

Tim W. Dornis | Sebastian Stober

Urheberrecht und Training generativer KI-Modelle

Technologische und juristische Grundlagen



Nomos

Recht und Digitalisierung | Digitization and the Law

Herausgegeben von | Edited by

Prof. Dr. Roland Broemel

Jun.-Prof. Dr. Lea Katharina Kumkar

Prof. Dr. Jörn Lüdemann

Prof. Dr. Rupprecht Podszun

Prof. Dr. Heike Schweitzer[†]

Band 19 | Volume 19

Tim W. Dornis | Sebastian Stober

Urheberrecht und Training generativer KI-Modelle

Technologische und juristische Grundlagen



Nomos

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage 2024

© Tim W. Dornis | Sebastian Stober

Publiziert von
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-7560-2305-9

ISBN (ePDF): 978-3-7489-4955-8

DOI: <https://doi.org/10.5771/9783748949558>



Onlineversion
Nomos eLibrary



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Über das Gutachten

Dieses Gutachten wurde erstellt im Auftrag der Initiative Urheberrecht (Authors' Rights Initiative).

Es handelt sich um eine interdisziplinäre Analyse durch Sebastian Stober und Tim W. Dornis. Die Untersuchung zu den technologischen Grundlagen (§ 2) wurde von Sebastian Stober verfasst. Tim W. Dornis verantwortet die Teile der Untersuchung zu den juristischen Fragen und Zusammenhängen (§§ 3, 4, 5 und 6).

Die Inhalte des Gutachtens geben ausschließlich die Ansichten und Einschätzungen der Autoren wieder. Die Autoren danken Julia Danevitch, Friederike Günther, Hendrik Meier und Pascal Turan Sierek herzlich für Unterstützung bei der Recherche sowie kritische Anmerkungen.

Über die Autoren

Sebastian Stober ist Professor für Künstliche Intelligenz an der Otto-von-Guericke-Universität Magdeburg.

Tim W. Dornis ist Professor für Bürgerliches Recht und Gewerblichen Rechtsschutz an der Leibniz Universität Hannover sowie Global Professor der NYU School of Law.

Über die Initiative Urheberrecht

Die Initiative Urheberrecht vertritt über ihre 45 Mitgliedsorganisationen die Interessen von rund 140.000 Urheber:innen und ausübenden Künstler:innen in den Bereichen Belletristik und Sachbuch, Bildende Kunst, Journalismus, Design, Musik, audio- und audiovisuelle Medien, Fotografie, Illustration, Schauspiel, Spieleentwicklung, Tanz und vielen mehr.

<https://urheber.info/wir>

Inhaltsverzeichnis

Executive Summary	13
§ 1. Einleitung und Übersicht	17
§ 2. Technologische Grundlegung	23
A. Maschinelles Lernen	23
I. Lernaufgaben	23
II. Parameter und Hyper-Parameter	25
III. Generalisierung und Modellkapazität	26
IV. Datenaugmentierung	27
B. Künstliche Neuronale Netze (KNNs)	27
I. Aufbau und Struktur	29
II. Embeddings und latent space	31
III. Training von KNNs	33
IV. Pre-Training und Fine-Tuning	35
V. Weiterverwendung von trainierten Modellen und catastrophic forgetting	36
VI. Reproduzierbarkeit eines Trainingsvorgangs	37
C. Generative KI-Modelle	39
I. Technische Grenzen der Trainierbarkeit	40
II. Lösung: Approximation	41
III. Large language models (LLMs) – Autoregressive Modelle	43
IV. Generative Adversarial Networks (GANs)	44
V. Variational Autoencoders (VAEs)	45
VI. Diffusionsmodelle	48
VII. Sampling und Konditionierung	49
VIII. Style transfer	51
D. Technische Betrachtungen zu Fragen des Urheberrechts	54
I. Datensammlung: Webscraping und Erstellung von Korpora zum Training	54
II. TDM: Anknüpfungspunkte und Abgrenzung	57

III. Datenverarbeitung und potentiellies Memorisieren beim Training	59
E. Ausblick	63
§ 3. Inhalt des Urheberrechts	67
A. Überblick	67
B. Verletzungshandlungen	68
I. Sammlung, Vorbereitung und Speicherung der Trainingsdaten	69
II. Training eines generativen KI-Modells	71
1. Meinungsstand	72
2. Korrekturen	74
a) Ausgangspunkt: Speicherung „im Innern“	74
b) Nicht gefordert: Unmittelbare Wahrnehmbarkeit	75
c) Method in flux: „Wahrnehmbarmachung“	77
3. Zwischenergebnis	79
III. Genese von KI-Output	80
1. Vervielfältigung und Bearbeitung: Entscheidung nach Einzelfall	80
2. Caveat: Keine „künstliche Doppelschöpfung“	83
3. Zwischenergebnis	85
§ 4. Schranken des Urheberrechts und Einwilligung	87
A. Überblick	88
B. Vorübergehende Vervielfältigungshandlungen während des KI-Trainings (§ 44 a UrhG, Art. 5 Abs. 1 InfoSoc-Richtlinie)	88
I. Vervielfältigung ist „vorübergehend“ sowie „flüchtig und begleitend“	89
II. Vervielfältigung ist „integraler und wesentlicher Teil eines technischen Verfahrens“	90
III. Vervielfältigung hat den Zweck einer „rechtmäßigen Nutzung“	91
IV. Vervielfältigung hat „keine eigenständige wirtschaftliche Bedeutung“	91
V. Zwischenergebnis	93

C. Text und Data Mining für Zwecke der wissenschaftlichen Forschung (§ 60d UrhG, Art. 3 DSM-Richtlinie)	94
D. Text und Data Mining (§ 44b UrhG, Art. 4 DSM-Richtlinie)	94
I. Zentrales Tatbestandsmerkmal: „Text und Data Mining“	95
1. Grundlegung: Syntax als Schutzgegenstand des Urheberrechts	96
2. Fehlkonzeption: Gleichsetzung von TDM und Training generativer KI-Modelle	98
3. Korrekturen: Wortlaut, System und Telos	101
a) Wortlaut	101
b) Systematisch-konzeptionelle Unanwendbarkeit der TDM-Schranke	103
aa) Klarstellung: Begriffliche Verwirrung um das „right to mine“	103
bb) Kritik und Analyse	105
(1) Missverständnis: Scheinbares Aussieben der Syntax	106
(2) Abgrenzung: TDM und Training generativer KI-Modelle	109
(a) Semantik/Syntax-Agnostik der Technologie	110
(b) Quantität der Datenextraktion: large-scale data extraction	110
(c) Technologieimmanenz: „copy expression for expression’s sake ...“	112
(d) Schließlich: Syntax-Repräsentation im Vektorraum	117
c) Zwischenergebnis	120
4. Historische Auslegung	121
a) Status: DSM-Richtlinie als Relikt der KI-Steinzeit	122
aa) Europäisches Gesetzgebungsverfahren	122
bb) Nationaler Gesetzgeber	125
b) Methodisches Axiom: Vorrang des objektiven Telos	127
c) Zwischenergebnis	129

5. Exkurs: Auslegung im Lichte der KI-Verordnung	129
a) Meinungsstand	129
b) Korrekturen	131
aa) Wortlaut	131
bb) System	132
cc) Telos	133
c) Zwischenergebnis	134
II. Weitere Tatbestandsmerkmale	135
1. „Vervielfältigungen“, nicht „Änderungen“	135
2. „Rechtmäßig zugängliche Werke“	136
3. „Erforderlichkeit“ und Löschpflicht	137
III. Opt-out des Rechteinhabers: Nutzungsvorbehalt	139
1. Verbot von Förmlichkeiten (Art. 5 Abs. 2 RBÜ)	139
2. Praktische Hindernisse	140
3. Scheitern der Abwicklung durch private ordering	141
IV. Schließlich: Unvereinbarkeit mit Dreistufentest	142
1. Meinungsstand	142
2. Kritik und Analyse	143
a) Überblick	143
b) Unheilbare Disruption: „Beeinträchtigung“ der „normalen Auswertung“	146
aa) Empirisch-quantitative und normative Perspektive	146
(1) Internationales Urheberrecht	147
(2) Europäisches Urheberrecht	149
bb) Anwendung: Wettbewerb durch generative KI-Modelle	150
cc) Exkurs: Opt-out und Beeinträchtigung der normalen Auswertung	152
c) Abwägung: „Ungebührliche Verletzung berechtigter Interessen“	153
aa) Relevante Interessen	153
bb) Anwendung: KI-Training	154
3. Zwischenergebnis	155
V. Intertemporalität und Schrankenlosigkeit: Training vor dem 7. Juni 2021	155

E. Sonstige gesetzliche Schranken und Einwilligungsfiktion	157
I. Vorübergehende Vervielfältigungshandlungen bei KI-Einsatz (§ 44a UrhG), Zitate (§ 51 UrhG) und unwesentliches Beiwerk (§ 57 UrhG)	157
II. Karikatur, Parodie und Pastiche (§ 51a UrhG)	158
III. Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch (§ 53 UrhG)	159
IV. Vorschaubilder-Doktrin des Bundesgerichtshofs: „Fingierte Einwilligung“	160
§ 5. Anwendbares Recht, internationale Zuständigkeit und Extraterritorialität	161
A. Internationales Privatrecht	161
I. Territorialitätsprinzip und Verletzungshandlung „im Inland“	162
II. KI-Training im Ausland und forum shopping	163
III. Blinder Fleck in der Diskussion: Öffentliche Zugänglichmachung nach §§ 15 Abs. 2 Nr. 2, 19a UrhG	164
B. Internationale Zuständigkeit und praktische Durchsetzung	165
I. Gerichtsstände der EuGVVO	165
II. Deliktgerichtsstand nach § 32 ZPO	166
III. Zwischenergebnis: Praktische Durchsetzung	167
C. KI-Verordnung: Mittelbare Durchsetzung des Urheberrechts	167
I. Vorab: Kategorisierung der KI-Modelle und -Systeme	167
II. Persönlicher und sachlicher Anwendungsbereich der KI-Verordnung	169
III. Schnittstelle: KI-Verordnung und Urheberrecht	170
IV. Streitpunkt: Räumliche Reichweite des europäischen Urheberrechts	172
V. Durchsetzung: Rechtsfolgen von Pflichtverstößen	174
VI. Zwischenergebnis	176
§ 6. Anschlussfragen	179
A. Diskussionsstand: Herrschende Narrative	179
I. Einzigartigkeit der menschlichen Kreativität	180

II. Schöne neue Welt unendlich gesteigerter Kreativität	180
III. Urheberrecht vs. KI-Innovation	182
B. Klarstellung #1: Farewell to human exceptionalism	183
C. Klarstellung #2: Ersticken im Überfluss algorithmisch recycelter Kreativität	186
D. Klarstellung #3: KI-Innovation vs. race to the bottom	189
§ 7. Zusammenfassung der Ergebnisse	193
Anhänge I bis IV	197
Literaturverzeichnis	203