

Jonathan Pukas

Werknutzung und künstliche Intelligenz

Analyse und Perspektiven der urheberrechtlichen Rahmenbedingungen für die Verwendung von Werken als Trainings- und Inputdaten künstlicher neuronaler Netze



Nomos

Datenrecht und neue Technologien

Herausgegeben von

Prof. Dr. Matthias Cornils

Prof. Dr. Martin Ebers

Prof. Dr. Mario Martini

Prof. Dr. Dr. Frauke Rostalski

Prof. Dr. Giesela Rühl

Prof. Dr. Björn Steinrötter

Prof. Dr. Domenik H. Wendt

Band 14

Jonathan Pukas

Werknutzung und künstliche Intelligenz

Analyse und Perspektiven der urheberrechtlichen Rahmenbedingungen für die Verwendung von Werken als Trainings- und Inputdaten künstlicher neuronaler Netze



Nomos

Open Access-Stellung gefördert durch den Publikationsfond der Martin-Luther-Universität Halle-Wittenberg. Gedruckt mit freundlicher Unterstützung der Studienstiftung ius vivum.

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

Zugl.: Halle-Wittenberg, Universität, Diss., 2025

1. Auflage 2025

© Jonathan Pukas

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): 978-3-7560-3383-6

ISBN (ePDF): 978-3-7489-6346-2

DOI: <https://doi.org/10.5771/9783748963462>



Onlineversion
Nomos eLibrary



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

Vorwort

Die vorliegende Dissertation ist im Zeitraum zwischen September 2021 und Januar 2025 entstanden. Literatur und Rechtsprechung ist ebenfalls bis Anfang Januar 2025 berücksichtigt. Denjenigen, die die Entwicklungen rund um künstliche Intelligenz in den letzten Jahren verfolgt haben, können erahnen, dass die Arbeit zwischen Beginn und Ende des Forschungszeitraums einen „KI-Quantensprung“ durchlebt hat. Die Aktualität des Themas habe ich stets als Chance begriffen, auch wenn sie mir sicherlich das ein oder andere Mal Kopfschmerzen bereitete.

Umso dankbarer bin ich, dass mich meine Familie und Freunde stets liebevoll und unterstützend durch die Promotionsphase begleitet haben. Danken möchte ich insbesondere meinen Eltern sowie auch meiner Partnerin *Hannah* für das Abfedern besonders vieler dieser Kopfschmerzen in der Endphase der Bearbeitung der Dissertation. Danken möchte ich darüber hinaus meinem Doktorvater *Prof. Dr. Malte Stieper*, der mich nicht nur während meiner Promotion betreut, sondern in mir seit Beginn des Studiums den kritischen Wissenschaftsgeist geformt hat, auf den ich heute zurückgreifen kann. Die Zeit als Hilfskraft am Lehrstuhl, in der gewissermaßen das akademische Fundament dieser Arbeit gelegt worden ist, habe ich sehr genossen. Mein besonderer Dank gilt daneben *Prof. Dr. Jan Bernd Nordemann* nicht nur für die zügige Erstellung des Zweitgutachtens, sondern vor allem für die stetige Förderung und das Begleiten meines Weges in den vergangenen Jahren. Ich danke außerdem der Stiftung der Deutschen Wirtschaft für die Aufnahme in ihr Promovierendenförderprogramm sowie der Studienstiftung *ius vivum* für die Gewährung eines großzügigen Druckkostenzuschusses. Mein Dank gilt außerdem dem Open Science Team der Universitäts- und Landesbibliothek Sachsen-Anhalt, die die Open Access-Veröffentlichung der Arbeit großzügig ermöglicht haben.

In inhaltlicher Hinsicht sollte die Rechtswissenschaft gerade bei der Einordnung zukunftsgewandter Problemfelder wie der Entwicklung und dem Einsatz künstlicher Intelligenz den Mut haben, den weiten Möglichkeitsraum der Rechtsentwicklung anzuerkennen und an dessen Ausgestaltung mitzuwirken. Rechtswissenschaft ist auch Rechtssetzungswissenschaft. Aus diesem Grund habe ich den Blick der vorliegenden Untersuchung auch auf die regulatorischen Bedürfnisse und die *lex ferenda* des Urheberrechts gelenkt. Allerdings möchte ich die Gelegenheit dieses Vorwortes nicht verstreichen lassen, um zu betonen, dass die vorliegende Arbeit keine

Rechtspolitik zum Zweck hat. Ganz im Gegenteil war ich stets bemüht, mit Blick auf die Möglichkeiten der Rechtssetzung einen Mechanismus zu entwickeln, der die regulatorischen Entscheidungen weg von rechtspolitischen Erwägungen hin zu einer Umgebung führt, die objektiv und in der juristischen Methodik bereits verankert ist. Mit mir allergrößter Sorgfalt habe ich die widerstreitenden Interessen analysiert und in Abwägung zueinander gestellt – teils auch unter Revision eigener, zuvor gefestigter Annahmen. Ich hoffe, mit der vorliegenden Arbeit einen fundierten Beitrag zur Rechtssetzungswissenschaft als Teil der Rechtswissenschaft leisten zu können.

De lege lata bleibt das Urheberrecht rund um die Entwicklung und den Einsatz künstlicher Intelligenz weiter im Fluss. Ein jüngst angestregtes Vorlageverfahren am Europäischen Gerichtshof (C-250/25) zu den auch in dieser Arbeit beleuchteten Fragestellungen gibt nur eine leise Vorahnung: Die Geschichte ist „to be continued“.

Inhaltsverzeichnis

Erster Teil: Abgrenzung des Untersuchungsgegenstands und dessen technische Grundlagen	23
A. Einleitung	23
B. Systeme künstlicher Intelligenz als Forschungsgegenstand	26
I. KI-Begriff der Untersuchung	26
II. Technische Grundlagen des maschinellen Lernens auf Basis von KNN	27
1. Maschinelles Lernen	27
2. Struktur und Arbeitsweise eines künstlichen neuronalen Netzes	28
3. Das Lernen: Training des künstlichen neuronalen Netzes	30
C. Forschungsfrage und Ziel der Arbeit	32
D. Gang der Untersuchung	33
Zweiter Teil: Regelungsbedürfnisse im Urheberrecht für die Verwendung von Werken als Trainings- und Inputdaten	35
A. Herleitung der Beurteilungskriterien und geeigneter Freistellungsmechanismen	35
I. Beurteilungskriterien für die Zweckmäßigkeit einer Freistellung	36
1. Der Schutz der Urheberinteressen als Hauptzweck des Urheberrechts	36
a) Die Urheberinteressen	36
b) Grundrechtlicher Schutz der Urheberinteressen	37
c) Schlussfolgerung: Freistellung wegen Unerheblichkeit für die Urheber	39
2. Freistellungen zum Ausgleich von Nutzer- und Urheberinteressen	39
a) Anknüpfungspunkt des Interessenausgleichs	39
aa) Grundrechte der Nutzer	39

bb) Berücksichtigung von Nutzerinteressen als sekundäres Regelungsziel des europäischen Urheberrechts	41
cc) Sozialbindung des Urheberrechts	42
b) Funktion des Ausgleichsmechanismus	43
3. Berücksichtigungsfähigkeit einer ökonomischen Analyse des Urheberrechts	45
a) Überblick über die Funktionsweise	45
b) Eignung der ökonomischen Analyse des Rechts für die Beantwortung der aufgeworfenen Fragestellungen	46
4. Techniksensibilität des Urheberrechts oder der „more technological approach“	50
5. Zusammenfassung: Maßstab für die regulatorische Analyse	51
II. Bestimmung des Freistellungsmechanismus	51
1. Zur Verfügung stehende Freistellungsmechanismen	51
a) Unvergütete Schutzbereichsbegrenzung: Unmittelbar begrenzende Ausgestaltung der Verwertungsrechte und vergütungsfreie Schrankenbestimmung	51
b) Gesetzliche Lizenz	52
c) Erweiterte kollektive Lizenz	53
d) Zusammenfassung	54
2. Entscheidungskriterien bei der Wahl zwischen den Freistellungsmechanismen	54
a) Angemessener Ausgleich für die Nutzungshandlung	54
b) Darlegungs- und Beweislast bei vergütungsfreier Nutzung	55
c) Schutz eingesetzter technischer Schutzmaßnahmen via § 95a UrhG	57
d) Notwendigkeit eines Opt-out-Mechanismus	60
e) Erhöhtes Bedürfnis nach Rechtssicherheit für Werknutzer	61
f) Weitere Aspekte und Zusammenfassung	62
B. Speicherung von urheberrechtlich geschützten Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern	63
I. Zweckmäßigkeit einer Freistellung	64
1. Beeinträchtigung der Urheberinteressen	64

2. Abwägung der Urheberinteressen mit den Interessen der Trainingsdatensammler	65
a) Zu berücksichtigende Trainingsdatensammler- und Allgemeininteressen	65
b) Beeinträchtigungsintensitäten der gegenläufigen Interessen	67
c) Erwägungen gegen eine Freistellung	68
d) Erwägungen zugunsten einer Freistellung	69
aa) Verhältnis zwischen der Speicherung zum Training und den Erzeugnissen des KI-Systems	69
bb) Kompensationswirkung einer Vergütungspflicht	71
cc) Art der Werkverwendung	71
dd) Innovationsbegünstigte und Bedeutung des Innovationsinteresses	72
ee) Praktische Erwägungen	73
ff) Ökonomischer Vorteil aus einer Freistellung	74
e) Schlussfolgerungen	75
II. Freistellungsmechanismus	75
1. Angemessener Ausgleich für die Nutzungshandlung	76
a) Rechtfertigungsansätze für eine vergütungsfreie Nutzung	76
b) Argumente für eine Vergütungspflicht	78
c) Schlussfolgerung	81
2. Notwendigkeit eines Opt-out-Mechanismus	82
a) Vorteile eines Opt-out-Mechanismus	82
b) Praktische Schwierigkeiten und weitere Nachteile eines Opt-out-Mechanismus	83
c) Schlussfolgerungen	84
3. Erhöhtes Rechtssicherheitsbedürfnis	85
4. Schlussfolgerung für den Freistellungsmechanismus	87
III. Unterscheidung zwischen kommerzieller und nicht-kommerzieller Tätigkeit	87
IV. Zwischenergebnis	89
C. Durchführung des Trainings auf Basis urheberrechtlich geschützter Trainingsdaten	90
I. Zweckmäßigkeit der Freistellung	91
II. Freistellungsmechanismus	91
1. Angemessene Vergütung der Urheber	92

2. Schutz technischer Schutzmaßnahmen über § 95a UrhG	94
3. Vergleich zum Werkgenuss	94
III. Zwischenergebnis	95
D. Speicherung von urheberrechtlich geschützten Inputdaten	95
I. Freistellungsbedürfnis	96
1. Beeinträchtigung der Urheberinteressen	96
2. Abwägung der Urheber- und Nutzerinteressen	97
3. Schlussfolgerung	99
II. Freistellungsmechanismus	99
1. Notwendigkeit eines angemessenen Ausgleichs	99
2. Schutz eingesetzter technischer Schutzmaßnahmen	101
III. Zwischenergebnis	103
E. Vorverarbeitung urheberrechtlich geschützter Inputdaten durch Normalisierung und Feature Encoding	104
I. Normalisierung der Inputdaten	104
1. Zweckmäßigkeit der Freistellung	104
2. Freistellungsmechanismus	106
3. Zwischenergebnis	107
II. Feature Encoding der normalisierten Inhalte	107
1. Zweckmäßigkeit einer Freistellung	108
2. Freistellungsmechanismus	109
III. Zwischenergebnis	111
F. Zusammenfassung der regulatorischen Bedarfsanalyse	111
Dritter Teil: Anwendbarkeit geltender urheberrechtlicher Bestimmungen und Abgleich mit den identifizierten Regelungsbedürfnissen	113
A. Speicherung von urheberrechtlich geschützten Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern	113

I. Anwendbares Recht	114
1. Hinreichender Inlandsbezug der Speicherung zur Anwendung deutschen Urheberrechts	114
2. Hilfsweise Geltung durch Extraterritorialität des Unionsurheberrechts nach der KI-VO	116
II. Eingriff in das urheberrechtliche Ausschließlichkeitsrecht	117
III. Eingreifen einer urheberrechtlichen Schrankenbestimmung	117
1. Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)	117
a) Vorübergehender Charakter der Vervielfältigung	118
b) Eigene wirtschaftliche Bedeutung der Vervielfältigung	119
c) Schlussfolgerung zur Anwendbarkeit der Schrankenbestimmung	121
2. Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)	121
a) Ziel der Informationsgewinnung	123
aa) Wortlaut als Ausgangspunkt der Auslegung	124
(1) Wortsinn des Begriffs der „Gewinnung“ im deutschen Regelungskontext	125
(2) Wortsinn des Begriffs der „Gewinnung“ im unionsrechtlichen Regelungskontext	127
(3) Schlussfolgerung aus dem Wortlaut der Bestimmungen	128
bb) Systematische Bezüge	128
cc) Technologische Neutralität des Urheberrechts	130
dd) Historische Auslegung	131
(1) Nationale Gesetzeshistorie	131
(2) Historie der Richtlinienbestimmung	132
ee) Zweck des zugrundeliegenden Richtlinienrechts	133
ff) Schlussfolgerungen	134
b) Bezugspunkt der Informationsgewinnung	135
c) Schlussfolgerungen für das KI-Training als Text und Data Mining im Sinne des § 44b Abs. 1 UrhG	136
d) Konformität mit dem Dreistufentest	138
aa) Bestimmter Sonderfall	139
bb) Beeinträchtigung der normalen Verwertung und ungebührliche Beeinträchtigung der Urheberinteressen	139

e) Zwischenergebnis zur Anwendbarkeit von § 44b UrhG	141
f) Beschränkung auf rechtmäßig zugängliche Werke	141
aa) Begriff der rechtmäßigen Zugänglichkeit	142
(1) Wortlaut und systematischer Bezug zu Art. 3 DSM-RL	142
(2) Regelungszweck	144
(3) Ergebnis der Auslegung	144
(4) Beispiele für rechtmäßige Zugängiskonstellationen	145
(5) Technische Schutzmaßnahmen sowie Geoblocking-Technologien und ihre Auswirkungen auf die rechtmäßige Zugänglichkeit	145
(6) Abgleich mit den identifizierten Regelungsbedürfnissen und Schlussfolgerungen	146
bb) Weitergabe des Zugangs oder Bereitstellung bereits angefertigter Vervielfältigungsvorlagen der Werke an Dritte	147
(1) Zugangsvermittlung im Interesse des Zugangsinhabers	148
(i) Datenverarbeiter als Vervielfältigungsdienstleister	149
(α) Voraussetzungen	149
(β) Aktiver Eingriff des Datenverarbeiters im vorliegenden Fall	150
(γ) Schlussfolgerungen	151
(ii) Eigener rechtmäßiger Zugang des Datenverarbeiters	151
(α) Strenge Wortlautanwendung	151
(β) Umkehrschluss aus § 53 Abs. 1 S. 2 UrhG	152
(γ) Ableitung eines allgemeinen Rechtsgedankens aus § 53 Abs. 1 S. 2 UrhG	153
(δ) Teleologische Erwägungen anhand des umgesetzten Art. 4 DSM-RL	154
(ε) Auslegungsergebnis	155
(iii) Zugangsweitergabe trotz vertraglicher Weitergabebeschränkung	156

(2) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzeugängen für Datenverarbeitung im Interesse des Zugangsinhabers	157
(3) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzeugängen für Datenverarbeitung im Eigeninteresse des Datenverarbeiters	160
(i) Rechtmäßiger Zugang der Datenverarbeiter	160
(ii) Begrenzung der Schrankenwirkung durch den Dreistufentest	161
(α) Argumente für eine ungebührliche Beeinträchtigung der Urheberinteressen	161
(β) Argumente gegen eine ungebührliche Beeinträchtigung der Urheberinteressen	162
(γ) Schlussfolgerungen	163
(δ) Abweichende Beurteilung bei objektiver Erkennbarkeit der zugangsgeschützten Herkunft genutzter Vervielfältigungsvorlagen	163
(4) Zugangsvermittlung für Datenverarbeitung im Interesse des Datenverarbeiters	165
cc) Fernwirkung einer fehlenden Zugangsberechtigung insbesondere bei der öffentlichen Wiedergabe von Trainingsdatensätzen	166
g) Kreis der Schrankenbegünstigten	168
aa) Kommerziell tätige Trainingsdatensammler	168
bb) Intermediäre Trainingsdatensammler	168
(1) Wortlaut der Regelung	169
(2) Systematische Anknüpfungspunkte	170
(3) Telos der Schrankenbestimmung	171
(4) Zwischenergebnis	174
(5) Dreistufentest	175
(i) Erwägungen zugunsten der betroffenen Urheber	175
(ii) Erwägungen zugunsten intermediärer Trainingsdatensammler	176
(6) Schlussfolgerungen	178

(7) Auswirkungen der Schrankenbestimmung auf die Geschäftsmodelle intermediärer Trainingsdatensammler	178
h) Löschungspflicht der Trainingsdatensammler	180
aa) Zeitpunkt des Eingreifens der Löschungspflicht	180
bb) Regulatorische Defizite des geltenden Urheberrechts	182
i) Opt-out-Mechanismus für Rechteinhaber	182
aa) Maschinenlesbarkeit des Nutzungsvorbehalts	183
(1) Notwendigkeit einer digitalen Hinterlegung der Informationen	183
(2) Bestehen eines Formalisierungs- und Strukturierungserfordernisses	184
(i) Natürlicher Sprachgebrauch des Begriffs „maschinenlesbar“ bzw. „machine-readable“	185
(ii) Begriffsverwendung im nationalen Regelungskontext	186
(iii) Einheitliches europäisches Begriffsverständnis	187
(α) Maschinenlesbarkeit in der PSI- Richtlinie	187
(β) Maschinenlesbarkeit in der Richtlinie zur Nachhaltigkeitsberichterstattung von Unternehmen	188
(γ) Maschinenlesbarkeit in der KI- Verordnung	188
(δ) Maschinenlesbarkeit in der Datenschutz-Grundverordnung	189
(ε) Schlussfolgerung zur Existenz eines einheitlichen europäischen Begriffsverständnisses	190
(iv) Telos der Regelung	190
(v) Systematik innerhalb des nationalen Regelungsgefüges	194
(vi) Schlussfolgerung	194
bb) Angemessenheit des Nutzungsvorbehalts	195
(1) Verhältnis zwischen Angemessenheits- und Maschinenlesbarkeitsvoraussetzung	195

(2) Richtlinienkonforme Auslegung des § 44b UrhG zur nationalen Anknüpfung der Angemessenheitsvoraussetzung	196
(3) Konkrete Anforderungen an die Vorbehaltserklärungen	196
(i) Ort der Vorbehaltserklärung bei online veröffentlichten Inhalten	197
(ii) Sprache der Vorbehaltserklärung	199
(iii) Mindestmaß an Struktur und Klarheit	199
cc) Ausdrücklichkeit des Nutzungsvorbehalts	200
dd) Notwendigkeit einer Beschränkung der Vorbehaltserklärung auf Nutzung zum Zwecke des Text und Data Minings	200
ee) Weitere formalisierte Einzelfälle	202
(1) Robots Exclusion Standard (robots.txt)	202
(2) Spawning-Protokoll (ai.txt)	204
(3) World Wide Web Consortium (W3C) TDM Reservation Protocol	205
(4) Technische Schutzmaßnahmen gemäß § 95a UrhG	206
ff) Werk- oder wiedergabebezogene Wirkung des Nutzungsvorbehalts	207
gg) Zeitliche Wirkung des Nutzungsvorbehalts	208
hh) Der Vorbehaltserklärende	210
(1) Einfache und ausschließliche Rechtsinhaber	210
(2) Inhaltliche, räumliche und zeitliche Beschränkungen des Nutzungsrechts	212
(i) Inhaltliche Beschränkungen und die Vorbehaltsbefugnis	212
(ii) Zeitliche sowie räumliche Beschränkungen und die Vorbehaltsbefugnis	214
(3) Vertragliche Schutzpflicht beim Auseinanderfallen von Wiedergabeherrschaft und Vorbehaltsbefugnis	216
ii) Fernwirkung des Nutzungsvorbehalts bei Bereitstellung von Vervielfältigungsvorlagen oder einer Zugangsvermittlung an Dritte	217
(1) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen	218

(2) Zugangsvermittlung	219
jj) Zwischenergebnis zum Nutzungsvorbehalt	220
j) Fazit zur Schrankenbestimmung für Text und Data Mining	221
3. Andere Schrankenbestimmungen	222
a) Schrankenbestimmung für Zitate (§ 51 UrhG)	222
b) Schrankenbestimmung für die Benutzung eines Datenbankwerks (§ 55a UrhG)	223
c) Schrankenbestimmung für die Nutzung als unwesentliches Beiwerk (§ 57 UrhG)	224
aa) Irrtümlich nicht erfasste Nutzungsvorbehalte (§ 44b Abs. 3 UrhG)	224
bb) Trainingsdatensätze aus primär nicht urheberrechtlich geschützten Inhalten	226
d) Schrankenbestimmung für Karikatur, Parodie und Pastiche (§ 51a UrhG)	228
e) Intermediäre Trainingsdatensammler als Bibliotheken (§ 60d UrhG)	229
4. Rolle der Gedächtnisinstitutionen für das Training von KI	231
a) Rechtmäßige Zugänglichkeit der Werke für Nutzer der Gedächtniseinrichtungen	231
b) Freistellung über § 60d UrhG zugunsten der Nutzer von Gedächtnisinstitutionen	233
5. Zwischenergebnis zum Eingreifen einer urheberrechtlichen Schrankenbestimmung	236
IV. Gesetzliche Vergütung für Urheber	237
V. Zwischenfazit zum Web Scraping von Werken als Trainingsdaten für KI	237
B. Durchführung des Trainings auf Grundlage urheberrechtlich geschützter Trainingsdaten	238
I. Vorverarbeitung der Trainingsdaten (Normalisierung und Feature Encoding)	238
1. Normalisierung der Trainingsdaten	239
a) Normalisierung als Vervielfältigung im Sinne des § 16 UrhG: Anwendung eines normativierten Vervielfältigungsbegriffs	239
aa) Der Begriff des Vervielfältigungsstücks nach dem UrhG	240

bb) Funktionsbezogene Auslegung des Vervielfältigungsbegriffs	240
(1) Herleitung einer Schutzbereichsbegrenzung aus der verwertungsrechtlichen Funktion des Vervielfältigungsrechts	242
(2) Wiedererkennbarkeit eigenschöpferischer Züge nach der Normalisierung	245
(3) Weitere funktionsorientierte Erwägungen zur Normalisierung von Werken	246
cc) Schlussfolgerungen	246
b) Anwendung urheberrechtlicher Schrankenbestimmungen zur Erlaubnis der Vervielfältigungshandlungen	247
aa) Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)	247
bb) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)	248
(1) Grundsätzliche Anwendbarkeit von § 44b UrhG	248
(2) Auswirkungen von Veränderungen der Werke auf die Anwendbarkeit der Schrankenbestimmung	248
c) Zwischenergebnis zur Normalisierung urheberrechtlich geschützter Trainingsdaten	250
2. Feature Encoding	251
a) Umwandlung in Tensoren als Vervielfältigung im Sinne des § 16 UrhG	251
aa) Wiedererkennbarkeitsgrenze	252
(1) Faktische oder normative Wiedererkennbarkeit	252
(2) Beurteilungsperspektive für die faktische Wiedererkennbarkeit schutzbegründender Elemente	254
(3) Technische Rekonstruierbarkeit	256
(4) Wiedererkennbarkeitsgrenze bei ausschließlich technisch bedingten Erzeugnissen ohne eigene Werkqualität	258
bb) Anwendung des Maßstabs der Wiedererkennbarkeit auf den konkreten Fall	259
b) Schlussfolgerung	261
3. Zwischenfazit zur Vorverarbeitung urheberrechtlicher geschützter Trainingsdaten	262

II. Trainingsanalyse der Werke selbst	262
III. Vervielfältigungen während der Trainingsanalyse	264
IV. Gewichtungswerte und andere Parameter des trainierten Modells	264
1. Objektive Identifizierbarkeit schutzbegründender Elemente in Parametern eines KNN	266
2. Möglichkeiten der Invertierung und die Wiedererkennbarkeit	267
3. Identifizierbarkeit schutzbegründender Elemente im Output von KI und die Wiedererkennbarkeit	268
4. Schlussfolgerungen	269
V. Fazit zur Durchführung des Trainings auf Basis von Werken	270
C. Speicherung urheberrechtlich geschützter Inputdaten	271
I. Eingriff in das Vervielfältigungsrecht	271
II. Anwendung urheberrechtlicher Schrankenbestimmungen	273
1. Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)	273
a) Vorübergehende Vervielfältigung	274
b) Wesentlicher und integraler Bestandteil eines technischen Verfahrens	274
c) Flüchtig oder begleitender Charakter der Vervielfältigungen	275
d) Rechtmäßige Nutzung	276
e) Keine eigenständige wirtschaftliche Bedeutung	277
aa) Eigene wirtschaftliche Bedeutung unmittelbar aus der Nutzung der Werke als Inputdaten	277
bb) Mittelbare eigene wirtschaftliche Bedeutung aus der Möglichkeit zur Sekundärverwertung von Inputdaten	278
cc) Schlussfolgerungen	280
f) Zwischenergebnis	281
g) Erfassung rechtswidrig verfügbarer Werke	281
2. Weitere Schrankenbestimmungen	282
a) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)	282
aa) Anwendbarkeit der Schrankenbestimmung	283
bb) Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen an dritte Datenverarbeiter	284

cc) Fernwirkung des Nutzungsvorbehalts bei Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen an dritte Datenverarbeiter	287
b) Anwendungsspezifische Schrankenbestimmungen	288
3. Verhältnis der Schrankenbestimmungen aus § 44a Nr. 2 UrhG und § 44b UrhG	288
4. Anwendung der Schrankenbestimmungen auf generative KI-Systeme	291
a) Technische Perspektive auf die Verarbeitung als Inputdaten	291
b) Erzeugung von Kreativinhalten im Output	292
c) Dreistufentest	293
d) Schlussfolgerungen	294
IV. Zwischenfazit zur Speicherung urheberrechtlich geschützter Inputdaten	295
D. Vorverarbeitung von urheberrechtlich geschützten Inputdaten durch Normalisierung und Feature Encoding	295
I. Normalisierung der Inputdaten	296
1. Eingriff in das urheberrechtliche Verwertungsrecht	296
2. Anwendung urheberrechtlicher Schrankenbestimmungen	297
a) Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)	297
aa) Grundvoraussetzungen des Schrankentatbestands	297
bb) Eigene wirtschaftliche Bedeutung der Vervielfältigungen	298
cc) Schlussfolgerungen zur grundsätzlichen Anwendbarkeit von § 44a Nr. 2 UrhG	299
dd) Verändernde Vervielfältigungen und das Änderungsverbot (§ 62 UrhG)	299
b) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)	301
3. Zwischenergebnis zur Normalisierung der Inputdaten	301
II. Feature Encoding	302
III. Zwischenfazit zur Vorverarbeitung urheberrechtlich geschützter Inputdaten	302
E. Zusammenfassung des dritten Teils	303

Vierter Teil: Perspektiven für das Urheberrecht	307
A. Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern	307
I. Reichweite der Schrankenbestimmung	308
II. Angemessene Vergütung für Urheber	309
1. Vergütungspflicht und Vergütungssystem	309
2. Vergütungshöhe und ihre Konkretisierung durch Verwertungsgesellschaften	311
3. Erwartung der Vervielfältigung als Voraussetzung für die Vergütung (§ 44c Abs. 1 UrhG-E)	313
4. Trainingsdatensammler als Vergütungsschuldner (§ 44c Abs. 2 UrhG-E)	314
III. Unionsrechtliche Perspektive auf den Regelungsvorschlag	316
1. Klarstellung zum KI-Training	317
2. Verzicht auf Löschungspflicht und Opt-Out-Mechanismus	317
3. Einführung einer Vergütungspflicht für das Sammeln von Trainingsdaten	318
a) Vergütungspflicht als solche	318
b) Trainingsdatensammler als Vergütungsschuldner	320
c) Zwischenergebnis	321
IV. Zusammenfassung und kritische Würdigung des Regelungsvorschlags	321
B. Speicherung urheberrechtlich geschützter Inputdaten	322
I. Unmittelbar begrenzende Ausgestaltung des urheberrechtlichen Vervielfältigungsrechts	323
II. Unionsrechtliche Perspektive auf den Regelungsvorschlag und Schlussfolgerungen	324
C. Normalisierung urheberrechtlich geschützter Trainings- und Inputdaten als Teil der Datenvorverarbeitung	324
I. Unmittelbar begrenzende Ausgestaltung des urheberrechtlichen Vervielfältigungsrechts	325
II. Unionsrechtliche Perspektive auf den Regelungsvorschlag und Schlussfolgerungen	326

Fünfter Teil: Zusammenfassung und Schlussfolgerungen aus der Untersuchung	327
Literaturverzeichnis	333

