eine nachvollziehbare Kontrollmöglichkeit über den Verbleib der Vervielfältigungen sowie klare Regeln der „Erforderlichkeit“, selbst wenn diese Bibliotheken, Museen und Archiven zur Aufbewahrung zur Verfügung gestellt werden.

In der Neufassung zeigt sich der politische Wille der deutschen Gesetzgeberin, zum vermeintlichen „Wohl der Gemeinschaft“ und der Forschung Urheber:innen in einen fortgesetzten und auf Dauer höchst schädlichen Nachteil mittels einer vergütungsfreien Schranke zu setzen. Gleichzeitig verkennt die Gesetzgeberin die wirtschaftlichen Folgen bereits aktiver Anwendungsgebiete von Produkten aus TDM und deren destruktive Auswirkungen auf die Zukunft der Urheber:innen.

Erläuterungen und Hintergrund

Seit Jahren sind lukrative Wirtschaftsprodukte im Einsatz, die aus urheberrechtlich geschützten Werken „gelernt“ haben und bestimmte Urheberleistungen nun allmählich ersetzen:

- (a) die automatisierte Erstellung von **Texten** (Börsennachrichten, Sport- und Wetternachrichten, journalistische Texte, automatisierte Kommentare und chat bots, die Kundengespräche oder über fake accounts in den sozialen Medien Meinungen simulieren, Polizeinachrichten, Unternehmenskommunikation);
- (b) die automatisierte Erstellung von **Übersetzungen** (Maschinelles Übersetzen MÜ; Suchergebnisse, Bücher, Gesetzestexte, Produktbeschreibungen, Übersetzungen auf Google, Facebook, Twitter);
- (c) imitative **Bild- und Videowerke** (deep fake³);
- (d) die automatisierte Erstellung von Melodien und kürzeren Musikwerken (**Werbejingles**);
- (e) sowie **Text-Analyseinstrumente** („Dictionaries“), die bspw. Wirtschaftsunternehmen von Universitäten gegen Gebühr zur Verfügung gestellt werden, um Verbesserungen von Stil, Wertausdruck, Text und Rhetorik für Produktbeschreibungen, Marketinginformationen, öffentliche Reden usw. vorzunehmen – oder sogar zum Vorsortieren und Aussieben von Literatur dienen (QualiFiction⁴).

Die Nutzbarkeit urheberrechtlich geschützter Werke für TDM stellt, im Hinblick auf die Entwicklung von Produkten künstlicher Intelligenz, folglich Lernmaterial und damit die wichtigste Grundlage für die eigentlich nicht sonderlich intelligenten, aber rechenstarken KI-Anwendungen und Software im Text-, Bild- und Musikbereich dar. Die Forschung arbeitet konsequent der Wirtschaft zu oder wird von dieser ohnehin finanziert. Gleichzeitig sind die Auswirkungen auf den Buch-, Bild- und Musiksektor absehbar, wo KI-Produkte schon jetzt zur kommerziellen Konkurrenz erwachsen – Konkurrenz für genau jene, durch deren kreative Leistungen sie nach dem derzeitigen Wunsch der Gesetzgeberin gratis ausgebildet werden sollen.

Die Enquete-Kommission Künstliche Intelligenz hat in ihrem derzeitigem Entwurf deutlich formuliert, dass urheberrechtlich geschützte Werke die Grundlage einer bedeutenden wirtschaftlichen Ausschöpfung seien: „Der Zugang zum Rohstoff Daten für die Anwendung von KI beeinflusst mithin also die *Wettbewerbssituation* auf digitalen Märkten“⁵ und betont: „Umfangreiche Regelungen und damit verbundene rechtliche Unklarheiten können den Zugang zu Daten zwecks wissenschaftlicher und

³ <https://mixed.de/geschichte-der-deepfakes-so-rasant-geht-es-mit-ki-fakes-voran/>, abgerufen am 26.2.2021

⁴ <https://www.qualifiction.info>, abgerufen am 26.2.2021

⁵ <https://dip21.bundestag.de/dip21/btd/19/237/1923700.pdf>, Seite 71, abgerufen am 12.2.2021,

wirtschaftlicher Nutzung erschweren, diese sind aber Voraussetzung für eine *wettbewerbsfähige* Anwendung des Maschinellen Lernens.“⁶

In dem fast 800-seitigen Reportentwurf lesen wir, wie Autor:innen, Übersetzer:innen und Urheber:innen gewissermaßen zu Organspender:innen der Wirtschaft werden: „Die im Einsatz entstehenden Daten und daraus mit Maschinellern erzeugten Modelle sind dann wiederum Teil der Datenökonomie, *können also ihrerseits weitere Einnahmequellen darstellen. KI-Systeme wie die maschinelle Übersetzung sind heute schon ein wichtiger Baustein der Globalisierung, wie man bei den großen Internetplattformen beobachten kann.* Die Sprachgrenzen sind dennoch im europäischen Binnenmarkt noch immer spürbar, und auch für KI und Datenverarbeitung stellt Vielsprachigkeit noch immer eine Herausforderung dar, wie etwa beim Erzeugen von Nutzermodellen im E-Commerce. *Sprachbarrieren werden teils als die größten Handelshemmnisse beschrieben. Hier kann die Wirtschaft von einer konsequenten Digitalisierung stark profitieren. Maßnahmen zur Reduzierung von Sprachbarrieren sind bereits in der Erprobung und sollten auch auf politischer Ebene bzw. von staatlicher Seite vorangetrieben werden, z. B. Übersetzung von Kundenanfragen und Produktbeschreibungen in leicht verständliche Sprache.*“⁷

Rechtliche Folgen

Aktuelle Bestrebungen der EU-Kommission zielen darauf ab, dass „technische Schöpfungen, die durch KI-Technologie erzeugt werden, im Rahmen des Rechtsrahmens für geistiges Eigentum geschützt werden müssen, um Investitionen in diese Form der Schöpfung zu fördern“⁸. Dies lässt die Deutung zu, dass jene Produkte, die aus gesetzlich unvergütetem TDM entstehen, ein Schutzrecht genießen sollen, während die menschlichen „Minen“, aus denen das Gold der Datenökonomie von Morgen geschürft wird, praktisch von der Gesetzgeberin im Vorfeld enteignet werden.

Urheber:innen: Bürger:innen zweiter Klasse?

Die Antwort lautet: nein! Entsprechend ist es in diesem vorliegenden Entwurf des Kabinetts nicht hinnehmbar, dass die Urheber:innen, durch deren Leistungen „wettbewerbsfähige“ KI-Produkte überhaupt erst entwickelt werden können, rechtlich und wirtschaftlich deutlich schlechter gestellt werden als die Verwerter und Vertreiber sowie jene Produkte, die aus ihren Werken entstehen. Schwache Künstliche Intelligenzen, ihre Analyse- und automatisierten Programme sind zukunftsweisende Technologien. Trotz ihres Nutzens und ihrer Vorteile kann man die weitere Entwicklung der Technologie nicht ausschließlich unter ethischen Aspekten durchleuchten und überwachen, sondern muss auch sicherstellen, dass die Entwicklung nicht nur wirtschaftliche Interessen einiger weniger bedient. In einer liberalen Marktwirtschaft sollte, ja *darf* es nicht erlaubt sein, Produkte von einer bestimmten Gruppe (hier Autor:innen, Übersetzer:innen und weitere schaffende Urheber:innen) kostenlos anderen zur Verfügung zu stellen, damit diese daraus einen Gewinn erzielen können. Dieser Sachverhalt kommt einer Enteignung gleich und zerstört die Basis des Systems einer fair konzipierten Gesellschaft, die unter dem Aspekt der Gleichbehandlung ethische wie demokratische Wertmaßstäbe entwickelt.

⁶ <https://dip21.bundestag.de/dip21/btd/19/237/1923700.pdf>

⁷ <https://dip21.bundestag.de/dip21/btd/19/237/1923700.pdf>, Seite 225, abgerufen am 16.02.2021

⁸ https://www.europarl.europa.eu/doceo/document/A-9-2020-0176_EN.pdf

Dies ist der Moment für Sie als Entscheidungsträger:innen, Weichen für die Zukunft zu stellen, die integer und gerecht gegenüber jener Leistung sind, die jegliche Künstliche Intelligenz erst möglich macht: Die menschliche Intelligenz.

Entsprechend fordern wir Autor:innen und Übersetzer:innen einen Nutzungsvorbehalt und eine Vergütungspflicht im TDM. Wir haben einen Anspruch auf eine angemessene Vergütung, genauso wie es die Verpächter landwirtschaftlicher Nutzflächen haben.

Wir stehen für Rückfragen und Erläuterungen zur Verfügung.

NAR Arbeitsgruppe Künstliche Intelligenz & Urheberrecht

*Carola Christiansen (Präsidentin Mörderische Schwestern e.V.), Janet Clark (Politische Beauftragte Mörderische Schwestern e.V.), Carlos Collado Seidel (PEN Zentrum Deutschland), Lena Falkenhagen (Bundesvorsitzende Verband deutscher Schriftstellerinnen und Schriftsteller), Carlos Fortea (Asociación Colegial de Escritores), Nina George (Präsidentin European Writers' Council), Tobias Kiwitt (Vorsitzender Bundesverband junger Autorinnen und Autoren), Gino Leineweber (EXIL-PEN), Tamara Leonhard (Geschäftsführerin Selfpublisher-Verband), Sabine Lipan (Bundeskongress Kinderbuch), Werner Richter (Vorsitzender IG Übersetzerinnen Übersetzer), Astrid Vehstedt (PEN Zentrum Deutschland).
IT-Fachreferentin: Patrizia Prudenzi (SYNDIKAT).*

Das Netzwerk Autorenrechte (NAR) (www.netzwerk-autorenrechte.de) repräsentiert 14 Verbände und über 15.500 AutorInnen und ÜbersetzerInnen aus Deutschland, Österreich und der Schweiz. Mitglieder: 42erAutoren, A*dS Autorinnen und Autoren der Schweiz, Autorinnenvereinigung e.V., Bundesverband junger Autoren und Autorinnen (BVJA), Bundeskongress Kinderbuch, IG Autorinnen Autoren, Mörderische Schwestern e.V., Phantastik-Autoren-Netzwerk (PAN) e.V., PEN-Zentrum Deutschland, PEN-Zentrum deutschsprachiger Autoren im Ausland, Selfpublisher-Verband e.V., SYNDIKAT – Verein für deutschsprachige Kriminalliteratur, Verband deutschsprachiger Übersetzerinnen und Übersetzer literarischer und wissenschaftlicher Werke e.V. (VdÜ), Verband deutscher Schriftstellerinnen und Schriftsteller (VS) in ver.di.
Kontakt: info@netzwerk-autorenrechte.de