

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 zukunftsgerichteter 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 Zugangskonstellationen	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 Werkzugängen für Datenverarbeitung im Interesse des Zugangsinhabers	157
(3)	Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzugängen 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 Rechtsinhaber	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

Erster Teil: Abgrenzung des Untersuchungsgegenstands und dessen technische Grundlagen

A. Einleitung

Weitreichende technische Veränderungen prägen das Urheberrecht seit jeher. So legte die Erfindung des mechanischen Buchdrucks im 15. Jahrhundert den methodischen Grundstein für das urheberrechtliche Ausschließlichkeitsrecht.¹ Der Einzug des Magnettonbands in die privaten Haushalte der Konsumenten hatte eine Reformierung des Systems der Privatkopierfreiheit zur Folge.² Die Relevanz und Marktmacht von Suchmaschinen und User Generated Content-Plattformen führte nicht zuletzt zur Einführung eines neuen Leistungsschutzrechts und der Begründung eines innovativen Haftungsregimes für sogenannte Diensteanbieter, dem UrhDaG.³ All diese Entwicklungen machen deutlich, dass eine besondere Wechselbeziehung zwischen Urheberrecht und technischer Innovation besteht. Grund hierfür ist unter anderem, dass ein wesentlicher Teil technischer Innovation die Informationsverarbeitung und -aufnahme durch Menschen beeinflusst. Trägermedium dieser Informationen sind oftmals aber urheberrechtlich geschützte Werke.

Auch die technischen Veränderungen durch die Entwicklung und Anwendung künstlicher neuronaler Netze (KNN) als Form des maschinellen Lernens sind seit der Veröffentlichung des Chatbots *ChatGPT* im November 2022 weitreichend in den gesellschaftlichen Alltag gerückt. Zwischen dem Urheberrecht und der Entwicklung sowie Anwendung von KNN besteht dabei ebenfalls eine besondere Wechselbeziehung. Denn für die Entwicklung der KNN, dem „Training“, sind große Datenmengen erforderlich.⁴ Bereits aufgrund der niedrigen urheberrechtlichen Schutzvoraussetzungen⁵

1 Dazu m. w. N. *Schack*, Urheber- und Urhebervertragsrecht Rn. 108 ff.

2 BT-Drs. IV/270, S. 71.

3 Durch die Richtlinie 2019/790/EU über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt (DSM-RL).

4 Dazu unter I. Teil B. II. 3.

5 Sogar einzelne Sätze oder Satzteile können urheberrechtlichen Schutz genießen, vgl. EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 47 – Infopaq.

ist davon auszugehen, dass einem wesentlichen Teil der verarbeiteten Trainingsdaten Schutz als Werke im Sinne des § 2 UrhG zukommt. Werke können darüber hinaus bei der Anwendung eines KNN als Input⁶ des Systems dienen. Das ist beispielsweise der Fall, wenn mit Sensoren verbundene KNN bei ihrem Einsatz im öffentlichen oder nicht-öffentlichen Raum befindliche Werke oder Werkteile erfassen. Dies ist unter anderem bei mit Kamerasensoren ausgestatteten autonomen Fahrzeugen denkbar.⁷ Auch das Web Scraping von Internetquellen kann Inputdaten für die Anwendung von KNN generieren. Es findet unter anderem im Kontext automatisierter Journalismus-Anwendungen statt.⁸ Hierbei liegt eine Betroffenheit von Urheberinteressen ebenfalls nahe.

Ein Rückgriff auf Algorithmen, die auf KNN basieren, hat sich im Verlauf der letzten Jahre zum bevorzugten Lösungsweg für die computerbasierte Bewältigung komplexer Anwendungsprobleme entwickelt.⁹ Häufig wird dabei allgemein von Künstlicher Intelligenz (KI) gesprochen. Besonders relevant sind vielschichtige, als „tiefe“ bezeichnete Ausführungen der KNN.¹⁰ Grund hierfür ist die erhebliche Leitungsfähigkeit der Systeme. Keine andere Technologie konnte bisher ohne wesentlichen menschlichen Einfluss realistisch anmutende Gemälde,¹¹ längere kohärente Texte¹² so-

6 Zur Funktionsweise genauer unter B. II. 2.

7 Beispielhafte Umgebungsaufnahme eines autonomen Fahrzeugs in *Paaf/Hecker*, Künstliche Intelligenz, S. 318 Abb. 8.28.

8 Siehe dazu unter anderem *Haim/Graefe*, in: Nuernbergk/Neuberger, Journalismus im Internet, S. 139; *Schippan*, ZUM 2024, 670; *Gräfe/Kahl*, MMR 2021, 121.

9 *Jordan/Mitchell*, Science 2015, 255 (255); *Drexel u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 3; *WIPO*, WIPO Technology Trends 2019: Artificial Intelligence, S. 31; siehe auch *Döbel u. a.*, Maschinelles Lernen, S. 8, hier wird Machine Learning als „Schlüsseltechnologie der künstlichen Intelligenz“ bezeichnet.

10 *Drexel u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 3; *Jordan/Mitchell*, Science 2015, 255 (255) („high impact area“); *WIPO*, WIPO Technology Trends 2019: Artificial Intelligence, S. 31 (Deep Learning sei der am schnellsten wachsende Zweig innerhalb des Machine Learning); vgl. auch *Döbel u. a.*, Maschinelles Lernen, S. 11, 16, 24 f. insbesondere auch mit einer Statistik zur stark zunehmenden wissenschaftlichen Forschung in diesem Bereich; *Baum*, in: Leupold/Wiebe/Glossner, MAH IT-Recht, Teil 9 Rn. 30, der den „Siegesszug“ des Deep Learning verkündet.

11 Ein durch KI erschaffenes Portrait des fiktiven Edmond de Belamy, siehe *SPIEGEL* v. 26.10.2018. Außerdem der „Next Rembrandt“, ein durch KI geschaffenes Gemälde in der malerischen Tradition Rembrandt van Rijns, siehe <https://www.nextrembrandt.com>.

12 So nun aber durch Textgeneratoren wie GPT-3, siehe *Romero*, Medium/towards data science v. 24.5.2021.

wie Musikstücke¹³ erschaffen oder bei einer Bildanalyse menschliche Fehlerquoten unterbieten¹⁴. Im Gegenteil war die Produktion insbesondere von Kreativinhalten bisher menschlichen Schöpfern vorbehalten. Die Leistungsfähigkeit der KNN spiegelt sich zudem auch wirtschaftlich wider.¹⁵ Es ist also die Basis geschaffen, damit das Urheberrecht auch durch KNN besonders geprägt werden kann. Vor diesem Hintergrund wird im Rahmen der vorliegenden Arbeit untersucht, wohin sich das Urheberrecht im Kontext der Entwicklung und Anwendung von KNN de lege ferenda bewegen wird.

Auch die Anwendung des geltenden Urheberrechts steht vor der Herausforderung, technische Veränderungen im Rahmen der gesetzlich gewährten Spielräume abzubilden. Dabei sollen zuweilen Interessen betroffener Urheber geschützt, an anderer Stelle aber auch Nutzerinteressen Rechnung getragen werden. In den letzten Jahren entlud sich das Spannungsverhältnis zwischen Urheberrecht und technischer Innovation insbesondere beim Umgang mit Verlinkungstechniken im Internet,¹⁶ aber auch im Kontext von Internetrekordern,¹⁷ Sharehosting-Plattformen,¹⁸ dem Schutz von Computerprogrammen¹⁹ oder dem E-Lending²⁰. Auch KNN nehmen nun ihren Platz in diesem Spannungsfeld ein. Angesichts dessen gilt es zu untersuchen, wie das Training und die Anwendung von KNN in den geltenden urheberrechtlichen Regelungskomplex eingeordnet werden können.

Besondere Rechtsfragen wirft demgegenüber der in der Anwendungsphase eines KNN generierte Output des Systems auf. Hier kann unter-

13 So beispielsweise die „Jukebox“, ein neuronales Netz von Open AI, siehe *Heaven*, MIT Technology Review 2021; außerdem direkt unter <https://openai.com/blog/jukebox>.

14 Vgl. dazu *He u. a.*, arXiv: 1502.01852 2015, S. 9.

15 Weltweit wird für den Absatz von Deep Learning-Software eine Umsatzsteigerung von 655 Mio. USD (2016) auf 35 Mrd. USD (2025) erwartet *Döbel u. a.*, *Maschinelles Lernen*, S. 24 mit weiteren Details und Nachweisen. Zur wirtschaftlichen Entwicklung bis 2023 siehe außerdem insgesamt *Maslej u. a.*, *The AI Index 2024 Annual Report* by Stanford University, S. 214 ff.

16 Beispielsweise siehe EuGH, Urt. v. 08.06.2016 - C-160/15, GRUR 2016, 1152 – GS Media; EuGH, Urt. v. 09.03.2021 - C-392/19, GRUR 2021, 706 – VG Bild-Kunst/SPK.

17 Insbesondere BGH, Urt. v. 22.04.2009 - I ZR 175/07, ZUM 2009, 765 – Online-Videorekorder; BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 – Internet-Radiorecorder.

18 EuGH, Urt. v. 22.06.2021 - C-682/18, C-683/18, GRUR 2021, 1054 – YouTube/Uploaded.

19 EuGH, Urt. v. 02.05.2012 - C-406/10, EuZW 2012, 584 – SAS Institute.

20 C-174/15, Urt. v. 10.11.2016, GRUR 2016, 1266 – VOB/Stichting.

sucht werden, ob und inwieweit dem Output urheberrechtlicher Schutz zukommt.²¹ Fraglich ist außerdem, wer für Urheberrechtsverletzungen insbesondere durch den Einsatz generativer KNN, also solcher Systeme, die Kreativinhalte produzieren, verantwortlich ist.²² Die Beantwortung der sich hieraus ergebenden, von Training und Input in ein KNN nur bedingt abhängigen Fragen würde den für die vorliegende Untersuchung zur Verfügung stehenden Umfang überschreiten. Sie werden deshalb im Rahmen dieser Arbeit nicht näher beleuchtet.

B. Systeme künstlicher Intelligenz als Forschungsgegenstand

Um die urheberrechtlichen Rahmenbedingungen für die Verwendung von Werken als Trainings- und Inputdaten für KNN analysieren und Perspektiven für das Regelungsregime aufzeigen zu können, muss zunächst der Forschungsgegenstand konkretisiert werden. Darüber hinaus bedarf es der Vorstellung einzelner technischer Prozesse und der Einführung einiger Fachbegriffe, deren Kenntnis für die Subsumtion des Forschungsgegenstands unter die urheberrechtlichen Bestimmungen notwendig sind.

I. KI-Begriff der Untersuchung

Wie bereits dargelegt wurde,²³ nehmen KNN innerhalb der KI-Forschung und Anwendung von umgangssprachlich als „KI“ bezeichneten Systemen eine besondere Rolle ein. Daher stehen die Forschung und Anwendung von „KI“ immer mehr synonym für die Forschung an und Anwendung von Systemen, die auf künstlichen neuronalen Netzen basieren. Die tatsächlichen Entwicklungen sollen auch in der nachfolgenden Untersuchung abgebildet werden. Demzufolge wird auch der Forschungsgegenstand der vorliegenden Arbeit auf Systeme beschränkt, die auf künstlichen neuronalen Netzen basieren und damit als Systeme maschinellen Lernens eingeordnet werden können. Der Beschränkung des Forschungsgegenstands entsprechend wer-

21 Dazu unter vielen hervorzuheben *Maamar*, Computer als Schöpfer.

22 Dazu unter anderem in *Pukas*, ZGE 2024, 38.

23 Unmittelbar oben unter A.

den unter Systemen „künstlicher Intelligenz“ im Rahmen der vorliegenden Untersuchung nur solche Systeme verstanden, die auf KNN aufbauen. Eine terminologische Bezugnahme zum Rechtsbegriff des „KI-Systems“ aus Art. 3 Nr. 1 VO (EU) Nr. 2024/1689 (KI-VO) ist demgegenüber nicht angestrebt. Denn der Anwendungsbereich der KI-VO geht über die Entwicklung und Nutzung von KNN hinaus. Die in der KI-VO normierte Legaldefinition des „KI-Systems“ übersteigt damit den Untersuchungsgegenstand der vorliegenden Arbeit.

II. Technische Grundlagen des maschinellen Lernens auf Basis von KNN

Hinsichtlich der technischen Grundlagen und Funktionalität von KI-Systemen im Sinne der vorliegenden Untersuchung lässt sich zwischen dem Grundprinzip maschinellen Lernens, der Struktur eines künstlichen neuronalen Netzes sowie dem Trainingsprozess, in dem der KI ihre Fähigkeiten zur Bewältigung einer Anwendungsaufgabe vermittelt werden, unterscheiden.

1. Maschinelles Lernen

„Maschinelles Lernen“ ist als Oberbegriff für eine Vielzahl algorithmischer Ansätze zu verstehen. Ziel des Systems ist, eine Aufgabe (Zielfunktion) durch das Sammeln von Erfahrungen während eines repetitiven Lernprozesses (Training) immer besser bewältigen zu können.²⁴ Bei Systemen, deren Arbeitsweisen konzeptionell auf maschinellen Lernverfahren beruhen, muss zur Bewältigung der Anwendungsaufgabe keine umfangreiche Wissensbasis für eine Erkenntnisableitung „top down“ angelegt werden. Vielmehr arbeiten maschinelle Lernverfahren „bottom up“: Die für die Aufgabe notwendigen Fähigkeiten werden aus zur Verfügung gestellten Beispieldaten durch das System selbst abgeleitet.²⁵ Auf diese Weise können in einer großen Zahl bestehender Inhalte verborgene Muster und Zusammenhän-

24 *Simon*, in: Michalski/Carbonell/Mitchell, Machine learning: an artificial intelligence approach, S. 25 (28); *Mitchell*, Machine Learning, S. 2.

25 *Morik*, in: Görz, Einführung in die künstliche Intelligenz, S. 243 (254 f.); *Görz/Braun/Schmid*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 1 (19); *Meyer*, ZRP 2018, 233 (235); *Alpaydın*, Machine learning, S. 16 f., 24 f.; *Rosengrün*, Künstliche Intelligenz zur Einführung, S. 27; vgl. außerdem *Sommer*, Haftung für autonome Systeme, S. 38.

ge, also statistische Regeln, extrahiert und zur Lösung eines Anwendungsproblems genutzt werden. Aus der Funktionsweise der algorithmischen Ansätze maschinellen Lernens folgt aber, dass alle auf einem maschinellen Lernverfahren basierenden Systeme auf einen Trainingsprozess angewiesen sind.²⁶

2. Struktur und Arbeitsweise eines künstlichen neuronalen Netzes

Künstliche neuronale Netze beruhen, wie auch ihre biologischen Vorbilder, auf „Neuronen“, also einer Vielzahl von Knotenpunkten, über die die Informationsverarbeitung abläuft.²⁷ Künstliche Neuronen nehmen dabei durch mathematische Abläufe ermittelte Zahlenwerte als Impulse (Input) auf und leiten diese an andere Neuronen weiter (Output).²⁸

In architektonischer Hinsicht sind die für diese Informationsübertragungen genutzten Netze streng strukturiert.²⁹ Sie lassen sich in verschiedene Schichten (Layer) aufteilen.³⁰ Jede Schicht besteht aus einer Vielzahl künstlicher Neuronen und hat eine eigene Aufgabe zu bewältigen. Die Schichten sind dabei miteinander verbunden. Über die Input-Schicht werden die Ausgangsinformationen, beispielsweise Sensordaten oder eine konkrete Arbeitsanweisung, in das KNN eingeleitet.³¹ Mithilfe der Output-Schicht gibt das KNN die ermittelten Werte am Ende des Rechenprozesses an den Systemnutzer aus.³² Die tatsächliche Mustererkennung findet auf Ebene der mittleren, sogenannten verborgenen Schichten statt.

26 Vgl. *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 429 (430 ff.); *Drexl u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 4 f.

27 *Rey/Wender*, Neuronale Netze, S. 16.

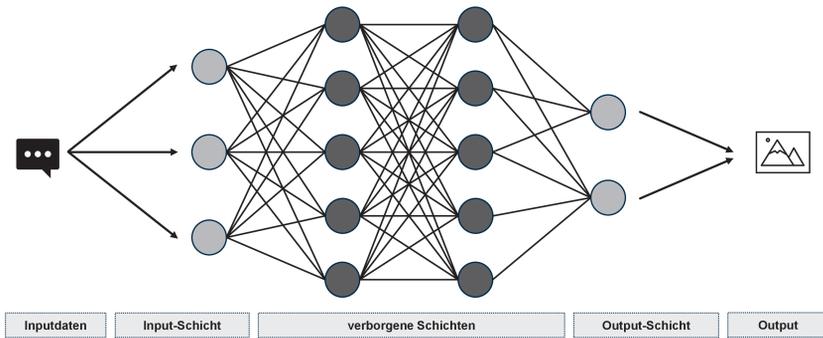
28 *Rey/Wender*, Neuronale Netze, S. 16 f.

29 *Ragni*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 227 (240).

30 Vgl. *Alpaydin*, Machine learning, S. 87; *Ertel*, Grundkurs Künstliche Intelligenz: eine praxisorientierte Einführung, S. 301.

31 *Rey/Wender*, Neuronale Netze, S. 17.

32 *Rey/Wender*, Neuronale Netze, S. 17; vgl. außerdem sehr anschaulich *Rosengrün*, Künstliche Intelligenz zur Einführung, S. 25 f.



Die Verbindung künstlicher Neuronen zueinander wird unter anderem durch eine Vielzahl anpassbarer, numerisch dargestellter Parameter beeinflusst, die die Berechnungsprozesse innerhalb des Algorithmus maßgeblich steuern.³³ Einige von ihnen, die Gewichtungswerte, werden automatisch beim Training des KNN eingestellt.³⁴ Sie haben einen besonders großen Einfluss auf die statistische Mustererkennung beim Training und den Ablauf der Informationsverarbeitung innerhalb des KNN.³⁵

Bevor Inhalte in einem künstlichen neuronalen Netz verarbeitet, also als Input in den Algorithmus eingegeben werden können, müssen sie für die algorithmische Nutzung vorbereitet werden. Dabei werden die Daten im Regelfall insbesondere auf eine vergleichbare Skala transponiert, also normalisiert und anschließend in eine numerische Darstellungsform überführt („Feature Encoding“).³⁶ Nur diese numerischen Repräsentationen können als Grundlage der algorithmischen Verarbeitung genutzt werden. Der Prozess wird zusammen mit anderen Verarbeitungsschritten, denen im Rahmen dieser Untersuchung keine Bedeutung zukommt, als Datenvorverarbeitung bezeichnet.

33 Mit weiteren Beispielen *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 429 (494); *Drexel u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 6; *Wahlster/Winterhalter*, Deutsche Normungsroadmap Künstliche Intelligenz des DIN e.V. und der DKE, S. 88.

34 *Drexel u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 6.

35 Vgl. *Rey/Wender*, Neuronale Netze, S. 17.

36 *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 429 (433); *Ertel*, Grundkurs Künstliche Intelligenz: eine praxisorientierte Einführung, S. 301 Abb. 9.26; mit sehr anschaulichen auch grafischen Beispielen *Paaf/Hecker*, Künstliche Intelligenz, S. 51 f., 89.

3. Das Lernen: Training des künstlichen neuronalen Netzes

Wie auch beim menschlichen Lernen³⁷ ist die Wiederholungstechnik Grundstein des künstlichen Lernens, also des Trainings einer KI. Dem KNN wird dabei durch wiederholte Aktivität die Möglichkeit gegeben, Muster und Zusammenhänge innerhalb der zur Verfügung gestellten Daten zu erkennen. Algorithmisch wird das Training durch eine systematische und repetitive Veränderung der flexiblen Parameter innerhalb des KNN bewerkstelligt.³⁸ Hierzu gehören insbesondere die Gewichtungswerte.³⁹ Die Veränderungen werden dabei so lange wiederholt, bis der Output des KI-Systems vordefinierte Qualitätsmerkmale aufweist.⁴⁰ Dies spricht indirekt dafür, dass die dem Trainingsmaterial zugrundeliegenden Muster und Zusammenhänge hinreichend genau im KNN repräsentiert worden sind. Zum Training eines KNN können dabei verschiedene Verfahren eingesetzt werden. Welcher Trainingsprozess von KI-Entwicklern gewählt wird, hängt von den verfügbaren Trainingsdaten, dem Anwendungsziel des KI-Systems sowie dem verwendeten Modelltyp ab.⁴¹

Um den repetitiven Trainingsprozess eines KNN durchzuführen, werden daher spezielle Trainingsdaten in ausreichender Menge benötigt. Als Schätzung dient die Veranschlagung, dass zehnmals so viele Trainingsdatenpaare benötigt werden, wie anpassungsbedürftige Parameter im KNN bestehen.⁴² Für fortgeschrittene KI-Systeme, die mitunter Milliarden oder sogar Bil-

37 Vgl. *Bechmann u. a.*, in: Anderhuber/Pera/Streicher, Waldeyer - Anatomie des Menschen, S. 945 (1107); *Heckmann/Dudel*, in: Schmidt/Lang/Heckmann, Physiologie des Menschen, S. 76 (93).

38 *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 429 (447); *Drexel u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 7.

39 *Mallot*, in: Görz, Einführung in die künstliche Intelligenz, S. 813 (826); *Otero*, GRUR Int. 2021, 1043 (1051); *Rey/Wender*, Neuronale Netze, S. 17.

40 Vgl. *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 429 (549); *Rey/Wender*, Neuronale Netze, S. 42.

41 Mit einer tabellarischen Übersicht *Döbel u. a.*, Maschinelles Lernen, S. 10; im Übrigen siehe *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 429 (432 ff.); *Drexel u. a.*, MPI for Innovation and Competition Research Paper Nr. 19-13, S. 8; *Rey/Wender*, Neuronale Netze, S. 28 f.

42 Vgl. zu der Faustregel beispielsweise *Alwosheel/Van Cranenburgh/Chorus*, Journal of Choice Modelling 2018, 167 (167), wobei Bestrebungen bestehen, leistungsfähige KI-Systeme auch auf Basis kleinerer Datensätze zu trainieren.

lionen Parameter beinhalten,⁴³ ist demzufolge eine erhebliche Menge an Trainingsdaten notwendig. Diese für das Training notwendigen Datensätze können auf verschiedenen Arten zusammengestellt werden. Praktisch relevant ist vor allem, die Daten aus frei verfügbaren Internetquellen ohne individuelle Erlaubnis der Betroffenen zusammenzutragen.⁴⁴ Dabei werden sie mit Hilfe sogenannter Webcrawler automatisch von Webseiten oder Datenbanken Dritter extrahiert.⁴⁵ Es zeigt sich auch, dass die Heterogenität von Webinhalten entscheidend für die Modellqualität von KNN ist.⁴⁶ Beispielsweise das Trainingsdatensatz *LAION-400M* beinhaltet Bild-Text-Paare, die zwischen 2014 und 2021 aus beliebigen Internetquellen heruntergeladen worden sind. Diese stammen aus dem *Common Crawl*, einem frei zugänglichen, nicht kommerziellen Datensatz, welches durch Web Scraping unspezifischer Internetquellen erstellt wurde.⁴⁷ Auch der dem Sprachmodell *GPT-3* zugrundeliegende Trainingsdatensatz korpus basierte zu großen Teilen auf diesem Datensatz. Darüber hinaus wurden auch andere frei zugängliche Texte wie insbesondere Wikipedia-Biträge abgespeichert.⁴⁸ Nach dem Web Scraping bedarf es auch für das Training auf Basis von Werken noch einer Vorverarbeitung und gegebenenfalls auch Annotation der Inhalte.

43 Das Sprachverarbeitungsmodell *GPT-4* über eine Billionen Parameter, siehe *Bak-tash/Dawodi*, arXiv:2305.03195 2023, S. 2. Bereits das Vorgängermodell *GPT-3* bewegt sich mit 175 Milliarden Parametern im Milliardenbereich, siehe *Brown u. a.*, arXiv:2005.14165 2020, S. 5. Ebenfalls das Sprachverarbeitungsmodell *PaLM*, welches Berechnungen auf Basis von 540 Milliarden Parametern vornimmt, siehe *Muller u. a.*, Annual report on European SMEs 2020/2021, S. 23. Selbst ältere Sprachverarbeitungsmodelle, die vor *GPT-3* veröffentlicht wurden, beinhalteten bereits eine Anzahl von Gewichtungswerten im dreistelligen Millionen- bis zweistelligen Milliardenbereich, siehe dazu auch *Brown u. a.*, arXiv:2005.14165 2020, S. 4 m. w. N. Zur Entwicklung der Parameteranzahl insgesamt siehe *Muller u. a.*, Annual report on European SMEs 2020/2021, S. 54.

44 *Longpre u. a.*, NAACL 2024, 3245 (3261).

45 Vgl. *Iglesias u. a.*, Intellectual Property and Artificial Intelligence - A literature review, S. 10; BT-Drs. 19/23700, S. 71; *IBM* hat durch Web-Scraping von Millionen von Fotografien vom Bildhosting-Dienst Flickr große mediale Empörung verursacht, siehe in den NBC News vom 12.3.2019: *Solon*, NBC News v.12.3.2019; auch der bereits erwähnte *Microsoft COCO*-Datensatz ist mit Hilfe von Web-Scraping (fachsprachlich auch als „harvesting“ bezeichnet) erstellt worden, siehe *Lin u. a.*, arXiv: 1405.0312 2015, S. 2.

46 *Longpre u. a.*, NAACL 2024, 3245 (3246, 3251).

47 Dazu <https://laion.ai/blog/laion-400-open-dataset> (zuletzt abgerufen am: 10.1.2024).

48 Zu *GPT-3* siehe *Brown u. a.*, arXiv:2005.14165 2020, S. 8 f. Vgl. außerdem *Hirschberg/Manning*, Science 2015, 261 (264).

C. Forschungsfrage und Ziel der Arbeit

In Bezug auf den Forschungsgegenstand soll untersucht werden, welche Rahmenbedingungen das Urheberrecht für die Verwendung von Werken als a) Trainings- und b) Inputdaten für auf KNN beruhende Systeme maschinellen Lernens, KI-Systeme im Sinne dieser Untersuchung, aufstellt. Dabei ist auch die notwendige Datenvorverarbeitung der als Berechnungsgrundlage der Algorithmen verwendeten Werke zu berücksichtigen. Weil die Generierung, also Speicherung von Werken zur Erstellung von Trainingsdatensätzen aus Internetquellen mittels Webcrawlern praktisch eine erhebliche Bedeutung hat, beschränkt sich die Untersuchung auf diese Art der Datenbeschaffung.

Fraglich ist infolgedessen, welche Rahmenbedingungen das geltende Urheberrecht für folgende Arbeitsschritte zur Verfügung stellt:

1. Speicherung von urheberrechtlich geschützten Trainingsdaten mit Hilfe von Webcrawlern (Web Scraping)
2. Training mit urheberrechtlich geschützten Inhalten als solches
3. Speicherung von urheberrechtlich geschützten Inputdaten für KI, beispielsweise mittels Sensortechnik wie Kameras oder Mikrofonen oder durch manuelle Eingabe der Werke
4. Vorverarbeitung von urheberrechtlich geschützten Inputdaten durch Normalisierung und Feature Encoding

Dabei sind hinsichtlich dieser Fallgruppen auch Perspektiven für die urheberrechtlichen Rahmenbedingungen aufzuzeigen. Untersucht werden muss deswegen auch, ob der bestehende urheberrechtliche Regelungskomplex, der auf die Verwendung von Werken als Trainings- und Inputdaten Anwendung findet, de lege ferenda angepasst werden sollte und wie Anpassungen gegebenenfalls aussehen könnten.

Gegenstand der nachfolgenden Untersuchung ist allerdings nur die Verwendung urheberrechtlich geschützter Werke, also persönlicher geistiger Schöpfungen im Sinne des § 2 Abs. 2 UrhG. Die Erfüllung dieser Kriterien wird für die Zwecke dieser Untersuchung jeweils vorausgesetzt. Der Untersuchung wird außerdem grundsätzlich ein Zweipersonenverhältnis zwischen Urheber und Werknutzer zugrunde gelegt. Das gilt insbesondere für die regulatorische Analyse bestehender Freistellungsinteressen, die im zweiten Teil der vorliegenden Arbeit stattfindet. Nichtsdestotrotz dürften eine Vielzahl der vorgenommenen Abwägungen auch auf das Verhält-

nis zwischen Verwerter, also sekundärem Rechtsinhaber, und Werknutzer übertragen werden können.

D. Gang der Untersuchung

Damit für die Verwendung von Werken als Trainings- und Inputdaten Perspektiven für das Urheberrecht aufgezeigt werden können, müssen zunächst urheberrechtliche Regelungsbedürfnisse für die Verwendung von Werken in den zu untersuchenden Fallgruppen identifiziert werden (zweiter Teil). Hierdurch wird zugleich ein Fahrplan für die Analyse der urheberrechtlichen Lex lata aufgestellt. Denn sie kann an den zuvor ermittelten Untersuchungsergebnissen ausgerichtet werden. Klärungsbedürftig ist dabei stets, ob die geltenden urheberrechtlichen Bestimmungen dem identifizierten Zielzustand der Rechtsordnung gerecht werden. Sind die Regelungsbedürfnisse identifiziert, schließt sich demzufolge die Analyse der Rahmenbedingungen an, die das geltende Urheberrecht zur Lösung der Fallkonstellationen zur Verfügung stellt (dritter Teil).

Die Ergebnisse aus der Analyse des geltenden Urheberrechts werden anschließend mit den identifizierten Regelungsbedürfnissen abgeglichen (dritter Teil). Aus diesem Vergleich ergibt sich, ob das Urheberrecht de lege lata einen interessengerechten Rechtsrahmen für die Verwendung von Werken als Trainings- und Inputdaten zur Verfügung stellt. Anderenfalls offenbaren sich Regelungsdefizite. Soweit sich Anpassungsbedürfnisse offenbart haben, werden im Anschluss Änderungsvorschläge erarbeitet und damit Perspektiven für die urheberrechtlichen Rahmenbedingungen aufgezeigt (vierter Teil). Hierdurch fügen sich die Abschnitte der vorliegenden Untersuchung zu einem kohärenten Gesamtsystem zusammen. Eine Zusammenfassung und Schlussfolgerungen aus der Analyse vervollständigen die Arbeit (fünfter Teil).

Zweiter Teil: Regelungsbedürfnisse im Urheberrecht für die Verwendung von Werken als Trainings- und Inputdaten

Untersucht werden muss zunächst, wie das urheberrechtliche Regelungsregime im Kontext der Verwendung von Werken als Trainings- und Inputdaten ausgestaltet sein sollte. Fraglich ist damit zum einen, ob die Verwendung urheberrechtlich geschützter Werke im Rahmen der untersuchten Fallgruppen den betroffenen Urhebern vorbehalten bleiben sollte, oder ob eine Freistellung der Sachverhalte zweckmäßig ist. Der Freistellungsbegriff im Sinne der vorliegenden Untersuchung umfasst jede Regelung, die es Trainingsdatensammlern oder Verwendern von KI ermöglicht, Werke für den jeweiligen Zweck ohne individuelle Zustimmung der betroffenen Urheber zu verwenden, womit die Nutzungen im Regelfall nicht mehr durch die Urheber kontrolliert werden können.⁴⁹ Falls eine Freistellung zweckmäßig ist, muss darüber hinaus geklärt werden, durch welchen Freistellungsmechanismus die Erlaubnis umgesetzt sein sollte.

A. Herleitung der Beurteilungskriterien und geeigneter Freistellungsmechanismen

Unklar ist allerdings, nach welchen Kriterien die Zweckmäßigkeit einer Freistellung zu beurteilen ist, welche Regelungsmechanismen für die Umsetzung eines Freistellungsbedürfnisses grundsätzlich zur Verfügung stehen und anhand welcher Anhaltspunkte eine Wahl zwischen den verschiedenen Rechtsinstituten erfolgen kann.

49 Angelehnt an EuGH, Urt. v. 16.11.2016 - C-301/15, GRUR 2017, 62 Rn. 33 – Soulier; einen engeren Freistellungsbegriff verwendet hingegen Stieper, siehe *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 6, er differenziert beispielsweise noch einmal zwischen Freistellungen und gesetzlichen Lizenzen.

I. Beurteilungskriterien für die Zweckmäßigkeit einer Freistellung

Ob eine Freistellung zweckmäßig ist, könnte von mehreren miteinander in Beziehung stehenden Faktoren abhängen.

1. Der Schutz der Urheberinteressen als Hauptzweck des Urheberrechts

Der Schutz der Schöpferinteressen ist Kernanliegen des Urheberrechts.⁵⁰ Das kommt bereits in § 1 UrhG zum Ausdruck. Umgekehrt ist der Schutz der Schöpferinteressen damit auch für regulatorische Überlegungen von wesentlicher Bedeutung.

a) Die Urheberinteressen

Abgesichert werden sollen zum einen die Vermögensinteressen der Urheber.⁵¹ Das Urheberrecht zielt dabei darauf ab, dem Schöpfer ein Wirtschaftsgut zu verschaffen, dass nach dessen schöpferischer Tätigkeit zu seinem Unterhalt beiträgt.⁵² Außerdem übt die Perspektive auf gerechte Entlohnung eine Anreizwirkung für schöpferische Leistung aus.⁵³ Aus diesen Gründen soll der Urheber vor einer unbefugten Auswertung seines Werkes geschützt werden.⁵⁴

50 Siehe auch *Schack*, Urheber- und Urhebervertragsrecht, S. 2 Rn. 2.

51 BT-Drs. IV/270, S. 28.

52 BGH, Urt. v. 18.05.1955 - I ZR 8/54, GRUR 1955, 492 (496) – Grundig-Reporter; *Schack*, Urheber- und Urhebervertragsrecht, S. 6 Rn. 11; *Ulmer*, Urheber- und Verlagsrecht, S. 24 f. Im europäischen Recht auch Richtlinie 2006/115/EG des Europäischen Parlaments und des Rates vom 12. Dezember 2006 zum Vermietrecht und Verleihrecht sowie zu bestimmten dem Urheberrecht verwandten Schutzrechten im Bereich des geistigen Eigentums (Vermiet- und Verleih-RL) ErwGr. 5.

53 Zum folgenorientierten Begründungsansatz siehe *Loewenheim*, in: Schricker/Loewenheim, Urheberrecht, Einleitung zum UrhG Rn. 9; *Fisher*, Chicago-Kent Law Review 1998, 1203 (1213 f.); sehr kritisch dazu aber *Kleinemenke*, Fair Use im deutschen und europäischen Urheberrecht?, S. 50 Fn. 97 m. w. N. Tatsächlich scheint sich zu bestätigen, dass der aus dem Urheberrechtsschutz resultierende Anreiz für Schöpfer geringer ist als angenommen, mit ersten vorsichtigen Ergebnissen hierzu *Handke/Girard/Mattes*, Fördert das Urheberrecht Innovation? Eine empirische Untersuchung, S. 105, 121.

54 Vgl. BT-Drs. IV/270, S. 27.

Darüber hinaus verkörpert jedes Werk aber auch die Persönlichkeit seines Urhebers.⁵⁵ Das Band, das zwischen Urheber und Werk geknüpft wird, muss neben den materiellen Interessen ebenfalls geschützt werden.⁵⁶ Daher sind auch ideelle Urheberinteressen zu berücksichtigen. Hierzu zählt insbesondere, sein Werk vor entstellenden Änderungen zu schützen.⁵⁷ Auch das Interesse eines Urhebers an der öffentlichen Anerkennung seiner Urheberschaft⁵⁸ sowie das Interesse daran, selbst darüber zu bestimmen, ob, wann und unter welchen Umständen sein Werk erstmalig öffentlich zugänglich gemacht wird,⁵⁹ ist zu berücksichtigen.

b) Grundrechtlicher Schutz der Urheberinteressen

Das Urheberrecht ist in seiner vermögensrechtlichen Ausprägung als geistiges Eigentum außerdem durch spezielle Grundrechtspositionen geschützt. Da die vermögensrechtliche Komponente des urheberrechtlichen Rechtsrahmens weitestgehend durch das Unionsrecht harmonisiert ist,⁶⁰ wird dieser Schutz vorrangig durch Art. 17 Abs. 2 EU-GRCh realisiert.⁶¹ Im Übrigen fällt das Urheberrecht als geistiges Eigentum aber auch in den Schutzbe-

55 EuGH, Urt. v. 01.12.2011 - C-145/10, GRUR 2012, 166 Rn. 88 – Painer; *Schack*, Urheber- und Urhebervertragsrecht, S. 4 Rn. 7; außerdem Richtlinie 93/98/EWG des Rates vom 29. Oktober 1993 zur Harmonisierung der Schutzdauer des Urheberrechts und bestimmter verwandter Schutzrechte (Schutzdauer-RL) ErwGr. 17.

56 Vgl. schon BT-Drs. IV/270, S. 44, 45.

57 So schon BT-Drs. IV/270, S. 43; eines der drei Pfeiler des Urheberpersönlichkeitsrechtes, siehe *Peukert*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 14 Rn. 1.

58 So schon RG, Urt. v. 08.04.1925 - I 251/24, RGZ 110, 393 (397); BGH, Urt. v. 16.06.1994 - I ZR 3/92, GRUR 1995, 671 (672) – Namensnennungsrecht des Architekten; BT-Drs. IV/270, S. 43.

59 „Grundnorm des Urheberschutzes“, so *Peukert*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 12 Rn. 1; außerdem bereits v. *Gierke*, Deutsches Privatrecht Band 1: Allgemeiner Teil u. Personenrecht, S. 790, 792.

60 Das gilt insbesondere für die Verwertungsrechte des Vervielfältigungs- und Verbreitungsrechts und des Rechts der öffentlichen Wiedergabe sowie der zugehörigen Schrankenbestimmungen, siehe v. *Ungern-Sternberg*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 15 Rn. 1; *Stieper*, in: Schricker/Loewenheim, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 25 ff. Siehe insgesamt zum Harmonisierungsstand des Urheberrechts *Nordemann*, in: Fromm/Nordemann, Urheberrecht, Einl. zum UrhG Rn. 43 ff.

61 Zum Urheberrecht als Teil des Schutzbereichs des Art. 17 EU-GRCh für alle EuGH, Urt. v. 29.07.2019 - C-469/17, GRUR 2019, 934 Rn. 57 – Afghanistan Papiere; *Jarass*, in: Jarass, Charta der Grundrechte der Europäischen Union, EU-GRCh Art. 17 Rn. 10. Selbst im mindestharmonisierten Bereich geben die Unionsgrundrechte den Mindeststandard vor, siehe EuGH, Urt. v. 26.02.2013 - C-399/11, NJW 2013, 1215 Rn. 60 –

reich des nationalen Art. 14 GG.⁶² Soweit ein Anwendungsspielraum für nationale Grundrechte verbleibt,⁶³ muss damit auch Art. 14 GG berücksichtigt werden. Hieraus folgt zum einen, dass bei einer Freistellung der Verwendung von Werken der unantastbare Kernbereich beziehungsweise Wesensgehalt des Urheberrechts gewahrt bleiben muss. Er ist sowohl in der europäischen als auch in der nationalen Grundrechtsdogmatik verankert.⁶⁴ Weiterhin muss aber auch die Bedeutung des Art. 17 EU-GRCh sowie unter Umständen des Art. 14 GG im grundrechtlichen Gefüge beachtet werden.

Als besondere Ausprägung des allgemeinen Persönlichkeitsrechts genießen die im Urheberpersönlichkeitsrecht gebündelten ideellen Interessen des Urhebers darüber hinaus auch über Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG besonderen grundrechtlichen Schutz.⁶⁵

Wenn untersucht wird, ob und inwiefern eine Freistellung zweckmäßig ist, muss damit auch berücksichtigt werden, welche Bedeutung die in der jeweiligen Sachverhaltskonstellation betroffenen Grundrechte haben, wie sie eingeschränkt werden können und in welchem Verhältnis sie zu anderen geschützten Rechtspositionen stehen.

Melloni; sowie insgesamt zum Anwendungsbereich der Unionsgrundrechte weitergehend *Stieper*, in: Schrickler/Loewenheim, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 32a.

- 62 Zum verfassungsrechtlichen Eigentumsbegriff vgl. BVerfG, Beschl. v. 10.06.2009 - 1 BvR 706/08 u.a., NJW 2009, 2033 Rn. 215 m. w. N.; zum Urheberrecht als Eigentum im Sinne des Art. 14 GG erstmals BVerfG, Beschl. v. 07.07.1971 - 1 BvR 765/66, GRUR 1972, 481 (483) – Kirchen- und Schulgebrauch.
- 63 Zu einem umfassenden Überblick über die Abgrenzungskriterien zur Differenzierung zwischen den Anwendungsbereichen der nationalen und europäischen Grundrechte siehe insgesamt BGH, Urt. v. 30.04.2020 - I ZR 115/16, GRUR 2020, 843 Rn. 45 ff. – Metall auf Metall IV m. w. N.
- 64 Zum Kernbereich des Art. 14 GG: BVerfG, Beschl. v. 02.03.1999 - 1 BvL 7–91, NJW 1999, 2877 (2878); *Axer*, in: Epping/Hillgruber, Beck'scher Online-Kommentar Grundgesetz, GG Art. 14 Rn. 87. Zum Wesensgehalt des Unionsgrundrechts, der über Art. 52 Abs. 1 S. 1 EU-GRCh gewährt wird, siehe im Kontext der Eigentumsgarantie *Jarass*, in: Jarass, Charta der Grundrechte der Europäischen Union, EU-GRCh Art. 17 Rn. 32a m. w. N. aus der Rechtsprechung des EuGH.
- 65 *Peukert*, in: Schrickler/Loewenheim, Urheberrecht, Vor UrhG §§ 12 ff. Rn. 6; *Specht-Riemenschneider*, in: Dreier/Schulze, Urheberrechtsgesetz, Vor UrhG §§ 12 ff. Rn. 5; *Krüger-Nieland*, in: FS Hauß, S. 215 (219 f., 223 f.). Der Fokus liegt dabei auf Art. 2 Abs. 1 GG, sodass im Hinblick auf die ideellen Urheberinteressen keine absoluten Einschränkungsvorbehalte bestehen, vgl. beispielsweise BVerfG, Beschl. v. 24.11.2009 - 1 BvR 213/08, ZUM 2010, 235 Rn. 73; anders aber *Schack*, GRUR 1985, 352 (353); *Schack*, Urheber- und Urhebervertragsrecht, S. 25 Rn. 48.

c) Schlussfolgerung: Freistellung wegen Unerheblichkeit für die Urheber

Aus dem Schutzziel des Urheberrechts folgt im Umkehrschluss, dass Handlungen wie unter anderem das Web Scraping von Werken als Trainingsdaten vom Ausschließlichkeitsrecht der Urheber ausgeklammert werden können, wenn die ideellen oder materiellen Urheberinteressen durch sie nicht beeinträchtigt sind. Dieser regulatorische Ansatz wurde beispielsweise auch bei der Einführung der Schrankenbestimmungen aus §§ 55, 56 UrhG (Vervielfältigungen durch Sendeunternehmen und in Geschäftsbetrieben) verfolgt.⁶⁶

2. Freistellungen zum Ausgleich von Nutzer- und Urheberinteressen

Zwar schützt das „Recht des Urhebers“, wie es schon in §1 UrhG zum Ausdruck kommt, im Grundsatz vorrangig die ideellen und materiellen Urheberinteressen. Neben den Interessen der betroffenen Urheber könnten bei der Evaluation von Freistellungsbedürfnissen aber auch verschiedene Interessen der betroffenen Nutzergruppen, also der Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen berücksichtigt werden müssen. Fraglich ist jedoch, wo der Anknüpfungspunkt zur Berücksichtigung dieser Nutzerinteressen liegt und wie eine Berücksichtigung methodisch abgebildet werden kann.

a) Anknüpfungspunkt des Interessenausgleichs

Um die betroffenen Nutzerinteressen einzubeziehen, könnten sich im Wesentlichen drei Anknüpfungspunkte sowohl aus der europäischen als auch aus der nationalen Rechtsordnung ergeben.

aa) Grundrechte der Nutzer

Möglich ist zum einen, dass auch die Interessen der Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen grundrechtlich geschützt sind. Wenn das Sammeln der urheberrechtlich geschützten Trai-

66 BT-Drs. IV/270, S. 32.

ningsdaten, die Durchführung des Trainingsvorgangs selbst oder die Anwendung eines KI-Systems auf Grundlage urheberrechtlich geschützten Inputs Teil der Berufsausübung oder der unternehmerischen Tätigkeit ist, wird der Schutzbereich der Berufsfreiheit aus Art. 15 Abs. 1 EU-GRCh sowie aus Art. 12 GG oder der unternehmerischen Freiheit aus Art. 16 EU-GRCh eröffnet.⁶⁷ Je nach Anwendungsumfeld der KI könnten auch die Wissenschafts- und Pressefreiheit (Art. 11, 13 EU-GRCh sowie Art. 5 Abs. 1 S. 2, Abs. 3 GG) betroffen sein. Beim Einsatz von KI kollidieren demzufolge immer wieder nicht nur gegenläufige Interessen von Urhebern und Nutzern, sondern auch Grundrechtspositionen auf unionsrechtlicher oder nationaler Ebene miteinander. Diese Grundrechtspositionen müssen in der Folge zu einem angemessenen, gerechten Ausgleich gebracht werden.⁶⁸ Das folgt auf unionsrechtlicher Ebene bereits aus Art. 52 Abs. 1 S. 2 EU-GRCh.⁶⁹ Gerecht bedeutet, dass nicht nur die Urheberinteressen, sondern auch die Belange der Trainingsdatensammler, KI-Entwickler und Anwender von KI-Systemen gebührend berücksichtigt werden müssen. Ziel des Ausgleichs ist also, alle involvierten, grundrechtlich geschützten und gegenläufigen Interessen bestmöglich zur Geltung zu bringen.⁷⁰ Dies wird auf national-verfassungsrechtlicher Ebene als praktische Konkordanz bezeichnet.⁷¹

67 Zum Schutzbereich des Art. 12 GG vgl. *Jarass*, in: *Jarass/Pieroth*, Grundgesetz für die Bundesrepublik Deutschland: Kommentar, GG Art. 12 Rn. 10 m. w. N.

68 Für die Unionsgrundrechte: EuGH, Urt. v. 29.01.2008 - C-275/06, GRUR 2008, 241 Rn. 66 - *Promusicae/Telefónica*; EuGH, Urt. v. 22.01.2013 - C-283/11, ZUM 2013, 202 Rn. 60 - *Sky Österreich*; EuGH, Urt. v. 27.03.2014 - C-314/12, NJW 2014, 1577 Rn. 63 - *UPC Telekabel*; EuGH, Urt. v. 26.04.2022 - C-401/19, GRUR 2022, 820 Rn. 66 - *Polen/Parlament und Rat*. Aus dem nationalen Recht: BVerfG, Beschl. v. 24.02.1971 - 1 BvR 435/68, NJW 1971, 1645 (1647) - *Mephisto*; BVerfG, Beschl. v. 03.11.1987 - 1 BvR 1257/84 u. a., NJW 1988, 325 (326) - *Herrnburger Bericht*; vgl. auch BVerfG, Beschl. v. 26.05.1970 - 1 BvR 83, 244, 345/69, NJW 1970, 1729 (1730) - *Dienstpflichtverweigerung*; *Epping*, Grundrechte, S. 33 Rn. 79.

69 Vgl. dazu auch *Streinz/Michl*, in: *Streinz*, EUV/AEUV, EU-GRCh Art. 52 Rn. 17; *Schwerdtfeger*, in: *Meyer/Hölscheidt*, Charta der Grundrechte der Europäischen Union, EU-GRCh Art. 52 Rn. 39.

70 Im Unionsrecht folgt das schon daraus, dass der EuGH von einem „Gleichgewicht“ der kollidierenden Grundrechte spricht, siehe EuGH, Urt. v. 29.01.2008 - C-275/06, GRUR 2008, 241 Rn. 66 - *Promusicae/Telefónica*; EuGH, Urt. v. 22.01.2013 - C-283/11, ZUM 2013, 202 Rn. 60 - *Sky Österreich*; EuGH, Urt. v. 27.03.2014 - C-314/12, NJW 2014, 1577 Rn. 63 - *UPC Telekabel*; EuGH, Urt. v. 26.04.2022 - C-401/19, GRUR 2022, 820 Rn. 66 - *Polen/Parlament und Rat*. Dies setzt eine bestmögliche Geltung aller kollidierender Grundrechtspositionen voraus. Für das nationale Recht siehe BVerfG, Beschl. v. 07.03.1990 - 1 BvR 266/86, 1 BvR 913/87, NJW 1990, 1982 (1983) - *Bundesflagge*.

Auch bei der Identifizierung von Freistellungsbedürfnissen kann dieser Ausgleichsmechanismus zur Anwendung gebracht werden. Mit seiner Hilfe lassen sich im konkreten Fall aufeinandertreffende, grundrechtlich geschützte Interessen von Urhebern sowie Trainingsdatensammlern, KI-Entwicklern oder Anwendern von KI-Systemen zum Ausgleich bringen.⁷²

Folge der Grundrechtskollision ist auf nationaler Ebene nicht nur ein Abwägungsraum für die Judikative bei Auslegung des geltenden Rechts. Vielmehr folgt aus ihm schon ein Auftrag an den nationalen Gesetzgeber, die widerstreitenden Interessen zum Ausgleich zu bringen.⁷³ Auf unionsrechtlicher Ebene könnte eine entsprechende Förderungs- und damit auch Ausgleichspflicht innerhalb des Kompetenzbereichs des europäischen Normgebers mitunter über Art. 51 Abs. 1 S. 2 EU-GRCh begründet werden.⁷⁴ Aus beidem folgt, dass der Ausgleich grundrechtlich abgesicherter Urheber- und Nutzerinteressen ein Leitmotiv für die Ermittlung normativer Regelungsbedürfnisse ist. Grundrechtspositionen der Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen bilden damit einen ersten Anknüpfungspunkt, über den ihre Interessen bei der Identifizierung von Regelungsbedürfnissen berücksichtigt werden können.

bb) Berücksichtigung von Nutzerinteressen als sekundäres Regelungsziel des europäischen Urheberrechts

Darüber hinaus könnten Nutzerinteressen aber auch dann berücksichtigt werden, wenn kein angemessener Ausgleich spezieller, kollidierender Grundrechtspositionen erforderlich ist. Denn Regelungsziel der InfoSoc-RL⁷⁵ des Unionsgesetzgebers, die sowohl die Verwertungsrechte als auch

71 Beispielsweise in Begriff geprägt von *Hesse*, Grundzüge des Verfassungsrechts der Bundesrepublik Deutschland, S. 142 Rn. 317; verwendet beispielsweise in BVerfG, Beschl. v. 18.02.2019 - 1 BvR 2556/17, GRUR 2019, 606 Rn. 9 – Loud; *Epping*, Grundrechte, S. 93.

72 Dazu auch allgemein *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 45 ff.

73 BVerfG, Beschl. v. 27.01.2015 - 1 BvR 471/10, 1 BvR 1181/10, NJW 2015, 1359 Rn. 98 – Kopftuchverbot.

74 *Schwerdtfeger*, in: Meyer/Hölscheidt, Charta der Grundrechte der Europäischen Union, EU-GRCh Art. 51 Rn. 68.

75 Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft (InfoSoc-RL).

einen wesentlichen Teil der urheberrechtlichen Schrankenbestimmungen harmonisiert und damit einen wesentlichen Grundstein des europäischen Urheberrechts verkörpert, sowie der DSM-RL⁷⁶ ist nicht nur der Schutz der urheberrechtlichen Verwertungsinteressen. Vielmehr dienen die Richtlinien gemäß ErwGr. 31 S.1 InfoSoc-RL sowie ErwGr. 6 S.1 DSM-RL auch dem Rechts- und Interessenausgleich zwischen Rechteinhabern und Nutzern. Beide Erwägungsgründe bringen einen allgemeinen Rechtsgedanken zum Ausdruck. Dieser wird auch in ErwGr. 1 verwaiste Werke-RL⁷⁷ anerkannt. Nach ihm bietet das europäische Urheberrecht einen Anknüpfungspunkt für die Berücksichtigung von Nutzerinteressen bei der Ausgestaltung des Rechtsrahmens unabhängig davon, ob diese Interessen durch spezielle Freiheitsrechte geschützt sind und damit in ein grundrechtlich geprägtes Spannungsverhältnis mit dem Urheberrecht eintreten.

cc) Sozialbindung des Urheberrechts

Ein dritter Bezugspunkt für die Berücksichtigung von Nutzerinteressen abseits des Ausgleichs spezieller, kollidierender Grundrechtspositionen könnte im nationalen Regelungszusammenhang zudem über Art.14 GG bestehen, soweit für ihn im konkreten Fall Anwendungsspielräume verbleiben. Er sieht in Art.14 Abs.1 S.2 GG einen Gesetzesvorbehalt vor. Dieser ermöglicht der Legislative, das Urheberrecht zur Verfolgung anderer Zwecke einzuschränken.⁷⁸ Hierbei muss aber stets die in Art.14 Abs.2 GG verankerte Sozialbindung des Eigentums berücksichtigt werden. Sie umfasst auch das Urheberrecht als geistiges Eigentum.⁷⁹

76 Richtlinie 2019/790/EU des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG (DSM-RL).

77 Richtlinie 2012/28/EU des Europäischen Parlaments und des Rates vom 25. Oktober 2012 über bestimmte zulässige Formen der Nutzung verwaister Werke (verwaiste Werke-RL).

78 Vgl. BVerfG, Beschl. v. 07.07.1971 - 1 BvR 276/71, GRUR 1972, 487 – Schulfunksendungen.

79 So schon das RG, Urt. v. 26.04.1933 - ZR I 301/32, RGZ 140, 264 (270) – Verlagsrecht an sämtlichen Ausgaben; im Weiteren *Götting*, in: Loewenheim, Handbuch des Urheberrechts, § 30 Rn.1; allgemein zur Sozialbindung und dem Gesetzesvorbehalt in Art.14 GG BVerfG, Urt. v. 10.03.1981 - 1 BvR 92/71, 1 BvR 96/71, NJW 1981, 1257 – Private Gondelbahn.

Von der Sozialbindung des Eigentums geht eine legitimierende, gleichzeitig aber auch eine begrenzende Wirkung für Einschränkungen des Urheberrechts aus.⁸⁰ Das Urheberrecht darf nur zugunsten öffentlicher, also solcher Interessen eingeschränkt werden, die für die gesamte oder zumindest einen Teil der Gesellschaft von Bedeutung sind.⁸¹ Die Sozialbindung ermöglicht eine Einschränkung allerdings auch dann, wenn die öffentlichen Interessen im Speziellen nicht unmittelbar grundrechtlich geschützt sind. Allgemeininteressen können auch Interessen der Gesamtheit aller Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen als Nutzergruppen der Allgemeinheit sein. Möglich ist deswegen, dass die Interessen der Urheber über Art. 14 Abs. 1 S. 2, Abs. 2 GG auch gegenüber nicht in speziellen Grundrechten widerspiegelten Interessen dieser Nutzergruppen zurücktreten müssen.

Aus Art. 14 Abs. 1 S. 2 GG kann dabei ein konkreter Regelungsauftrag an den Gesetzgeber abgeleitet werden.⁸² Er ist angehalten, proaktiv den Ausgleich zwischen kollidierenden Urheber- und öffentlichen Interessen vorzunehmen, um die Sozialbindung des geistigen Eigentums zu verwirklichen.⁸³ Bei der Analyse, wie die urheberrechtlichen Regelungen ausgestaltet sein sollen, bilden der Gesetzesvorbehalt des Art. 14 Abs. 1 S. 2 GG und die Sozialbindung des Urheberrechts damit einen dritten, national geprägten Anknüpfungspunkt, über den objektivierte Interessen der Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen berücksichtigt werden können.

b) Funktion des Ausgleichsmechanismus

Klärungsbedürftig bleibt, über welchen Mechanismus der Ausgleich von objektivierten Urheber- und Nutzerinteressen vorgenommen werden kann.

80 BVerfG, Beschl. v. 02.03.1999 - 1 BvL 7-91, NJW 1999, 2877 (2878); für das Urheberrecht: *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 43.

81 *Götting*, in: Loewenheim, Handbuch des Urheberrechts, § 30 Rn.1; oft auch als „Allgemeininteressen“ bezeichnet, vgl. beispielsweise BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (46) – Kirchenmusik; *Schack*, Urheber- und Urhebervertragsrecht, S. 48 Rn. 93; BT-Drs. IV/270, S. 30.

82 BVerfG, Urt. v. 06.12.2016 - 1 BvR 2821/11, 1 BvR 321/12, 1 BvR 1456/12, NJW 2017, 217 Rn. 218 – Atomausstieg; BVerfG, Beschl. v. 12.06.1979 - 1 BvL 19/76, NJW 1980, 985 (987).

83 BVerfG, Beschl. v. 12.06.1979 - 1 BvL 19/76, NJW 1980, 985 (987).

Erster Anknüpfungspunkt zur Berücksichtigung der Interessen von Trainingsdatensammlern, KI-Entwicklern und Anwendern von KI-Systemen sind ihre Grundrechtspositionen, die mit der Grundrechtsposition der Urheber kollidieren können. Aus einer europarechtlichen Perspektive darf der Unionsgesetzgeber in diesen Fällen in jedes der betroffenen Grundrechte nur so weit eingreifen, wie es verhältnismäßig ist. Dies folgt unmittelbar aus Art. 52 Abs. 1 S. 2 EU-GRCh.⁸⁴ Auch beim Ausgleich kollidierender Grundrechtspositionen auf national-verfassungsrechtlicher Ebene kommt es im Rahmen der Anwendung des Grundsatzes der praktischen Konkordanz darauf an, inwieweit die Einschränkung des zurückstehenden Grundrechts zwingend erforderlich,⁸⁵ also verhältnismäßig ist.⁸⁶ Der auf beiden Stufen des europäischen Mehrebenensystems zum Tragen kommende Verhältnismäßigkeitsgrundsatz schließt dabei eine umfassende Abwägung der in Widerspruch zueinander stehenden Grundrechtspositionen ein.⁸⁷ Der Ausgleich der Urheberinteressen mit den grundrechtlich geschützten Interessen der bezeichneten Nutzergruppen käme infolgedessen vor allem durch eine umfassende Interessenabwägung zustande.

Eine Konkretisierung der Funktionsweise des Interessenausgleichsmechanismus, den ErwGr. 31 S. 1 InfoSoc-RL sowie ErwGr. 6 S. 1 DSM-RL in Bezug nehmen, findet in den jeweiligen Erwägungsgründen darüber hinaus nicht statt. Dort, wo die Durchführung des Interessenausgleichs von der Normsetzungs- auf die Anwendungsphase verlagert wurde, wie es insbesondere beim in Art. 5 Abs. 5 InfoSoc-RL verankerten Dreistufentest der Fall ist, wird jedoch deutlich, dass der bezweckte Interessenausgleich auch gemäß der Richtlinienkonzeption im Wesentlichen durch eine umfassende Interessenabwägung gewährleistet wird.⁸⁸ Im Rahmen des im Richtlinienrecht zum Ausdruck kommenden Interessenausgleichs kommt es demzufolge

84 Vgl. dazu auch EuGH, Urt. v. 22.01.2013 - C-283/11, ZUM 2013, 202 Rn. 47 ff. – Sky Österreich; EuGH, Urt. v. 27.03.2014 - C-314/12, NJW 2014, 1577 Rn. 63 – UPC Telekabel; *Streinz/Michl*, in: Streinz, EUV/AEUV, EU-GRCh Art. 52 Rn. 17; *Schwerdtfeger*, in: Meyer/Hölscheidt, Charta der Grundrechte der Europäischen Union, EU-GRCh Art. 52 Rn. 39.

85 BVerfG, Beschl. v. 26.05.1970 - 1 BvR 83, 244, 345/69, NJW 1970, 1729 (1730) – Dienstpflichtverweigerung; *Hesse*, Grundzüge des Verfassungsrechts der Bundesrepublik Deutschland, S. 142 Rn. 318.

86 *Hesse*, Grundzüge des Verfassungsrechts der Bundesrepublik Deutschland, S. 142 Rn. 318.

87 Für die unionsrechtliche Dogmatik *Kingreen*, in: Calliess/Ruffert, EUV/AEUV, EU-GRCh Art. 52 Rn. 70 ff. Für das nationale Recht siehe *Grabitz*, AÖR 1973, 568 (575).

88 Zum Dreistufentest im Detail siehe 3. Teil A. III. 2. d).

ge ebenfalls auf eine freie Gegenüberstellung der betroffenen Urheber- und Nutzerinteressen an.

Auch Einschränkungen des Urheberrechts, die aus der national angelegten Sozialbindung des Immaterialgüterrechts folgen, müssen sich am Verhältnismäßigkeitsgrundsatz messen lassen.⁸⁹ In diesen Fällen müssten die gegenläufigen Urheber- und Nutzerinteressen demzufolge ebenfalls in eine umfassende Abwägung zueinander gestellt werden.

Unabhängig davon, über welchen Anknüpfungspunkt die jeweilig betroffenen Nutzerinteressen berücksichtigt werden können, ist die umfassende Abwägung mit den Interessen betroffener Urheber damit zentraler Ausgleichsmechanismus. Hieraus folgt für die vorliegende Untersuchung, dass die umfassende Interessenabwägung als Ausgleichsinstrument angewendet werden kann, um Freistellungsbedürfnisse zu identifizieren. Die urheberrechtliche Freistellung einer der im Folgenden zu untersuchenden Sachverhaltskonstellationen ist neben der Nichtbeeinträchtigung der Urheberinteressen demzufolge auch dann zweckmäßig, wenn eine umfassende Interessenabwägung dazu führt, dass den Belangen der betroffenen Nutzergruppe im konkreten Fall Vorrang vor den Urheberinteressen zu gewähren ist.⁹⁰

3. Berücksichtigungsfähigkeit einer ökonomischen Analyse des Urheberrechts

Auch die ökonomische Analyse des Urheberrechts könnte darüber hinaus zur Identifizierung von Freistellungsbedürfnissen geeignet sein.

a) Überblick über die Funktionsweise

Nach der volkswirtschaftlich geprägten ökonomischen Analyse des Rechts müsse die Rechtsordnung das Ziel verfolgen, einen Zustand der Effizienz herbeizuführen.⁹¹ Er ist erreicht, wenn alle wirtschaftlichen Mittel innerhalb der Gesellschaft so verteilt sind, dass die Bedürfnisse aller bestmöglich, also aus einem gesamtwirtschaftlichen Blickwinkel optimal, befriedigt

89 BVerfG, Beschl. v. 08.04.1987 - 1 BvR 564, 684, 877, 886, 1134, 1636, 1711/84; BVerfG, Beschl. v. 26.04.1995 - 1 BvL 19/94 u.a., DtZ 1995, 323 (324).

90 Vgl. insbesondere *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 96; *Peukert*, Urheberrecht, § 2 Rn. 5; BT-Drs. IV/270, S. 30.

91 *Schäfer/Ott*, Lehrbuch der ökonomischen Analyse des Zivilrechts, S. XI.

werden.⁹² In einem solchen Zustand bedeutete jede Besserstellung eines Beteiligten unweigerlich die Schlechterstellung einer anderen Person.⁹³

Um zu bestimmen, ob eine rechtliche Regelung dieses Effizienzziel fördert, sind daher stets ihre Folgen in den Blick zu nehmen.⁹⁴ Wichtiger Maßstab zur Beurteilung der Folgen einer Regelung ist das Pareto-Kriterium. Hiernach ist ein Zustand einem anderen vorzuziehen, wenn er zumindest von einem Beteiligten präferiert, gleichzeitig aber von keinem anderen Beteiligten abgelehnt wird.⁹⁵ Ob der wirtschaftlich rational handelnde Mensch die Anpassung der Rechtsordnung gegenüber dem Verbleib im Ist-Zustand vorzieht, hängt maßgeblich davon ab, welche wirtschaftlichen Folgen die Rechtsänderung hätte. Die anzustellende Kosten-Nutzen-Abwägung muss stets positiv ausfallen.⁹⁶

Neben Pareto-Superiorität können Anpassungen der Rechtsordnung das Effizient-Ziel auch fördern, wenn sie das Kaldor/Hicks-Kriterium erfüllen. Dazu muss es der wirtschaftlich bevorteilten Gruppe möglich sein, die Benachteiligten monetär zu entschädigen, ohne selbst ihren erlangten Vorteil vollständig aufzubrechen.⁹⁷

b) Eignung der ökonomischen Analyse des Rechts für die Beantwortung der aufgeworfenen Fragestellungen

Unklar ist aber, ob mit Hilfe der ökonomischen Analyse des Urheberrechts aussagekräftige Antworten auf die zu untersuchenden Fragestellungen gefunden werden können.

Hierfür spricht zunächst, dass die ökonomische Rechtsanalyse einen objektiven Maßstab verspricht, anhand derer Regelungsbedürfnisse zur Erreichung von Allokationseffizienz identifiziert werden können.⁹⁸ Darüber hinaus führt die ökonomische Analyse des Rechts zu einer Vereinfachung

92 Schäfer/Ott, Lehrbuch der ökonomischen Analyse des Zivilrechts, S. XI.

93 Schäfer/Ott, Lehrbuch der ökonomischen Analyse des Zivilrechts, S. 14.

94 Vgl. Eidenmüller, Effizienz als Rechtsprinzip, S. 3 ff.

95 Eidenmüller, Effizienz als Rechtsprinzip, S. 48; Schäfer/Ott, Lehrbuch der ökonomischen Analyse des Zivilrechts, S. 13.

96 Kirchgässner, Homo oeconomicus, S. 12.

97 Eidenmüller, Effizienz als Rechtsprinzip, S. 51.

98 Vgl. Maamar, Computer als Schöpfer, S. 92 („quasi-mathematische Argumentation“).

komplexer Sachverhalte in eine Modellstruktur.⁹⁹ Sie hilft dabei, die vielschichtigen Sachverhalte Lösungen zuzuführen.¹⁰⁰

Allerdings müssen bei der Ausgestaltung der Rechtsordnung auch der Grundrechtsschutz der Urheber, diverse Nutzergrundrechte und weitere Allgemeininteressen berücksichtigt werden.¹⁰¹ Das Modelldenken verhindert jedoch, dass diese Bandbreite normativ abzubildender Interessen in die Beurteilung eines Freistellungsinteresses mit einbezogen werden können.¹⁰² Die vielschichtigen Sachverhalte auf das rechtsökonomische Effizienzziel zu reduzieren, welches selbst weder in den Primärrechtsakten der Europäischen Union noch in der nationalen Verfassung¹⁰³ abgebildet ist, wird der tatsächlichen Komplexität der zu untersuchenden Sachverhaltskonstellationen nicht gerecht. Zwar könnten diese Aspekte durch eine spätere Korrektur des ökonomischen Analyseergebnisses zugunsten anderer, höherwertiger und insbesondere nicht ökonomischer Interessen berücksichtigt werden.¹⁰⁴ Gerade wenn das Ergebnis der ökonomischen Analyse des Urheberrechts aber nur insoweit für maßgeblich erachtet wird, als es mit anderen, höherwertigen Interessen nicht kollidiert, ist die zuvor vorgenommene Analyse der Regelungsbedürfnisse auf Basis ökonomischer Modelle nicht von Bedeutung.¹⁰⁵ Es käme vielmehr nur auf die sich anschließende Interessenabwägung an.

Darüber hinaus haben urheberrechtliche Regelungen, die auf konkrete, eng abgegrenzte Arbeitsschritte wie die hier in Rede stehenden bezogen sind, nur sektoral begrenzte Auswirkungen auf die zu erwartenden Kosten sowie die generierbaren Vorteile für Konsumenten, also Nutzer wie Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen. Gleiches gilt für die Kosten-Nutzen-Rechnung der Urheber. Aus solchen nanoökonomischen Beobachtungen lassen sich aber keine Auswirkungen

99 *Bisges*, ZUM 2014, 930 (930); *Maamar*, Computer als Schöpfer, S. 92.

100 *Bischoffshausen*, Die ökonomische Rechtfertigung der urheberrechtlichen Schutzfrist, S. 259 spricht von einer „Verringerung der Komplexität juristischer Problemstellungen“; Auch *Maamar*, Computer als Schöpfer, S. 92.

101 Dazu bereits oben unter 2. Teil A. I. 1. sowie 2.

102 So bereits auch *Fezer*, JZ 1988, 223 (224).

103 *Eidenmüller*, Effizienz als Rechtsprinzip, S. 444 f.; siehe außerdem BVerfG, Urt. v. 01.03.1979 - 1 BvR 532, 533/77, 419/78, 1 BvL 21/78, NJW 1979, 699 (702) – Mitbestimmung: Das BVerfG sieht im Grundgesetz keine Wirtschaftsordnung angelegt; vgl. außerdem *Fezer*, JZ 1988, 223 (224).

104 So *Schäfer/Ott*, Lehrbuch der ökonomischen Analyse des Zivilrechts, S. XVIII; auch *Maamar*, Computer als Schöpfer, S. 138.

105 Ähnlich auch *Fezer*, JZ 1988, 223 (228).

auf die Innovations- und Verbreitungsanreize für Urheber, die Höhe der noch verbleibenden Urheberrente insgesamt oder ein sektoral hervorgerufenes Marktversagen ableiten. Es fehlt insbesondere auch an einer entsprechenden Datenlage.¹⁰⁶ Eine volkswirtschaftliche Prognose, ob beispielsweise die Freistellung des Web Scrapings von Werken als Trainingsdaten spezifisch den Anreiz für Urheber verringert, neue Werke zu schaffen oder die Urheberrente erheblich mindert, lässt sich demzufolge nicht abgeben. Vorhersagen zu Auswirkungen auf Indikatoren, die die ökonomische Analyse der urheberrechtlichen Rechtsnormen maßgeblich mitprägen, sind in den hier betreffenden Fällen damit nicht möglich.

In diesem Zusammenhang muss auch berücksichtigt werden, dass leistungsfähige KI-Systeme erst seit wenigen Jahren auf dem Massenmarkt zur Verfügung stehen.¹⁰⁷ Es sind noch keine empirischen Untersuchungen betreffend die ökonomischen Auswirkungen von urheberrechtlichen Regelungen auf die Erreichung der Allokationseffizienz im Kontext des Einsatzes von KI durchgeführt worden.¹⁰⁸ Ohne eine empirische Datengrundlage ist

106 Allein eine im Jahr 2015 veröffentlichte Studie deutet an, dass die konkrete Ausgestaltung des urheberrechtlichen Schutzes keine Auswirkungen auf die Schaffenskraft der Urheber haben könnte und demzufolge überhaupt keinen Innovationsanreiz ausübe. Hierüber abschließende Aussagen zu treffen, sei mangels ausreichender Datenlage aber noch nicht möglich gewesen. Siehe *Handke/Girard/Mattes*, *Fördert das Urheberrecht Innovation? Eine empirische Untersuchung*, S. 105, 121.

107 Bildklassifikationsfähigkeiten von KI-Systemen haben sich in den letzten zehn Jahren um 27,7 % verbessert (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 81.), Fähigkeiten der Objekterkennung um 26 % seit 2015 (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 88.), der beste Q→AR-Score, ein Maßstab zur Bestimmung der Leistungsfähigkeit von Systemen für optisches Bildverständnis, hat sich seit 2018 um 66 % erhöht (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 95.), der SuperGLUE-Score zur Bestimmung der Leistungsfähigkeit von KI-Systemen zum Verständnis englischer Sprache um 8 % seit 2019, außerdem wurde in diesem Bereich im Jahr 2020 die Human Baseline, die durchschnittliche menschliche Leistungsfähigkeit, überschritten (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 100.), die Genauigkeit des Leseverständnisses von KI-Systemen nahm seit 2020 um rund 17 % zu (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 101.), die Fehlerquote von Spracherkennungssystemen verbesserte sich seit 2017 von über 7 % auf 0,14 % (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 109.), im Bereich der verwendeten Rechenhardware verdreifachte sich seit 2021 zugleich der Median der FLOP/s-Geschwindigkeit (*Maslej u. a.*, *The AI Index 2023 Annual Report by Stanford University*, S. 118.).

108 Das ergibt sich aus zwei systematischen Auswertungen der rechtsökonomischen Forschung, siehe *Peukert/Windisch*, *Center for Law & Economics Working Paper Series ETH Zürich 2023*; *Margoni*, 21 for 2021: *Computational Uses – CREATE*. Au-

die rechtsökonomische Analyse der aufgeworfenen Fragestellungen damit aber nicht belastbar.¹⁰⁹

Zwar könnte dem entgegengehalten werden, eine spätere Änderung der zur Analyse genutzten Hypothesen und Prognosen beeinträchtigt nicht die Tauglichkeit der ökonomischen Analyse des Rechts als Instrument für die Entscheidungsfindung als solche.¹¹⁰ Vielmehr könnten die Analyseergebnisse bei einer Änderung der zugrundeliegenden empirischen Grundannahmen in der Zukunft angepasst werden.¹¹¹ Allerdings ist Ziel dieser Arbeit, bereits im Zeitpunkt der Untersuchung tragfähige Antworten auf die regulatorischen Fragestellung herauszuarbeiten. Eine Verweis auf die Reversibilität der regulatorischen Erwägungen ist in diesem Rahmen damit nicht ausreichend. Insgesamt ist die ökonomische Analyse des Urheberrechts mithin kein Instrument, welches im Rahmen dieser Untersuchung dafür geeignet ist, regulatorische Freistellungsbedürfnisse zu identifizieren.¹¹²

Weil das Effizienzziel auf die Maximierung der Wohlfahrt aller Gesellschaftsmitglieder gerichtet ist, erfüllt es aber dennoch die Voraussetzungen aus der Sozialbindung des geistigen Eigentums gemäß Art. 14 Abs. 1 S. 2, Abs. 2 GG. Damit kann es dennoch als Allgemeininteresse im Rahmen der freien, umfassenden Interessenabwägung berücksichtigt werden. Das gilt auch für das Coase Theorem, nach dem geringe beziehungsweise keine Transaktionskosten für die Übertragung von Gütern auf einem Markt dazu führen, dass dieser sich wohlfahrtsökonomisch autonom reguliert.¹¹³

ßerdem wurde zuletzt am 9.10.2024 eine ergebnislose Stichwortsuche zu „Artificial Intelligence“ und „Künstliche Intelligenz“ bei *copyright evidence* durchgeführt, einer enzyklopädischen Datenbank von empirischen Studien zum Urheberrecht, siehe <https://www.copyrightevidence.org> (zuletzt abgerufen am 9.10.2024).

109 Ähnlich auch *Hennemann*, Interaktion und Partizipation, S. 317; grundsätzlich auch *Maamar*, Computer als Schöpfer, S. 93, der im Anschluss aber dennoch eine rechtsökonomische Analyse vornimmt.

110 *Maamar*, Computer als Schöpfer, S. 93.

111 *Maamar*, Computer als Schöpfer, S. 93.

112 Anders *Maamar*, Computer als Schöpfer, S. 93.

113 *Coase*, *The Journal of Law and Economics* 1960, 1 (10, 15 f.); mit Beispielen *Schäfer/Ott*, Lehrbuch der ökonomischen Analyse des Zivilrechts, S. 80 ff.; zu einer kritischen Einordnung *Eidenmüller*, Effizienz als Rechtsprinzip, S. 91 ff.

4. Techniksensibilität des Urheberrechts oder der „more technological approach“

Fraglich ist schließlich, inwiefern auch technische Notwendigkeiten beim Einsatz von KI im Rahmen der regulatorischen Analyse berücksichtigt werden können.

Zum einen könnte in diesem Zusammenhang ein *more technological approach*¹¹⁴ genommen werden. Hiernach muss das Urheberrecht besonders für die systemischen Notwendigkeiten beim Einsatz von Technik aufgeschlossen sein.¹¹⁵ Nutzungen urheberrechtlich geschützter Werke, die beim Einsatz von KI aus einer technischen Perspektive notwendig oder jedenfalls unvermeidlich sind, sollten demnach ermöglicht werden, insofern der Einsatz von KI als solches förderungswürdig ist. Neben der Freistellung aufgrund einer fehlenden Beeinträchtigung der Urheberinteressen oder einer umfassenden Abwägung der Urheber- und Nutzerinteressen wäre dies demzufolge ein dritter Ansatzpunkt, um Freistellungsbedürfnisse für die zu untersuchenden Fallgruppen zu identifizieren.

Gegen einen solchen *more technological approach* spricht allerdings, dass Aufgabe des Urheberrechts nicht primär die Steuerung des Einsatzes von Technik, sondern der Schutz der Urheber und ihrer Werke sowie der Ausgleich von Urheber- mit Nutzerinteressen ist.¹¹⁶ Auch wenn die Reflexion technischer Möglichkeiten und Notwendigkeiten im urheberrechtlichen Kontext notwendig ist, können aus technischen Abläufen nicht unmittelbar normative Entscheidungen abgeleitet werden.¹¹⁷

Nichtsdestotrotz muss das Urheberrecht aber techniksensibel ausgestaltet und angewendet werden.¹¹⁸ Das bedeutet, dass technische Besonderheiten bei der regulatorischen Ausgestaltung ebenfalls mitzubehütenden sind.¹¹⁹ In die umfassende Abwägung von Interessenabwägung können demzufolge auch technische Bedürfnisse mit einfließen.¹²⁰

114 Podszun/Grünberger, ZGE 2014, 269.

115 Podszun/Grünberger, ZGE 2014, 269 (270).

116 Hofmann, ZGE 2016, 482 (501).

117 Godt, ZGE 2014, 279 (281); sich anschließend auch Hofmann, ZGE 2016, 482 (511).

118 Hofmann, ZGE 2016, 482 (511); de la Durantaye/Hofmann, ZUM 2021, 873 (877).

119 Becker, ZGE 2016, 239 (267); Hofmann, ZGE 2016, 482 (511); de la Durantaye/Hofmann, ZUM 2021, 873 (877); so schlussendlich auch Grünberger, ZUM 2015, 273 (275 f.).

120 Becker, ZGE 2016, 239 (267).

5. Zusammenfassung: Maßstab für die regulatorische Analyse

Alles in allem sollten die zu untersuchenden Fallgruppen, also das Web Scraping von Werken zum Training von KI, die eigentliche Durchführung des Trainingsvorgangs auf Basis von Werken, die Speicherung urheberrechtlich geschützter Inputdaten für die Anwendung von KI-Systemen sowie die Vorverarbeitung der als Inputdaten erfassten Werke vom Ausschließlichkeitsrecht betroffener Urheber ausgenommen sein, wenn die Freistellung regulatorisch zweckmäßig ist. Das ist zum einen der Fall, wenn in den jeweiligen Sachverhaltskonstellationen keine Urheberinteressen berührt werden. Zum anderen ist eine Freistellung aber auch dann zweckmäßig, wenn die Nutzer- die betroffenen Urheberinteressen im Rahmen einer umfassenden Abwägung überwiegen. Hierbei können auch ökonomische, wohlfahrtsorientierte Aspekte berücksichtigt werden.

II. Bestimmung des Freistellungsmechanismus

Für die Umsetzung eines identifizierten Freistellungsbedürfnisses kommt es weiterhin darauf an, welche geeigneten Freistellungsmechanismen im urheberrechtlichen Regelungsregime überhaupt zur Verfügung stehen, welche Wirkungen und Unterschiede ihnen mit Blick auf die betroffenen Urheber jeweils zukommen und anhand welcher Beurteilungskriterien der interessengerechte Regelungsmechanismus identifiziert werden kann.

1. Zur Verfügung stehende Freistellungsmechanismen

Damit ein Regelungsmechanismus für eine Freistellung im Sinne der vorliegenden Untersuchung in Betracht kommt, muss er Trainingsdatensammeln, KI-Entwicklern oder Anwendern von KI-Systemen ermöglichen, die betroffenen Werke ohne individuelle Zustimmung der Urheber zu verwenden.

- a) Unvergütete Schutzbereichsbegrenzung: Unmittelbar begrenzende Ausgestaltung der Verwertungsrechte und vergütungsfreie Schrankenbestimmung

Möglich ist dabei zum einen, die Freistellung ohne angemessene Vergütung betroffener Urheber auf Ebene des urheberrechtlichen Schutzbereichs

umzusetzen. Folge einer solchen Begrenzung wäre, dass die betreffende Sachverhaltskonstellation wie beispielsweise das Web Scraping von Werken zum Training von KI nicht vom Ausschließlichkeitsrecht der Urheber umfasst würde. Damit bestünde kein Verbot der Urheber, sodass die jeweilige Handlung erlaubnisfrei vorgenommen werden dürfte. Individuelle Nutzungsrechte einzuholen, wäre demzufolge nicht mehr nötig. Eine vergütungsfreie Begrenzung des urheberrechtlichen Schutzbereichs kann dabei einmal mit Hilfe einer vergütungsfreien Schrankenbestimmung (§§ 44a ff. UrhG) oder durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte in den §§ 15 ff. UrhG erfolgen.

b) Gesetzliche Lizenz

Eine Freistellung könnte darüber hinaus auch im Wege einer gesetzlichen Lizenz umgesetzt werden. Hierbei handelt es sich um eine Art von Schrankenbestimmung, durch die die betreffende Werknutzung zwar freigestellt, zugunsten der Urheber allerdings vergütungspflichtig ausgestaltet wird. Nutzer werden damit in die Lage versetzt, die gewünschten Werke beispielsweise zum Training ihrer KI zu speichern, ohne Individuallizenzen einholen zu müssen. Trotzdem müssen sie den Urhebern im Gegenzug eine dem wirtschaftlichen Wert der Nutzung entsprechende angemessene Vergütung gewähren.¹²¹

Die Bemessung dieser Vergütung kann pauschalisiert erfolgen, wenn eine genauere Abrechnung mit unzumutbarem Aufwand oder zu hohen Transaktionskosten verbunden oder sogar unmöglich wäre.¹²² Insbesondere Trainingsdatensammler speichern mit Hilfe von Webcrawlern automatisiert große Datenmengen. Dabei haben sie zumeist keinen genauen Überblick über die gespeicherten Inhalte. Auch eine manuelle nachträgliche Durchsicht ist mit Rücksicht auf die Größe typischer Trainingsdatensätze unzumutbar.¹²³ Eine gesetzliche Lizenz für das Web Scraping von Werken als Trainingsdaten könnte damit insbesondere mit einem pauschal veran-

121 Zur Vergütungshöhe: *Pflüger*, Gerechter Ausgleich und angemessene Vergütung, S.145; insbesondere zu anderen bei der Bestimmung der Vergütungshöhe berücksichtigungsfähigen Aspekten: *Lauber-Rönsberg*, Urheberrecht und Privatgebrauch, S.112 ff.

122 So z. B. bei der Vergütung zur Nutzung von Werken als Unterrichts- und Lehrmedien, § 60a UrhG, siehe *Hagemeier*, in: *Ahlberg/Götting/Lauber-Rönsberg*, Beck'scher Online-Kommentar Urheberrecht, UrhG § 60h Rn. 10.

123 Vgl. zur Größe der Trainingsdatensätze oben unter 1. Teil B. II. 3.

schlagten gesetzlichen Vergütungsanspruch kombiniert werden, der über Verwertungsgesellschaften abgewickelt wird.

c) Erweiterte kollektive Lizenz

Eine Freistellung der zu untersuchenden Sachverhaltskonstellationen ist darüber hinaus auch über einen erweiterten kollektiven Lizenzmechanismus (ECL) möglich.¹²⁴ Hierbei handelt es sich um einen Lizenzmechanismus, der Verwertungsgesellschaften ermöglicht, Lizenzen auch zur Nutzung solcher Werke zu erteilen, die nicht Teil ihres Repertoires sind.¹²⁵ Bei Umsetzung der Freistellung im Wege eines ECL-Mechanismus müssten Nutzer wie Trainingsdatensammler zwar immer noch eine Lizenz für die Nutzung von Werken beispielsweise zum Training von KI abschließen. Da die Lizenzvereinbarung aber kollektiv mit einer Verwertungsgesellschaft getroffen und für eine unbestimmte Zahl von Werken einer Vielzahl von Urheber gelten würde, erfüllt der ECL-Mechanismus dennoch die Voraussetzungen, die an eine Freistellung im Sinne der vorliegenden Untersuchung gestellt werden.

ECL können beispielsweise durch Kulturerbe-Einrichtungen wie öffentlichen Bibliotheken zur Nutzung nicht verfügbarer Werke gemäß § 52 VGG abgeschlossen werden. Außenstehende Urheber, deren Werke von der erweiterten Wirkung der Lizenzvereinbarung betroffen sind, haben in Folge einer ECL die gleichen Rechte und Pflichten wie solche Urheber, die mit der Verwertungsgesellschaft Wahrnehmungsverträge geschlossen haben. Das betrifft insbesondere auch die Verteilung der Einnahmen aus der Verwertung der Werke (vgl. § 23 VGG). Damit eine Lizenzvereinbarung von der erweiterten Wirkung erfasst wird, müssen aber von der Verwertungsgesellschaft und Lizenznehmern einige Voraussetzungen erfüllt werden. Um ECL-Vereinbarungen abzuschließen, bedarf es insbesondere einer repräsentativen Verwertungsgesellschaft (vgl. § 51b VGG). Nur ihre Tarife haben sich als so angemessen erwiesen, dass es gerechtfertigt ist, sie auch auf Werke außenstehender Rechtsinhaber anzuwenden.¹²⁶

Um ein Leerlaufen der Freistellungsregelung zu verhindern, wenn keine repräsentative Verwertungsgesellschaft besteht, muss der ECL-Mechanismus mit einer subsidiär eingreifenden Schrankenbestimmung verbunden

124 In diese Richtung auch *Rau*, ZUM 2021, 793 (796).

125 Zum Mechanismus genauer *Conrad*, ZUM 2022, 436; *Staats*, ZUM 2019, 703; *Staats*, ZUM 2020, 728; umfassend außerdem *Trumpke*, Exklusivität und Kollektivierung.

126 *Bergström*, GRUR Ausl 1962, 364 (372).

werden. Ein entsprechender Mechanismus ist in den §§ 52 ff. VGG i. V. m. § 61d UrhG auch für die Nutzung vergriffener Werke vorgesehen. Nutzungen auf Basis der Schrankenbestimmung des § 61d UrhG erfolgen dabei vergütungsfrei.

d) Zusammenfassung

Im Ergebnis stehen zur Freistellung des Web Scrapings von urheberrechtlich geschützten Trainingsdaten, der eigentlichen Durchführung des KI-Trainings, der Speicherung von Werken als Inputdaten für die Anwendung von KI-Systemen sowie der Vorverarbeitung dieser urheberrechtlich geschützten Inputdaten vier Freistellungsmechanismen zur Verfügung. Neben einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte und einer unvergüteten Schrankenbestimmung wäre eine Freistellung auch im Wege einer vergütungspflichtigen Schrankenbestimmung (gesetzliche Lizenz) und mit über einen ECL-Mechanismus möglich.

2. Entscheidungskriterien bei der Wahl zwischen den Freistellungsmechanismen

Fraglich ist weiterhin, welche Entscheidungskriterien die Wahl zwischen den zur Verfügung stehenden Freistellungsmechanismen beeinflussen.

a) Angemessener Ausgleich für die Nutzungshandlung

Als Grundlage der Entscheidung dient zum einen, dass zwei der vier Freistellungsmechanismen (gesetzliche Lizenz, erweiterte kollektive Lizenz) eine (teilweise) Vergütung der betroffenen Urheber vorsehen. Die Frage danach, ob den betroffenen Urhebern für die in der jeweiligen Sachverhaltskonstellation maßgebliche Handlung wie beispielsweise der Speicherung ihrer Werke zum KI-Training eine Vergütung zustehen soll, kann damit als erstes Kriterium zur Wahl des geeigneten Freistellungsmechanismus herangezogen werden. Grundsätzlich sollte dem Urheber dabei eine Vergütung für die Freistellung zuerkannt werden, wenn nicht ein gesteigertes öffentli-

ches Interesse ausnahmsweise eine vergütungsfreie Nutzung rechtfertigt.¹²⁷ Ob das der Fall ist, muss anhand einer umfassenden Abwägung ermittelt werden.¹²⁸ Insbesondere, aber nicht nur ausgleichsbedürftig sind demnach Einschränkungen der für die Urheber besonders wirtschaftlich relevanten Kernverwertungsmöglichkeiten.¹²⁹ Insgesamt genügt es aber, wenn die Auslegung und Anwendung einer Freistellungsregelung genügend Spielräume lässt, über die die Vergütungsinteressen betroffener Urheber angemessen berücksichtigt werden können.¹³⁰ Ausreichend ist weiterhin, wenn „unter dem Strich“¹³¹ eine angemessene Vergütung der Urheber für die Schöpfungsleistung gewährleistet bleibt.¹³²

b) Darlegungs- und Beweislast bei vergütungsfreier Nutzung

Ist eine Vergütung der Urheber hingegen nicht nötig, stehen für die Freistellung eine vergütungsfreie Schrankenbestimmung sowie eine unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte zur Verfügung. Eine Differenzierung zwischen beiden Freistellungsmechanismen könnte zum einen anhand der Darlegungs- und Beweislastverteilung vorgenommen werden, die jeweils aus den Freistellungsmechanismen resultiert. Hierfür ist notwendig, dass die Freistellungen im Wege einer Schrankenbestimmung einerseits und einer unmittelbar begrenzenden Ausgestaltung der Verwertungsrechte andererseits hinsichtlich der Darlegungs- und Beweislast unterschiedliche Wirkung entfalten.

Gegen eine unterschiedliche Wirkung von Schrankenbestimmung und unmittelbar begrenzender Ausgestaltung der Verwertungsrechte in Bezug auf die herbeigeführte Darlegungs- und Beweislastverteilung spricht, dass

127 BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (48) – Kirchenmusik; vgl. auch BVerfG, Ur. v. 11.10.1988 - 1 BvR 743/86 u. a., NJW 1992, 1307 (1308) – Justizvollzugsanstalten.

128 Vgl. *Lauber-Rönsberg*, Urheberrecht und Privatgebrauch, S. 112.

129 *Lauber-Rönsberg*, Urheberrecht und Privatgebrauch, S. 112.

130 BVerfG, Ur. v. 31.05.2016 - 1 BvR 1585/13, GRUR 2016, 690 Rn. 79 f. – Metall auf Metall.

131 BVerfG, Ur. v. 31.05.2016 - 1 BvR 1585/13, GRUR 2016, 690 Rn. 79 f. – Metall auf Metall.

132 BVerfG, Beschl. v. 07.07.1971 - 1 BvR 765/66, GRUR 1972, 481 (483) – Kirchen- und Schulgebrauch; BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (46) – Kirchenmusik; BVerfG, Ur. v. 31.05.2016 - 1 BvR 1585/13, GRUR 2016, 690 – Rn. 80 – Metall auf Metall.

beide Freistellungsmechanismen das Ausschließlichkeitsrecht der Urheber gesetzlich ausgestalten und damit vom Zeitpunkt seiner Entstehung an in seiner Gestalt formen.¹³³ Sowohl die Schrankenbestimmungen als auch die begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte wirken also tatbestandlich auf das Ausschließlichkeitsrecht der Urheber ein.¹³⁴ Demzufolge besteht auf einer materiell-rechtlichen Ebene grundsätzlich Wirkungsgleichheit zwischen beiden Rechtsinstituten. Das könnte eine weitere Differenzierung überflüssig machen.¹³⁵

Neben der materiell-rechtlichen Ausgestaltung ist aber auch die Durchsetzung der Freistellungsregelungen zu berücksichtigen. Nach einem zivilprozessualen Grundsatz hat hierbei jede Partei die für sie günstigen Umstände zu beweisen.¹³⁶ In Bezug auf das Urheberrecht bedeutet das für Urheber, dass sie darlegen und beweisen müssen, dass ihre Rechtsposition überhaupt betroffen ist. Damit sind sie für Tatsachen betreffend die Werkqualität des streitgegenständlichen Inhalts¹³⁷ und darüber hinaus auch dafür beweispflichtig, dass ihre Verwertungsrechte berührt sind.¹³⁸ Hieraus folgt aber auch, dass Urheber bei einer auf Ebene der Verwertungsrechte durchgeführten Freistellung grundsätzlich beweispflichtig dafür wären, dass die streitige Handlung nicht in den Anwendungsbereich der Freistellungsregelung und damit in den Schutzbereich des Ausschließlichkeitsrechts fällt. Im Gegensatz hierzu trägt der Nutzer die Beweislast

133 Zur Rechtsnatur der urheberrechtlichen Schrankenbestimmungen *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 129 ff., insb. S. 131, 149 f.

134 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 149 f.; aA. nur vereinzelt insbesondere in der Rechtsprechung, siehe beispielhaft LG Köln, Urt. v. 13.08.2020 - 14 O 77/19, GRUR-RR 2020, 482 Rn. 40 – Feindliche Übernahme.

135 Ähnlich das Bundesverfassungsgericht, welches von einer bloßen „Frage der Gesetzestechnik“ spricht, siehe BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (45) – Kirchenmusik.

136 *Prütting*, in: Krüger/Rauscher, Münchener Kommentar zur ZPO, ZPO § 286 Rn. 114 f.

137 So beispielsweise BGH, Urt. v. 27.05.1981 - I ZR 102/79, GRUR 1981, 820 (822) – Stahlrohrstuhl II.

138 Beispielsweise zur öffentlichen Wiedergabe OLG Frankfurt a. M., Urt. v. 20.01.2015 - 11 U 95/14, ZUM-RD 2015, 296 (298); v. *Ungern-Sternberg*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 15 Rn. 405 m. w. N.; zur Wiedererkennbarkeit eines Werkes innerhalb eines derivativen Erzeugnisses *Leistner*, GRUR 2019, 1008 (1010).

für das Vorliegen einer für ihn günstigen Schrankenbestimmung.¹³⁹ Demzufolge hat die Ausgestaltung des Freistellungsmechanismus Auswirkungen darauf, welchem der Beteiligten die Darlegungs- und Beweislast für die Freistellungsvoraussetzungen obliegt.¹⁴⁰

Dies gilt selbst dann, wenn einer nicht beweisbelasteten Partei eine sekundäre Darlegungslast auferlegt wird. Denn eine Pflicht zum vollständigen Gegenbeweis ergibt sich aus der sekundären Darlegungslast nicht.¹⁴¹ Vielmehr sind substantiierte Ausführungen ausreichend, die darlegen, warum die vom Prozessgegner (hier: Urheber) vorgetragene(n) Tatsachen nicht vorliegen können.¹⁴²

Alles in allem kann die Verteilung der Darlegungs- und Beweislast damit zur Wahl zwischen einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte und der Freistellung mittels einer unvergüteten Schrankenbestimmung herangezogen werden. Anzumerken bleibt allerdings, dass die gezielte Verteilung der Darlegungs- und Beweislast im Kontext konkreter Regelungsentwürfe nur über präzise, beweislastsensible Normausgestaltungen verwirklicht werden kann. Erweist es sich beispielsweise als interessengerecht, den Urheber für das Nicht-Vorliegen einer Freistellungsvoraussetzung darlegungs- und beweispflichtig zu machen, darf die Freistellungsregelung nicht in Form einer Regel-Ausnahme-Formulierung („dies gilt nicht, wenn“; „es sei denn“) umgesetzt werden. Denn diese würde zu einer Umkehrung der Beweislast zurück auf den Werknutzer führen.

c) Schutz eingesetzter technischer Schutzmaßnahmen via § 95a UrhG

Neben der Darlegungs- und Beweislast könnte bei der Wahl zwischen einer Freistellung durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte oder eine urheberrechtliche Schrankenbestimmung auch darauf abgestellt werden, ob technische Schutzmaßnahmen gegen die betreffende Verwendung der Werke von § 95a UrhG geschützt werden sollen. Denn ein solcher Schutz über § 95a UrhG könnte

139 St. Rspr., siehe bspw. BGH, Urt. v. 30.03.2017 - I ZR 19/16, GRUR 2017, 1233 Rn. 15 – Loud; Leistner, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 97 Rn. 28; Flechsig, GRUR 1993, 532 (536).

140 Zu dieser Verschiebung grundsätzlich schon Dreier, GRUR Int. 2015, 648 (650).

141 BGH, Urt. v. 30.03.2017 - I ZR 19/16, GRUR 2017, 1233 Rn. 15 – Loud.

142 Vgl. Stadler, in: Musielak/Voit, ZPO, ZPO § 138 Rn. 10a.

voraussetzen, dass die technische Schutzmaßnahme auf die Abwehr von Handlungen abzielt, die als Teil der urheberrechtlichen Verwertungsrechte im Sinne der §§ 15 ff. UrhG dem Ausschließlichkeitsrecht der Urheber zugewiesen sind.

Hiergegen – und im Umkehrschluss für einen weiten, nicht an die §§ 15 ff. UrhG geknüpften Schutzbereich des § 95a UrhG – spricht zunächst, dass sich technische Maßnahmen nach dem Wortlaut des § 95a Abs. 2 S.1 UrhG nicht gegen *Werknutzungen*, sondern nur gegen *Handlungen* richten müssen, die vom Rechtsinhaber nicht genehmigt worden sind. Zudem bezieht die Regelung in § 95a Abs.2 S.2 UrhG ausdrücklich auch zugangshindernde Maßnahmen ein.¹⁴³ Der Zugang zum Werk (und die anschließende Rezeption des Werkes) sind aber für sich genommen nicht Teil des urheberrechtlichen Ausschließlichkeitsrechts. Von einem über die urheberrechtlichen Nutzungsrechte hinausgehenden Verständnis des § 95a UrhG, welches auch den urheberrechtlich nicht relevanten Werkzugang erfasst, ging auch der historische Gesetzgeber bei der Einführung des § 95a UrhG aus.¹⁴⁴

Allerdings ist die Verwendung des Rechtsbegriffs „Handlung“ darauf zurückzuführen, dass der historische Gesetzgeber eine möglichst richtliniennahe Umsetzung anstrebte.¹⁴⁵ Im § 95a UrhG zugrundeliegenden Art. 6 InfoSoc-RL ist von Maßnahmen gegen *Handlungen* die Rede, die vom Rechtsinhaber nicht genehmigt worden sind. Beispielsweise in ErwGr. 16 InfoSoc-RL ist jedoch von einer „Haftung für Handlungen im Netzwerk-Umfeld“ die Rede. Die *Haftung* kann aber nur entstehen, wenn ein Haftungstatbestand erfüllt ist. Hierfür bedarf es einer urheberrechtlich relevanten Nutzung. Dies spricht dafür, dass der Handlungsbegriff im hier verwendeten Kontext nur urheberrechtlich relevante Nutzungshandlungen umfasst. Nach Art.3 Abs.3 InfoSoc-RL gilt der Erschöpfungsgrundsatz zudem nicht für *Handlungen der öffentlichen Wiedergabe*, womit eine dem Urheber zugewiesene Werknutzung gemeint sein muss. Beides zeigt, dass der Handlungsbegriff innerhalb der InfoSoc-RL synonym zur urheberrechtlich relevanten Werknutzung zu verstehen ist. Daraus, dass das Gesetz den Schutz technischer Schutzmaßnahmen nach § 95a Abs.2 S.1 UrhG auf Maßnahmen erstreckt, die sich gegen *Handlungen* richten, die vom Urheber nicht genehmigt worden sind, kann demzufolge nicht geschlossen

143 Stieper, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 462.

144 BT-Drs. 15/38, S. 26.

145 BT-Drs. 15/38, S. 26.

werden, dass § 95a UrhG einen über die urheberrechtlichen Verwertungsrechte hinausgehenden Schutzbereich eröffnet.

Im Gegenteil verfolgt die InfoSoc-RL das Ziel, ein hohes Schutzniveau für Urheber zu gewährleisten. Dieses Schutzziel ist auch in Art. 6 Abs. 3 S. 2 InfoSoc-RL verankert. Zweck des hohen Schutzniveaus ist, dass Urhebern eine angemessene Vergütung zuteilwird, auf Basis derer sie weiterhin schöpferisch tätig sein können.¹⁴⁶ Eine solche Vergütungsleistung ist aber nur für urheberrechtlich relevante Werknutzungen gerechtfertigt. Denn nur sie berühren die ausschließlichen Verwertungsrechte, welche den Urhebern insbesondere durch Art. 2 bis 4 InfoSoc-RL zur Verwirklichung des Schutzniveaus zugewiesen worden sind.

Darüber hinaus müssen technische Maßnahmen, die von § 95a UrhG geschützt werden sollen, nach dem Wortlaut der Regelung solche Handlungen verhindern oder erschweren, die vom Urheber nicht genehmigt worden sind (Art. 6 Abs. 3 S. 1 InfoSoc-RL respektive § 95a Abs. 2 S. 1 UrhG). Dies setzt denklogisch aber eine Zustimmungsbedürftigkeit durch den Urheber voraus. Diese kann nur dann vorliegen, wenn eine urheberrechtlich relevante Werknutzung stattfindet oder jedenfalls stattfinden soll.¹⁴⁷

Alles in allem sprechen sowohl teleologische als auch systematische und grammatikalische Gründe dafür, dass der Umgehungsschutz aus § 95a UrhG nur gegen solche Nutzerhandlungen geltend gemacht werden kann, mittels derer ein Werk in einer Art und Weise genutzt wird, die dem Urheber über die Verwertungsrechte aus §§ 15 ff. UrhG zugewiesen worden ist.¹⁴⁸ Der Umgehungsschutz ist außerdem auf vorgelagerte¹⁴⁹ Handlungen

146 ErwGr. 4, 9, 10 InfoSoc-RL.

147 *Wandtke/Ohst*, in: *Wandtke/Bullinger*, Urheberrecht, UrhG § 95a Rn. 43; *Arlt*, in: *Hoeren/Sieber/Holznapel*, Handbuch Multimedia-Recht, Teil 7.7 (Schutz und Identifizierung durch technische Schutzmaßnahmen) Rn. 25; *Peukert/Leistner*, in: *Loewenheim*, Handbuch des Urheberrechts, § 40 (Schutz technischer Maßnahmen) Rn. 3; ähnlich wohl *Rieber-Mohn*, IIC 2006, 182 (186 f.), wobei der Autor ein vertragliches Zustimmungsrecht als ausreichend erachtet; a. A. *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 463, nach dem es für einen Schutz über § 95a UrhG ausreicht, wenn jedenfalls auch urheberrechtlich relevante Nutzungshandlungen erfasst werden.

148 EuGH, Urt. v. 23.01.2014 - C-355/12, GRUR 2014, 255 Rn. 25, 31 – Nintendo/PC Box; *Czychowski*, in: *Fromm/Nordemann*, Urheberrecht, UrhG § 95a Rn. 14; *Götting*, in: *Schricker/Loewenheim*, Urheberrecht, UrhG § 95a Rn. 45; *Wandtke/Ohst*, in: *Wandtke/Bullinger*, Urheberrecht, UrhG § 95a Rn. 43; *Arlt*, in: *Hoeren/Sieber/Holznapel*, Handbuch Multimedia-Recht, Teil 7.7 (Schutz und Identifizierung durch technische Schutzmaßnahmen) Rn. 25; *Peukert/Leistner*, in: *Loewenheim*, Handbuch des Urheberrechts, § 40 (Schutz technischer Maßnahmen) Rn. 3; *Specht-*

zu erstrecken, mit denen das Ziel verfolgt wird, das betreffende Werk in urheberrechtlich relevanter Weise zu nutzen.¹⁵⁰ Hierzu gehört auch der Zugang zum Werk, der in Art. 6 Abs. 3 S. 2 InfoSoc-RL respektive § 95a Abs. 2 S. 2 UrhG ausdrücklich mitumfasst wird.¹⁵¹ Umgehungen zur Ermöglichung urheberrechtlich freier Verwendungen eines Werkes können über § 95a UrhG hingegen nicht untersagt werden. Besteht ein Bedürfnis danach, technische Schutzmaßnahmen über § 95a UrhG zu schützen, die gegen die jeweilige Verwendung von Werken in den zu untersuchenden Sachverhaltskonstellationen wie beispielsweise dem Web Scraping von Werken als KI-Trainingsdaten eingesetzt worden sind, spricht dies infolgedessen dafür, dass die Freistellung durch eine urheberrechtliche Schrankenbestimmung umgesetzt sein sollte.

d) Notwendigkeit eines Opt-out-Mechanismus

Sollen Urheber als Ausgleich für eine freigestellte Werknutzung eine Vergütung erhalten, kann die Freistellung im Wege einer gesetzlichen Lizenz oder mit Hilfe eines ECL-Mechanismus umgesetzt werden. Fraglich ist vor diesem Hintergrund aber, welche Unterschiede sich im Weiteren zwischen den beiden Rechtsinstituten ergeben, die eine Wahl des besser geeigneten Freistellungsmechanismus möglich machen.

Zum einen könnte die Notwendigkeit eines Opt-out-Mechanismus, den betroffener Urheber auslösen können, Einfluss darauf haben, welcher der beiden Regelungsmechanismen für eine Freistellung gewählt werden sollte. Unter einem Opt-out-Mechanismus wird dabei eine Regelung verstanden,

Riemenschneider, Diktat der Technik, S. 199, 203; ähnlich wohl *Rieber-Mohn*, IIC 2006, 182 (186 f.), wobei der Autor ein vertragliches Zustimmungsrecht als ausreichend erachtet; BT-Drs. 15/38, S. 26; a. A. *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 463, nach dem es für einen Schutz über § 95a UrhG ausreicht, wenn jedenfalls auch urheberrechtlich relevante Nutzungshandlungen erfasst werden; ebenso im norwegischen Urheberrecht, siehe dazu *Rieber-Mohn*, IIC 2006, 182 (191).

149 Für den Werkzugang *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 462 f.

150 Vgl. EuGH, Urt. v. 23.01.2014 - C-355/12, GRUR 2014, 255 Rn. 25, 31 – Nintendo/PC Box.

151 A. A. *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 462. Nach ihm spricht die Erfassung des Werkzugangs dafür, dass insgesamt auch urheberrechtlich freie Handlungen erfasst sind, solange die technische Schutzmaßnahme umgangen wird.

die es Urhebern ermöglicht, die Freistellungswirkung durch einseitige Erklärung zu beenden. Ein Opt-out-Mechanismus kann dabei grundsätzlich sowohl im Rahmen gesetzlicher Lizenzen¹⁵² als auch im Rahmen eines ECL-Mechanismus¹⁵³ vorgesehen werden.

Üblicherweise werden gesetzliche Vergütungsansprüche für Massennutzungen verwertungsgesellschaftenpflichtig ausgestaltet. Zu diesen Massennutzungen gehört auch das Web Scraping von Werken als Trainingsdaten für KI.¹⁵⁴ Sollten gesetzliche Lizenzen mit einem Opt-out-Mechanismus verknüpft werden, müsste die Opt-out-Erklärung allerdings unmittelbar am Werkstück, beispielsweise in den Metadaten einer Bilddatei, abgegeben werden. Denn eine Abwicklung der Nutzungshandlungen selbst über die Verwertungsgesellschaften, beispielsweise durch Abschluss eines Lizenzvertrags, findet nicht statt. Die Nutzung wäre vielmehr eo ipso erlaubnisfrei gestellt. Damit bestünde allerdings die Gefahr, dass Urheber von der Verwertungsgesellschaft eine aufgrund der Massennutzung pauschalisierte Ausschüttung erhielten, obwohl sie eine Vorbehaltserklärung zur Beendigung der Freistellungswirkung am Werkstück abgegeben haben.

Im Rahmen des erweiterten kollektiven Lizenzmechanismus wird der Opt-out hingegen direkt gegenüber der lizenzierenden Verwertungsgesellschaft erklärt, die die ECL-Vereinbarung abschließt (vgl. § 51 Abs. 2 VGG). Infolgedessen wäre die Verwertungsgesellschaft in diesem Fall in die Lage versetzt, eine ungerechtfertigte Vergütung vorbehaltserklärender Urheber zu verhindern. Wenn die Einrichtung eines Opt-out-Mechanismus zugunsten betroffener Urheber erforderlich ist, spricht dies im Ergebnis dafür, eine Freistellung im Wege eines ECL-Mechanismus umzusetzen.

e) Erhöhtes Bedürfnis nach Rechtssicherheit für Werknutzer

Ist kein Opt-out-Mechanismus zugunsten betroffener Urheber erforderlich, so spricht dies umgekehrt allerdings noch nicht ohne Weiteres dafür, eine Freistellung im Wege einer gesetzlichen Lizenz umzusetzen. Denn ein Freistellungsmechanismus ohne Opt-out-Mechanismus kann sowohl als ECL (ohne Widerspruchsmöglichkeit) als auch durch eine gesetzliche Lizenz

152 Beispielsweise der Rechteevorbehalt in § 49 Abs. 1 S. 1 UrhG.

153 Beispielhaft die Widerrufsmöglichkeiten des Rechtsinhabers in den erweiterten kollektiven Lizenzmechanismen, die durch Umsetzung der DSM-RL im VGG implementiert worden sind, siehe §§ 51 Abs. 2, 52 Abs. 2 VGG.

154 Dazu bereits siehe oben unter 2. Teil A. II. 2. a).

(ohne Vorbehaltsmöglichkeit) umgesetzt werden. Bei der Wahl zwischen beiden Regelungsmechanismen könnte es in diesem Fall vielmehr darauf ankommen, ob die Trainingsdatensammler, KI-Entwickler oder Anwender von KI-Systemen ein besonders ausgeprägtes Bedürfnis nach Rechtssicherheit haben.

Gesetzliche Lizenzen sind als Rechtsnormen abstrakt-generell ausgestaltet. Dies geht mit einer Einschränkung der Rechtssicherheit in Bezug auf die konkrete Einzelfallanwendung einher. ECL beruhen hingegen auf Lizenzvereinbarungen, die im Rahmen der gesetzlichen Spielräume für den konkreten Einzelfall individualisiert werden können. Im direkten Vergleich beider Rechtsinstitute gewährleistet der ECL-Mechanismus damit mehr Rechtssicherheit.¹⁵⁵ Wenn ein besonderes ausgeprägtes Rechtssicherheitsbedürfnis besteht, ist daher eine Freistellung im Wege eines ECL-Mechanismus vorzuziehen. Das ist beispielsweise dann der Fall, wenn das Sammeln von Werken als Trainingsdaten oder andere Sachverhaltskonstellationen ein hohes wirtschaftliches Risiko für die Freistellungsbegünstigten mit sich bringen.

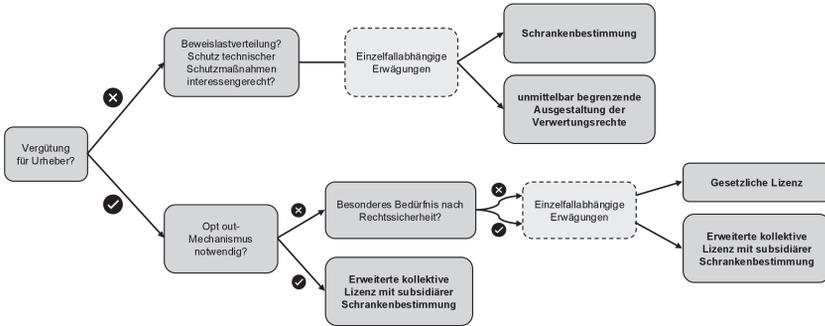
Bestehen keine überdurchschnittlichen Anforderungen an die von der Freistellungsregelung gewährleistete Rechtssicherheit, so ist die Werknutzung hingegen auf Basis einer gesetzlichen Lizenz niederschwelliger möglich. Denn ihre Freistellungswirkung tritt bereits von Gesetzes wegen ein. In diesen Fällen sollte die Freistellung daher im Wege einer gesetzlichen Lizenz umgesetzt werden.

f) Weitere Aspekte und Zusammenfassung

Neben den herausgearbeiteten Entscheidungskriterien, die im Regelfall die Wahl zwischen den zur Verfügung stehenden Freistellungsmechanismen ermöglichen, können im Einzelfall aber auch andere Aspekte berücksichtigt werden.

155 *Trumpke*, Exklusivität und Kollektivierung, S. 578.

Alles in allem ist die Wahl des für eine Freistellung geeigneten Regelungsmechanismus anhand des folgenden Musters vorzunehmen:



B. Speicherung von urheberrechtlich geschützten Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern

Bevor ein KNN mit Werken aus Internetquellen trainiert werden kann, müssen die Werke zunächst abgespeichert werden. Dies geschieht insbesondere mit Hilfe automatisierter Webcrawler. Der Prozess wird auch als Web Scraping bezeichnet. Fraglich ist zunächst, ob das Web Scraping der Werke zur Verwendung als Trainingsdaten ohne individuelle Zustimmung betroffener Urheber möglich, also urheberrechtlich freigestellt sein sollte. Hiervon würden Entwickler von KI-Systemen und intermediäre Trainingsdatensammler profitieren, deren Tätigkeit darauf ausgerichtet ist, Trainingsdatensätze für Dritte zu erstellen und in Verkehr zu bringen. Damit ein regulatorisches Freistellungsbedürfnis besteht, muss es sich als zweckmäßig erweisen, das Web Scraping von Werken als Trainingsdaten vom Ausschließlichkeitsrecht betroffener Urheber auszuklammern.¹⁵⁶

¹⁵⁶ Zur Herleitung der regulatorischen Entscheidungskriterien insgesamt siehe unter 2. Teil A.

I. Zweckmäßigkeit einer Freistellung

Die Freistellung des Web Scrapings von Werken als Trainingsdaten ist zweckmäßig, wenn die Interessen der Urheber, deren Werke für das Training einer KI abgespeichert werden, nicht berührt sind oder wenn eine umfassende Abwägung von Urheber- und Nutzerinteressen ergibt, dass den Nutzerinteressen beim Web Scraping von Werken als Trainingsdaten Vorrang zu gewähren ist.

1. Beeinträchtigung der Urheberinteressen

Die Speicherung von bereits im Internet verfügbar gemachten Werken geht nicht mit einer Veränderung der Inhalte einher. Mangels identifizierender Werkrezipienten beim rein-technischen KI-Training wird durch das sachverhaltstypische Weglassen der Urhebernennung im Trainingsprozess auch die Bekanntheit des Urhebers auf dem Kreativmarkt nicht geschmälert. Aus alledem folgt, dass vom Web Scraping weder das Erstveröffentlichungs- noch das Integritätsinteresse sowie das Interesse des Urhebers an der Anerkennung seiner Urheberschaft betroffen wird. Das Web Scraping beeinträchtigt daher keine geschützten ideellen Urheberinteressen.

Gegen eine Beeinträchtigung der urheberrechtlichen Verwertungsinteressen spricht weiterhin, dass für das Web Scraping notwendigerweise bereits Zugang zu den betroffenen Werken bestehen muss, womit die Werke bereits im Rahmen der Zugänglichmachung einmal verwertet worden sind.¹⁵⁷ Allerdings setzen insbesondere sämtliche Beschränkungen des Urheberrechts, die die Vervielfältigung von Werken durch Nutzer erlaubnisfrei stellen, denkllogisch Zugang zu Repräsentationen dieser Werke voraus.¹⁵⁸ Hieraus folgt allerdings nicht zwingend, dass durch die Herstellung weiterer Werkstücke keine Verwertungsinteressen der betroffenen Urheber mehr beeinträchtigt sind. Vielmehr handelt es sich um zwei voneinander unabhängige Verwertungshandlungen: In einem ersten Schritt kann der Zugang zum Werk monetarisiert werden. Dies geschieht beispielsweise im Rahmen der Kinovorführung eines Filmwerks, für die Nutzer eine Vergütung entrichten. Im Anschluss hieran hat aber auch die Vervielfältigung des betreffenden Werkes weiterhin einen wirtschaftlichen Wert. Kinobesuchern

157 So *Raue*, CR 2017, 656 (661); *Raue*, GRUR 2017, II (14).

158 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 58.

wird nicht gestattet, im Anschluss an die Filmvorführung nun auch ein Vervielfältigungsstück des Filmwerks anzufertigen. Im Gegenteil müssten sie hierfür weitere Nutzungsrechte oder ein bereits angefertigtes Vervielfältigungsstück erwerben.

Eine Freistellung, nach der individuelle Nutzungsrechtseinräumungen durch jeden betroffenen Urheber für die Speicherung der Trainingsdaten nicht mehr notwendig wären, hätte vielmehr zur Folge, dass die Urheber mangels Nachfrage nicht mehr in der Lage wären, ihre Werke – unmittelbar oder mittelbar über Trainingsdatenanbieter – auf dem Markt für Trainingsdaten zu monetarisieren. Auch wenn es sich hierbei nicht um die übliche Kernverwertung urheberrechtlich geschützter Inhalte handelt, würde den Urhebern hiermit eine Möglichkeit zur Verwertung ihrer Werke genommen. Die Perspektive der Urheber auf eine den Lebensunterhalt und weitere schöpferische Tätigkeit sichernde Amortisation der Werkschöpfung würde infolgedessen geschmälert.

Hieraus folgt, dass das Web Scraping von Werken als Trainingsdaten die Verwertungsinteressen der Urheber beeinträchtigt.¹⁵⁹ Eine urheberrechtliche Freistellung kann demzufolge nicht bereits wegen einer Nichtbeeinträchtigung der Urheberinteressen erfolgen.

2. Abwägung der Urheberinteressen mit den Interessen der Trainingsdatensammler

Eine umfassende Abwägung von Urheber- und Nutzerinteressen könnte allerdings zur Folge haben, dass den Nutzerinteressen beim Web Scraping von Werken als Trainingsdaten Vorrang vor den Urheberinteressen zu gewähren ist.

a) Zu berücksichtigende Trainingsdatensammler- und Allgemeininteressen

Dem betroffenen Verwertungsinteresse der Urheber steht das Ziel der KI-Entwickler und intermediären Trainingsdatensammler gegenüber, mit Hilfe

159 So auch *Dregelies*, GRUR 2024, 1484 (1485); *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S.150. Für das allgemeine Text- und Data-Mining anders hingegen *Hauk/Pflüger*, ZUM 2020, 383 (385); *Raue*, CR 2017, 656 (661); *Raue*, GRUR 2017, 11 (14); *Schack*, ZUM 2016, 266 (269); *Schack*, ZUM 2017, 802 (806); *Schack*, Urheber- und Urhebervertragsrecht, S. 311 Rn. 627; *Stieper*, GRUR 2020, 1 (4); *Stieper*, ZUM 2021, 776 (778); *Raue*, ZUM 2019, 684 (686), die davon ausgehen, dass die urheberrechtlichen Verwertungsrechte nicht berührt werden.

von Web Scraping-Technologien urheberrechtlich geschützte Trainingsdaten zu generieren. In diesem Ziel könnten sich verschiedene Interessen bündeln, die zugunsten der Nutzer in die Abwägung eingestellt werden müssen. Gemein ist allen Fallkonstellationen aber, dass die für das Training genutzten Werke während des eigentlichen Lernvorgangs nur als Trägermedium für Muster und Zusammenhänge genutzt werden. Sie sollen aus den Werken extrahiert werden. Auch bei der menschlichen Informationserfassung werden Werke als Trägermedium, also nur um ihrer Informationsinhalte Willen, konsumiert. Infolgedessen ist zugunsten der Trainingsdatensammler zu berücksichtigen, dass das Ziel, Daten zum Training einer KI zu generieren, strukturelle Ähnlichkeiten zum menschlichen Informationsinteresse aufweist.¹⁶⁰

Aufgrund des besonders hohen Daten- und Ressourcenaufwands¹⁶¹ findet die Entwicklung von KI außerdem regelmäßig durch gewerblich handelnde Akteure statt. Damit ist das Web Scraping regelmäßig als Teil der Berufs- beziehungsweise unternehmerischen Freiheit der Datensammler geschützt, die in Art. 15 f. EU-GRCh sowie auf nationaler Ebene in Art. 12 Abs. 1, 2 GG verankert ist. Ebenso sind im Einzelfall auch Überschneidungen mit anderen Freiheitsrechten wie insbesondere der Wissenschaftsfreiheit (Art. 5 Abs. 3 GG) möglich.

Unabhängig davon, in welchem Zusammenhang das Web Scraping im konkreten Fall stattfindet, verfolgen sowohl KI-Entwickler als auch intermediäre Trainingsdatensammler direkt oder mittelbar das Ziel, durch das Training eines neuen KI-Systems den technologischen Fortschritt im Bereich des maschinellen Lernens zu fördern. Ziel der Trainingsdatensammler ist demzufolge auch die Innovationsförderung. Auch dieses Innovationsinteresse kann zugunsten der Trainingsdatensammler in die Abwägung aufgenommen werden.¹⁶² Seine Berücksichtigung im urheberrechtlichen Regelungsregime zeigt sich bereits in einer anderen urheberrechtlichen Schrankenbestimmung, § 60c UrhG.¹⁶³

160 Auch *Heesen/Jüngels*, RuZ 2021, 45 (53); *Raue*, CR 2017, 656 (661); *Hofmann*, WRP 2024, 11 Rn. 23.

161 Dazu bereits oben unter 1. Teil B II. 3.

162 So auch *Hofmann*, WRP 2024, 11 Rn. 22; *Trapova/Gervassis*, Kluwer Copyright Blog v. 16.3.2022; *Vesala*, IIC 2023, 351 (355); vgl. allgemein für das Text und Data-Mining auch *de la Durantaye*, GRUR 2017, 558 (561 f.); *Raue*, GRUR 2017, 11 (15); *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 125.

163 Vgl. auch ErwGr. 5, 8 DSM-RL, auch der Richtlinienggeber bezieht das Innovationsinteresse in seine Erwägungen ein.

Eine Freistellung des von Trainingsdatensammlern vorgenommenen Web Scrapings von Werken als Trainingsdaten kann also erfolgen, wenn die Abwägung des durch das Web Scraping betroffenen Verwertungsinteresses der Urheber mit dem von den KI-Entwicklern bzw. intermediären Trainingsdatensammlern verfolgten Innovationsinteresse und insbesondere ihrer Berufsfreiheit ergibt, dass den dargelegten Interessen der Datensammler im konkreten Fall Vorrang vor dem Verwertungsinteresse der Urheber zu gewähren ist.

b) Beeinträchtigungsintensitäten der gegenläufigen Interessen

Im Hinblick auf die betroffenen Urheberinteressen lässt sich hierfür zunächst feststellen, dass die Freistellung des Web Scrapings den Urhebern nicht die Möglichkeit nimmt, ihre Werke auszustellen, öffentlich wiederzugeben und Vervielfältigungsstücke der Werke in Verkehr zu bringen. Daher nimmt das Web Scraping den betroffenen Urhebern nicht die Möglichkeit, ihre Kernverwertung, also diejenigen Verwertungsvorgänge, bei denen die menschliche Rezeption der Werke in ihrer Eigenschaft als geistige Schöpfungen im Mittelpunkt steht, vorzunehmen.¹⁶⁴ Dies wäre aufgrund der gleichbleibenden Verwertungsform selbst dann der Fall, wenn auch kommerzielle Datensammler von einer Freistellungsregelung erfasst wären.¹⁶⁵ Dementsprechend sind die Verwertungsinteressen der Urheber vom Web Scraping zum Zwecke des Trainings einer KI nicht wesentlich betroffen.

Ähnliches könnte allerdings auch für die Ziele der KI-Entwickler oder intermediären Trainingsdatensammler gelten. Denn die mit dem Web Scraping im Regelfall ausgeübte Berufs- bzw. unternehmerische Freiheit im Sinne der Art. 15, 16 EU-GRCh ist nur in ihrer Ausformung als Ausübungsfreiheit berührt. Der Eingriff ist damit nicht intensiv.¹⁶⁶ Das ist auch bei der Kollision mit anderen Grundrechten wie dem über Art. 17 EU-GRCh geschützten Urheberrecht zu berücksichtigen.

164 Wie hier auch *Jager*, *Artificial Creativity?*, S. 347. So auch schon für das allgemein Text- und Data-Mining ohne spezifischen Bezug zu KI *de la Durantaye*, *Allgemeine Bildungs- und Wissenschaftsschranke*, S. 240; *European Copyright Society*, *General Opinion on the EU Copyright Reform Package*, S. 5.

165 So im Kontext des Dreistufentests auch *Hacker*, *ZGE* 2020, 239 (269); a. A. ebenfalls zum Dreistufentest allerdings *Jager*, *Artificial Creativity?*, S. 371; zum unspezifischen Text und Data-Mining außerdem *Triaille/Meeûs d'Argenteuil/Francquen*, *Study on the legal framework of text and data mining (TDM)*, S. 99 f.

166 *Calliess*, in: *Calliess/Ruffert*, *EUV/AEUV, EU-GRCh Art. 15 Rn. 16*.

Das Innovationsinteresse ist darüber hinaus nicht ausdrücklich grundrechtlich abgebildet. Damit sind an seinen Schutz ebenfalls keine besonders hohen Anforderungen zu stellen. Zwar könnten im Einzelfall auch andere Grundrechte – insbesondere die Kommunikationsfreiheiten oder die Wissenschaftsfreiheit aus Art. 11, 13 EU-GRCh – betroffen sein. Da auch in diesen Fällen nur spezifische Formen der Freiheitsausübung, nicht aber die Freiheiten in ihrem Wesensgehalt berührt sind, wiegen auch die hier zu berücksichtigenden Nutzerinteressen im Verhältnis zum über Art. 17 EU-GRCh geschützten Verwertungsinteresse der Urheber nicht überproportional schwer. Bereits aus der Interessenlage als solcher ergibt sich demzufolge noch kein Überwiegen der Interessen der Urheber auf der einen oder der Trainingsdatensammler auf der anderen Seite.

c) Erwägungen gegen eine Freistellung

Dafür, das Web Scraping urheberrechtlich geschützter Trainingsdaten dem Ausschließlichkeitsrecht der Urheber zuzuweisen, spricht, dass das Sammeln der urheberrechtlich geschützten Trainingsdaten nicht mit dem Informationsvorgang selbst, also dem Rezipieren eines Werkes, vergleichbar ist.¹⁶⁷ Aufgrund etwaiger Parallelen hierzu kann daher zwar möglicherweise der eigentliche Trainingsvorgang der KI,¹⁶⁸ nicht aber das Web Scraping der Inhalte mit dem Werkgenuss gleichgesetzt und deswegen urheberrechtlich freigestellt werden.

Weiterhin werden beispielsweise Nachrichtenartikel als Trainingsdaten genutzt, um Anwendungen des Algorithmischen Journalismus – einer speziellen Art von Textgeneratoren, die journalistische Inhalte erzeugen – anzulernen. Gleiches gilt für Bildwerke, die genutzt werden, um bildgenerierende KI zu trainieren. Trainingsdatensammler verfolgen also mitunter das Ziel, ein KI-System herzustellen oder die Herstellung eines KI-Systems zu ermöglichen, das Konkurrenzprodukte zu den Werken produzieren kann, die zum Training des KI-Systems genutzt worden sind.¹⁶⁹ Mit dem Einsatz solcher generativer KI-Systeme wird die Möglichkeit der Urheber, ihre Schöpfungs-

167 In dieser Hinsicht präzise *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 20.

168 Dazu ausführlich unter 3. Teil B. II.

169 Dies stellt auch *de la Durantaye*, ZUM 2023, 645 (649) heraus. Ebenfalls im Kontext des Dreistufentests *Schack*, NJW 2024, 113 Rn. 8; *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 154 f., letztere an dieser Stelle aber etwas apodiktisch.

leistungen insbesondere auch durch Kernverwertungsarten zu amortisieren, beeinträchtigt. Dies kann mittelbar auch bei der Freistellung des Web Scrapings von Werken für das KI-Training berücksichtigt werden.

In diesem Kontext könnten zum Training genutzte Werke im Einzelfall sogar im Output des trainierten KI-Systems identifizierbar sein. In der Folge bestünde die Gefahr, dass der Output der KI Urheberrechte der vom Training betroffenen Urheber verletzt.¹⁷⁰

Gegen eine Freistellung spricht darüber hinaus, dass die Mehrzahl derjenigen Urheber, die das Web Scraping trotz einer urheberrechtlichen Freistellung verhindern wollen, nicht über die technischen Fähigkeiten verfügen, diesem Bedürfnis mit Hilfe gegen das Web Scraping gerichteter technischer Schutzmaßnahmen gerecht zu werden. Effektiver Schutz bestünde für betroffene Urheber damit nur, wenn sie ihre Werke überhaupt nicht im freien Internet verfügbar machen. Da gerade hier aber auch beispielsweise durch Lizenzentnahmen für die öffentliche Zugänglichmachung der Werke oder Zugangsentgelte wesentliches Amortisationspotential liegt, könnten Urheber ihr Interesse nicht durchsetzen, ohne im Ergebnis doch mittelbar ihre Verwertungsmöglichkeiten zu beeinträchtigen, bei denen die Nutzung der Werke um ihren schöpferischen Gehalt Willen im Vordergrund steht. Diese ist Teil der Kernverwertung.

d) Erwägungen zugunsten einer Freistellung

Zugunsten einer Freistellung des Web Scrapings urheberrechtlich geschützter Trainingsdaten könnten allerdings ebenfalls eine Vielzahl verschiedener Argumente angeführt werden.

aa) Verhältnis zwischen der Speicherung zum Training und den Erzeugnissen des KI-Systems

Für eine Freistellung spricht zunächst, dass beim Training der KI selbst nach aktuellem Stand von Wissenschaft und Technik nicht vollständig ausgeschlossen werden kann, dass es in Einzelfällen zur Wiedererkennbarkeit von Trainingsdaten im Output kommt. Das Risiko für eine Identifizierbarkeit von Trainingsdaten im Output ist daher nicht restlos vermeidbar und

170 So im Kontext des Dreistufentests *Vesala*, IIC 2023, 351 (366 f.).

gleichzeitig erheblich einzelfallabhängig. Das Sammeln von Werken zum Training von KI insgesamt zu verhindern, wird diesem Risikograd nicht gerecht. Es würde vielmehr zu einem überschießenden Schutz der Urheber führen. Zudem zeigt sich erst mit der Anwendung der KI, ob verwendete Trainingsdaten im Einzelfall im Output identifizierbar sind und ob es hierdurch zu Urheberrechtsverletzungen kommt. Die Generierung von Output durch KI geht aber auf einen zweiten Willensentschluss und demzufolge auf eine eigenständige natürliche Handlung zurück. Sie wird häufig auch von anderen Beteiligten durchgeführt als das Training. Sie ist deswegen vom Training einer KI abgrenzbar. Die Output generierenden Nutzer entrichten für die Generierung des Outputs zudem unter Umständen eine gesonderte Vergütung. Der Markt für die Anwendung eines KI-Systems und die Generierung von Output, an dem KI-Nutzer als Nachfrager teilnehmen, ist vom Markt für das Training von KI, dessen Teilnehmer KI-Entwickler sind, zu trennen. Die Generierung von Output ist deswegen wirtschaftlich klar vom Training abgrenzbar und selbstständig verkehrsfähig. Sie ist aus diesen Gründen ein eigenständiger Nutzungssachverhalt.¹⁷¹ Ihre Rechtmäßigkeit ist daher auch eigenständig zu beurteilen. Gegen im Einzelfall auftretende Urheberrechtsverletzungen im Output von KI kann im jeweiligen Einzelfall eigenständig vorgegangen werden. Gleichzeitig kann es für die Freistellung der Datensammlung zum Training von KI aus einer zeitlich-logischen Perspektive nicht darauf ankommen, ob es zu einem späteren Zeitpunkt durch eine andere Nutzungshandlung, die Generierung des Outputs, zu Urheberrechtsverletzungen kommt.¹⁷²

Inwieweit KI-Systeme auch unabhängig von Urheberrechtsverletzungen in konkretem Output in der Lage sind, Konkurrenzprodukte zu den betroffenen Werken zu generieren, hängt darüber hinaus von der Zielfunktion des trainierten KI-Systems sowie der Qualität des Outputs und der Nachfrage auf dem Kreativmarkt ab. Die mögliche Beeinträchtigung der Urheber ist dementsprechend einzelfallabhängig. Daraus folgt, dass die *Möglichkeit* zur Generierung von Konkurrenzprodukten nicht per se zum Ausbleiben

171 Außerdem eine eigenständige Nutzungsart im urhebervertragsrechtlichen Sinne (§ 31 Abs.1 UrhG), vgl. zu den Kriterien für eine Abgrenzung eigenständiger Nutzungsarten Wiebe, in: Spindler/Schuster, Recht der elektronischen Medien, UrhG § 31 Rn. 9 m. w. N.

172 So im Kontext der Anwendung des Dreistufentests aber Vesala, IIC 2023, 351 (366f.).

einer Freistellungsregelung führen sollte.¹⁷³ Es bedarf vielmehr einer Betrachtung des jeweiligen Einzelfalls, insbesondere über den Dreistufentest bei der Anwendung einer Freistellungsregelung im konkreten Einzelfall.¹⁷⁴ Die Generierung von Output ist – wie bereits gezeigt werden konnte – ein eigenständiger, vom Training abgrenzbarer Nutzungssachverhalt, sodass er auch hinsichtlich der möglichen Konkurrenzsituation und dessen Auswirkungen vom Trainingsprozess losgelöst beurteilt werden sollte. Dies spricht im Ergebnis ebenfalls für eine einzelfallbezogene, am konkret generierten Output ansetzende, nicht aber eine systemische Betrachtung im Kontext des Trainingsprozesses.

bb) Kompensationswirkung einer Vergütungspflicht

Zudem könnte die Gewährung einer angemessenen Vergütung für die Nutzung der Werke zum Training von KI¹⁷⁵ die wirtschaftlichen Folgen für betroffene Urheber abmildern und dazu beitragen, den Substitutionsdruck zu senken. Dies gilt insbesondere deswegen, weil die wirtschaftliche Bedeutung der Werkverwendung beim Training von KI auch bei der Bestimmung der Vergütungshöhe berücksichtigt wird.¹⁷⁶

cc) Art der Werkverwendung

Zu berücksichtigen ist darüber hinaus, dass mit dem Training von KI keine Nutzung der schöpferischen Elemente verwendeter Werke *als Werk* verbunden ist. Es findet vielmehr ein rein-technischer Vorgang statt, im Rahmen dessen die Werke als Informationsträger ausgewertet werden.¹⁷⁷

Auch wenn das Sammeln urheberrechtlich geschützter Trainingsdaten zudem nicht selbst mit dem menschlichen Informationsvorgang, also dem

173 A. A., aber ohne eingehende Begründung *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 151, 154 f.

174 Für eine Einzelfallbetrachtung im Kontext des Dreistufentests auch *Leistner*, GRUR 2024, 1665 (1670); *Novelli u. a.*, Computer Law & Security Review 2024, (10).

175 Dazu siehe unter 2. Teil B. II. 1.

176 Zur Vergütungshöhe: *Pflüger*, Gerechter Ausgleich und angemessene Vergütung, S. 145; insbesondere zu anderen bei der Bestimmung der Vergütungshöhe berücksichtigungsfähigen Aspekten: *Lauber-Rönsberg*, Urheberrecht und Privatgebrauch, S. 112 ff.

177 *Hofmann*, WRP 2024, II Rn. 25; *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260).

Werkgenuss, verglichen werden kann, besteht darüber hinaus ein untrennbarer Zusammenhang zwischen der beim KI-Training stattfindenden Musteranalyse und dem Sammeln der Werke mit Hilfe eines Webcrawlers. Besteht ein freihaltebedürftiges Verarbeitungsinteresse, beispielsweise das Interesse am Werkgenuss oder aber auch das Interesse an der Analyse der den Inhalten zugrundeliegenden Muster und Zusammenhänge, muss auch eine hierfür zunächst notwendige Vervielfältigung möglich sein. Nur so kann die rechtmäßige Benutzung des betreffenden Werkes ermöglicht werden. Diese Kausalität kommt bereits in der urheberrechtlichen Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen insbesondere im Arbeitsspeicher eines Computers (§ 44a UrhG) zum Ausdruck.¹⁷⁸ In gleicher Weise muss auch eine Vervielfältigungshandlung, die für den „automatisierten Werkgenuss“, also das Training einer KI,¹⁷⁹ erforderlich ist, zugunsten der Werknutzer freigestellt sein.¹⁸⁰

dd) Innovationsbegünstigte und Bedeutung des Innovationsinteresses

Für eine Freistellung des Web Scrapings urheberrechtlich geschützter Trainingsdaten spricht weiterhin, dass von den technischen Innovationen, die durch umfangreich verfügbare Trainingsdatensätze ermöglicht werden, nicht nur die Innovatoren profitieren. Vielmehr kann das Innovationsinteresse als gesamtgesellschaftliches Anliegen verstanden werden. Insbesondere von generativen KI-Systemen, also solcher KI, die kreative Leistungsergebnisse produziert, profitieren auch wiederum Urheber selbst. Sie können die KI-Systeme als Werkzeug im Prozess des Werkschaffens einsetzen und Effizienzgewinne erzielen. Ungehinderter Zugang zu Trainingsdaten unabhängig von ihrer Werkqualität ist eine wichtige Voraussetzung dafür, die Entwicklung der KI insgesamt voranzubringen.¹⁸¹ Andersherum droht durch einen mangelnden Zugang zu qualitativ hochwertigen Trainingsdaten, insbesondere verkörpert in urheberrechtlich geschützten Werken wie

178 Stieper, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 8, vgl. auch S. 56 ff.

179 Dazu siehe unter 2. Teil C. II. 1.

180 Hilty/Moscon, Modernisation of the EU Copyright Rules, S. 20; vgl. ebenfalls Hofmann, WRP 2024, 11 Rn. 23.

181 COM(2018) 237 final, S. 12; vgl. außerdem *European Commission u. a., Study on copyright and new technologies: copyright data management and artificial intelligence (SMART 2019/0038)*, S. 208 f.; Hofmann, WRP 2024, 11 Rn. 22 m. w. N.

zum Beispiel wissenschaftlicher Fachliteratur, ein Qualitätsdefizit bei der Entwicklung von KI-Systemen. Denn die Güte des Systems verhält sich proportional zur Güte der verwendeten Trainingsdaten.¹⁸² Defizitäre KI-Systeme haben hingegen Effizienzverluste zur Folge, sodass ihre Entwicklung gesellschaftlich nicht erwünscht ist.¹⁸³

Aus einer ökonomischen Perspektive muss auch berücksichtigt werden, dass es Anwendungsszenarien gibt, für die der Markt keine geeigneten Trainingsdaten zur Verfügung stellt. Dies kann einmal auf die Art der benötigten Trainingsdaten, zum Beispiel umgangssprachliche Formulierungen, zum anderen aber auch auf die Menge der benötigten Trainingsdaten zurückzuführen sein.¹⁸⁴ Ohne eine Freistellung der Speicherung der benötigten, urheberrechtlich geschützten Trainingsdaten wären Innovationen in den jeweiligen Anwendungsbereichen erschwert. Das sollte vor allem generative KI-Systeme betreffen. Dies widerspräche aber dem Interesse an einer Förderung technischer Innovationen, das sowohl im UrhG¹⁸⁵ als auch durch den Richtliniengeber¹⁸⁶ anerkannt worden ist.

ee) Praktische Erwägungen

Aus einer praktischen Perspektive für eine Freistellung des Web Scrapings von Werken zum Training von KI spricht zudem, dass beim Web Scraping von Inhalten aus dem Grenzbereich der urheberrechtlichen Schutzfähigkeit, beispielsweise beim Sammeln von Zeitungsartikeln zum Training einer Algorithmische Journalism-Anwendung, weder von KI-Entwicklern oder anderen Trainingsdatensammlern noch durch den Webcrawler in jedem Fall beurteilt werden kann, ob der abgespeicherte Inhalt als Werk Schutz genießt und daher die urheberrechtlichen Rahmenbedingungen beachtet werden müssen.¹⁸⁷ Die mit dieser Rechtsunsicherheit einhergehenden Ab-

182 *de la Durantaye*, ZUM 2023, 645 (647); *Margoni/Kretschmer*, GRUR Int. 2022, 685 (687); aus einer etwas universelleren Perspektive *Hacker*, ZGE 2020, 239 (243); im Anschluss an die vorgenannten *Hofmann*, WRP 2024, 11 Rn. 22.

183 *Hofmann*, WRP 2024, 11 Rn. 22.

184 Für den Textgenerator „GPT-3“ waren Trainingsdaten mit einem Umfang von 300 Milliarden Wörtern notwendig, siehe *Brown u. a.*, arXiv:2005.14165 2020, S. 8.

185 Beispielsweise in §§ 44b, 60c, 60d UrhG.

186 ErwGr. 5 DSM-RL.

187 Vgl. *European Commission u. a.*, Study on copyright and new technologies: copyright data management and artificial intelligence (SMART 2019/0038), S. 273;

schreckungseffekte können nur durch eine Freistellung der Nutzung auch urheberrechtlich geschützter Inhalte beseitigt werden.

ff) Ökonomischer Vorteil aus einer Freistellung

Selbst wenn quantitativ und qualitativ hinreichende Trainingsdaten auf dem Trainingsdatenmarkt verfügbar wären, müssten KI-Entwickler oder intermediäre Trainingsdatensammler schließlich einen nicht überschaubaren Aufwand betreiben, um Lizenzvereinbarungen mit jedem betroffenen Urheber abzuschließen. Die für Trainingsdatensammler hierdurch entstehenden Transaktionskosten wären sehr hoch und könnten von vielen Trainingsdatensammlern auch nicht erbracht werden.¹⁸⁸ Schließlich sind bei der Wiedergabe von Werken im Internet auch oft die betreffenden Rechteinhaber nicht bekannt, womit auch ein individueller Lizenzabschluss nicht möglich wäre.¹⁸⁹ Das gilt vor allem für Alltagswerke der sogenannten kleinen Münze. Auch das hätte eine Innovationshemmung zur Folge.¹⁹⁰ Insbesondere wäre aber eine die ökonomische Effizienz verwirklichende, selbstgesteuerte Güterallokation durch Transaktionen zwischen den Urhebern und Trainingsdatensammlern bei einer fehlenden Kenntnis über die betroffenen Urheber nicht möglich. Gerade aufgrund der im Zusammenhang zu KI bestehenden Prognoseunsicherheiten, die eine manuelle Güterzuordnung ausschließen,¹⁹¹ sollte dieser Allokationsmechanismus aus einer ökonomischen Perspektive aber möglich sein.

Prohibitiv hohe Transaktionskosten und Probleme bei der Identifizierung von Rechteinhabern führen darüber hinaus dazu, dass eine unterbleibende Freistellung des Web Scrapings von Werken zum Training zu einer

Ebenso im Zusammenhang zur im Entwurf der KI-VO vorgeschlagenen Transparenzpflicht *de la Durantaye*, ZUM 2023, 645 (656).

188 *Bomhard*, InTeR 2023, 174 (175); *Hofmann*, WRP 2024, II Rn.22; *Novelli u. a.*, Computer Law & Security Review 2024, (10); *Raue*, ZUM 2021, 793 (796); *Steinbrecher*, MMR 2019, 639 (640); *European Commission u. a.*, Study on copyright and new technologies: copyright data management and artificial intelligence (SMART 2019/0038), S. 273; *Senftleben u. a.*, JIPITEC 2022, 67 Rn.12; *Senftleben*, IIC 2023, 1535 (1346 f.).

189 *de la Durantaye*, ZUM 2023, 645 (656).

190 So auch *Hacker*, ZGE 2020, 239 (261); *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, S. 238; *Vesala*, IIC 2023, 351 (355); vgl. außerdem *The Intellectual Property Office (GB)*, Modernising Copyright: A modern, robust and flexible framework, S. 37.

191 Dazu siehe oben unter 2. Teil A. I. 3. b).

lose-lose-Situation für Urheber und Trainingsdatensammler führt: Mangels Freistellung können Trainingsdatensammler kein KI-Training auf Basis der Werke durchführen, gleichzeitig erzielen Urheber mit ihren Werken aber auch keine Vergütung durch das Training von KI auf Basis ihrer Werke, weil prohibitiv hohe Transaktionskosten individuelle Lizenzabschlüsse verhindern.¹⁹² Damit wären die Amortisationsinteressen der Urheber durch eine vergütungspflichtige Freistellung sogar besser geschützt als durch eine individuell lizenzliche Lösung.

e) Schlussfolgerungen

Insgesamt sprechen bessere Gründe für eine Freistellung des Web Scrapings von Werken zum KI-Training. Vereinzelt Argumente gegen eine Freistellung konnten hingegen entkräftet werden. Dementsprechend ist den Interessen der Trainingsdatensammler vor den Verwertungsinteressen betroffenen Urheber Vorrang zu gewähren. Infolgedessen ist eine Freistellung des Web Scrapings von Werken als Trainingsdaten insgesamt zweckmäßig und sollte aus einer regulatorischen Perspektive daher erfolgen.¹⁹³

II. Freistellungsmechanismus

Die Zweckmäßigkeit der Freistellung des Web Scrapings von Werken als Trainingsdaten wirft im Weiteren allerdings die Frage auf, durch welchen Regelungsmechanismus eine solche Freistellung konkret umzusetzen ist.

192 *Senfileben*, IIC 2023, 1535 (1548).

193 Zustimmung *Hacker*, ZGE 2020, 239 (261, 269); *Hofmann*, WRP 2024, 11 Rn. 18; *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260); *Trapova/Gervassis*, Kluwer Copyright Blog v. 16.3.2022; grundsätzlich auch *Jäger*, Artificial Creativity?, S. 371, wenn auch mit Bedenken hinsichtlich einer Erfassung kommerzieller Handlungen; in Bezug auf allgemeine Text- und Data-Mining-Aktivitäten auch *Schack*, ZUM 2016, 266 (269); *Schack*, ZUM 2017, 802 (806, 807); *Schack*, Urheber- und Urhebervertragsrecht S. 311 Rn. 627; *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, S. 237 ff.; *Cocoru/Boehm*, An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, S. 7; *European Copyright Society*, General Opinion on the EU Copyright Reform Package, S. 5; *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 125. Ebenfalls ErwGr. 8-II DSM-RL.

1. Angemessener Ausgleich für die Nutzungshandlung

Zur Beantwortung dieser Fragestellung muss zunächst herausgearbeitet werden, ob den Urhebern ein angemessener Ausgleich für das Web Scraping ihrer Werke gezahlt werden sollte.

a) Rechtfertigungsansätze für eine vergütungsfreie Nutzung

Gegen einen finanziellen Ausgleich spricht, dass die Kernverwertungsmöglichkeiten der Urheber durch das Web Scraping der Werke, um sie als Trainingsdaten zu verwenden, nicht unmittelbar beeinträchtigt werden.¹⁹⁴ Die Werke werden zudem nicht *als Werk*, also ihrer geistig-schöpferischen Gehalte wegen, sondern lediglich als Informationsträger ausgewertet.¹⁹⁵ Darüber hinaus profitieren anwendungsbezogen nicht nur die KI-Entwickler, sondern auch andere Beteiligtegruppen und mitunter sogar Urheber selbst von KI-Neuentwicklungen. Mit umfangreichen Trainingsdatensätzen wird zudem auch die Entwicklung der KI als solche vorangebracht.¹⁹⁶ Nicht zuletzt wird verhindert, dass qualitativ minderwertige KI-Systeme Schäden für die Allgemeinheit verursachen.¹⁹⁷ Das mit dem Sammeln der urheberrechtlich geschützten Trainingsdaten verfolgte Innovationsinteresse ist damit auch von gesamtgesellschaftlicher Relevanz.¹⁹⁸ Es ist demzufolge ein öffentliches Interesse, dass eine vergütungsfreie Nutzung grundsätzlich rechtfertigen kann.

Weiterhin gegen einen angemessenen Ausgleich der betroffenen Urheber spricht, dass beim Web Scraping vorrangig solche Quellen untersucht werden, die im *Surface Web*, also dem Teil des Internets verfügbar sind, der ohne weiteres für Webcrawler und damit auch für die Öffentlichkeit verfügbar ist.¹⁹⁹ Damit sind auch die für das KI-Training durch Webcrawler vervielfältigten Werke überwiegend bereits öffentlich verfügbar gewesen.

194 Dazu bereits siehe oben unter 2. Teil B. I. 1.

195 Hofmann, WRP 2024, II Rn. 25; Lennartz/Kraetzig, GRUR 2024, 1258 (1260).

196 COM(2018) 237 final, S.12; vgl. außerdem *European Commission u. a., Study on copyright and new technologies: copyright data management and artificial intelligence (SMART 2019/0038)*, S. 208 f.

197 Dregelies, GRUR 2024, 1484 (1484).

198 In diese Richtung auch Dregelies, GRUR 2024, 1484 (1484).

199 Vgl. zu Web Crawling und dem Hidden Web Gupta/Bhatia, IJCTT 2014, III (III); zum Begriff des Surface Web siehe Lewandowski, in: Kühlen/Lewandowski/Semmar/Womser-Hacker, Grundlagen der Informationswissenschaft, S. 391 (395).

Da demzufolge zumeist bereits Zugang zum Werk bestand, könnte hieraus geschlussfolgert werden, dass durch das Web Scraping für die betroffenen Urheber im Regelfall kein ausgleichsbedürftiger Schaden entsteht.²⁰⁰ In diesem Kontext spricht ebenfalls gegen eine angemessene Vergütung, dass Urheber von im Internet abrufbaren Werken auch die Möglichkeit haben, ihre Werke mit Hilfe digitaler Rechtemanagementsysteme gegen den Zugriff Dritter zu schützen. Hierzu gehört auch, den Zugriff und in der Folge auch das KI-Training auf Basis der Werke von einer angemessenen Vergütung abhängig zu machen.²⁰¹ Mit Blick auf den vom Urheberrecht zur Verfügung gestellten Rechtsrahmen könnten für Trainingsdatensammler, die Werke nur kurzzeitig zum Training von KI abspeichern, Abgrenzungsschwierigkeiten zwischen der vergütungsfreien Schrankenbestimmung des § 44a UrhG für vorübergehende Vervielfältigungshandlungen und einer vergütungspflichtigen Freistellung des Web Scrapings von Werken als Trainingsdaten entstehen, die zu Rechtsunsicherheit führen würden.²⁰² Weiterhin besteht die Gefahr, dass die jeweils anteilige Anrechnung des mit der KI erzielten Gewinnes auf die große Zahl betroffener Urheber mitunter dazu führen könnte, dass die Vergütungshöhe für einzelne Urheber unerheblich und zudem praktisch schwer zu bestimmen ist.²⁰³

Schließlich könnte einer Vergütungspflicht für das Web Scraping urheberrechtlich geschützter Werke zum Training von KI zuletzt entgegengehalten werden, dass viele der betroffenen Urheber aufgrund der niedrigen urheberrechtlichen Schutzworaussetzungen²⁰⁴ und seiner Formfreiheit²⁰⁵ ihre Werke wie beispielsweise Beiträge in sozialen Netzwerken oder kurze Blogposts gar nicht als Werke verwerten. Das führt dazu, dass sie weder in einer Verwertungsgesellschaft organisiert sind, noch ein Interesse daran haben, ihre urheberrechtlichen Befugnisse aktiv wahrzunehmen. Infolge-

200 So für das allgemeine, nicht Trainingsdaten-spezifische Text und Data-Mining *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, S. 243; *de la Durantaye*, GRUR 2017, 558 (562), mit Verweis auf BT-Drs. 18/12329, S. 41. Ähnlich *Stieper*, ZUM 2021, 776 (778), nachdem der Schaden für Rechtsinhaber auch in nicht wissenschaftsbezogenen Fällen minimal sei.

201 *Hofmann*, WRP 2024, 11.

202 *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44b Rn. 14; BT-Drs. 19/27426, S. 88.

203 *Jager*, Artificial Creativity?, S. 371.

204 Selbst ein Satz oder einzelne Satzteile können urheberrechtlichen Schutz begründen, siehe EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 47 – Infopaq.

205 Art. 5 Abs. 2 S. 1 RBÜ, dazu ferner *de la Durantaye/Hofmann*, ZUM 2021, 873 (876 f.).

dessen besteht ein Risiko, dass die von Verwertungsgesellschaften eingezogene Vergütung nicht zu einer Entschädigung der tatsächlich betroffenen Urheber führt.²⁰⁶ Insgesamt sprechen damit einige Gründe dafür, vom Web Scraping betroffenen Urhebern keinen finanziellen Ausgleich für das Web Scraping ihrer Werke zum Training von KI zu gewähren.²⁰⁷

b) Argumente für eine Vergütungspflicht

Allerdings müssen für das KI-Training gesammelte Inhalte vor dem eigentlichen Trainingsverfahren zumindest annotiert, häufig aber auch in qualitativer Hinsicht aufbereitet werden. Dieser Prozess ist ressourcen- und insbesondere zeitintensiv.²⁰⁸ Darüber hinaus müssen Datensätze unter Umständen auch nach Abschluss des Trainingsverfahrens aufbewahrt werden, um Anforderungen an die Nachvollziehbarkeit der durchgeführten Forschung und Entwicklung zu erfüllen sowie ein Weitertraining des KI-Modells im Anwendungsprozess zu ermöglichen.²⁰⁹ Intermediäre Trainingsdatensammler speichern die Werke außerdem, um Trainingsdatensätze zu erstellen, die nicht nur vorübergehend im Sinne des § 44a UrhG auf dem Trainingsdatenmarkt angeboten werden. Die für das KI-Training gesammelten Daten werden damit regelmäßig nicht nach einer unerheblichen Zeit automatisch wieder gelöscht. Das betrifft auch Werke, die für das Training von KI gespeichert worden sind. Die Voraussetzungen der urheber-

206 *Triaille/Meeùs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 103 f.

207 So für das allgemeine Text und Data-Mining *de la Durantaye*, GRUR 2017, 558 (562); *Heesen/Jüngels*, RuZ 2021, 45 (53); *Raue*, CR 2017, 656 (661); *Schack*, ZUM 2016, 266 (282); *Schack*, GRUR 2021, 904 (907); *Stieper*, ZUM 2021, 776 (778); *Stieper*, GRUR 2020, 1 (4); *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, S. 243.

208 Mit einem anschaulichen Beispiel *Paafß/Hecker*, Künstliche Intelligenz, S. 15 f.; *Tenyks*, The Unseen Cost of “Low Quality” Large Datasets.

209 Teilweise ist in Konzepten für das sogenannte Data Lifecycle Management in Bezug auf die KI-Entwicklung und den Umgang mit Trainingsdaten überhaupt keine Löschung der gesammelten Trainingsdaten vorgesehen, siehe beispielsweise *Mishra*, Medium/towards data science v. 12.2.2022. In der Norm ISO/IEC 8183 zum Data Lifecycle Management bei der Entwicklung von KI wird die Löschung beispielsweise neben der Archivierung und Umwidmung mit anschließender Weiterverwendung als eine der Möglichkeiten für das sogenannte Data Decommissioning genannt, siehe dort unter Punkt 6.10. Die Norm empfiehlt weiterhin, bestimmte, selbst festzulegende Datenkategorien für ein späteres Systemaudit aufzubewahren.

rechtlichen Schrankenbestimmung für vorübergehende Vervielfältigungen (§ 44a UrhG) sind praktisch demzufolge im Regelfall nicht erfüllt.²¹⁰ Die befürchteten Abgrenzungsprobleme bestehen demzufolge zwar beim Web Scraping von sonstigen Inhalten zum Zwecke der Informationsextraktion, nicht aber beim Web Scraping von Werken, die als Trainingsdaten für KI verwendet werden sollen.

Auch daraus, dass bereits Zugang zum Werkstück besteht, kann nicht geschlussfolgert werden, dass Urheber durch das Web Scraping keinen ausgleichsbedürftigen Schaden erleiden würden. Denn zum einen führt die bereits bestehende Zugänglichkeit eines Werkes nicht dazu, dass Nutzer auf dieser Basis nicht weitere Verwertungshandlungen vornehmen können, die im Folgenden zu einer Intensivierung der Beeinträchtigung der Urheberinteressen führen. Aus diesem Grund unterliegt zum Beispiel auch die Privatkopiefreiheit Begrenzungen (§ 53 UrhG), auch wenn hierfür denklogisch bereits Zugang zu einem – nicht offensichtlich rechtswidrig hergestellten – Werkstück bestehen muss.²¹¹

In diesem Zusammenhang kann zugunsten der Urheber außerdem berücksichtigt werden, dass das Innovationsinteresse der Allgemeinheit im Gegensatz zum von Art. 17 EU-GRCh beziehungsweise subsidiär Art. 14 GG umfassten Urheberrecht nicht von einem eigenen, speziellen Freiheitsgrundrecht geschützt ist. Nicht die Vergütungsfreiheit, sondern der Werkzugang ist außerdem relevant für das Vorantreiben von Innovationen. Zugang kann aber im Zusammenspiel mit einer angemessenen Vergütung der Urheber erreicht werden.

KI-Entwicklung wird wegen des hohen Datenaufwandes im Übrigen insbesondere von gewerblich handelnden Akteuren durchgeführt. Wirtschaftliche Individualinteressen bei der Entwicklung von KI können jedoch aus sich heraus keine vergütungsfreie Nutzung rechtfertigen. Dass eine Verquickung von Allgemeininteressen wie dem Interesse an Innovationen durch KI-Entwicklung mit starken wirtschaftlichen Individualinteressen im Regelfall zu einer Vergütung derjenigen Urheber führt, die von der Freistellungsregelung betroffen sind, zeigen in systematischer Hinsicht auch die Schrankenbestimmungen zur Herstellung von Sammlungen für den

210 Zu den Voraussetzungen der Schrankenbestimmung aus § 44a UrhG vgl. EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 62 ff. – Infopaq; zu § 44a UrhG und dem Training von KI vgl. ebenfalls Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44a Rn. 4. Im Detail zu § 44a UrhG siehe auch 3. Teil A. III. 1.

211 Stieper, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 58.

religiösen Gebrauch (§ 46 UrhG) sowie zur Herstellung von Schulbüchern (§ 60b UrhG).

Insbesondere bei der Entwicklung generativer KI-Systeme werden urheberrechtlich geschützte Werke außerdem als Trainingsdaten verwendet, damit nach Abschluss des Trainingsprozesses mit den verwendeten Werken ähnliche oder identische Erzeugnisse generiert werden können. Die später generierten Erzeugnisse sind damit grundsätzlich geeignet, die zum Training verwendeten Werke zu ersetzen. Ziel der Generierung des Outputs generativer KI ist typischerweise außerdem eine menschliche Rezeption der erzeugten Inhalte. Damit beeinträchtigen KI-Entwickler oder -Anwender, die mit solchen Systemen am Markt auftreten, die Kernverwertungsmöglichkeiten betroffener Urheber. Gerade in diesen Fällen sollten Urheber nicht die vergütungsfreie Nutzung ihrer Werke dulden müssen.²¹² Weil Trainingsdatensammler mit den Anbietern solcher KI-Systeme auf vertraglicher Ebene unmittelbar oder mittelbar verbunden sind und sie durch die Datenaggregation einen konkreten Einfluss darauf haben, welche Urheber von den Auswirkungen des generativen KI-Systems betroffen sind, kann diese Beeinträchtigung mittelbar auch den Trainingsdatensammlern zugerechnet werden.

Zugunsten der Urheber muss in diesem Zusammenhang auch berücksichtigt werden, dass sich der wirtschaftliche Trainingsvorteil im später folgenden Vorteil fortsetzt, der aus der Anwendung eines KI-Systems generierbar ist. Allein durch die Anwendung generativer KI-Systeme, die nicht unwesentlich auf Grundlage urheberrechtlich geschützter Werke trainiert werden, ergibt sich eine Produktivitätssteigerung von 2,4 bis 4,4 Billionen US-Dollar jährlich.²¹³ Diese Wertschöpfung geht auch auf die Nutzung urheberrechtlich geschützter Werke zurück. Urheber sollten an ihr daher auch beteiligt werden. Dass eine solche Beteiligung entlang der Wertschöpfungskette dem urheberrechtlichen Regelungskomplex systemimmanent ist, zeigen nicht zuletzt auch die Regelungen zum Urhebervertragsrecht und insbesondere § 32a UrhG.

Unabhängig vom Output des KI-Systems können gewerblich tätige Trainingsdatensammler durch die Vermarktung ihres Produktes, beispielsweise einer KI-Anwendung, außerdem Einkünfte erzielen. Gleiches gilt für

212 *European Commission u. a.*, Study on copyright and new technologies: copyright data management and artificial intelligence (SMART 2019/0038), S.207; ebenso *Schack*, NJW 2024, 113 Rn. 24, 26; sich anschließend außerdem *Senfleben*, IIC 2023, 1535 (1538); vgl. zudem *Wandtke*, MMR 2017, 367 (368).

213 *McKinsey*, The economic potential of generative AI: The next productivity frontier.

intermediäre Trainingsdatensammler, die Trainingsdatensätze an KI-Entwickler veräußern. Der Trainingsdatenmarkt hatte im Jahr 2024 ein Volumen von knapp unter drei Milliarden US-Dollar.²¹⁴ Teil hiervon sind auch urheberrechtlich geschützte Werke. Lediglich die Urheber, deren Werke die Entwicklung der KI als Trainingsgrundlage ermöglicht haben, würden bei der vergütungslosen Freistellung des Web Scrapings nicht an der Generierung von Einnahmen auf Basis ihrer eigenen Werke teilhaben. Dieser Value-Gap²¹⁵ könnte durch eine angemessene Ausgleichszahlung der wirtschaftlichen Profiteure, also der KI-Entwickler und intermediären Trainingsdatensammler, geschlossen werden. Das Werke auch Teil des Trainingsdatenmarktes sind, spricht dafür, sie an dieser Wertschöpfung auch partizipieren zu lassen.

Selbst wenn darüber hinaus ein gewisses Risiko bestünde, dass eine von Verwertungsgesellschaften eingezogene Vergütung nicht zu einer Entschädigung der tatsächlich betroffenen Urheber führte, ändert das aber nichts daran, dass betroffenen Urhebern aus den oben genannten Gründen ein Ausgleich für die Werknutzungen zustehen muss. Keine Vergütungspflicht vorzusehen, weil mitunter praktische Schwierigkeiten bei der Verteilung der Einnahmen bestehen, bringt betroffene Urheber in eine umso schlechtere Stellung, wobei das Problem selbst, die Wahrnehmung der Urheberrechte und insbesondere der Vergütungsansprüche, nicht adressiert wird. Gleiches gilt für den Fall, in dem Urheber wegen des geringen Einfluss ihres einzelnen Werkes auf das Training einer KI insgesamt nur eine unwesentliche Vergütung erhalten können.

c) Schlussfolgerung

Alles in allem sprechen damit durchgreifende Gründe für eine vergütungspflichtige Freistellung des Web Scrapings von Werken zum Training von KI. Das Innovationsinteresse der Trainingsdatensammler ist infolgedessen kein gesteigertes öffentliches Interesse, das eine vergütungsfreie Nutzung der Werke zum Training von KI ausnahmsweise rechtfertigt.²¹⁶ Demzufol-

214 *Fortune Business Insights*, AI training data set market.

215 Im hier verwendeten Sinne eine Diskrepanz zwischen den aus dem Umgang mit dem Werk erzielten Einnahmen der KI-Wirtschaft und der Vergütung der Urheber, deren Werke für das Training einer KI verwendet wurden, siehe zum Begriff m. w. N. *Wimmers/Barudi*, GRUR 2017, 327 (328).

216 Vgl. zu einer Vergütungspflicht auch *de la Durantaye*, ZUM 2023, 645 (659).

ge ist die regulatorisch zweckmäßige Freistellung des Web Scrapings von urheberrechtlich geschützten Trainingsdaten nur gegen eine angemessene Vergütung der betroffenen Urheber vorzusehen.²¹⁷

2. Notwendigkeit eines Opt-out-Mechanismus

Fraglich ist weiterhin, ob die Freistellung des Web Scrapings im Wege einer gesetzlichen Lizenz oder mit Hilfe eines ECL-Mechanismus vorgenommen werden sollte. Dafür sind Aspekte der Rechtssicherheit und die Notwendigkeit eines Opt-out-Mechanismus entscheidend.

a) Vorteile eines Opt-out-Mechanismus

Für einen Opt-out-Mechanismus, mit dessen Hilfe betroffene Urheber das Web Scraping ihrer Werke verhindern können, spricht zunächst, dass Urheber hierdurch in die Lage versetzt werden, ihre Sekundärinteressen vor der Beeinträchtigung durch Trainingsdatensammler zu schützen. Enthält beispielsweise ein Schriftwerk autobiografische Züge, so möchte der Urheber unter Umständen vermeiden, dass sein Werk zum Training einer KI verwendet wird. So kann er verhindern, dass das KI-System persönlichkeitsrelevante Inhalte aus seinem Werk nachbildet. Dieses Ziel wird erreicht, indem die Freistellung des Web Scrapings der betreffenden Werke durch Erklärung des Urhebers (rückwirkend) beendet wird.

217 Auch befürwortet von Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 14; Dregelies, GRUR 2024, 1484 (1486); Geiger, IIC 2024, 1129 (1152); Geiger, in: FS Hilty, S. 67 (79 f.); Kowala, IIC 2024, 1604 (1618); Papastefanou, CR 2023, 1 Rn. 49; Schack, NJW 2024, 113 Rn. 24, 26 (vor dem Hintergrund des Dreistufentests); für kommerzielles KI-Training auch Wulf/Löbeth, GRUR 2024, 737 (740 f.); im Grundsatz für eine Vergütungspflicht, aber mit anderer Ausgestaltung (Anknüpfung an den Betrieb des KI-Systems) auch Senfleben, IIC 2023, 1535 (1549 ff.); Jacques/Flynn, GRUR Int 2024, 1137 (1141, 1148), dazu weitergehend siehe unten unter 4. Teil A. II. 4.; weiterhin im Hinblick auf allgemeines Text und Data-Mining auch Spindler, ZGE 2018, 273 (294); Wandtke, MMR 2017, 367 (368); Hilty/Moscon, Modernisation of the EU Copyright Rules, S. 142. anders aber Becker, GRUR 2024, 505 (507); Hofmann, WRP 2024, II Rn. 20; Konertz/Schönhof, WRP 2024, 289 Rn. 35; Raue, ZUM 2021, 793 (797); Jager, Artificial Creativity?, S. 371: nicht kalkulierbar, für einzelne Urheber zu vernachlässigen; in Bezug auf allgemeines Text und Data Mining, aber zur neuen Schrankenregelung auch Stieper, GRUR 2020, I (4).

Darüber hinaus eröffnet ein Opt-out-Mechanismus den vom Web Scraping betroffenen Urhebern aber auch die Möglichkeit, statt einer gesetzlichen Freistellung Individuallizenzen für die Nutzung ihrer Werke zum Training von KI zu erteilen.²¹⁸ Dies ist insbesondere für Urheber von Bedeutung, die nur bereit sind, Lizenzverträge zur Nutzung ihrer Werke für KI-Training zu einem höheren als dem marktüblichen Preis abzuschließen. Letztlich schützt ein Opt-out-Mechanismus damit auch das Selbstbestimmungsrecht der Urheber.²¹⁹

Ein Opt-out-Mechanismus würde schließlich den Vorteil mit sich bringen, dass Urhebern als *Ultima Ratio* ein Werkzeug zur Verfügung gestellt würde, mit dessen Hilfe sie in die Lage versetzt wären, die Nutzung ihrer Werke als Trainingsdaten für solche KI-Systeme zu verhindern, die ihre Kernverwertung besonders intensiv beeinträchtigen.

b) Praktische Schwierigkeiten und weitere Nachteile eines Opt-out-Mechanismus

Allerdings muss auch berücksichtigt werden, dass der Abschluss individueller Lizenzvereinbarungen für das automatisierte Web Scraping von urheberrechtlich geschützten Trainingsdaten häufig *de facto* nicht möglich oder aufgrund der Zahl der betroffenen Rechtsinhaber jedenfalls unzumutbar ist.²²⁰ Da es praktisch demzufolge kaum zu Individuallizenzen kommen wird, besteht für Urheber infolgedessen kein Spielraum, im Rahmen dessen von der erweiterten Selbstbestimmung Gebrauch gemacht werden könnte. Im Gegenteil führen Schwierigkeiten bei der Identifizierung von Trainingsdatensammlern, die Werke ohne Erlaubnis der Urheber abspeichern, dazu, dass die trotz einer Opt-out-Erklärung stattfindende Speicherung von Werken zum Training von KI nur mit erheblichem Aufwand verhindert und durch Schadensersatzforderungen monetär ausgeglichen werden kann. Eine gesetzlich ausgestaltete und pauschalisierte Vergütung erreicht, unabhängig von konkreten Nutzungsnachweisen, alle Urheber.²²¹ Sie sind

218 ErwGr. 18 S. 2 DSM-RL.

219 Raue, ZUM 2021, 793 (796).

220 Dazu bereits ausführlicher siehe oben unter 2. Teil B. I. 2.

221 Alle bereits entstandenen und zukünftigen Ansprüche aus der Hersteller- und Betreibervergütung werden bei Meldung eines Werkes durch Urheber pauschal und einmalig abgegolten, siehe beispielsweise § 3 Nr. 12 Verteilungsplan *VG Wort*.

mit ihr demzufolge bessergestellt als durch einen Opt-out-Mechanismus.²²² Weiterhin ist der Schutz sonstiger, beispielsweise allgemein-persönlichkeitsbezogener Interessen abseits der geistigen und persönlichen Beziehung des Urhebers zu seinem Werk nicht Zweck des Urheberrechts.

Abgesehen davon spricht gegen einen Opt-out-Mechanismus außerdem, dass eine Vorbehaltserklärung insbesondere bei digitalen Werkwiedergaben (im Internet) mit geringem Aufwand abgegeben werden kann.²²³ Neben den faktischen Hürden, die die praktische Wirksamkeit der Vorbehaltserklärung betreffen, droht der Freistellungsmechanismus damit de jure leer zu laufen.²²⁴

Urheber erklären einen Nutzungsvorbehalt zudem ex ante. Mit Beginn einer öffentlichen Wiedergabe der Werke haben sie also noch keine Kenntnis darüber, für welche Arten von KI-Systemen die Werke als Trainingsdaten später einmal genutzt werden sollen. Dies erschwert ihnen die effektive und zielgerichtete Anwendung des Opt-out-Mechanismus als Ultima Ratio zur Verhinderung von Beeinträchtigungen ihrer Kernverwertungsarten. Hinzu kommt, dass gerade urheberrechtlich geschützte Inhalte „kleiner Münze“ wie Beiträge in sozialen Netzwerken oft über Plattformen von Drittanbietern und nicht im Herrschaftsbereich der Urheber selbst öffentlich zugänglich gemacht werden. Insofern für die Erklärung des Opt-outs in diesen Fällen die Mitwirkung der Plattformbetreiber und Infrastrukturanbieter notwendig ist, käme es durch den Opt-out-Mechanismus infolgedessen nicht zur Stärkung der Selbstbestimmung der Urheber. Zudem ist ein Ausgleich zwischen den Interessen betroffener Urheber und Nutzer bereits im Rahmen der Identifizierung des Freistellungsinteresses vorgenommen worden. Daher ist ein Opt-out-Mechanismus zur Wahrung der Urheberinteressen nicht erforderlich.²²⁵

c) Schlussfolgerungen

Alles in allem sprechen demzufolge bessere Gründe dafür, bei der Freistellung des Web Scrapings von Werken zum Training von KI keinen Opt-out-Mechanismus für Urheber vorzusehen.²²⁶

222 Auch dazu bereits siehe oben unter 2. Teil B. II. 1. So auch *Schack*, NJW 2024, 113 Rn. 25.

223 *Hacker*, ZGE 2020, 239 (259).

224 Ähnlich auch *Hacker*, ZGE 2020, 239 (259); *Hofmann*, WRP 2024, 11 Rn. 18.

225 A. A. hingegen *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, VGG § 52 Rn. 1.

3. Erhöhtes Rechtssicherheitsbedürfnis

Trainingsdatensammler könnten beim Web Scraping der Werke weiterhin ein erhöhtes Rechtssicherheitsbedürfnis haben. Dann wäre für sie eine Lizenzlösung, also eine Freistellung im Wege eines ECL-Mechanismus, vorzuzugwürdig. Für die Beurteilung des Rechtssicherheitsbedürfnisses sind unter anderem die Folgen eines möglichen Rechtsverstoßes von Bedeutung. Sie beeinflussen maßgeblich das mit dem Web Scraping Werken zum Training von KI einhergehende wirtschaftliche Risiko.

Für ein erhöhtes Rechtssicherheitsbedürfnis der Trainingsdatensammler spricht zunächst, dass das Web Scraping von Werken als Trainingsdaten typischerweise eine Vielzahl von Urhebern betrifft. Bestehen hinsichtlich der zu erfüllenden Freistellungsvoraussetzungen strukturelle Defizite bei der Speicherung der Werke, droht Trainingsdatensammlern, sich Ansprüchen von einer Mehrzahl von Rechtsinhabern ausgesetzt zu sehen. Das führt zu einer relativen Erhöhung des prozessualen Kostenrisikos für Trainingsdatensammler.

Allerdings ist das Prozessrisiko, dem sich Trainingsdatensammler aussetzen, wegen der erschwerten Aufdeckungsmöglichkeiten für Urheber ohnehin begrenzt. Darüber hinaus folgen auf das Web Scraping der Werke stets die Vorverarbeitung und ggf. Annotation des Trainingsmaterials. Sobald die vorbereitenden Maßnahmen abgeschlossen sind, kann der eigentliche Trainingsvorgang beginnen. Die Nutzung von Trainingsdaten folgt strukturell infolgedessen stets dem gleichen Muster. Der Sachverhalt kann wegen seiner konzeptuellen Uniformität damit auch in einer abstrakt-generellen

226 So auch *Geiger*, in: FS Hilty, S. 67 (79 f.); *Hofmann*, WRP 2024, 11 Rn. 20; *Chiou*, JIPITEC 2019, 398 Rn. 23; *Margoni/Kretschmer*, GRUR Int. 2022, 685 (686); *Schack*, NJW 2024, 113 Rn. 25; *Trapova/Gervassis*, Kluwer Copyright Blog v. 16.3.2022; im Ergebnis auch *Brockmeyer*, Text und Data Mining, S. 166 ff.; in diese Richtung, jedoch lediglich bezogen auf allgemeines Text und Data-Mining im nicht-kommerziellen, wissenschaftlichen Kontext außerdem *Triaille/Meeüs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 107 ff.; anders *Hacker*, ZGE 2020, 239 (259): „sachgerechter Ausgleich von Verwertungs- und Informationsinteressen“.

Auch praktisch kann in einem ersten Fall die Einführung eines Opt-outs beobachtet werden: Nach Protesten einiger der die Plattform *Artstation* nutzenden Künstler, die herausfanden, dass die dort hochgeladenen, digitalen Vervielfältigungsstücke ihrer Werke ohne ihr Wissen zum Training eines KI-Bildgenerators genutzt wurden, führte der Plattformbetreiber *Epic Games* eine Funktion ein, mit Hilfe derer die Künstler der Nutzung ihrer Werke zum KI-Training beim Hochladen manuell widersprechen können, siehe *Bünthe*, heise online v. 16.12.2022.

Rechtsnorm abgebildet werden. Es besteht daher kein Konkretisierungsbedarf durch einzelfallspezifische Lizenzverträge.

Weiterhin können die urheberrechtlichen Beseitigungs-, Unterlassungs- und Löschungsansprüche (§§ 97 Abs. 1, 98 UrhG) nur gegen konkrete Urheberrechtsverletzungen gerichtet werden. Die rechtsverletzende Vervielfältigung eines urheberrechtlich geschützten Trainingsdatums führt daher zu Ansprüchen gegen den Trainingsdatensammler in Bezug auf das konkrete Werk als Trainingsdatum. Die trainierte KI²²⁷ oder ein vollständiger Trainingsdatensatz sind von den spezifischen Beseitigungs- oder Löschungsansprüchen damit nicht betroffen.²²⁸ Im Falle einer Schadensersatzhaftung (§ 97 Abs. 2 UrhG) wäre möglicherweise nur der Verkehrswert des betreffenden Werkes als Trainingsdatum zu ersetzen (Lizenzanalogie).²²⁹ Soll hingegen der Verletzergewinn abgeschöpft werden, kommt es maßgeblich darauf an, welcher Gewinn tatsächlich auf der streitigen Urheberrechtsverletzung, also der Verletzung der Rechtsposition am jeweiligen Werk, beruht.²³⁰ Insbesondere Gewinn, der aufgrund anderer Umstände erzielt worden ist, muss dem Trainingsdatensammler verbleiben.²³¹ Da eine KI mit einer großen Menge von Trainingsdaten angelernt wurde und einzelne, auf rechtsverletzende Weise aggregierte Werke dementsprechend nur geringen Einfluss auf den Trainingsdatensatz insgesamt und damit auch das trainierte KI-System haben, kann der betroffene Urheber nur einen kleinen Teil des mit dem KI-System erzielten Gewinns herausverlangen. Gleiches gilt für das Verhältnis zwischen einzelnen, urheberrechtsverletzend gespeicherten Werken und den Trainingsdatensätzen, in denen Werke in urheberrechtsverletzender Weise enthalten sind. Selbst wenn ein Trainingsdatensammler solchen Ansprüchen ausgesetzt ist, sind seine wesentlichen Investitionen daher nicht bedroht.

Für Trainingsdatensammler ist das wirtschaftliche Risiko bei der Verletzung von Urheberrechten an als Trainingsdaten genutzten Werken alles in allem daher nicht außergewöhnlich hoch. Auch aufgrund der schematisierbaren Nutzungshandlung besteht beim Web Scraping von Werken zum Training von KI infolgedessen kein besonderes Rechtssicherheitsbedürfnis für Trainingsdatensammler, welches eine Freistellung im Wege eines ECL-

227 Dazu im Detail unter 3. Teil B. IV.

228 Wie hier auch *Linke*, in: Kuschel/Asmussen/Golla, Intelligente Systeme - intelligentes Recht: GRUR Junge Wissenschaft Hamburg 2020/2021, S. 179 (197).

229 Vgl. BGH, Urt. v. 22.03.1990 - I ZR 59/88, GRUR 1990, 1008 (1010) - Lizenzanalogie.

230 BGH, Urt. v. 30.01.1959 - I ZR 82/57, GRUR 1959, 379 (380) - Gasparone; *Wolff/Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 97 Rn. 76 m. w. N.

231 BGH, Urt. v. 24.07.2014 - I ZR 27/13, GRUR 2015, 269 Rn. 21 - K-Theory.

Mechanismus nahelegt. Stattdessen kann die Freistellung als abstrakt-generelle Rechtsnorm umgesetzt werden.

4. Schlussfolgerung für den Freistellungsmechanismus

Die Freistellung des Web Scrapings von Werken für das Training von KI sollte als abstrakt-generelle Freistellungsregelung realisiert sein. Trotz der Zweckmäßigkeit der Freistellung ist jedoch ein angemessener Ausgleich für die Speicherung der Werke zugunsten betroffener Urheber zu gewährleisten. Ein Opt-out-Mechanismus zur Beendigung der Freistellungswirkung sollte den Urhebern allerdings nicht zur Verfügung gestellt werden. Insgesamt ist die Freistellung des Web Scrapings von Werken zum Training von KI im Urheberrecht daher als gesetzliche Lizenz, also vergütungspflichtige Schrankenbestimmung, vorzusehen.²³²

III. Unterscheidung zwischen kommerzieller und nicht-kommerzieller Tätigkeit

Die Freistellung des Web Scrapings von urheberrechtlich geschützten Trainingsdaten könnte weiterhin unterschiedslos sowohl kommerzielle als auch nicht-kommerzielle Verarbeitungszwecke erfassen. Dagegen spricht zwar, dass das Web Scraping im nicht-kommerziellen Umfeld wegen der einge-

232 So auch *Geiger*, IIC 2024, 1129 (1152 f.); anders aber *Margoni/Kretschmer*, GRUR Int. 2022, 685 (689), die eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte fordern. Ebenso wohl auch *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260). So zum allgemeinen Text und Data-Mining mit einigen Einschränkungen grundsätzlich auch *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, S. 214, 240 f.; *Triaille/Meeûs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 99 ff.; außerdem der Unionsgesetzgeber, siehe ErwGr. 8, 18 DSM-RL; eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte zugunsten des allgemeinen Text und Data-Mining fordern hingegen *Schack*, ZUM 2016, 266 (269, 282); *Schack*, ZUM 2017, 802 (806); *Schack*, GRUR 2021, 904 (907); *Schack*, Urheber- und Urhebervertragsrecht, S. 311 Rn. 627; *European Copyright Society*, General Opinion on the EU Copyright Reform Package, S. 5; sogar noch weitergehend (Ausschluss jeder digitalen Vervielfältigung aus dem urheberrechtlichen Schutzbereich und Fokus auf die nachfolgenden Verwertungshandlungen) *Cocoru/Boehm*, An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, S. 7.

schränkten Verwertung der gesammelten Trainingsdaten oder entwickelten KI-Systeme eine geringere Belastung für betroffene Urheber bedeutet.²³³ Andererseits besteht die Befürchtung, dass eine auch auf kommerzielle Handlungen ausgeweitete Freistellung nicht mit dem Dreistufentest vereinbar ist, insbesondere weil kein bestimmter Sonderfall mehr freigestellt wäre, so wie es Art. 5 Abs. 5 InfoSoc-RL voraussetzt.²³⁴ Allerdings würden sich für kommerziell tätige Trainingsdatensammler und KI-Entwickler Wettbewerbsnachteile im Vergleich zu anderen, für sie günstigeren Rechtsordnungen ergeben. Diese würden zu einer Abwanderung der KI-Industrie führen.²³⁵ Abgrenzungsschwierigkeiten zwischen kommerzieller und nicht-kommerzieller Tätigkeit würden weiterhin Rechtsunsicherheiten hervorrufen, die Abschreckungseffekte hervorruft.²³⁶ Insbesondere ergab die Abwägung von Nutzer- und Urheberinteressen ein umfassendes Freistellungsinteresse aber auch ohne Einschränkungen hinsichtlich der Kommerzialität des Web Scrapings. Das zeigt, dass das Freistellungsbedürfnis auch für kommerzielle Nutzungszwecke uneingeschränkt besteht. Auch aus dieser Perspektive bedarf es daher keiner Einschränkung auf nicht-kommerzielle Tätigkeiten.²³⁷ Neben alledem verdeutlicht das U.S.-Unternehmen *OpenAI*, dass es keinen Garanten dafür gibt, dass ein im nicht-kommerziellen, wissenschaftlich orientierten Kontext trainiertes KI-System nach Abschluss der urheberrechtlich relevanten Nutzungshandlungen und einer Neuausrichtung des Unternehmensziels nicht doch zu kommerziellen Zwecken eingesetzt wird. Gleiches gilt für Ausgründungen aus Hochschulen, die aus innovationspolitischer Sicht auch wünschenswert sind. Dann ist eine Beschränkung der Freistellungswirkung auf die Tätigkeit nicht-kommerzieller Trainingsdatensammler ohnehin obsolet.

233 Vgl. auch *Triaille/Meeüs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 102 f.

234 *Jager*, Artificial Creativity?, S. 371; vorläufiges Ergebnis zum allgemeinen Text und Data-Mining bei *Triaille/Meeüs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 102 f.

235 Vgl. *Raue*, GRUR 2017, 11 (15 f.); *Hargreaves u. a.*, Standardisation in the area of innovation and technological development, notably in the field of text and data mining, S. 64; *Senfleben*, IIC 2023, 1535 (1348).

236 Im Kontext von Text und Data-Mining zu sonstigen Datenanalysezwecken *Cocoru/Boehm*, An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, S. 8; *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 129.

237 Für das allgemeine Text und Data-Mining ähnlich auch *European Copyright Society*, General Opinion on the EU Copyright Reform Package, S. 5.

Transaktionskosten müssen zudem auch für kommerzielle Trainingsdatensammler gesenkt werden.²³⁸ Auch im Rahmen ihrer Tätigkeit besteht ein erhebliches Innovationspotential.²³⁹ Auch wenn sich die Freistellungswirkung auf kommerziell tätige Trainingsdatensammler erstreckt, ist die Freistellung darüber hinaus dennoch inhaltlich auf das Web Scraping von Werken zum KI-Training beschränkt. Das ist eine abgegrenzte Nutzungsart. Sie kann nicht deswegen kein Sonderfall im Sinne des Dreistufentests mehr sein, weil mehr Werknutzer im Sonderfall von der Freistellungswirkung profitieren. Ob ein Sonderfall vorliegt, richtet sich nutzungsbezogen nach der Reichweite der Freistellung und nicht nach dem von der Regelung angesprochenen Adressatenkreis.²⁴⁰ Demzufolge wäre eine Ausweitung der Freistellungswirkung auf kommerziell tätige Werknutzer auch mit dem Dreistufentest vereinbar.

Insgesamt überwiegen damit die Gründe dafür, das Web Scraping von Werken zum Training von KI gleichermaßen für kommerziell und nicht-kommerziell tätige Trainingsdatensammler freizustellen.²⁴¹

IV. Zwischenergebnis

Das Web Scraping von Werken, um sie nach einer Aufbereitung als Trainingsdaten für die Entwicklung von KI zu verwenden, kann sowohl von KI-Entwicklern selbst als auch von intermediären Trainingsdatensammlern durchgeführt werden. Letztere stellen Trainingsdatensätze zusammen, die auf dem Markt für Trainingsdaten erworben werden können.

Das Web Scraping beeinträchtigt die Verwertungsinteressen betroffener Urheber nur geringfügig. Gleichzeitig verfolgen Trainingsdatensammler schutzwürdige Innovationsinteressen. Abschlüsse von Lizenzverträgen zur Nutzung der Werke sind aufgrund praktischer Schwierigkeiten allerdings nicht möglich. Demzufolge sollte das Web Scraping von Werken zum Trai-

238 Zum allgemeinen Text und Data-Mining *Raue*, CR 2017, 656 (656).

239 So auch für das allgemeine Text und Data-Mining *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 129; ebenfalls der Richtliniengeber, siehe ErwGr. 18 UAbs. 1 DSM-RL.

240 Vgl. ähnlich *Loewenheim*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 44a Rn. 14, der von „eng umrissenen Tatbeständen“ spricht.

241 So auch *Hacker*, ZGE 2020, 239 (269); a. A. allerdings *Jager*, Artificial Creativity?, S. 371; anders zum unspezifischen Text und Data-Mining außerdem *Triaille/Meeûs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 99 f.

ning von KI urheberrechtlich freigestellt sein. Die Freistellung ist im Wege einer vergüteten Schrankenbestimmung umzusetzen. Von der Schrankenbestimmung sollten sowohl kommerzielle als auch nicht-kommerzielle Trainingsdatensammler profitieren. Ein Opt-out-Mechanismus, mit dem Urheber die Freistellungswirkung beenden können, ist hingegen nicht vorzusehen.

C. Durchführung des Trainings auf Basis urheberrechtlich geschützter Trainingsdaten

Nach der Speicherung der für das Training bestimmten Werke mittels Webcrawlern wird auf Basis der Werke der eigentliche Trainingsvorgang durchgeführt. Dieser beginnt im Regelfall mit einer Vorverarbeitung der zu verwendenden, urheberrechtlich geschützten Trainingsdaten. Einzelfallabhängig werden hierbei erhebliche Änderungen an den für das Training bestimmten digitalen Werkexemplaren vorgenommen. An die Vorverarbeitung der Werke schließt sich der eigentliche Trainingsvorgang im engeren Sinne, also die Analyse der Werke durch das KI-System, an. Ziel des Trainingsvorgangs ist eine Konfiguration der im KNN enthaltenen numerischen Parameter. Diese erfolgt in einer Art und Weise, die im Anschluss eine bestmögliche Qualität des vom KI-System generierten Outputs gewährleistet.

Klärungsbedürftig ist auch für die Durchführung dieses Trainingsprozesses auf Basis von Werken, ob und inwiefern regulatorische Freistellungsbedürfnisse bestehen. Entscheidend ist damit die Zweckmäßigkeit der Freistellung. Insofern eine Freistellung des Trainings erfolgen sollte, stehen hierfür verschiedene Freistellungsmechanismen bereit.²⁴²

242 Zur Herleitung der regulatorischen Entscheidungskriterien insgesamt siehe unter 2. Teil A.

I. Zweckmäßigkeit der Freistellung

Eine Freistellung des eigentlichen KI-Trainingsprozesses ist zweckmäßig, wenn hierdurch keine Urheberinteressen beeinträchtigt werden oder die betroffenen Urheberinteressen die gegenläufigen Nutzerinteressen im Rahmen einer Abwägung überwiegen. Für die an dieser Stelle vorzunehmende Analyse und Gegenüberstellung der Urheber- und Nutzerinteressen kann dabei insgesamt auf das zur Speicherung urheberrechtlich geschützter Trainingsdaten Ausgeführte verwiesen werden.²⁴³ Die Interessenlage ist insoweit identisch. Daraus folgt, dass die Durchführung des Trainings auf Basis urheberrechtlich geschützter Werke zwar die Verwertungsinteressen betroffener Urheber beeinträchtigt. Dennoch überwiegen die verfolgten Nutzerinteressen in einer Gesamtabwägung. Damit erweist sich eine Freistellung der Durchführung des Trainings auf Basis urheberrechtlich geschützter Werke als zweckmäßig.

II. Freistellungsmechanismus

Die Wahl eines geeigneten Freistellungsmechanismus richtet sich zum einen danach, ob den betroffenen Urhebern als Kompensation für die Freistellung des KI-Trainings auf Basis ihrer Werke eine angemessene Vergütung gewährt werden sollte. Sofern dies der Fall ist, ist eine Freistellung im Wege einer gesetzlichen Lizenz oder eines ECL-Mechanismus umzusetzen. Falls eine Vergütung nicht notwendig ist, ist die Freistellung demgegenüber als unvergütete Schrankenbestimmung oder durch unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte interessengerecht. Die Wahl zwischen gesetzlicher Lizenz und ECL-Mechanismus richtet sich insbesondere danach, ob den Urhebern ein Opt-out-Mechanismus zur Verfügung zu stellen ist, mittels dessen sie die Freistellungswirkung beenden können. Von Bedeutung ist außerdem, ob seitens der KI-Entwickler ein besonderes Rechtssicherheitsbedürfnis besteht. Die Wahl zwischen unvergüteter Schrankenbestimmung und unmittelbar begrenzender Ausgestaltung der urheberrechtlichen Verwertungsrechte richtet sich insbesondere danach, inwiefern ein Schutz eingesetzter technischer Schutzmaßnahmen gegen das Training von KI über § 95a UrhG interessengerecht erscheint. Auch die Beweislastverteilung zwischen Urheber und Nutzer ist maßgeblich.²⁴⁴

243 Dazu siehe oben unter 2. Teil B. I.

1. Angemessene Vergütung der Urheber

Den Urhebern ist für eine Beschränkung ihrer absoluten Rechtsposition am Werk grundsätzlich eine angemessene Vergütung zu gewähren, wenn nicht ein gesteigertes öffentliches Interesse ausnahmsweise eine vergütungsfreie Nutzung rechtfertigt.²⁴⁵ Ob das der Fall ist, muss anhand einer Abwägung ermittelt werden.²⁴⁶ Dabei ist ausreichend, dass „unter dem Strich“ eine angemessene Vergütung der Urheber gewährleistet bleibt.²⁴⁷

Mit dem KI-Training wird ein öffentliches Innovationsinteresse verfolgt, sodass eine vergütungsfreie Verwendung der Werke nach einer umfassenden Interessenabwägung grundsätzlich möglich ist.²⁴⁸ Für eine Vergütung der Urheber bei der Verwendung ihrer Werke zum KI-Training spricht, dass hinsichtlich der Interessenlage kaum Unterschiede zur Generierung, also Speicherung der Werke mit Hilfe von Webcrawlern bestehen, die von Urhebern ebenfalls nur gegen Zahlung einer angemessenen Vergütung zu dulden ist.²⁴⁹

Allerdings werden die als Trainingsdaten genutzten Werke beim KI-Training mit dem Ziel analysiert, die numerischen Parameter des KI-Systems zu konfigurieren. Dabei wird die Gesamtheit aller den Trainingsdaten zugrundeliegenden Muster und Zusammenhänge, also Informationen, auf die Parameter des KI-Systems übertragen. Kern des KI-Trainings ist also eine inhalts- und informationsbezogene Analyse der Trainingsdaten. Die Analyse der in einem Werk enthaltenen Informationen ist aber auch Wesensmerkmal des Werkgenusses.²⁵⁰ Dass die Informationsaufnahme dabei nicht unmittelbar durch einen Menschen stattfinden muss, sondern auch durch Maschinen erfolgen kann, zeigt der urheberrechtliche Schutz von Computerprogrammen, deren „Rezeption“ im ersten Schritt durch Compu-

244 Zur Herleitung der Differenzierungskriterien siehe oben unter 2. Teil A. II. 2.

245 BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (48) – Kirchenmusik; vgl. auch BVerfG, Ur. v. 11.10.1988 - 1 BvR 743/86 u. a., NJW 1992, 1307 (1308) – Justizvollzugsanstalten.

246 Vgl. *Lauber-Rönsberg*, Urheberrecht und Privatgebrauch, S. 112.

247 BVerfG, Beschl. v. 07.07.1971 - 1 BvR 765/66, GRUR 1972, 481 (483) – Kirchen- und Schulgebrauch; BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (46) – Kirchenmusik; BVerfG, Ur. v. 31.05.2016 - 1 BvR 1585/13, GRUR 2016, 690 – Rn. 80 – Metall auf Metall.

248 Vgl. zur Begründung des öffentlichen Innovationsinteresses oben unter 2. Teil B. II. 1.

249 Dazu 2. Teil B. II. 1.

250 So auch *Sucker*, Der digitale Werkgenuss im Urheberrecht, S. 39 f., 42 (Werkgenuss im engen Sinne).

ter erfolgt.²⁵¹ Der (automatisierte) Werkgenuss und das KI-Training sind als inhaltsbezogene Analysen der Werke demzufolge miteinander vergleichbar.²⁵² Hiergegen spricht auch nicht, dass das KI Training den Werkgenuss sowohl quantitativ als auch qualitativ übersteigt.²⁵³ Denn dass die *Art und Weise* zweier Werkverwendungen zu strukturellen Ähnlichkeiten zwischen beiden Verwendungsarten führt, kann nicht dadurch in Frage gestellt werden, in welchem *Maß* die Verwendung der Werke stattfindet. Ein überdurchschnittlich aufmerksamer Vielleser wird ebenso wenig gegenüber einem durchschnittlichen Rezipienten schlechter gestellt.

Die Rezeption urheberrechtlich geschützter Werke ist jedoch vergütungsfrei möglich. Dies spricht dafür, auch die eigentliche Durchführung des KI-Trainings auf Basis von Werken vergütungsfrei zu stellen.

Darüber hinaus konnte gezeigt werden, dass bereits die vorgelagerte Generierung, also Speicherung der urheberrechtlich geschützten Trainingsdaten mit Hilfe von Webcrawlern, nur gegen einer angemessene Vergütung der betroffenen Urheber freigestellt sein sollte.²⁵⁴ Wenn im Folgenden auch die eigentliche Durchführung des KI-Trainings nur gegen eine Vergütung der Urheber erlaubnisfrei gestellt würde, drohte eine Doppelvergütung des wirtschaftlich einheitlichen Sachverhalts zulasten der KI-Entwickler, die sowohl die Werke mit Hilfe von Web Scraping in Trainingsdatensätzen speichern als auch das KI-Training durchführen.

Außerdem ist durch die Entschädigung für die Speicherung der Werke bereits „unter dem Strich“ eine angemessene Vergütung betroffener Urheber für die Verwendung der Werke zum Training von KI gewährleistet. Insgesamt sprechen in einer Gesamtabwägung damit durchgreifende Gründe dafür, dass das gesteigerte öffentliche Interesse der KI-Entwickler ausnahmsweise eine vergütungsfreie Durchführung des KI-Trainings rechtfertigt. Demzufolge sollte die Freistellung des eigentlichen Trainingsvorgangs

251 *Kleinkopf*, Text- und Data-Mining, S. 177 f.

252 Ebenso *Kögel*, DSRITB 2023, 285 (287); *Lennartz/Kraetzig*, GRUR 2024, 1258 (1259); vgl. zum allgemeinen Text und Data-Mining und dem Werkgenuss weiterhin *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 3; *Raue*, ZUM 2019, 684 (686); *Cocoru/Boehm*, An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, S. 8; im Hinblick auf allgemeines Text und Data-Mining kritisch, letztendlich aber zustimmend allerdings *Kleinkopf*, Text- und Data-Mining, S. 172 ff., 181; a. A. hingegen *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 103 ff.

253 *Schack*, NJW 2024, 113 Rn. 9; in diese Richtung wohl auch *Dornis*, GRUR 2021, 784 (103 ff.).

254 Dazu siehe oben unter 2. Teil B. II. 1.

entweder im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte oder einer unvergüteten Schrankenbestimmung umgesetzt werden.

2. Schutz technischer Schutzmaßnahmen über § 95a UrhG

Die Wahl zwischen der Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte und einer vergütungsfreien Schrankenbestimmung richtet sich weiterhin insbesondere nach dem Bedürfnis zum Schutz technischer Schutzmaßnahmen über § 95a UrhG, die gegen das KI-Training eingesetzt worden sind. Gegen einen derartigen Schutz spricht insbesondere, dass das KI-Training neue technische Anwendungen hervorbringt und eng mit dem öffentlichen Innovationsinteresse verbunden ist. Der Schutz technischer Schutzmaßnahmen sollte allerdings nur für solche Maßnahmen gelten, die nicht die Entwicklung neuer technischer Geräte behindern.²⁵⁵ Außerdem sichert der Schutzmechanismus umfassend die Verwertung von Werken im digitalen Umfeld ab.²⁵⁶ Das KI-Training berührt aber nicht die Kernverwertungsmöglichkeiten betroffener Urheber, weil das Werk nicht als Werk, also um seines schöpferischen Gehalts Willen, verarbeitet wird.

Im Ergebnis sind technische Schutzmaßnahmen zur Verhinderung des KI-Trainings dementsprechend nicht in den Schutzbereich des § 95a UrhG einzubeziehen. Das spricht für eine Umsetzung des regulatorischen Freistellungsinteresses im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte.

3. Vergleich zum Werkgenuss

Bereits gezeigt werden konnte, dass die eigentliche Durchführung des KI-Trainings mit dem Werkgenuss vergleichbar ist.²⁵⁷ Er wird nicht von den urheberrechtlichen Verwertungsrechten erfasst.²⁵⁸ Diese teleologische

255 ErwGr. 48 InfoSoc-RL.

256 Zur Differenz zwischen urheberrechtlichem Schutzbereich und Schutzzumfang des § 95a UrhG siehe *Specht-Riemenschneider*, GRUR 2019, 253 (258).

257 Dazu m. w. N. siehe unmittelbar oben unter 2. Teil C. II. 1.

258 *Heerma*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 15 Rn. 10 m. w. N. außerdem bereits BT-Drs. IV/270, S. 28. Nach *Sucker* gilt das jedenfalls für den analogen Werk-

Parallele spricht ebenfalls dafür, die Durchführung des KI-Trainings auf Basis urheberrechtlich geschützter Trainingsdaten durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte (in den §§ 15 ff. UrhG) freizustellen.

III. Zwischenergebnis

Alles in allem bestehen regulatorische Freistellungsinteressen, aufgrund derer die Durchführung des KI-Trainings auf Basis von Werken urheberrechtlich freigestellt sein sollte. Die vergütungsfreie Freistellung ist im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte umzusetzen. Die eigentliche Durchführung des KI-Trainings muss jedoch von der Datenaggregation abgegrenzt werden, für die vorangehend bereits das Interesse an einer Freistellung im Wege einer vergüteten Schrankenbestimmung identifiziert werden konnte.

D. Speicherung von urheberrechtlich geschützten Inputdaten

Freistellungsbedürftig könnte darüber hinaus auch die Speicherung von urheberrechtlich geschützten Inputdaten sein.²⁵⁹ Diese können beispielsweise mittels Sensortechnik wie Kameras oder Mikrofonen oder durch Webcrawler erfasst, aber auch manuell in ein KI-System eingegeben werden. Hierzu gehört auch die Verwendung von geschützten Werken als Prompt eines generativen KI-Systems. Von der Speicherung urheberrechtlich geschützter Trainingsdaten unterscheidet sie sich dadurch, dass in der Anwendungsphase des KI-Systems, in der die Inputdaten verarbeitet werden, nur vorübergehende Speicherungen der Werke notwendig sind.

genuss, während der digitale Werkgenuss mittelbar durch die zwangsläufig stattfindenden Vervielfältigungshandlungen vom urheberrechtlichen Ausschließlichkeitsrecht erfasst wird, siehe *Sucker*, Der digitale Werkgenuss im Urheberrecht, S. 47 ff., 89. Diese – zutreffende – folgenreorientierte Betrachtungsweise hat aber keinen Einfluss darauf, dass der eigentliche menschliche Analysevorgang, also der Werkgenuss im engeren Sinne, auch im digitalen Kontext nicht dem Ausschließlichkeitsrecht der Urheber zugewiesen ist.

259 Zur Herleitung der regulatorischen Entscheidungskriterien insgesamt siehe unter 2. Teil A.

I. Freistellungsbedürfnis

Die Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten ist dann vorzunehmen, wenn sie sich insgesamt als zweckmäßig erweist.

1. Beeinträchtigung der Urheberinteressen

Eine Zweckmäßigkeit der Freistellung kann sich zum einen daraus ergeben, dass die Interessen der Urheber, deren Werke als Inputdaten erfasst und verarbeitet werden, im vorliegenden Fall nicht berührt sind. In Betracht kommen dabei neben dem materiellen Verwertungsinteresse auch die immateriellen Anerkennungs-, Integritäts- und Erstveröffentlichungsinteressen betroffener Urheber.

Weil allein durch die Speicherung eines Werkes zur Verwendung als Inputdatum keine Umgestaltung stattfindet, scheidet eine Beeinträchtigung des Integritätsinteresses der Urheber aus. Mangels menschlicher Wahrnehmung der vervielfältigten Inputdaten im KI-System ist auch nicht das mit dem Amortisationsbedürfnis verknüpfte Interesse der Urheber betroffen, öffentlich als Schöpfer des Werkes anerkannt zu werden. Da die Speicherung von als Inputdaten zu verwendenden Werken nur vorübergehend ist und keine Verfügbarmachung der Inhalte gegenüber Dritten erfolgt, ist auch das Erstveröffentlichungsinteresse der Urheber nicht berührt. Insgesamt sind die immateriellen Urheberinteressen von der Speicherung urheberrechtlich geschützter Inputdaten demzufolge nicht beeinträchtigt.

Fraglich bleibt aber, ob die Speicherung der Werke zur Verwendung als Inputdaten die Vermögensinteressen der Urheber beeinträchtigt. Das ist der Fall, wenn sie die Perspektive der Urheber auf eine Amortisation ihrer Schöpfungsleistung schmälert.

Inputdaten verwirklichen ihren Zweck in der Anwendungsphase der KI, ermöglichen also den operativen Einsatz des KI-Systems. Je nach Aufgabenstellung und den hierfür benötigten Inputdaten müssen sie sofort oder zeitnah verarbeitet werden. Das ist beispielsweise bei KI-Systemen der Fall, die audiovisuelle Daten für die Steuerung autonomer Fahrzeuge nutzen. Gleiches gilt für Anwendungen im nachrichtenbezogenen Tätigkeitsbereich, die auf eine zeitnahe Verarbeitung aktueller journalistischer Veröffentlichungen zur automatisierten Erstellung von Pressespiegeln angewiesen sind. Hierbei ist im Vorhinein auch nicht bekannt, welche Werke für die Anwendung

des KI-Systems benötigt werden. Dies hängt insbesondere vom konkreten Anwendungsszenario und der Umgebung des KI-Systems ab. Somit ist auch ein vorbeugender individueller Lizenzerwerb nicht möglich. Eine individuelle Lizenzierung der Nutzung erfasster Werke kann in diesen Anwendungsfällen daher regelmäßig nicht erfolgen. Mangels Nachfrage wird es in solchen Anwendungsszenarien daher nicht zu Lizenzabschlüssen für die Nutzung von Werken als Inputdaten kommen. Bei der Verwendung von Werken als Inputdaten für solche KI-Systeme wird die Amortisationsmöglichkeit der Urheber und damit ihr Verwertungsinteresse demzufolge nicht unmittelbar beeinträchtigt.

Allerdings können Inputdaten selbst in den beschriebenen Fallkonstellationen und auch darüber hinaus jedenfalls nach der Anwendung des KI-Systems zu anderen Zwecken weiterverwendet werden. Denkbar ist insbesondere eine Verwendung der Daten zum Training eines KI-Systems.²⁶⁰ Da den Inhalten im Rahmen solcher Sekundärverwertungen eine wirtschaftliche Bedeutung zukommen kann, ist bereits bei der Speicherung der urheberrechtlich geschützten Inputdaten das Verwertungspotenzial der erfassten Werke mit in den Blick zu nehmen. Aus diesem Grund beeinträchtigt die Speicherung urheberrechtlich geschützter Inputdaten in den skizzierten Fallkonstellationen die Verwertungsinteressen der Urheber mittelbar.

Darüber hinaus ist nicht auszuschließen, dass in anders gelagerten, voraussehbaren und nicht zeitkritischen Fallkonstellationen wie der gezielten Anwendung generativer KI-Systeme Spielräume für individuellen Lizenzerwerb zur Nutzung spezifisch ausgewählter Werke als Inputdaten bestehen. Daraus würde eine unmittelbare Beeinträchtigung der Verwertungsinteressen betroffener Urheber folgen. Demzufolge ergibt sich die Zweckmäßigkeit einer Freistellung der Speicherung urheberrechtlich geschützter Inputdaten nicht schon aus einer fehlenden Beeinträchtigung der Urheberinteressen.

2. Abwägung der Urheber- und Nutzerinteressen

Eine Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten bei der Anwendung von KI-Systemen könnte aber trotzdem zweckmäßig sein, wenn die Nutzerinteressen die betroffenen Verwertungsinteressen der Urheber im Rahmen einer umfassenden Abwägung überwiegen.

260 Dazu siehe unter 1. Teil B. II. 3.

Zugunsten der KI-Nutzer ist das Interesse an einer reibungslosen Anwendung des KI-Systems in die Abwägung einzustellen. Dieses Interesse kann, je nach spezifischem Anwendungsfall, auch durch die Berufsfreiheit (Art. 12 GG) oder andere Freiheitsrechte geschützt sein.

Für die Urheber spricht, dass innerhalb der Sphäre der KI-Anbieter oder -Nutzer eine Kontrolle über die Verwendung der als Inputdaten erfassten Werke nicht mehr möglich ist. Es besteht damit die Gefahr, dass ein Schutz vor Sekundärverwertungen nach der Speicherung der Werke als Inputdaten nicht mehr effektiv sichergestellt werden kann. Gerade diese Sekundärverwertungen können allerdings die Verwertungsinteressen der Urheber beeinträchtigen.

Jedoch dürfte die praktisch häufigste Beeinträchtigung der wirtschaftlichen Interessen der Urheber aus einer sich an die Verwendung der Werke als Inputdaten anschließenden Nutzung zum KI-Training folgen. Gerade für diesen Zweck bestehen aber regulatorische Freistellungsinteressen,²⁶¹ sodass ein vorgelagerter Schutz der Urheber in Bezug auf diesen Fall nicht als notwendig erscheint.

Darüber hinaus ist insbesondere bei der unkontrollierten Erfassung von Inputdaten aus Bereichen, in denen nicht ausschließlich urheberrechtlich geschützte Werke verarbeitet werden, weder vorherseh- noch kontrollierbar, ob, zu welchem Umfang und welche Werke konkret verarbeitet werden. Wäre die Speicherung von Werken als Inputdaten in diesen Fällen nicht erlaubnisfrei möglich, hätte das zur Folge, dass der Einsatz der KI-Systeme nur unter Inkaufnahme von Rechtsverletzungen oder insgesamt nicht möglich wäre. Eine solche prohibitive Wirkung kann vom Urheberrecht allerdings nicht ausgehen. Andersherum verhält sich das Urheberrecht, wie beispielsweise ErwGr. 31 InfoSoc-RL oder Art. 3 f. DSM-RL für neue Medien oder automatisierte Text und Data Mining-Prozesse zum Ausdruck bringen, offen gegenüber dem Einsatz neuer Technologien.²⁶²

Im Übrigen spricht zugunsten der Nutzer, dass die Verwertungsinteressen der Urheber in einigen Fallkonstellationen nur mittelbar betroffen sind. Selbst wenn eine unmittelbare Interessenbeeinträchtigung stattfindet,

261 Siehe oben unter 2. Teil B. I. 2.

262 Auch der EuGH betont regelmäßig die Funktion gewisser urheberrechtlicher Bestimmungen, sowohl die Entwicklung als auch den Einsatz neuer Technologien zu ermöglichen und zu gewährleisten, siehe u.a. EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 164 - Football Association Premier League u. Murphy; EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 24 - PRCA/NLA. In diesem Sinne schreibt auch Hofmann: „Die moderne Infrastruktur hängt am Urheberrecht“, siehe Hofmann, GPR 2017, 176 (176).

ist die Verwendung von Werken als Inputdaten darüber hinaus nur rein-technischer Natur. Die Kernverwertungsarten, bei denen die menschliche Rezeption der Werke in ihrer Eigenschaft als geistige Schöpfungen im Mittelpunkt steht, werden nicht berührt. Insofern die Kernverwertungsarten nachfolgend durch den Output des KI-Systems beeinträchtigt werden, ist dies auf Ebene des Outputs und für den jeweiligen Einzelfall zu untersuchen. Die Verwendung des generierten Outputs ist von der Verwendung der Werke als Inputdaten zu trennen. Alles in allem wiegen die Interessen der KI-Anbieter beziehungsweise -Nutzer bei der Verwendung von Werken als Inputdaten damit schwerer als die Beeinträchtigung der Verwertungsinteressen der Urheber.

3. Schlussfolgerung

Aus der umfassenden Abwägung von Urheber- und Nutzerinteressen ergibt sich, dass den Nutzerinteressen vor dem Verwertungsinteresse betroffener Urheber Vorrang zu gewähren ist. Damit ist eine urheberrechtliche Freistellung der Speicherung von Werken als Inputdaten insgesamt zweckmäßig.

II. Freistellungsmechanismus

Geklärt werden muss weiterhin, mit Hilfe welches Mechanismus die Freistellung der Speicherung urheberrechtlich geschützter Inputdaten umgesetzt sein sollte. Neben einer vergütungsfreien Schrankenbestimmung und einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte kann hierfür auch auf eine gesetzliche Lizenz oder einen ECL-Mechanismus zurückgegriffen werden.

1. Notwendigkeit eines angemessenen Ausgleichs

Fraglich ist zunächst, ob den betroffenen Urhebern ein angemessener Ausgleich für die Verwendung ihrer Werke als Inputdaten gezahlt werden sollte. Die Notwendigkeit eines angemessenen Ausgleichs für betroffene Urheber spricht dabei für die Freistellung mittels einer gesetzlichen Lizenz oder eines ECL-Mechanismus. Soll ein angemessener Ausgleich hingegen nicht gewährt werden, spricht dies weiterhin für die Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen

Verwertungsrechte oder einer vergütungsfreien Schrankenbestimmung. Grundsätzlich ist dabei jede Einschränkung des Verfügungsrechts der Urheber ausgleichspflichtig, sofern nicht ausnahmsweise gesteigerte öffentliche Interessen eine vergütungsfreie Nutzung der Werke und damit auch die Einschränkung des urheberrechtlichen Verwertungsrechts rechtfertigen.²⁶³

Ohne Inputdaten kann eine KI nicht in der Anwendungsphase eingesetzt werden. Sie werden zur Erzeugung eines Outputs benötigt. Inputdaten sind damit stets kausal für die Funktionsfähigkeit von KI. Das gilt demzufolge auch für die vorgelagerte Speicherung der als Inputdaten verwendeten Inhalte. Die Förderung von KI-Innovationen ist darüber hinaus ein öffentliches Anliegen.²⁶⁴ Innovationen zu fördern, welche anschließend nicht angewendet werden können, würde die Innovationsförderung allerdings ad absurdum führen. Die Anwendung von KI auf Basis von urheberrechtlich geschützten Inputdaten ist demzufolge in das öffentliche Innovationsinteresse mit einzubeziehen. Es kann grundsätzlich eine vergütungsfreie Nutzung rechtfertigen.

Klärungsbedürftig ist im Weiteren aber, ob das öffentliche Innovationsinteresse eine vergütungsfreie Speicherung der urheberrechtlich geschützten Inputdaten tatsächlich rechtfertigt.²⁶⁵ Das ist im Rahmen einer umfassenden Abwägung zu ermitteln.²⁶⁶ Dabei ist maßgeblich, ob „unter dem Strich“²⁶⁷ eine angemessene Vergütung der betroffenen Urheber für ihre Schöpfungsleistung verbleibt.²⁶⁸ In systematischer Hinsicht gegen eine Vergütung der Urheber spricht, dass das Urheberrecht auch an anderer Stelle (§§ 44a, 69d UrhG) dazu dient, die normale Benutzung eines Werkes sicherzustellen.²⁶⁹ Weil der Werkzugang in diesen Fällen rechtmäßig erfolgt,

263 BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (48) – Kirchenmusik; vgl. auch BVerfG, Urt. v. 11.10.1988 - 1 BvR 743/86 u. a., NJW 1992, 1307 (1308) – Justizvollzugsanstalten.

264 Dazu siehe oben unter 2. Teil B. II. 1.

265 Vgl. zum Kriterium BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (48) – Kirchenmusik; vgl. auch BVerfG, Urt. v. 11.10.1988 - 1 BvR 743/86 u. a., NJW 1992, 1307 (1308) – Justizvollzugsanstalten.

266 Vgl. *Lauber-Rönsberg*, Urheberrecht und Privatgebrauch, S. 112.

267 BVerfG, Urt. v. 11.10.1988 - 1 BvR 743/86 u. a., NJW 1992, 1307 (176) – Justizvollzugsanstalten; BVerfG, Urt. v. 31.05.2016 - 1 BvR 1585/13, GRUR 2016, 690 Rn. 74 – Metall auf Metall.

268 BVerfG, Beschl. v. 07.07.1971 - 1 BvR 765/66, GRUR 1972, 481 (483) – Kirchen- und Schulgebrauch; BVerfG, Beschl. v. 25.10.1978 - 1 BvR 352/71, GRUR 1980, 44 (46) – Kirchenmusik.

269 Vgl. zu § 44a: *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 8 f.

kann die rein akzessorische Nutzungssicherung vergütungsfrei erfolgen. Ähnlich könnten die Dinge bei einer Freistellung der Speicherung urheberrechtlich geschützter Inputdaten liegen: Denn auch mit ihr wird der Zweck verfolgt, die urheberrechtlich grundsätzlich nicht beschränkte Funktionalität der KI, also ihre normale Benutzung, sicherzustellen. Die Vervielfältigung der urheberrechtlich geschützten Werke ist nur technisch bedingt.²⁷⁰

In einigen Anwendungsfällen ist die Speicherung urheberrechtlich geschützter Inputdaten darüber hinaus nur zufällig und beiläufig. Da es in diesen Fällen mitunter nicht auf die schöpferischen Eigenschaften des Werkes ankommt, schmälert dies den wirtschaftlichen Vorteil, der aus der Verwendung des Werkes als solches gezogen wird. Er bildet die Basis für die Bestimmung einer angemessenen Vergütung.²⁷¹

Insgesamt sprechen damit wesentliche Gründe für eine vergütungsfreie Nutzung der betroffenen Werke. Das öffentliche Innovationsinteresse bei der Speicherung urheberrechtlich geschützter Inputdaten rechtfertigt damit ausnahmsweise eine vergütungsfreie Nutzung. Demzufolge ist für die Speicherung urheberrechtlich geschützter Inputdaten kein angemessener Ausgleich für Urheber vorzusehen.

2. Schutz eingesetzter technischer Schutzmaßnahmen

Mangels Notwendigkeit einer angemessenen Vergütung könnte die Freistellung der Speicherung urheberrechtlich geschützter Inputdaten im Wege einer nicht vergüteten Schrankenbestimmung oder durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte in den §§ 15 ff. UrhG umgesetzt werden.

Bei der Wahl zwischen beiden Regelungsmechanismen ist zum einen entscheidend, ob Urhebern, die ihre Werke trotz der Freistellungsregelung mit Hilfe technischer Maßnahmen gegen die Verwendung als Inputdaten schützen wollen, der Schutz des § 95a UrhG zukommen soll. Der Schutzbereich der Regelung geht über den originären Schutzbereich des urheberrechtlichen Ausschließlichkeitsrechts hinaus und sichert Urhebern umfas-

270 Das war auch schon bei § 44a UrhG maßgeblich, siehe BT-Drs. 15/38, S. 18.

271 Zur Vergütungshöhe bei vergütungspflichtigen Schrankenbestimmungen: *Pflüger*, Gerechter Ausgleich und angemessene Vergütung, S. 145. Für die Tarife der Verwertungsgesellschaften ergibt sich das aus § 39 Abs. 1 S. 1 VGG.

send die Verwertung ihrer Werke im digitalen Umfeld.²⁷² Dennoch muss auch beim Schutz von technischen Schutzmaßnahmen der Grundsatz der Verhältnismäßigkeit beachtet werden.²⁷³ Deswegen sind nur solche Maßnahmen schutzfähig, die nicht den normalen Betrieb elektronischer Geräte und deren technische Entwicklung beeinträchtigen.²⁷⁴ Andersherum dürfen keine Handlungen untersagt werden, deren wirtschaftlicher Zweck nicht in der Umgehung technischer Schutzvorkehrungen besteht.²⁷⁵ Von Bedeutung ist außerdem, dass ein Nutzer zunächst den Aufwand erbringen muss, eine Nutzung durch die Zurverfügungstellung geeigneter Umgehungsmittel vom Urheber ermöglichen zu lassen (§ 95b UrhG), wenn er technisch geschützte Werke insbesondere auf Basis urheberrechtlicher Schrankentatbestände rechtmäßig nutzen möchte. Der Einsatz technischer Schutzmaßnahmen verlagert demzufolge die Aktionslast auf die Werknutzer.²⁷⁶

Gegen einen Schutz technischer Schutzmaßnahmen, die die Nutzung von Werken als Inputdaten für KI verhindern, spricht, dass die Verwertungsinteressen der Urheber in einigen Fallkonstellationen nur mittelbar betroffen sind. Selbst wenn es zu einer unmittelbaren Beeinträchtigung kommt, wird diese stets nur durch eine rein-technische Nutzung der Werke hervorgerufen. Da hierbei nicht die Rezeption der Inhalte bezweckt wird, sind von der Speicherung der Werke als Inputdaten nicht die Kernverwertungsarten der Urheber betroffen. Vor dem Hintergrund des Grundsatzes der Verhältnismäßigkeit erscheint es ungerechtfertigt, ihnen dafür den weitreichenden Schutz des § 95a UrhG zuzubilligen.

Darüber hinaus kann eine KI nicht ohne Inputdaten angewendet werden.²⁷⁷ Sind KI-Systeme auf eine Verarbeitung von Werken auf Inputebene angewiesen, beeinträchtigen die technischen Schutzmaßnahmen demzufolge den normalen Betrieb dieser elektronischen KI-Systeme. Dass der Betrieb solcher KI-Systeme nicht grundsätzlich rechtlich missbilligt wird, zeigt jedoch bereits die Existenz der KI-VO.

Mit dem Einsatz von KI-Systemen, welche (unter anderem) Werke als Inputdaten verarbeiten, werden darüber hinaus eigene wirtschaftliche Zwecke verfolgt. Diese sind in der jeweiligen Zielfunktion und dem Anwendungsumfeld der KI-Systeme konkretisiert. Der wirtschaftliche Zweck

272 Zur Differenz zwischen urheberrechtlichem Schutzbereich und Schutzzumfang des § 95a UrhG siehe *Specht-Riemenschneider*, GRUR 2019, 253 (258).

273 ErwGr. 48 S. 3 InfoSoc-RL.

274 ErwGr. 48 S. 1 InfoSoc-RL.

275 ErwGr. 48 S. 3 InfoSoc-RL.

276 So ebenfalls *Specht-Riemenschneider*, GRUR 2019, 253 (259).

277 Zur technischen Funktionalität siehe oben unter 1. Teil B. II. 2.

der Verarbeitungshandlungen besteht demzufolge nicht in der Umgehung eingerichteter technischer Schutzvorrichtungen, sondern richtet sich nach dem jeweiligen Zweck des KI-Systems.

Außerdem verlagern technische Schutzmaßnahmen die Aktionslast auf den Werknutzer. Sie müssten gegebenenfalls den Anspruch aus § 95b UrhG durchsetzen. Da die Speicherung und Verarbeitung urheberrechtlich geschützter Inputdaten in diversen Anwendungskonstellationen wie auch das KI-Training automatisch ablaufen, besteht für eine Vielzahl von KI-Anwendern aber keine Möglichkeit, dieser Aktionslast nachzukommen. Allgemein werden KI-Systeme oftmals eingesetzt, um Effizienz- und Temposteigerungen bei der Bewältigung von Anwendungsproblemen zu erzielen. Aus beiden Gründen wäre in einer Vielzahl der Fallkonstellationen nicht mit einer Geltendmachung des Anspruchs aus § 95b UrhG zu rechnen. Diese Verlagerung der Aktionslast auf die KI-Anwender und dessen Legitimation durch § 95a UrhG sind daher nicht interessengerecht.

Aus alledem folgt, dass ein Rechtsschutz für technische Schutzmaßnahmen, deren Zweck es ist, die Speicherung von Werken als Inputdaten für KI zu verhindern, nicht gewährt werden sollte. Das spricht im Ergebnis für die Freistellung des Sachverhalts durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte.

III. Zwischenergebnis

Eine Freistellung der Speicherung urheberrechtlich geschützter Inputdaten ist aufgrund eines Überwiegens der Nutzerinteressen zweckmäßig. Da für diese Freistellung keine Ausgleichsleistung für Urheber vorzusehen ist und von Urhebern eingesetzte technische Schutzmaßnahmen nicht vom Schutz des § 95a UrhG profitieren sollen, ist die Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte umzusetzen. Hierfür kommt es *de lege lata* insbesondere auf die Reichweite des Vervielfältigungsrechts aus § 16 UrhG an.

E. Vorverarbeitung urheberrechtlich geschützter Inputdaten durch Normalisierung und Feature Encoding

Freistellungsbedürfnisse könnten sich weiterhin auch für die Vorverarbeitung urheberrechtlich geschützter Inputdaten durch Normalisierung und Feature Encoding ergeben.²⁷⁸ Dabei werden die Daten auf eine vergleichbare Skala transponiert (normalisiert) und anschließend in eine numerische Darstellungsform überführt („Feature Encoding“).

I. Normalisierung der Inputdaten

Analysiert werden muss dabei zum einen die Normalisierung der als Inputdaten erfassten Werke. Ob ein regulatorisches Freistellungsbedürfnis besteht, richtet sich auch hierbei nach der Zweckmäßigkeit der Freistellung.²⁷⁹ Für eine Umsetzung der Freistellung stehen im Folgenden eine vergütete oder vergütungslose Schrankenbestimmung, der ECL-Mechanismus oder die unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte zur Verfügung.²⁸⁰

1. Zweckmäßigkeit der Freistellung

Eine Freistellung der Normalisierung von bei der Anwendung eines KI-Systems als Inputdaten erfassten Werken ist zweckmäßig, wenn durch die Normalisierung keine Urheberinteressen beeinträchtigt werden oder die bei der Normalisierung verfolgten Nutzerinteressen die betroffenen Urheberinteressen überwiegen.

Zwar müssen die Inputdaten oft aus dem situationsspezifischen Anwendungsumfeld generiert werden. Damit besteht für Urheber in vielen Fällen kein Markt, um ihre Werke als Inputdaten für KI zu verwerten.²⁸¹ Nichtdes-

278 Zur Bedeutung beider technischen Begriffe siehe oben unter 1. Teil B. II. 2.

279 Zur Herleitung der regulatorischen Entscheidungskriterien insgesamt siehe unter 2. Teil A.

280 Zur detaillierten Herleitung und Beschreibung der Entscheidungskriterien siehe oben unter 2. Teil A. II. 2.

281 Dazu bereits oben unter 2. Teil D. I. 1.

totrotz können auch normalisierte Inputdaten sekundär verwertet werden. Dies umfasst insbesondere auch das Training eines KI-Systems auf Basis der durch die Anwendung der KI gewonnenen Datensätze. Hierfür macht es keinen Unterschied, ob die Daten normalisiert worden sind oder nicht. Ohnehin muss auch für das Training im Regelfall eine Normalisierung der Daten stattfinden. Normalisierte Trainingsdatenkorpora werden mitunter auch bereits öffentlich angeboten.²⁸² Auch die Herstellung normalisierter Kopien der Inputdaten greift demzufolge mittelbar in die Verwertungsinteressen der Urheber ein.

Darüber hinaus könnte die Normalisierung auch die Integritätsinteressen der als Inputdaten erfassten Urheber beeinträchtigen. Denn bei der Normalisierung der Werke werden beispielsweise Abkürzungen in Textinhalten aufgelöst, Begriffe auf ihren Wortstamm zurückgeführt und die Textinhalte tokenisiert. Bildinhalte können je nach Bedarf skaliert, zugeschnitten und insbesondere hinsichtlich der wiedergegebenen Farbschemata angepasst werden. Die Normalisierung der urheberrechtlich geschützten Trainingsdaten geht also mit einer Veränderung der Inhalte einher. Allerdings wird dem Urheber mit dem Schutz des Integritätsinteresses zum einen das Fortbestehen seines Werkes zugesichert.²⁸³ Dieses ist durch die Normalisierung der Werke aber nicht gefährdet. Darüber hinaus soll verhindert werden, dass der Urheber mit Inhalten identifiziert wird, die in der jeweiligen Form nicht auf ihn zurückgehen.²⁸⁴ Neben der Veränderung wird dafür allerdings vorausgesetzt, dass das veränderte Werkstück auch in der Öffentlichkeit wahrgenommen wird oder zumindest wahrgenommen werden kann.²⁸⁵ Zwischen der Normalisierung der urheberrechtlich geschützten Inputdaten und dem Ablauf des Algorithmus auf Basis der vorverarbeiteten Werke gibt es jedoch kein Publikum, welches die durch die Normalisierung umgearbeiteten Darstellungen der Werke rezipieren könnte. Infolgedessen besteht auch keine Gefahr, dass Urheber in diesem Rahmen veränderte Inhalte zugerechnet werden. Demzufolge berührt die die Normalisierung urheberrechtlich geschützter Inputdaten nicht die Integritätsinteressen der Urheber.

282 Beispielsweise der MINT-Datensatz, siehe <http://yann.lecun.com/exdb/mnist/> (zuletzt abgerufen am 18.10.2023).

283 Vgl. BGH, Urt. v. 21.02.2019 - I ZR 98/17, ZUM 2019, 508 Rn. 33 - HHole (for Mannheim).

284 BGH, Urt. v. 01.10.1998 - I ZR 104/96, ZUM 1999, 146 (148) - Treppenhausegestaltung.

285 Vgl. auch *Peukert*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 14 Rn. 1, 13.

Insgesamt beeinträchtigt die Normalisierung urheberrechtlich geschützter Inputdaten die Verwertungs-, nicht aber die Integritätsinteressen betroffener Urheber. Um die Zweckmäßigkeit der Freistellung der Datennormalisierung zu bestimmen, bedarf es dennoch einer Abwägung der betroffenen Urheber- und Nutzerinteressen.

Die Abwägung unterscheidet sich dabei inhaltlich nicht von der Abwägung, die für die originäre Speicherung der urheberrechtlich geschützten Inputdaten bei der Anwendung von KI-Systemen durchgeführt wurde.²⁸⁶ Denn aus den gleichen Gründen, wie auch die Erfassung und Speicherung von Werken als Inputdaten freigestellt sein sollte, muss auch die Normalisierung freigestellt sein, die einer Verwendung der Werke als Input für KI vorausgeht. Wäre das nicht der Fall, würde eine Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten vielmehr ins Leere laufen. Im Ergebnis führt ein Überwiegen der Nutzerinteressen daher dazu, dass auch die Normalisierung der als Inputdaten generierten Werke urheberrechtlich freigestellt werden sollte.²⁸⁷

2. Freistellungsmechanismus

Auch bei der Wahl eines geeigneten Freistellungsmechanismus für die Normalisierung urheberrechtlich geschützter Inputdaten kommt es darauf an, ob den Urhebern für die Verwendung ihrer Werke als KI-Input eine angemessene Vergütung gezahlt werden muss. Ist das nicht der Fall, kann im Weiteren danach differenziert werden, welcher der Beteiligten im Verletzungsprozess die Beweislast tragen sollte. Entscheidungserheblich ist außerdem, ob ein Schutz technischer Schutzmaßnahmen über § 95a UrhG interessengerecht ist.²⁸⁸

Auch die Normalisierung ist technisch bedingt und erzeugt für KI-Anwender unmittelbar keinen eigenen wirtschaftlichen Vorteil. Die Verwertungsinteressen der Urheber sind nicht in ihrem Kernbereich betroffen. Im Wesentlichen kann betreffend des Freistellungsmechanismus daher wieder auf die Erwägungen zur Speicherung der Inputdaten zurückgegriffen werden.²⁸⁹ Das öffentliche Innovationsinteresse rechtfertigt daher ausnahmsweise eine vergütungsfreie Nutzung auch bei der Normalisierung von als

286 Dazu siehe unmittelbar oben unter 2. Teil D. I.

287 Zu einem ähnlich gelagerten Problem vgl. auch *Raue*, GRUR 2017, 11 (15, 17).

288 Im Detail zur Herleitung der Entscheidungskriterien siehe oben unter 2. Teil A. II.

289 Dazu oben 2. Teil D. II.

Inputdaten erfassten Werken für die Anwendung von KI.²⁹⁰ Die Normalisierung der Inputdaten geht darüber hinaus zumeist mit einer Veränderung der Werke einher. Technische Schutzmaßnahmen gegen diese Veränderungen beeinträchtigen demzufolge direkt die Funktionsfähigkeit von KI. Wegen der Automatisierung der Datenverarbeitung bei der Anwendung von KI ist auch keine Verlagerung der Aktionslast auf die Werknutzer interessengerecht. Daher sind technische Schutzmaßnahmen gegen die Normalisierung von als Inputdaten erfassten Werken auch nicht in den Schutzbereich des § 95a UrhG einzubeziehen.²⁹¹ Insgesamt sollte die Normalisierung von Inputdaten daher durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte freigestellt sein.

3. Zwischenergebnis

Eine Freistellung der Normalisierung urheberrechtlich geschützter Inputdaten ist, wie auch im Rahmen der zuvor stattfindenden Speicherung der Inhalte, aufgrund überwiegender Nutzerinteressen zweckmäßig. Da diese Freistellung ohne Ausgleichsleistung für Urheber vorzusehen ist und von Urhebern eingesetzte technische Schutzmaßnahmen nicht vom Schutz des § 95a UrhG profitieren sollen, ist die Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte umzusetzen.

II. Feature Encoding der normalisierten Inhalte

An die Normalisierung der urheberrechtlich geschützten Inputdaten schließt sich das Feature Encoding der Inhalte an. Dabei werden sie in ein numerisches Format, im Regelfall einen Tensor, transformiert. Auch das Feature Encoding könnte aus einer urheberrechtsregulatorischen Perspektive freigestellt werden müssen, wenn sich die Freistellung nach den bereits hergeleiteten Kriterien als zweckmäßig erweist.²⁹²

290 Vgl. oben 2. Teil D. II. 1.

291 Vgl. oben 2. Teil D. II. 2.

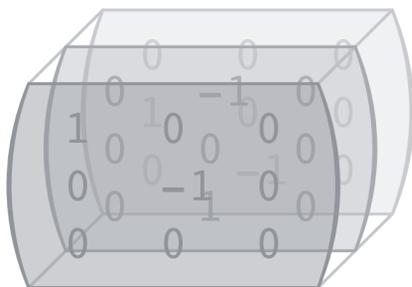
292 Zur Herleitung der regulatorischen Entscheidungskriterien insgesamt siehe unter 2. Teil A.

1. Zweckmäßigkeit einer Freistellung

Eine urheberrechtliche Freistellung des Feature Encodings ist zweckmäßig, wenn durch den Verarbeitungsschritt von keine Urheberinteressen berührt werden oder die Nutzer- die gegenläufigen Urheberinteressen im Rahmen einer Gesamtabwägung überwiegen.

Fraglich ist dabei allem voran, ob durch das Feature Encoding der erfassten Werke die Verwertungsinteressen der Urheber beeinträchtigt werden. Hierfür spricht zunächst, dass sowohl bei der Speicherung als auch bei der anschließenden Normalisierung der Inhalte eine Beeinträchtigung der Verwertungsinteressen der Urheber stattfindet. Auch das Feature Encoding ist Teil dieses Prozesses, sodass die bereits angestellten Erwägungen auch auf das Feature Encoding übertragen werden könnten.

Allerdings werden die Werke beim Feature Encoding in ein numerisches Repräsentationsformat überführt. Die gebildeten Tensoren können unterschiedliche Komplexität aufweisen und insbesondere auch mehrdimensional sein. Beispielsweise werden dreidimensionale Tensoren benutzt, um Bildinhalte numerisch darzustellen. Jede Ebene des Tensors repräsentiert dabei einen Kanal des RGB-Farbmodells, mit dem sich die Farbwerte der einzelnen Pixel beschreiben lassen. Jedem Pixel des Bildes ist übereinanderliegend auf jeder Ebene des Tensors ein Zahlenwert zugeordnet, der den Pixelwert des Farbkanals im RGB-Spektrum repräsentiert. Ein solcher Tensor könnte wie folgt aussehen:



Für einen Rezipienten besteht bei diesen rein-numerischen Darstellungen keine Möglichkeit mehr, den Inhalt oder die Formgebung eines Werkes, also die jeweiligen schutzbegründenden Elemente, zu identifizieren. Selbst wenn Tensor und Ursprungswerk nebeneinander abgebildet sein sollten, sind die durchschnittlichen Rezipienten nicht mehr in der Lage, die nu-

merischen Darstellungen den Ursprungswerken zuzuordnen. In Folge des Feature Encodings können die Werke demzufolge nicht mehr *als Werk* verwertet werden. Eine Rezeption der schöpferischen Merkmale scheidet vielmehr aus. Damit berührt das Feature Encoding nicht die Kernverwertungsmöglichkeiten der Urheber.

Im Gegensatz zur Speicherung und Normalisierung der Inputdaten könnte darüber hinaus auch keine Sekundärverwertung der Inhalte mehr realisierbar sein. Denn die Sekundärverwertung von Inputdaten wird maßgeblich davon geprägt, dass die als Inputdaten generierten Werke später als Trainingsdaten weitergenutzt werden können. Als potenzielle Trainingsdaten kommt ihnen auch ein wirtschaftlicher Wert zu. Bereits in Tensoren umgewandelte, urheberrechtlich geschützte Inputdaten können aufgrund mangelnder menschlicher Wahrnehmbarkeit jedoch nicht mehr manuell annotiert werden. Eine solche Annotation durch Menschen ist aktuell aber zumeist noch notwendig.²⁹³ Ohne menschliche Wahrnehmbarkeit der verkörperten Informationen ist die Vorverarbeitung der Daten insgesamt nur eingeschränkt möglich. Das betrifft insbesondere auch eine Klassifizierung und Vorsortierung der erfassten Inhalte. Aus alledem folgt, dass nach dem Feature Encoding der generierten Inputdaten keine Sekundärverwertung der umgewandelten Werke mehr möglich ist. Damit sind durch das Feature Encoding insgesamt keine Verwertungsinteressen der Urheber betroffen. Wie bei der Normalisierung der erfassten Werke fehlt es außerdem an einem Rezipientenkreis, der für eine Beeinträchtigung des Integritätsinteresses des Urhebers notwendig ist.²⁹⁴

Insgesamt werden durch das Feature Encoding der Werke keine Urheberinteressen beeinträchtigt, sodass eine Freistellung bereits danach zweckmäßig ist.

2. Freistellungsmechanismus

Mithilfe welches Regelungsmechanismus die Freistellung erfolgte sollte, richtet sich auch hier im Grundsatz nach den bereits herausgearbeiteten Kriterien.²⁹⁵

293 Vgl. *Paaß/Hecker*, Künstliche Intelligenz, S. 45, 143, 463, die von einer manuellen Annotation der Trainingsdaten ausgehen; *Rasmussen/Kirk/Moeslund*, *Sensors* 2022, 1596 S. 1 („traditional time-consuming manual annotation approach“).

294 Dazu siehe oben unter 2. Teil E. I. 1.

295 Dazu also insgesamt siehe oben unter 2. Teil A. II.

Fraglich ist also zunächst, ob Urhebern für das Feature Encoding der Werke eine angemessene Vergütung zu gewähren ist. Dann ist die Freistellung im Wege einer gesetzlichen Lizenz oder eines ECL-Mechanismus umzusetzen, wobei es bei der Wahl zwischen beiden Rechtsinstituten auf die Notwendigkeit eines Opt-out-Mechanismus sowie das Rechtssicherheitsbedürfnis der KI-Anwender ankommt. Ist hingegen keine Vergütung notwendig, kommt es insbesondere auf die Beweislastverteilung zwischen Urheber und KI-Anwender und das Schutzbedürfnis für technische Schutzmaßnahmen gegen das Feature Encoding der als Inputdaten gespeicherten Werke an. Mit Hilfe dieser Kriterien kann ermittelt werden, ob die Freistellung im Wege einer unvergüteten Schrankenbestimmung oder einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte erfolgen soll.

Wie auch schon die Speicherung und Normalisierung urheberrechtlich geschützter Inputdaten ist auch das Feature Encoding obligatorischer Teil des Funktionsablaufs eines KI-Systems. Um die Anwendung von KI insgesamt zu ermöglichen und dem hiermit einhergehenden Innovationsinteresse Rechnung zu tragen, muss folglich nicht nur die Speicherung der als Inputdaten generierten Werke und ihre Normalisierung, sondern auch die Vorverarbeitung ebendieser Daten gleichermaßen ermöglicht werden. Da der Markt für Urheber durch das Feature Encoding der Werke nicht beeinträchtigt wird, würde jede Ausgleichzahlung zudem eine Überkompensation bedeuten. Demzufolge sollte auch das Feature Encoding vergütungsfrei vorgenommen werden können.²⁹⁶

Klärungsbedürftig ist damit nur noch, ob eine Freistellung des Feature Encodings durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte oder eine vergütungsfreie Schrankenbestimmung umgesetzt werden sollte. Das Rechtsschutzregime aus § 95a UrhG, welches dabei in den Blick zu nehmen ist, dient dem Schutz der urheberrechtlichen Verwertungsinteressen.²⁹⁷ Im Umkehrschluss bedeutet das, dass ein Schutz technischer Schutzmaßnahmen vor Umgehung über § 95a UrhG nur dann gerechtfertigt ist, wenn die Verwertungsinteressen der von der Freistellung betroffenen Urheber berührt sind. Das ist beim Feature Encoding der Eingabedaten allerdings nicht der Fall. Ein Schutz von gegen das Feature Encoding der Werke eingesetzten technischen Maßnahmen über § 95a UrhG ist infolgedessen nicht interessengerecht. Das bedeutet, dass auch die Freistellung des Feature Encodings, also der Umwandlung urhe-

296 Im Detail dazu siehe oben unter 2. Teil D. II. 1.

297 Dazu ausführlich siehe oben unter 2. Teil A. II. 2. c).

berrechtlich geschützter Inputdaten in numerische Repräsentationen, aus einer regulatorischen Perspektive im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte umgesetzt werden sollte.²⁹⁸

III. Zwischenergebnis

Aus der regulatorischen Bedarfsanalyse folgt, dass sowohl die Normalisierung als auch das Feature Encoding urheberrechtlich geschützter Inputdaten, also deren Vorverarbeitung, im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte freigestellt sein sollten. De lege lata ist hierfür insbesondere die Reichweite des Vervielfältigungsrechts aus § 16 UrhG entscheidend.

F. Zusammenfassung der regulatorischen Bedarfsanalyse

Alles in allem muss bei der Identifizierung von Regelungsbedürfnissen für die Verwendung von Werken als Trainings- sowie Inputdaten für KI-Systeme untersucht werden, ob die genannten Fallgruppen urheberrechtlich freigestellt, also dem Ausschließlichkeitsrecht betroffener Urheber entzogen, oder ob Urhebern weiterhin ermöglicht werden soll, ihr Verbotungsrecht in den betreffenden Fällen auszuüben. Zentral für diese Beurteilung ist eine umfassende Abwägung von Urheber- und Nutzerinteressen. Die rechtökonomische Analyse bietet hingegen keinen geeigneten Ansatzpunkt für eine regulatorische Bedarfsanalyse. Trotzdem können ökonomische Argumente auch im Rahmen der freien Abwägung berücksichtigt werden.

Ist eine Freistellung zweckmäßig, stellt das Urheberrecht vier Mechanismen zur Verfügung, mit Hilfe derer die identifizierten Regelungsbedürfnisse für den Einsatz von KI umgesetzt werden können. Zum einen ist die Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der

298 Zu einem ähnlich gelagerten Problem vgl. auch *Raue*, GRUR 2017, 11 (17). Er befürwortet im Kontext des Text und Data-Mining die Freistellung der Transformation der gesammelten, urheberrechtlich geschützten Inhalte in ein analysefähiges Format.

urheberrechtlichen Verwertungsrechte oder einer vergütungslosen Schrankenbestimmung möglich. Darüber hinaus könnte sie aber auch im Wege einer gesetzlichen Lizenz oder mit Hilfe eines ECL-Mechanismus umgesetzt werden. Neben Fragen der Beweislastverteilung und der Schutzbedürftigkeit eingesetzter technischer Schutzmaßnahmen über § 95a UrhG kommt es bei der Wahl des geeigneten Freistellungsmechanismus insbesondere darauf an, ob den Urhebern, die in der jeweiligen Fallgruppe vom KI-Einsatz betroffen sind, eine angemessene Vergütung für die Verwendung ihrer Werke gewährt werden sollte. Berücksichtigungsfähig ist darüber hinaus, ob ein Opt-out-Mechanismus zugunsten der Urheber interessengerecht ist, mit Hilfe dessen die Freistungswirkung einer Regelung beendet werden kann. Ein besonderes Bedürfnis nach Rechtssicherheit fällt ebenfalls ins Gewicht. Im Einzelfall können aber auch andere Aspekte berücksichtigt werden.

Mit Hilfe des skizzierten, regulatorischen Untersuchungsmechanismus kann ermittelt werden, dass das Web Scraping von Werken zur Verwendung als Trainingsdaten durch eine vergütete Schrankenbestimmung ermöglicht sein sollte. Hierbei sind kommerzielle und nicht-kommerzielle Trainingsdatensammler gleichermaßen zu privilegieren. Ein Opt-out-Mechanismus für Urheber ist hingegen nicht vorzusehen. Auch die auf die Datenaggregation folgende, eigentliche Durchführung des KI-Trainings auf Basis der gesammelten Werke sollte aus einer regulatorischen Perspektive urheberrechtlich freigestellt sein. Zweckmäßig ist hierbei eine Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte, die den Trainingsvorgang selbst aus dem urheberrechtlichen Ausschließlichkeitsrecht ausklammert.

Die Analyse der urheberrechtlichen Regelungsbedürfnisse ergibt darüber hinaus, dass sowohl die Speicherung urheberrechtlich geschützter Inputdaten als auch die Vorverarbeitung im Sinne der Normalisierung und des Feature Encodings der erfassten Inhalte durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte urheberrechtlich freigestellt sein sollten. Infolgedessen kommt weder eine Vergütung der Urheber für die betreffenden Werkverwendungen noch ein Schutz von Verhinderungsmaßnahmen über § 95a UrhG in Betracht. De lege lata wird es für eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte insbesondere auf den Schutzbereich des Vervielfältigungsrechts aus § 16 UrhG ankommen.

Dritter Teil: Anwendbarkeit geltender urheberrechtlicher Bestimmungen und Abgleich mit den identifizierten Regelungsbedürfnissen

Im Anschluss an die Identifizierung von Regelungsbedürfnissen, die im zweiten Teil der vorliegenden Arbeit erfolgt, ist klärungsbedürftig, welche Rahmenbedingungen das geltende Urheberrecht für die Verwendung von Werken als Trainings- und Inputdaten zur Verfügung stellt.

A. Speicherung von urheberrechtlich geschützten Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern

Fraglich ist zunächst, ob die hinsichtlich des Web Scrapings von urheberrechtlich geschützten Trainingsdaten identifizierten Regelungsbedürfnisse vom geltenden Urheberrecht erfüllt werden. Damit das der Fall ist, muss das Web Scraping von Werken als Trainingsdaten von einer urheberrechtlichen Schrankenbestimmung gedeckt sein. Weiterhin darf diese Schrankenbestimmung Urhebern keinen Opt-out-Mechanismus zur Verfügung stellen. Notwendig ist außerdem, dass sowohl nicht-kommerziell als auch kommerziell tätige Trainingsdatensammler in ihren persönlichen Anwendungsbereich fallen. Den betroffenen Urhebern muss durch die Schrankenbestimmung außerdem eine angemessene Vergütung zugesichert werden.²⁹⁹

Damit das Web Scraping urheberrechtlich geschützter Trainingsdaten von einer Schrankenbestimmung freigestellt werden kann, muss überhaupt ein Eingriff in die urheberrechtlichen Verwertungsrechte stattfinden. Fraglich ist zunächst aber auch, welche Voraussetzungen erfüllt sein müssen, damit sich die Rechtmäßigkeit des Web Scrapings von Werken für das Training von KI überhaupt nach deutschem beziehungsweise europäischem Urheberrecht richtet. Dies kann bei grenzüberschreitenden Sachverhaltskonstellationen mit Auslandsbezug Schwierigkeiten hervorrufen. Ein sol-

299 Dazu ausführlich siehe oben unter 2. Teil B. II. 1.

cher liegt beispielsweise dann vor, wenn der Trainingsdatensammler seinen Sitz nicht in Deutschland hat.

I. Anwendbares Recht

Bei der Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern kommt ein Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG in Betracht. Das deutsche Urheberrecht ist aus einer international-privatrechtlichen Perspektive gemäß Art. 8 Abs. 1 VO (EG) 864/2007 (Rom II-VO)³⁰⁰ anwendbar, wenn Schutz vor einer Rechtsverletzung in Deutschland beansprucht wird. Darüber hinaus bedarf es materiell-rechtlich aber auch einer Nutzungshandlung innerhalb des Geltungsbereichs des UrhG, die sich an den im deutschen Urheberrecht verankerten Ausschließlichkeitsrechten messen lassen muss.

1. Hinreichender Inlandsbezug der Speicherung zur Anwendung deutschen Urheberrechts

Für einen Eingriff in das Vervielfältigungsrecht aus § 16 UrhG wird materiell-rechtlich ein hinreichender Inlandsbezug der betreffenden Handlung vorausgesetzt.³⁰¹ Hierfür kann zwar sprechen, dass die natürliche Handlung im deutschen Hoheitsgebiet vorgenommen wird. Erforderlich ist das aber nicht.³⁰² Ausreichend für einen hinreichenden Inlandsbezug ist vielmehr, wenn sich die Speicherung der Werke als Trainingsdaten auf das deutsche Inland auswirkt und damit die Interessen der Urheber in Deutschland beeinträchtigt.³⁰³

300 Verordnung (EG) Nr. 864/2007 des Europäischen Parlaments und des Rates vom 11. Juli 2007 über das auf außervertragliche Schuldverhältnisse anzuwendende Recht („Rom II“).

301 BGH, Urt. v. 16.06.1994 - I ZR 24/92, GRUR 1994, 798 (799) – Folgerecht bei Auslandsbezug; vgl. zum Datenbankherstellerrecht aber unter Zugrundelegung dieser Prinzipien auch EuGH, Urt. v. 18.10.2012 - C-173/11, ZUM 2013, 119 Rn. 34 ff. – Football Dataco/Premier League; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, Vor UrhG §§ 120 ff. Rn. 99.

302 *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, Vor UrhG §§ 120 ff. Rn. 100.

303 Zum Verbreitungsrecht: BGH, Urt. v. 15.02.2007 - I ZR 114/04, GRUR 2007, 871 Rn. 31 – Wagenfeld-Leuchte.

Ein hinreichender Inlandsbezug bei möglichen Vervielfältigungshandlungen liegt zum einen dann vor, wenn der Erfolg der Handlung, also die Herstellung des Vervielfältigungsstücks, in Deutschland eintritt.³⁰⁴ Bei digitalen Speicherungen von Werken ist das grundsätzlich der Ort, an dem der Computer beziehungsweise Server steht, auf dem die Werke abgespeichert werden.³⁰⁵ Der Anwendungsbereich des deutschen Urheberrechts bei der Speicherung von Werken zum KI-Training mit Hilfe von Webcrawlern ist also zum einen dann eröffnet, wenn der für die Speicherung genutzte Server in Deutschland lokalisiert ist. Die Handlung, welche die Speicherung auslöst, also die Steuerung des Web Scraping-Algorithmus, muss dabei hingegen nicht im Inland ausgeführt worden sein.³⁰⁶

Andersherum wird die Verwertung von Werken durch Urheber als Trainingsdaten im Inland auch dann beeinträchtigt, wenn im Inland ansässige Trainingsdatensammler für ihre Web Scraping-Aktivitäten Server im Ausland nutzen. Denn auch in diesen Fällen wird die Möglichkeit der Urheber eingeschränkt, in Deutschland Vorteile aus der wirtschaftlichen Verwertung ihrer Werke als Trainingsdaten durch die im Inland ansässigen Trainingsdatensammler zu generieren. Auch wenn im Inland ansässige Trainingsdatensammler für die technische Verarbeitung des Web Scrapings Server im Ausland nutzen, sind sie in diesem Fall Marktteilnehmer desselben inländischen Marktes wie die hiervon im Inland betroffenen Urheber. Deswegen liegt ein hinreichender Inlandsbezug bei der Speicherung von Werken zum KI-Training mit Hilfe von Webcrawlern auch dann vor, wenn die Speicherung zwar auf im Ausland befindlichen Servern erfolgt, aber die Trainingsdatensammler, welche die Speicherung veranlassen, im Geltungsbereich des deutschen UrhG ihren Sitz haben.³⁰⁷

Kein hinreichender Inlandsbezug liegt demgegenüber vor, wenn weder der für die Speicherung genutzte Server noch der Sitz des Trainingsda-

304 Raue, in: Dreier/Schulze, Urheberrechtsgesetz, Vor UrhG §§ 120 ff. Rn. 104.

305 BGH, Urt. v. 29.04.2010 - I ZR 69/08, GRUR 2010, 628 Rn. 17 – Vorschaubilder I; Katzenberger/Metzger, in: Schricker/Loewenheim, Urheberrecht, UrhG Vor §§ 120 ff. Rn. 144; Hoeren, in: Hoeren/Sieber/Holznapel, Handbuch Multimedia-Recht, Teil 7.8 (Kollisionsrechtliche Anknüpfungen in internationalen Datenbanken) Rn. 16; konkret für das Sammeln von Trainingsdaten Baumann, NJW 2023, 3673 (3675 f.); v. Welser, GRUR-Prax 2023, 516 Rn. 39.

306 Raue, in: Dreier/Schulze, Urheberrechtsgesetz, Vor UrhG §§ 120 ff. Rn. 104.

307 Schack, NJW 2024, 113 Rn. 13; Heine, GRUR-Prax 2024, 87 Rn. 8; allgemein für die Veranlassung einer Speicherung als hinreichender Inlandsbezug Katzenberger/Metzger, in: Schricker/Loewenheim, Urheberrecht, UrhG Vor §§ 120 ff. Rn. 144; Raue, in: Dreier/Schulze, Urheberrechtsgesetz, Vor UrhG §§ 120 ff. Rn. 110.

tensammlers im Inland verortet werden kann. Hierauf hat auch keinen Einfluss, ob das trainierte KI-System im Inland in Verkehr gebracht wird.³⁰⁸

2. Hilfsweise Geltung durch Extraterritorialität des Unionsurheberrechts nach der KI-VO

Auch ohne hinreichenden Inlandsbezug und damit ohne urheberrechtlich relevante Nutzungshandlung in Deutschland ist zumindest der Rechtsrahmen des europäischen Urheberrechts, dessen Umsetzung das deutsche UrhG weitreichend prägt, für Anbieter von KI-Modellen mit allgemeinem Verwendungszweck (vgl. Art. 3 Nr. 3, 53 KI-VO) auch dann zu beachten, wenn sie ihr KI-Modell auf dem europäischen Markt in Verkehr bringen wollen. Dies ergibt sich aus Art. 53 lit. c KI-VO.³⁰⁹ Die Nichtbeachtung der produktsicherheitsrechtlichen Vorschrift löst zwar keine urheberrechtlichen Rechtsbehelfe nach den §§ 97 ff. UrhG aus. Weiterhin wird durch die Regelung keine Fiktion einer Nutzungshandlung im Inland begründet,³¹⁰ womit insbesondere gesetzliche Vergütungsansprüche des nationalen Urheberrechts nicht begründet werden. Dennoch verleiht die Regelung den unionsurheberrechtlichen Rahmenbedingungen für das Speichern von Werken als KI-Trainingsdaten eine indirekte Normgeltung auch unabhängig vom Sitzland des Trainingsdatensammlers und Standort des für die Speicherung genutzten Servers. Die indirekte Wirkung erstreckt sich gemäß Art. 53 lit. c KI-VO insbesondere, aber nicht nur auf die Beachtung wirksamer Nutzungsvorbehalte im Sinne des Art. 4 Abs. 3 DSM-RL, welcher in § 44b Abs. 3 UrhG national umgesetzt worden ist. Darüber hinaus müssen beispielsweise auch andere Voraussetzungen unionsrechtlicher Schrankentatbestände inklusive des in Art. 5 Abs. 5 InfoSoc-RL verankerten Dreistufentests beachtet werden. Das gilt auch für die Übrigen Voraussetzungen des Art. 4 DSM-RL.

308 Zu möglichen Vervielfältigungen in den Parametern des trainierten KI-Systems siehe im Detail unter 3. Teil B. IV.

309 Hierzu im Detail siehe *Stieper/Denga*, GRUR 2024, 1473 (1478 ff.).

310 *Stieper/Denga*, GRUR 2024, 1473 (1479).

II. Eingriff in das urheberrechtliche Ausschließlichkeitsrecht

Werden Werke mit Hilfe von Webcrawlern erfasst und gespeichert, werden sie in ihrer im Internet verfügbaren, wahrnehmbarer Form auf einem Datenträger fixiert und infolgedessen vervielfältigt.³¹¹ Diese Fixierung beeinträchtigt auch die Verwertungsinteressen der Urheber.³¹² Das dem KI-Training vorgelagerte Web Scraping von Werken greift demzufolge in das ausschließliche Vervielfältigungsrecht der Urheber aus § 16 UrhG ein.³¹³ Die Freistellung der infolgedessen grundsätzlich erlaubnispflichtigen Handlung könnte im Weiteren damit durch eine urheberrechtliche Schrankenbestimmung erfolgen.

III. Eingreifen einer urheberrechtlichen Schrankenbestimmung

Fraglich ist allerdings, ob das Web Scraping von Werken zum KI-Training in den Anwendungsbereich einer der Schrankenbestimmungen fällt, die in den §§ 44a ff. UrhG normiert sind.

1. Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)

Denkbar ist zum einen, dass das Web Scraping zum Training von KI von der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG) freigestellt ist. Hierfür ist notwendig, dass die Vervielfältigungen der Werke für das KI-Training als integraler und wesentlicher Teil

311 Allgemeine Meinung, zum Download als Vervielfältigungshandlung *Loewenheim*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 16 Vervielfältigungsrecht Rn. 21 m. w. N.

312 Siehe dazu unter 2. Teil B. I. 1.

313 Die präzise Abgrenzung zwischen freier Inhaltsanalyse und erlaubnispflichtiger Vervielfältigungshandlung betonen auch *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 20; etwas missverständlich, dem Grunde nach aber ebenfalls *Kleinkopf*, Text- und Data-Mining, S. 178 ff., insbesondere 181; *Kögel*, InTeR 2023, 179 (180); *Raue*, IIC 2018, 379 (381); im Übrigen *Hargreaves u. a.*, Standardisation in the area of innovation and technological development, notably in the field of text and data mining, S. 39, 52 ff.; *Triaille/Meeüs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 31; *Wiss. Dienst des Deutschen Bundestages*, KI und ML, eine urheberrechtliche Betrachtung (BT WD 10 – 3000 – 67/18), S. 7.

eines technischen Verfahrens vorübergehender und flüchtiger oder begleitender Natur sind. Darüber hinaus darf ihnen keine eigene wirtschaftliche Bedeutung zukommen.

a) Vorübergehender Charakter der Vervielfältigung

Vorübergehend ist eine Vervielfältigung dann, wenn sie nach Abschluss des technischen Verfahrens, für das sie einen integralen und wesentlichen Bestandteil darstellt, automatisch wieder gelöscht wird.³¹⁴ Dabei darf das Vervielfältigungsstück insgesamt nur für eine kurze Zeit bestehen.³¹⁵ Zwar könnte das KI-Training ohne Trainingsdaten nicht durchgeführt werden, sodass die von den Trainingsdatensammlern hergestellten Vervielfältigungsstücke integraler und wesentlicher Bestandteil eines technischen Verfahrens sind.³¹⁶ Allerdings müssen die für das Training einer KI vorgesehenen Daten vor dem Trainingsvorgang vorverarbeitet, insbesondere normalisiert und gegebenenfalls annotiert werden.³¹⁷ Die Vervielfältigungen können nach ihrer Herstellung daher im Regelfall nicht direkt zum Training von KI verwendet werden. Vielmehr nimmt die Vorbereitung der Trainingsdatensätze (die Annotation der Daten erfolgt beispielsweise in der Regel manuell) eine gewisse Zeit in Anspruch. Die Trainingsdaten werden demzufolge nicht nur für eine kurze Zeit vervielfältigt.³¹⁸ Zudem erfolgt mitunter eine dauerhafte Speicherung der genutzten Trainingsdaten auch über den Abschluss eines konkreten KI-Trainingsprojekts hinaus.³¹⁹ Eine automatisierte Löschung der Trainingsdaten ist dem KI-Training demzufolge jedenfalls grundsätzlich nicht systemimmanent.³²⁰ Im Herrschaftsbereich intermediärer Trainingsdatensammler schließt sich an die Datensammlung darüber hinaus nicht zeitlich unmittelbar das Training eines KI-Systems an. Die Vervielfältigungen sind demzufolge nicht vorübergehend im Sinne des § 44a UrhG.³²¹

314 EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 60 ff. – Infopaq; EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 26 – PRCA/NLA.

315 Vgl. EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 66 – Infopaq; außerdem Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44a Rn. 4; Wiebe, in: Spindler/Schuster, Recht der elektronischen Medien, UrhG § 44a Rn. 3.

316 Vgl. EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

317 Dazu siehe oben in Teil A. So für das nicht KI-spezifische Text und Data-Mining auch Raue, IIC 2018, 379 (381); Raue, ZUM 2021, 793 (795).

318 Auch Käde, CR 2024, 599 Rn. 17.

319 Dazu bereits siehe oben unter Fn. 203.

320 Darauf stellt auch Käde, CR 2024, 599 Rn. 17 ab. Außerdem siehe oben Fn. 203.

b) Eigene wirtschaftliche Bedeutung der Vervielfältigung

Das Eingreifen der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a UrhG könnte darüber hinaus auch deshalb ausscheiden, weil den Vervielfältigungen der Werke beim Speichern für das Training von KI eine eigene wirtschaftliche Bedeutung zukommt. Die angefertigten Vervielfältigungen dienen unmittelbar dem Training von KI. In Werken enthaltene Informationen als Berechnungsgrundlage auch für Algorithmen zu verwenden, ist kein Teil des urheberrechtlichen Ausschließlichkeitsrechts. Damit qualifiziert sich der eigentliche Trainings-/Analysevorgang als urheberrechtlich freie, rechtmäßige Nutzung im Sinne des § 44a Nr. 2 UrhG.³²² Die Vervielfältigungen, die beim Web Scraping urheberrechtlich geschützter Trainingsdaten vorgenommen werden, dienen dieser rechtmäßigen Nutzung. Aus ihr ergibt sich der wirtschaftliche Trainingsvorteil, also der Wert des Trainings des KI-Systems.

Auch wenn die Berechnungen durch das KI-System effektiver und präziser sind als menschliche Berechnungen und Musteranalysen auf Grundlage von Werken, wird der Vorteil aus dem effizienten mathematischen Ablauf des Algorithmus, also der Trainingsvorteil, dennoch unmittelbar aus der rechtmäßigen Nutzung, also der Verwendung der Werke als Berechnungsgrundlage eines Algorithmus, gewonnen. Die Produktivitätssteigerung im Vergleich zur menschlichen Berechnung und Musteranalyse begründet für sich genommen demzufolge noch keine eigenständige wirtschaftliche Bedeutung der Vervielfältigungshandlungen.³²³ Die wirtschaftliche Bedeutung der rechtmäßigen Nutzung wird zwar potenziert, aber nicht von der rechtmäßigen Nutzung emanzipiert.

Fraglich ist allerdings, ob sich eine eigene wirtschaftliche Bedeutung der Vervielfältigungen daraus ergibt, dass der späteren, mit dem Training ermöglichten Anwendung der KI eine eigenständige wirtschaftliche Bedeutung zukommt.³²⁴ Hiergegen spricht allerdings, dass der Vorteil aus der Anwendung der KI unmittelbar aus dem Vorteil folgt, der mit dem Trai-

321 So für generative KI-Systeme auch *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 88 ff.; *Kleinkopf*, Text- und Data-Mining, S. 184; außerdem *Spindler*, ZGE 2018, 273 (278) für allgemeines Text und Data-Mining; in diese Richtung auch *Triaille/Meeûs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 46.

322 Dazu ausführlich auch siehe unten 3. Teil B. II.

323 So zu Produktivitätssteigerung bei rechtmäßigen Nutzungen EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 50 – Infopaq II.

324 So *Leistner*, GRUR 2024, 1665 (1666); *Schack*, NJW 2024, 113 Rn. 6.

ning des Systems einhergeht. Eine KI wird gerade um ihrer Anwendung Willen trainiert. Die Vervielfältigungen zum Training der KI können nicht hinweggedacht werden, ohne dass auch die spätere Anwendung des KI-Systems entfele. Der wirtschaftliche Anwendungsvorteil ist demzufolge lediglich eine Fortentwicklung des zunächst realisierten Trainingsvorteils. Eine Eigenständigkeit *neben* dem Vorteil aus der rechtmäßigen Nutzung, also dem Training der KI, besteht demzufolge nicht.³²⁵

Vervielfältigungen von Werken als Trainingsdaten für KI ermöglichen allerdings nicht nur das Training eines konkreten KI-Systems, sondern auch die Veräußerung der Inhalte auf dem Trainingsdatenmarkt.³²⁶ Mit der Speicherung der Werke wird damit eine über das konkrete KI-Training hinausgehende, eigenständige Nutzungsmöglichkeit eröffnet. Der wirtschaftliche Vorteil, der durch die Veräußerung der Werke auf dem Trainingsdatenmarkt geschaffen wird, ist dabei von dem wirtschaftlichen Vorteil aus der rechtmäßigen Nutzung, also der Analyse der Werke im Rahmen des KI-Trainings, zu trennen.³²⁷ Das wird schon dadurch deutlich, dass eine Veräußerung von Werken als Trainingsdaten auch möglich ist, ohne selbst den wirtschaftlichen Vorteil aus der Trainingsanalyse eines spezifischen KI-Systems und dessen späterer Anwendung zu ziehen. Am Trainingsprozess Beteiligte können für sich wahlweise beide (Vorteil aus der Trainingsanalyse zum einen sowie der Veräußerung auf dem Trainingsdatenmarkt zum anderen) oder auch nur einen dieser wirtschaftlichen Vorteile realisieren. Die Speicherung der Werke zum Training von KI verschafft damit zumindest potenziell einen wirtschaftlichen Vorteil, der über den Vorteil hinausgeht, der mit einem konkreten KI-System erzielt werden kann.³²⁸ Damit kommt der Erfassung und Speicherung von Werken als Trainingsdaten neben der

325 A. A. *Leistner*, GRUR 2024, 1665 (1666); *Schack*, NJW 2024, 113 Rn. 6. Ebenfalls keine eigenständige wirtschaftliche Bedeutung allein durch die spätere Anwendung der Analyseergebnisse sehen wohl *Spindler*, GRUR 2016, 1112 (1115); *Raue*, ZUM 2021, 793 (795).

326 Vgl. auch *Hofmann*, WRP 2024, 11 Rn.14. In diese Richtung ebenfalls *Dregelies*, GRUR 2024, 1484 (1485); *Leistner*, GRUR 2024, 1665 (1666).

327 A. A. *Hofmann*, ZUM 2024, 166 (169), der beide wirtschaftlichen Vorteile einheitlich behandelt und infolgedessen eine eigene wirtschaftliche Bedeutung der Vervielfältigungshandlungen ablehnt. Zu einem anderen Sachverhalt, aber einem vergleichbaren Problem wie hier auch der EuGH, siehe EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 52 – Infopaq II.

328 Wie hier auch *Heinze/Wendorf*, in: Ebers/Heinze/Krüger/Steinrötter, Künstliche Intelligenz und Robotik, § 9 KI und Urheberrecht Rn.17; *Dregelies*, GRUR 2024, 1484 (1485); ähnlich außerdem *Spindler*, GRUR 2016, 1112 (1115).

Durchführung des Trainingsvorgangs eine darüberhinausgehende, eigene wirtschaftliche Bedeutung zu.³²⁹

c) Schlussfolgerung zur Anwendbarkeit der Schrankenbestimmung

Aus all diesen Gründen fällt das Web Scraping urheberrechtlich geschützter Trainingsdaten nicht in den Anwendungsbereich der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG.³³⁰

2. Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)

Das Web Scraping von urheberrechtlich geschützten Trainingsdaten könnte allerdings von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfasst und damit urheberrechtlich freigestellt sein. Die Regelung stellt Vervielfältigungen von einzelnen oder mehreren rechtmäßig zugänglichen Werken für eine automatisierte Analyse frei, sofern damit das Ziel der Informationsgewinnung insbesondere über Muster, Trends und Korrelationen, sogenanntes Text und Data Mining, verfolgt wird (§ 44b Abs. 1, Abs. 2 S. 1 UrhG). Während § 44b UrhG den Kreis der Schrankenbe-

329 Vgl. zur Voraussetzung EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 175, 177 – Football Association Premier League; EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 50 – Infopaq II; im Kontext nicht KI-bezogenen Text und Data-Minings halten das auch *Spindler*, GRUR 2016, 1112 (1116); *Spindler*, ZGE 2018, 273 (278); *Triaille/Meeûs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 47 für denkbar; wie hier zu KI-Trainingsdaten auch *Heinze/Wendorf*, in: Ebers/Heinze/Krüger/Steinrötter, Künstliche Intelligenz und Robotik, § 9 KI und Urheberrecht Rn. 17; *Siglmüller/Gassner*, RD 2023, 124 Rn. 10; im Ergebnis auch *Leistner*, GRUR 2024, 1665 (1666); für das Training generativer KI-Systeme *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 88 ff.; in diese Richtung auch *Chiou*, JIPITEC 2019, 398 Rn. 16; noch offengelassen in *Hofmann*, WRP 2024, 11 Rn. 14; eine eigene wirtschaftliche Bedeutung der Vervielfältigung dann aber ablehnend in *Hofmann*, ZUM 2024, 166 (169); tendenziell ablehnend auch *Maamar*, ZUM 2023, 481 (485); *Vesala*, IIC 2023, 351 (364 f.); *Kögel*, DSRITB 2023, 285 (289).

330 So auch *Kleinkopf*, Text- und Data-Mining, S. 184, 186; *Dregelies*, GRUR 2024, 1484 (1486); im Ergebnis außerdem auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 31 ff.; *Wiss. Dienst des Deutschen Bundestages*, KI und ML, eine urheberrechtliche Betrachtung (BT WD 10 – 3000 – 67/18), S. 12; so in Bezug auf nicht KI-Trainingsdatenspezifischen Text und Data-Minings auch *Raue*, IIC 2018, 379 (381); *Spindler*, ZGE 2018, 273 (278).

günstigen für das Text und Data Mining dabei nicht einschränkt, stellt die auf § 44b aufbauende Schrankenbestimmung aus § 60d UrhG zwar weitere Nutzungshandlungen frei, privilegiert aber nur nicht-kommerziell tätige oder vergleichbare Werknutzer, die wissenschaftliche Forschung durchführen (vgl. § 60d Abs. 2, 3 Nr. 2 UrhG), sowie Kulturerbe-Einrichtungen, wozu öffentliche Bibliotheken, Museen und Archive gehören (vgl. § 60d Abs. 3 Nr. 1 UrhG). Um die Freistellungsbedürfnisse für das Web Scraping urheberrechtlich geschützter Trainingsdaten zu erfüllen, bedarf es allerdings einer Schrankenbestimmung, die gleichermaßen sowohl nicht-kommerziell tätige als auch gewinnwirtschaftlich orientierte Trainingsdatensammler privilegiert. Auch die Beschränkung auf wissenschaftliche oder auf die Erhaltung des Kulturerbes bezogene Zwecke deckt sich nicht mit dem im zweiten Teil der Untersuchung herausgearbeiteten Regelungsbedarf.³³¹ Untersucht werden muss damit vorrangig, ob die allgemeine Text und Data Mining-Schrankenbestimmung aus § 44b UrhG geeignet ist, die identifizierten Freistellungsbedürfnisse zu erfüllen.

Damit das Web Scraping von Werken zur Verwendung als Trainingsdaten von der Schrankenbestimmung des § 44b UrhG umfasst wird, muss das KI-Training unter den Text und Data Mining-Begriff des § 44b Abs. 1 UrhG subsumiert werden können. Bei der Auslegung des Tatbestandsmerkmals ist dabei auch die DSM-RL zu berücksichtigen, deren Umsetzung § 44b UrhG dient.³³² Die Schrankenbestimmung ist dort in Art. 4 DSM-RL geregelt. Die in § 44b Abs. 1 UrhG umgesetzte Legaldefinition des Text und Data Mining findet sich in Art. 2 Nr. 2 DSM-RL. Gemäß dieser Bestimmung ist unter Text und Data Mining die Technik der automatisierten Analyse von Texten und Daten in digitaler Form zu verstehen, mit deren Hilfe Informationen unter anderem – aber nicht ausschließlich – über Muster, Trends und Korrelationen gewonnen werden können.

Unzweifelhaft ist zunächst, dass beim Training mit Werken eine automatisierte Analyse, also eine nicht händische Nutzbarmachung digitaler Werke stattfindet. Weder § 44b Abs. 1 UrhG noch Art. 2 Nr. 2 DSM-RL machen Angaben dazu, welcher Teil der Werke analysiert werden darf. Gegenstand der Analyse können demzufolge auch schutzbegründende Elemente sein.³³³

331 Dazu siehe insgesamt 2. Teil B. I. 2.

332 BT-Drs. 19/27426, S. 87.

333 Anders für das Training generativer KI-Systeme mit Werken wohl *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 108, wobei nicht klar ist, an welchem Punkt die einschränkende Auslegung angeknüpft wird.

a) Ziel der Informationsgewinnung

Fraglich ist allerdings, ob beim Training einer KI eine tatbestandsmäßige Informationsgewinnung insbesondere über Muster, Trends und Korrelationen im Sinne des § 44b Abs.1 UrhG stattfindet. Denn KI ist eine Blackbox,³³⁴ die zwar die den Werken zugrundeliegenden Muster und Zusammenhänge nutzt, um ihre Parameter anzupassen.³³⁵ Informationen über die aus den Trainingsdaten extrahierten und im System abgebildeten Muster werden aber nicht in einer menschlich wahrnehmbaren Form ausgegeben. Sie sind vielmehr subsymbolisch, also in numerischer Form im KI-System hinterlegt.³³⁶ Welche Muster eine KI bei der Analyse berücksichtigen soll, wird auch nicht vom Entwickler festgelegt.³³⁷ Umgekehrt ist eine Entschlüsselung der subsymbolisch hinterlegten Muster und Zusammenhänge selbst für KI-Experten nicht ohne weiteres und nur im Einzelfall möglich.³³⁸ Obwohl computerwissenschaftliche Forschung das Ziel verfolgt, diese Blackbox aufzulösen (*explainable AI*),³³⁹ ist dies nach bisherigem Forschungsstand noch nicht gelungen.³⁴⁰ Darüber hinaus greift das KI-System regelmäßig Muster aus den Trainingsdaten auf, die vom Entwickler nicht

334 Zum Begriff und dem Gegensatz, sogenannter „Whitebox“-Modelle, siehe: *Bauckhage/Fürnkranz/Paaß*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 571 (575 ff.); ganz grundlegend *Seiffert/Stürmer*, in: Görz, Einführung in die künstliche Intelligenz, S. 863 (899 f.); außerdem *Döbel u. a.*, Maschinelles Lernen, S. 30.

335 *Hofmann* stellt für die Informationsgewinnung auch auf die Anpassung der Parameter des KNN ab, siehe *Hofmann*, WRP 2024, II Rn. 15.

336 *Döbel u. a.*, Maschinelles Lernen, S. 11.

337 *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 509 (592); *Rey/Wender*, Neuronale Netze, S. 98; vgl. auch *Döbel u. a.*, Maschinelles Lernen, S. 46.

338 *Bauckhage/Fürnkranz/Paaß*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 571 (575); *Puppe*, in: Beck/Kusche/Valerius, Digitalisierung, Automatisierung, KI und Recht, S. 121 (126); *Döbel u. a.*, Maschinelles Lernen, S. 33; *Konertz/Schönhof*, Das technische Phänomen „Künstliche Intelligenz“ im allgemeinen Zivilrecht, S. 51; *Gausling*, ZD 2019, 335 (335 f.) außerdem mit einer kurzen Darstellung der Möglichkeit, Licht in die Black-Box zu bringen.

339 *Menzel*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 601 (661); *Bauckhage/Fürnkranz/Paaß*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 571 (576 ff.); *Döbel u. a.*, Maschinelles Lernen, S. 30, 37.

340 *Bauckhage/Fürnkranz/Paaß*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 571 (592); *Ragni*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 227 (228).

antizipiert worden sind.³⁴¹ Der KI fehlt es am sozialen, den Werkkontext einbeziehenden Verständnis, das die Relevanz von Mustern anhand ihres sozialen Sinngehalts einordnet.

Auch wenn ein KI-System beim Trainingsvorgang auf gewisse Weise die Muster, Zusammenhänge und Korrelationen verarbeitet, die den eingegebenen Werken zugrunde liegen, stellt sich damit dennoch die Frage, ob das Training der KI aufgrund des beschriebenen Blackbox-Phänomens in einer tatbestandlichen Informationsgewinnung mündet, wie § 44b Abs.1 UrhG es voraussetzt.³⁴² Auf den natürlichen Wortsinn des „Text und Data Mining“-Begriffs kommt es hingegen nicht an.³⁴³ Denn das Text und Data Mining ist in § 44b Abs.1 UrhG legaldefiniert. Die Bedeutung des normativen Rechtsbegriffs bestimmt sich damit allein nach den Bedeutungsmerkmalen, die in der Legaldefinition vorgegeben worden sind. Entscheidend ist demzufolge, wie diese Bedeutungsmerkmale verstanden werden können. Das muss anhand der gängigen Auslegungsmethoden ermittelt werden.

Ausschlaggebend für das Verständnis des Informationsgewinnungsbegriffs im Sinne des § 44b Abs.1 UrhG ist, ob der Begriff des *Gewinnens* von Informationen notwendig macht, dass die Informationen, also die den Werken zugrunde liegenden Muster und Zusammenhänge, bei der Gewinnung für einen Menschen wahrnehmbar gemacht werden. Demgegenüber könnte auch jede Art der Nutzbarmachung der Informationen ohne eine menschliche Wahrnehmarmachung genügen. Das ist beim Anpassen der Parameter eines KNN im Rahmen des KI-Trainings der Fall.³⁴⁴

aa) Wortlaut als Ausgangspunkt der Auslegung

Ausgangspunkt der Auslegung ist der Wortsinn des Gewinnungsbegriffs.³⁴⁵ Das gilt auch im europäischen Recht.³⁴⁶ Der Wortsinn des Gewinnungsbe-

341 *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 509 (592); zur schönen Geschichte des „klugen Hanses“, einem rechnenden Pferd, siehe *Bauckhage u. a.*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 509 (572 f.).

342 Auch *Linke*, in: Kuschel/Asmussen/Golla, Intelligente Systeme - intelligentes Recht: GRUR Junge Wissenschaft Hamburg 2020/2021, S. 179 (188 ff.).

343 Anders jedoch *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 94 ff.

344 Ähnlich *Kleinkopf*, Text- und Data-Mining, S. 233, 238.

345 Zur Auslegungsregel als allg. M. für alle *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 163 f.

griffs ist damit sowohl für das in § 44b Abs. 1 UrhG verortete Tatbestandsmerkmal als auch für den in Art. 2 Nr. 2 DSM-RL verwendeten Gewinnungsbegriff heranzuziehen, der wiederum im Rahmen richtlinienkonformer Auslegung in § 44b Abs. 1 UrhG berücksichtigt werden muss.

(1) Wortsinn des Begriffs der „Gewinnung“ im deutschen Regelungskontext

Der Begriff *Gewinnen* kann in verschiedenen Zusammenhängen verwendet werden, beispielsweise im Sport („einen Pokal gewinnen“) oder, um einen meinungsbildenden Vorgang zu beschreiben („jemanden für etwas gewinnen“). Darüber hinaus ist der Begriff auch für die Beschreibung der Rohstoffförderung nutzbar („Kohle oder Eisenerz gewinnen“).³⁴⁷

Die gesetzliche Bezeichnung des Datenverarbeitungsprozesses als Text und Data *Mining* (zu Deutsch: „Abbau/Bergbau/Förderung“ von Text und Daten) macht deutlich, dass im Normtext auf die letztgenannte Bedeutungsvariante Bezug genommen wird. Damit ist der natürliche Wortsinn des Tatbestandsmerkmals der Informationsgewinnung anhand der Bedeutung des Wortes *gewinnen* im Kontext der Rohstoffförderung zu bestimmen.

Das *Gewinnen* beschreibt in diesem Sinnzusammenhang einen Prozess, bei dem ein in einem Ausgangsobjekt enthaltenes Produkt zum Zwecke der Nutzung aus dem Ausgangsobjekt herausgeholt wird.³⁴⁸ Synonym können auch die Wörter *produzieren* oder *erzeugen* verwendet werden.³⁴⁹ Bei der Eisengewinnung wird darüber hinaus beispielsweise nicht schon mit dem Abbau des Eisenerzes, in dem das Eisen enthalten ist, sondern erst mit der Verarbeitung des Erzes im Hochofen zu Roheisen von einer *Eisengewinnung* gesprochen.³⁵⁰ Dies spricht dafür, dass der Begriff des *Gewinnens* voraussetzt, dass durch den Verarbeitungsprozess ein abgrenzbarer Gegenstand extrahiert worden ist, der als selbstständige Einheit wahrgenommen werden kann.

346 EuGH, Urt. v. 05.07.2012 - C-49/11, EuZW 2012, 638 Rn. 32 – Content Services; EuGH, Urt. v. 10.03.2005 - C-336/03, MMR 2005, 364 Rn. 21 – easycar.

347 Zu den verschiedenen Bedeutungsvarianten siehe *Dudenredaktion (Bibliographisches Institut)*, Duden: das Bedeutungswörterbuch, S. 455.

348 *Dudenredaktion (Bibliographisches Institut)*, Duden: das Bedeutungswörterbuch, S. 455.

349 *Dudenredaktion (Bibliographisches Institut)*, Duden: das Bedeutungswörterbuch, S. 455.

350 Vgl. beispielsweise die Terminologie in *Bachmann*, Archäologie in Deutschland 1993, 35 (35, 38).

Auch die Etymologie des Gewinnungsbegriffs könnte ein solches Verständnis nahelegen. Denn das Wort geht auf das mittelhochdeutsche *Gewinnen* und das althochdeutsche *Giwinnan* zurück, die beschreiben, dass durch Arbeit zu etwas gelangt worden oder etwas geschaffen worden ist.³⁵¹ Damit ist der Begriff *gewinnen* auch etymologisch mit der Extrahierung eines Produkts, also eines als selbstständigen Gegenstands wahrnehmbaren Erzeugnisses, verknüpft. Auch das spricht daher für das dargelegte Begriffsverständnis.

Der natürliche Sinn des Wortes *gewinnen* spricht demzufolge dafür, dass auch bei der *Gewinnung von Informationen* im Sinne des § 44b Abs. 1 UrhG Informationen als abgrenzbarer Gegenstand extrahiert werden müssen, so dass sie im Folgenden als selbstständige Einheit angesehen werden können. Dies hat zur Folge, dass die Informationen als durch die automatisierte Analyse gewonnenes Produkt jedenfalls selbstständig wahrnehmbar sein müssen. Eine implizite Symbolisierung der Muster und Zusammenhänge in den numerischen Parametern eines KNN genügt diesen Anforderungen nicht.

Berücksichtigt werden muss darüber hinaus auch ein besonderer juristischer Sinngehalt des als Rechtsbegriffs verwendeten Wortes *Gewinnen*. Voraussetzung dafür ist, dass der Begriff bereits im UrhG geprägt worden ist.³⁵² Das ist in § 69e Abs. 2 UrhG der Fall. Er beschränkt die Weiterverwendung von Informationen, die beim Dekompilieren eines Computerprogrammes zur Herstellung von Interoperabilität gewonnen worden sind. Eine solche Weiterverwendung der Informationen zu nicht erlaubten Zwecken wie beispielsweise der Vermarktung ähnlicher Software kann jedoch überhaupt nur dann erfolgen, wenn die erlangten Informationen menschlich wahrnehmbar geworden sind. Infolgedessen setzt auch § 69e Abs. 2 UrhG eine Wahrnehmbarkeit der gewonnenen Informationen voraus. Grundsätzlich streitet eine Vermutung für eine einheitliche Auslegung von Rechtsbegriffen innerhalb desselben Regelungswerkes, solange keine besonderen Gründe ein abweichendes Begriffsverständnis nahelegen.³⁵³ Für ein von § 69e UrhG abweichendes Begriffsverständnis spricht allerdings, dass § 69e UrhG Teil der in den §§ 69a ff. UrhG normierten Sonderbestimmungen für Compu-

351 *Dudenredaktion (Bibliographisches Institut)*, Duden: das Herkunftswörterbuch, S. 338.

352 Zur methodischen Grundlage vgl. *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 142 f.

353 *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 142; *Reimer*, Juristische Methodenlehre, S. S. 147 Rn. 288 in Fn. 326.

terprogramme ist. Die aus der Bestimmung folgenden Einschränkungen der Dekompilierungsfreiheit stehen dabei dem Schutz von Knowhow näher als dem originären Urheberrechtsschutz.³⁵⁴

Für das Erfordernis der Wahrnehmbarmachung der den Werken zugrundeliegenden Informationen spricht weiterhin allerdings, dass es begrifflich einen Unterschied zwischen einer *Verarbeitung* und einer *Gewinnung* von Informationen gibt. Der Verarbeitungsbegriff ist in Art. 4 Nr. 2 DS-GVO legaldefiniert. Es genügt danach jede Verarbeitung von Informationen. Mit dem Gewinnungsbegriff muss demgegenüber ein anderer Sachverhalt adressiert worden sein.

Auch wenn aus § 69e UrhG kein allgemeines Verständnis des Rechtsbegriffs der Informationsgewinnung abgeleitet werden kann, legen jedenfalls der natürliche Sprachgebrauch sowie der terminologische Unterschied zwischen der *Verarbeitung* und *Gewinnung* von Informationen nahe, dass das *Gewinnen* von Informationen eine menschliche Wahrnehmbarkeit der extrahierten Inhalte voraussetzt.

(2) Wortsinn des Begriffs der „Gewinnung“ im unionsrechtlichen Regelungskontext

Auch die unionsrechtliche Legaldefinition aus Art. 2 Nr. 2 DSM-RL setzt voraus, dass die Analysetechnik jedenfalls dazu geeignet sein muss, Informationen beispielsweise über Muster oder Zusammenhänge zu gewinnen. Damit stellen sich die aufgeworfenen Auslegungsprobleme auch auf Ebene des europäischen Rechts. Der Gewinnungsbegriff, der wie in § 44b Abs. 1 UrhG auch in der deutschen Fassung der DSM-RL verwendet wird, spricht seinem natürlichen Sprachgebrauch nach dafür, dass die Richtlinie in Art. 2 Nr. 2 DSM-RL eine Wahrnehmbarmachung der extrahierten Informationen voraussetzt.³⁵⁵ Auch in der englischen, italienischen und französischen Sprachfassung (EN: „generate information“ – etwas „erzeugen“, „schaffen“, „herstellen“; IT: „di generare informazioni“ – Bedeutung wie in englischer Sprachfassung; FR: „dégager des informations“ – „bergen“, aus etwas „etwas herausziehen“, etwas „herausarbeiten“) werden Begriffe verwendet, die ihrem natürlichen Wortsinn nach die Extrahierung eines abgrenzbaren Gegenstands, der als selbstständige Einheit angesehen und daher wahrnehmbar sein muss, voraussetzen.

354 Grützmacher, in: Wandtke/Bullinger, Urheberrecht, UrhG § 69e Rn. 1 ff.

355 Im Detail zum natürlichen Sprachgebrauch siehe unmittelbar oben unter 3. Teil A. III. 2. a) aa).

Ein besonderer juristischer Sprachgebrauch der Formulierung findet sich, der deutschen Rechtslage entsprechend, nur in der Computerprogramm-RL³⁵⁶. Ein Schluss auf ein universell gültiges juristisches Verständnis des Begriffs der Informationsgewinnung ist hieraus allerdings nicht möglich.³⁵⁷

(3) Schlussfolgerung aus dem Wortlaut der Bestimmungen

Der natürliche Wortsinn des Gewinnungsbegriffs spricht sowohl aus einer europäischen als auch aus einer national-rechtlichen Perspektive im Ergebnis dafür, dass eine menschliche Wahrnehmbarmachung der Informationen erforderlich ist, damit von einer *Gewinnung* dieser Informationen im Sinne des § 44b Abs. 1 UrhG respektive Art. 2 Nr. 2 DSM-RL gesprochen werden kann.

bb) Systematische Bezüge

Abweichend vom natürlichen Wortsinn könnte die Stellung des § 44b UrhG im Normgefüge aber gegen das Erfordernis der menschlichen Wahrnehmbarkeit der durch die automatisierte Analyse gewonnenen Informationen sprechen. Denn § 44b UrhG steht im unmittelbaren systematischen Zusammenhang zu § 44a UrhG. Dieser stellt technisch bedingte, vorübergehende Vervielfältigungshandlungen frei, die beispielsweise für den digitalen Werkgenuss oder die Übermittlung in einem Netz zwischen Dritten benötigt werden. Dieser Regelungszusammenhang macht deutlich, dass auch § 44b UrhG vorrangig das Ziel verfolgt, einen technischen Verarbeitungsprozess zu ermöglichen. Dann muss für die Anwendbarkeit der Schrankenbestimmung die technische, informationsgestützte Analyse der Werke, nicht aber die menschliche Wahrnehmbarkeit der in den Werken enthaltenen Informationen entscheidend sein. Sollte die Norm – was eine Wahrnehmbarmachung der Informationen voraussetzen würde – vorrangig die menschliche Informationsaufnahme ermöglichen, hätte sie systematisch im Zusammenhang mit § 53 UrhG umgesetzt werden können. Dieser schützt allem das

356 Richtlinie 2009/24/EG des Europäischen Parlaments und des Rates vom 23. April 2009 über den Rechtsschutz von Computerprogrammen.

357 Im Detail zum natürlichen Sprachgebrauch siehe unmittelbar oben unter 3. Teil A. III. 2. a) aa).

menschliche Informationsinteresse.³⁵⁸ § 53 UrhG hätte dann den privaten Werkgenuss, § 60c den wissenschaftlichen Werkgenuss, „§ 53a UrhG“ den privaten automatisierten Werkgenuss mittels Text und Data Mining und § 60d UrhG den wissenschaftlichen automatisierten Werkgenuss mittels Text und Data Mining freigestellt. Dass dies aber nicht der Fall ist, spricht systematisch dafür, dass es für das Text und Data Mining im Sinne des § 44b UrhG auf die *automatisierte informationsbezogene Analyse*, nicht aber auf eine Informationsgewinnung im Sinne einer Wahrnehmbarmachung der Informationen ankommt.

Für eine weite Auslegung des Tatbestandsmerkmals, die keine Wahrnehmbarkeit der Informationen voraussetzt und im Ergebnis damit auch das Training von KI unter den Text und Data Mining-Begriff fasst, spricht zudem aus Perspektive des in § 44b Abs.1 UrhG umgesetzten Art.2 Nr.2 DSM-RL auch, dass die europäische KI-Verordnung Anbieter von KI-Basismodellen durch Art.53 Nr.1 lit.c KI-VO verpflichtet, beim Training ihrer KI-Systeme europäisches Urheberrecht und insbesondere auch Art.4 Abs.3 DSM-RL zu beachten.³⁵⁹ Hieraus kann geschlussfolgert werden, dass Vervielfältigungen zum Training von KI in den Anwendungsbereich von Art.4 DSM-RL fallen.³⁶⁰ Dies ist vorrangig bei einer weiten Auslegung des Text und Data Mining-Begriffs möglich, der keine Wahrnehmbarkeit der extrahierten Informationen voraussetzt. Gegen einen solchen systematischen Schluss spricht nicht, dass Art.53 Nr.1 lit.c KI-VO erst nach Art.4 DSM-RL eingefügt worden ist.³⁶¹ Denn bei der systematischen Auslegung wird die Stellung einer Rechtsnorm im gesamten Regelungssystem in den

358 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 58; a. A. hingegen BT-Drs. 16/1828, S. 20 f.

359 Der Bezug zum Training der KI ergibt sich deutlich aus ErwGr. 104, 105 KI-VO.

360 So auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 48; *Leistner*, GRUR 2024, 1665 (1668); *Stieper/Denga*, GRUR 2024, 1473 (1475); *Bomhard, David/Siglmüller*, RD 2024, 45 Rn. 25; *Buchalik/Gehrmann*, Computer und Recht, 145 Rn. 59; a. A. *Schippan*, ZUM 2024, 670 (676 f.); *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 129 ff. Beide Gegenstimmen argumentieren widersprüchlich, dass sich Art. 53 Nr. 1 lit. c KI-VO nur auf Text und Data Mining „beim“ KI-Training beziehe, was nicht bedeuten würde, dass KI-Training und Text und Data Mining gleichzusetzen seien. Vorangehend lehnen sie aber ab, dass beim KI-Training Text und Data Mining stattfände, womit die Vervielfältigungen für das Training von KI nicht unter § 44b Abs. 1 UrhG fallen würden.

361 A. A. *Schippan*, ZUM 2024, 670 (677). Wie hier auch *Leistner*, GRUR 2024, 1665 (1669 f.).

Blick genommen.³⁶² Hierfür kommt es auf den Ist-Zustand der Rechtsnormen an. Die zeitliche Reihenfolge, in der die Regelungen in Kraft getreten sind, ist aus der auf die Systematik bezogenen Perspektive hingegen nicht von Bedeutung.

cc) Technologische Neutralität des Urheberrechts

Auch der Grundsatz der Technikneutralität des Urheberrechts, nach dem die urheberrechtlichen Rechtsnormen soweit wie möglich technologisch neutral auszugestaltet sind,³⁶³ könnte darüber hinaus gegen die Notwendigkeit einer Wahrnehmbarmachung der herausgearbeiteten Informationen sprechen. Denn die verschiedenen Verfahren maschinellen Lernens unterscheiden sich insbesondere zwischen Blackbox-Verfahren, auf denen Systeme wie KNN beruhen, und sogenannten Whitebox-Methoden. Bei Letzteren sind die extrahierten Muster und Korrelationen im Anschluss an den Trainingsprozess unmittelbar wahrnehmbar. Eine menschliche Auswertung der durch das Training extrahierten Informationen ist auf dieser Basis möglich. Bei einem engen Verständnis des Tatbestandsmerkmals der *Informationsgewinnung*, das eine Wahrnehmbarkeit der extrahierten Informationen voraussetzt, würde also das Training von Whitebox-Modellen unter den Text und Data Mining-Begriff des § 44b UrhG fallen. Blackbox-Modelle wie KNN könnten hingegen nicht von der Schrankenbestimmung profitieren. Ob § 44b UrhG Vervielfältigungshandlungen zum Training von KI privilegiert, kann wegen der Technikneutralität des Urheberrechts aber nicht davon abhängen, welche Art des maschinellen Lernens eingesetzt wird. Maßgeblich muss sein, dass sich sämtliche dieser Technologien die in den Werken enthaltenen Informationen, also Muster und Zusammenhänge, zunutze machen.

Black- und Whitebox-Modelle sind im Übrigen strukturell vergleichbar. Beide Verfahren gehören zur Klasse des maschinellen Lernens. Bei einer normativen Differenzierung zwischen Black- und Whitebox-Modellen droht damit ein Wertungswiderspruch. Sowohl die Technikneutralität des Urheberrechts als auch das Bedürfnis nach einer Vermeidung von Wertungswidersprüchen sprechen daher im Ergebnis dafür, im Rahmen des

362 *Riesenhuber*, in: *Riesenhuber, Europäische Methodenlehre*, §10 Die Auslegung Rn. 22, 24.

363 *de la Durantaye/Hofmann*, ZUM 2021, 873 (877); BT-Drs. 19/27426, S. 74.

§ 44b UrhG keine Wahrnehmbarmachung der *gewonnenen* Informationen vorauszusetzen.

dd) Historische Auslegung

Auch die Vorstellungen sowohl des nationalen Gesetzgebers als auch des europäischen Richtliniengebers bei der Einführung der Schrankenbestimmung für Text und Data Mining könnten gegen das Erfordernis einer Wahrnehmbarkeit der verarbeiteten Informationen sprechen.

(1) Nationale Gesetzeshistorie

Zwar ist die historische Auslegung insbesondere bei der Konkretisierung von Tatbestandsmerkmalen aus älteren Gesetztestexten im nationalen Recht nur von geringer Bedeutung.³⁶⁴ Bei jungen Regelungen ist der objektiv ermittelbare Wille des Normgebers aber noch gegenwärtig. Den in den Gesetzesmaterialien zum Ausdruck kommenden Regelungsabsichten ist in diesen Fällen daher noch eine erhebliche Bedeutung beizumessen.³⁶⁵ Das gilt gerade dann, wenn die übrigen Auslegungsmechanismen kein klares Ergebnis hervorbringen können.³⁶⁶ Da § 44b UrhG erst im Juni 2021 in Umsetzung der DSM-RL eingeführt wurde und bisher kein klares Auslegungsergebnis herausgearbeitet werden konnte, kann bei der Auslegung des Begriffs der Informationsgewinnung auf die Gesetzgebungsmaterialien und damit auf die Regelungsabsicht des historischen Gesetzgebers Bezug genommen werden.

Der Gesetzgeber verfolgt mit der Regelung unter anderem einen innovationspolitischen Zweck.³⁶⁷ Gerade, weil das Erfordernis der Wahrnehmbarmachung der genutzten Informationen einer Anwendung des § 44b UrhG auf die Beschaffung von Trainingsdaten für KI entgegenstehen würde, spricht das gegen diese Auslegung des Tatbestandsmerkmals.

364 Walz, ZJS 2010, 482 (486); noch relativierender BAG, Urt. v. 06.04.2011 - 7 AZR 716/09, NZA 2011, 905 (908).

365 BVerfG, Urt. v. 11.06.1980 - 1 PBvU 1/79, NJW 1981, 39 (43); Walz, ZJS 2010, 482 (485 f.).

366 BVerfG, Urt. v. 11.06.1980 - 1 PBvU 1/79, NJW 1981, 39 (43).

367 Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 1; BT-Drs. 19/27426, S. 60.

Weiterhin setzte die vor dem Inkrafttreten des Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarkts im Jahr 2021 geltende Schrankenbestimmung für Text und Data Mining, § 60d UrhG a. F., lediglich voraus, dass eine Vielzahl von Werken automatisiert ausgewertet werden sollen. Das Ziel der Analyse wurde nicht weiter eingegrenzt. Damit hätte das Training einer KI unter den Tatbestand subsumiert werden können.³⁶⁸ Es erscheint vor dem innovationspolitischen Hintergrund der Norm unwahrscheinlich, dass mit der Einführung des § 44b UrhG eine Schlechterstellung der KI-Entwickler und Trainingsdatensammler erfolgen sollte. Auch das spricht dagegen, dass die Informationsgewinnung im Sinne des § 44b Abs. 1 UrhG eine menschliche Wahrnehmbarmachung der in den Werken enthaltenen Informationen voraussetzt.

Darüber hinaus geht der Gesetzgeber selbst davon aus, dass § 44b UrhG als Basis für maschinelles Lernen genutzt werden kann.³⁶⁹ Hierunter können auch die KNN gefasst werden. Es ergeben sich keine Hinweise darauf, dass der Gesetzgeber nur ausgewählte Arten maschinellen Lernens, namentlich Whitebox-Verfahren, adressieren wollte. Insgesamt sprechen die Gesetzgebungsmaterialien des nationalen Gesetzgebers damit aus mehreren Gründen gegen ein Erfordernis der Wahrnehmbarmachung der den analysierten Werken zugrundeliegenden Informationen.

(2) Historie der Richtlinienbestimmung

Für das Erfordernis der Wahrnehmbarkeit und damit gegen eine Anwendung des Text und Data Mining-Begriffs auf das Training von KI spricht hingegen, dass KI in den Materialien zum europäischen Legislativprozess³⁷⁰ kein einziges Mal erwähnt wurde.³⁷¹ Weiterhin hat der Richtliniengeber den Informationsgewinn in ErwGr. 8 S. 1 DSM-RL als „Erkenntnisgewinn“ umschrieben. Damit ein menschlicher Erkenntnisgewinn möglich ist, ist es aber erforderlich, dass die Informationen für Menschen wahrnehmbar gemacht worden sind.

Allerdings hat die genannte Formulierung ihren Ursprung bereits im ersten Richtlinienvorschlag der EU-Kommission (COM/2016/0593 final).

368 Anders wohl *Kleinkopf*, Text- und Data-Mining, S. 238, die davon ausgeht, § 44b Abs. 1 UrhG sei „technologieneutraler als die Vorgängernorm (§ 60d UrhG a.F.)“.

369 BT-Drs. 19/27426, S. 60.

370 Rechtssetzungsverfahren 2016/0280/COD.

371 Hierauf recurriert auch *Schack*, NJW 2024, 113 Rn. 8.

Hierin war die Schrankenbestimmung für Text und Data Mining noch auf den Zweck wissenschaftlicher Forschung beschränkt (Art. 3 Nr. 1 DSM-RL-E). Damit bestand im Richtlinienentwurf tatsächlich noch eine engere Bindung der Schrankenbestimmung an den (wissenschaftlich betriebenen) Erkenntnisgewinn. Nach der Erweiterung der Schrankenbestimmung auf allgemeines Text und Data Mining, in deren Zusammenhang auch Formulierungen in den Richtlinienentwurf eingefügt wurden, die insbesondere die Bedeutung der Schrankenbestimmung für Innovationen und die Entwicklung neuer Anwendungen und Technologien hervorheben (ErwGr. 5, 18 DSM-RL, siehe angenommenen Text des Europäischen Parlaments aus erster Lesung, Dokument P8_TA-PROV(2019)0231),³⁷² wurde die betreffende Passage in ErwGr. 8 S.1 DSM-RL allerdings nicht mehr angepasst. Vor dem Hintergrund dieser Regelungshistorie ist der Umschreibung der Informationsgewinnung mit dem Begriff des Erkenntnisgewinns in ErwGr. 8 S. 1 DSM-RL daher keine besonders hohe Bedeutung beizumessen. Es handelt sich vielmehr um eine redaktionelle Ungenauigkeit.

ee) Zweck des zugrundeliegenden Richtlinienrechts

Die DSM-RL verfolgt weiterhin den Zweck, mit der Schrankenbestimmung für Text und Data Mining zu Innovationen anzuregen und neue Technologien zu ermöglichen.³⁷³ KI und insbesondere KNN, die Gegenstand dieser Untersuchung sind, sind dabei wesentlicher Teil der technischen Innovationen. Auch das Innovationspotential ist in diesem Bereich hoch. Die in der Richtlinie angelegte innovationsfreundliche Auslegung des Tatbestandsmerkmals spricht damit entgegen der Richtlinienhistorie gegen das Erfordernis einer Wahrnehmbarmachung der bei der automatisierten Analyse extrahierten Informationen. Der innovationsfördernde Zweck der Regelung besteht außerdem unabhängig davon, ob durch das innovative Erzeugnis Konkurrenzprodukte genutzter Werken generiert werden.³⁷⁴ Allgemein zielt die DSM-RL auch auf die Ermöglichung digitaler Nutzungen ab.³⁷⁵

372 Diese Bedeutung wird auch von *de la Durantaye*, ZUM 2023, 645 (651) hervorgehoben.

373 ErwGr. 18 DSM-RL.

374 So aber *Schack*, NJW 2024, 113 Rn. 8. Ähnlich wie hier auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 43 ff.

375 Vgl. Art. 1 Abs. 1, ErwGr. 5 DSM-RL.

ff) Schlussfolgerungen

Alles in allem sprechen aus einer nationalen Perspektive systematische Erwägungen, der in den Gesetzgebungsmaterialien zum Ausdruck kommende Wille des historischen Gesetzgebers und das Prinzip der Technikneutralität des Urheberrechts für eine weite Auslegung des Tatbestandsmerkmals der Informationsgewinnung, nach der keine Wahrnehmbarmachung der extrahierten Informationen erforderlich ist. Aus europäischer Perspektive streiten insbesondere der in den Erwägungsgründen zum Ausdruck kommende Zweck der Regelung sowie der systematische Bezug zu Art. 53 Nr. 1 lit. c KI-VO für ein solches Verständnis des Informationsgewinnungsbegriffs. Gleiches gilt daher auch für die in Art. 2 Nr. 2 DSM-RL normierte, unionsrechtliche Legaldefinition des Text und Data Minings. Das muss richtlinienkonform auch bei Auslegung des § 44b Abs. 1 UrhG berücksichtigt werden.

Allerdings ist sowohl die aus nationaler Perspektive vorgenommene als auch die richtlinienkonforme Auslegung des Tatbestandsmerkmals der Informationsgewinnung dadurch limitiert, dass der natürliche Wortsinn des Gewinnungsbegriffs voraussetzt, dass eine Extraktion der Informationen als ein selbstständig in Erscheinung tretendes und daher durch Menschen wahrnehmbares Produkt erforderlich macht. Der natürliche Wortsinn bildet aber sowohl auf nationaler als auch auf europäischer Ebene die Grenze jeder Auslegung.³⁷⁶ Dort, wo diese Grenze überschritten wird, findet keine Auslegung, sondern eine Umdeutung des Regelungsgehalts statt.³⁷⁷ Aufgrund der Wortlautgrenze muss der Rechtsbegriff der Informationsgewinnung im Sinne des § 44b Abs. 1 UrhG daher so ausgelegt werden, dass er eine Wahrnehmbarmachung der extrahierten Informationen voraussetzt. Damit Informationen im Sinne der Norm *gewonnen* werden, müssen sie demzufolge als selbstständige Einheit abgrenzbar und infolgedessen aus sich heraus menschlich wahrnehmbar sein.

Die in den Trainingsdatensätzen verborgenen Muster und Zusammenhänge beeinflussen als Folge des Trainingsvorgangs die Einstellung der Parameter des KI-Systems. Sie werden in den Algorithmen jedoch nur numerisch, subsymbolisch und abgebildet. Demzufolge sind die beim Training in den Parametern eines KI-Systems verarbeiteten Muster und Zusammenhänge in den justierten Parametern nicht aus sich heraus für Menschen

376 *Riesenhuber*, in: *Riesenhuber*, Europäische Methodenlehre, §10 Die Auslegung Rn. 19; *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 144, 165, 166.

377 *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 143.

wahrnehmbar und verständlich. Bei der im Trainingsprozess stattfindenden Anpassung der numerischen Parameter eines KI-Systems auf Basis von Werken werden infolgedessen keine Informationen im Sinne des § 44b Abs. 1 UrhG gewonnen.³⁷⁸

b) Bezugspunkt der Informationsgewinnung

Neben den Anforderungen an das Ziel der Informationsgewinnung bei der automatisierten Analyse von Werken und dem Erfordernis der menschlichen Wahrnehmbarkeit der extrahierten Informationen ist weiterhin aber auch klärungsbedürftig, an welchen Bezugspunkt die Informationsgewinnung anknüpft, woraus die Informationen also gewonnen werden müssen, damit Text und Data Mining im Sinne des § 44b Abs. 1 UrhG stattfindet.

Zwar geht § 44b Abs. 1 UrhG davon aus, dass die Informationen „daraus“, also aus den untersuchten Werken gewonnen werden müssen. Damit wäre es notwendig, dass unmittelbar beim Training einer KI Informationen aus den genutzten Trainingsdaten gewonnen werden müssen.

Art. 2 Nr. 2 DSM-RL verlangt allerdings nur, dass die Informationen „mit Hilfe“ der Analyse der Werke gewonnen werden. Daraus folgt, dass die automatische Analyse der Inhalte zwar kausal für die Informationsgewinnung sein muss. Die gewonnenen Informationen müssen aber nicht aus den analysierten Daten hervorgehen, so wie § 44b Abs. 1 UrhG es voraussetzen scheint. Auch die englische, italienische, niederländische, spanische und französische Legaldefinition des Text und Data Mining im Sinne des Art. 2 Nr. 2 DSM-RL beschränken sich darauf, einen Kausalzusammenhang zwischen automatischer Analyse der Daten und Informationsgewinnung vorauszusetzen.³⁷⁹ § 44b Abs. 1 UrhG muss demzufolge insoweit richtlinienkonform ausgelegt werden, als dass für die Tatbestandserfüllung ebenfalls nur eine Kausalbeziehung zwischen automatisierter Analyse und Informationsgewinnung notwendig ist. Daraus folgt, dass keine Identität zwischen dem Trägermedium der Informationen und dem Analysegegenstand beste-

378 A. A. Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 11a; Hofmann, WRP 2024, 11 Rn. 15; Hofmann, ZUM 2024, 166 (170 f.); implizit auch Konertz/Schönhof, WRP 2024, 289 Rn. 30.

379 Ähnlich auch Kleinkopf, Text- und Data-Mining, S. 230; verfehlt daher Jager, Artificial Creativity?, S. 351, der eine Informationsgewinnung direkt aus der Datenanalyse, also dem Training des KNN voraussetzt und unter anderem deswegen die Anwendbarkeit des § 44b UrhG auf das KI-Training ablehnt.

hen muss. Weiterhin müssen die automatisierte Analyse sowie die Informationsgewinnung nicht auf derselben Handlung beruhen oder gleichzeitig stattfinden. Die richtlinienkonforme Auslegung des § 44b Abs. 1 UrhG führt vielmehr dazu, dass auch die Subsumtion zweiaktiger Vorgänge unter den Tatbestand ermöglicht wird, solange eine Kausalbeziehung zwischen automatisierter Analyse und Informationsgewinnung besteht.

c) Schlussfolgerungen für das KI-Training als Text und Data Mining im Sinne des § 44b Abs. 1 UrhG

Für das Training einer KI und dessen Subsumtion unter § 44b Abs. 1 UrhG bedeuten die aus der richtlinienkonformen Auslegung des § 44b Abs. 1 UrhG folgenden Anforderungen an den Bezugspunkt der Informationsgewinnung, dass der Anknüpfungspunkt der tatbestandlich notwendigen Gewinnung wahrnehmbarer Informationen weg von der Konfiguration der Parameter des KNN im Trainingsprozess und hin zur späteren Generierung des Outputs des KI-Systems verschoben werden kann. Es ist also ausreichend, wenn Informationen nicht bereits im Rahmen des eigentlichen Lernprozesses, also aus der automatisierten Analyse der Werke, sondern erst in der Anwendungsphase des Systems gewonnen werden.³⁸⁰ Denn jedes KI-Training, also jede automatisierte Analyse von Inhalten, findet statt, damit in der Anwendungsphase eines KI-Systems anhand neuen Inputs neue Informationen, der KI-Output, gewonnen werden können. Es besteht demzufolge die notwendige Kausalbeziehung zwischen der Analyse der Werke beim eigentlichen Training und der späteren Erzeugung des Outputs als Informationsgewinnung. Deutlich wird das beispielhaft und insbesondere bei Klassifikations-Systemen, die nach dem Training in der Lage sind, eingegebene Objekte vordefinierten Gruppen zuzuordnen.

Unter Informationen werden zudem sämtliche in Daten verkörperte semantische Inhalte verstanden.³⁸¹ In welcher Form die Verkörperung erfolgt (Schriftdaten, Bilddaten, ...), wird gesetzlich nicht präzisiert. § 44b Abs. 1 UrhG macht außerdem keine Angaben dazu, ob die gewonnenen Informationen auch wahr sein müssen. Die Inhalte können daher auch fiktional

380 Wie hier auch *Vesala*, IIC 2023, 351 (356); a. A. hingegen *Jager*, *Artificial Creativity?*, S. 351; eine Verbindung zwischen Trainingsanalyse und Generierung des Outputs stellt darüber hinaus auch *Schippan* her, wobei im Ergebnis abweichend, siehe *Schippan*, ZUM 2024, 670 (674).

381 *Kleinkopf*, *Text- und Data-Mining*, S. 70; *Martini u. a.*, MMR-Beil. 2021, 3 (3).

sein.³⁸² Hiergegen spricht auch nicht, dass es sich bei fiktionalen Information nicht um „nützliche Erkenntnisse“ handelt.³⁸³ Denn § 44b Abs. 1 UrhG setzt *expressis verbis* nur die Gewinnung von Informationen, nicht aber von Erkenntnissen mit wissenschaftlichem oder gesellschaftlichem Wert voraus. Eine strengere Auffassung ist mit dem Regelungswortlaut nicht in Einklang zu bringen. Darüber hinaus könnte *de jure* nicht konkretisiert werden, wann eine Information nützlich genug ist, sodass ihre Gewinnung unter § 44b Abs. 1 UrhG subsumiert werden kann. Eine Beschränkung insbesondere auf wissenschaftliche Inhalte wie Klassifikationen ist deswegen nicht angezeigt.³⁸⁴ Selbst die Erzeugung kreativer Leistungsergebnisse wie Bilder oder Texte ist, wie auch ErwGr. 8 S. 1 DSM-RL nahelegt, eine Erzeugung von Informationen im Sinne der Schrankenbestimmung.³⁸⁵ Daher kann auch das Training generativer KI, die in der Anwendungsphase Kreativinhalte ausgibt, unter die Legaldefinition des § 44b Abs. 1 UrhG gefasst werden.³⁸⁶

Für die Subsumtion des Trainings von KI unter den Text und Data Mining-Begriff des § 44b Abs. 1 UrhG lässt sich damit schlussfolgern, dass eine automatisierte Analyse mehrerer digitaler oder digitalisierter Werke im Sinne des § 44b Abs. 1 UrhG stattfindet, *mit deren Hilfe* zu einem späteren Zeitpunkt Informationen in Form des KI-Outputs gewonnen werden sollen.³⁸⁷ Infolgedessen fällt das KI-Training trotz des vorangehend herausgearbeiteten engen Verständnisses der „Informationsgewinnung“ unter den Text und

382 Vgl. z. B. *Zech*, Information als Schutzgegenstand, S. 247 f.

383 So aber *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 57 f., 101 f.

384 Anders allerdings *Jager*, Artificial Creativity?, S. 349, der von einem engeren Informationsbegriff auszugehen scheint. Außerdem *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 101 f., wobei hier ebenfalls keine vertiefte Auseinandersetzung mit dem Gesetzeswortlaut stattfindet.

385 A. A. *Shippan*, ZUM 2024, 670 (675), aber ohne nähergehende Begründung; außerdem *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 101 ff.

386 A. A. *Jager*, Artificial Creativity?, S. 349; *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 94 ff.

387 So auch *Heinze/Wendorf*, in: Ebers/Heinze/Krüger/Steinrötter, Künstliche Intelligenz und Robotik, § 9 KI und Urheberrecht Rn. 21; im Kontext der DSM-RL *Meys*, GRUR Int. 2020, 457 (464 f.); *Linke*, in: Kuschel/Asmussen/Golla, Intelligente Systeme - intelligentes Recht: GRUR Junge Wissenschaft Hamburg 2020/2021, S. 179 (189 f.). Ähnlich *Käde*, Kreative Maschinen und Urheberrecht, S. 230, die die richtlinienkonforme Auslegung aber noch im Tatbestandsmerkmal der „Informationsgewinnung“ vornimmt.

Data Mining-Begriff des § 44b Abs. 1 UrhG.³⁸⁸ Auf den Anwendungszweck der KI, insbesondere einer Nicht-Kommerzialität des Handelns, kommt es hingegen nicht an.³⁸⁹ Gegen dieses Auslegungsergebnis spricht auch nicht, dass der Richtliniengeber nicht ausdrücklich an das KI-Training als Fall des Text und Data Mining gedacht zu haben scheint.³⁹⁰ Denn es liegt im Wesen abstrakt-genereller Rechtsnormen, dass sie auch auf konkrete Einzelfälle angewendet werden können, die vom Gesetzgeber nicht vorhergesehen worden sind. Einer Rechtsfortbildung bedarf es demzufolge nicht.

d) Konformität mit dem Dreistufentest

Fraglich ist allerdings, ob eine Freistellung des Web Scrapings von Werken als Trainingsdaten für KI über § 44b UrhG mit dem Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL vereinbar ist. Danach dürfen die unionsrechtlich harmonisierten Schrankenbestimmungen nur in bestimmten Sonderfällen angewandt werden, in denen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beein-

388 So wohl auch, aber ohne nähere Begründung: *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 1; *Dregelies*, GRUR 2024, 1484 (1486); *Raue*, ZUM 2021, 793 (796); *Käde*, Kreative Maschinen und Urheberrecht, S. 91; jedenfalls ausdrücklich in diesem Sinne auch *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 11; *Leistner*, GRUR 2024, 1665 (1668); *Kleinkopf*, Text- und Data-Mining, S. 238. Bezogen auf die Reichweite des Art. 4 DSM-RL ebenfalls *Heinze/Wendorf*, in: Ebers/Heinze/Krüger/Steinrötter, Künstliche Intelligenz und Robotik, § 9 KI und Urheberrecht Rn. 21; *Linke*, in: Kuschel/Asmussen/Golla, Intelligente Systeme - Urheberrecht: GRUR Junge Wissenschaft Hamburg 2020/2021, S. 179 (189 f.); *Iglesias u. a.*, Intellectual Property and Artificial Intelligence - A literature review, S. 10; *Kleinkopf*, Text- und Data-Mining, S. 230; *Morgan*, Responsible AI, S. 273; *Chiou*, JIPITEC 2019, 398 Rn. 22; *Ducato/Strowel*, CRIDES Working Paper Series 2021, (19); implizit *Hacker*, ZGE 2020, 239 (269); *Margoni/Kretschmer*, GRUR Int. 2022, 685 (685, 686); *Senfleben u. a.*, JIPITEC 2022, 67 Rn. 10; *Spindler*, CR 2019, 277 Rn. 2; wohl ebenfalls *Würtenberger/Freischem*, GRUR 2019, 1140 (1140); *European Commission u. a.*, Study on copyright and new technologies: copyright data management and artificial intelligence (SMART 2019/0038), S. 198, 216 ff.; a. A. aber *Schippa*, ZUM 2024, 670 (675); *Jager*, Artificial Creativity?, S. 349 ff. Eine differenzierte Analyse des Regelungsbereichs von Art. 4 DSM-RL respektive § 44b UrhG im Hinblick auf KI-Training hat bisher allerdings nur bei *Heinze/Wendorf*, *Linke*, *Jager* und *Kleinkopf* stattgefunden. Ausdrücklich offengelassen in LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 40.

389 A. A. *Jager*, Artificial Creativity?, S. 349.

390 *Jager*, Artificial Creativity?, S. 350.

trächtig und die berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden.

aa) Bestimmter Sonderfall

Die Schrankenbestimmung des § 44b UrhG ist nur auf einen eng abgegrenzten Sachverhalt, nämlich Vervielfältigungen von Werken für eine automatisierte Analyse der Inhalte mit dem Ziel der Informationsgewinnung, also Text und Data Mining (§ 44b Abs. 1 UrhG), anwendbar. Daher betrifft die Schrankenbestimmung bereits aus sich heraus nur einen bestimmten Sonderfall.³⁹¹ Sie erfüllt demzufolge die erste Anforderung des Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL. Eine weitergehende Einschränkung des Anwendungsbereichs der Schrankenbestimmung zur Wahrung des Dreistufentests ist nicht notwendig.³⁹²

bb) Beeinträchtigung der normalen Verwertung und ungebührliche Beeinträchtigung der Urheberinteressen

Darüber hinaus darf die Freistellung der Speicherung urheberrechtlich geschützter Trainingsdaten über § 44b Abs. 2 UrhG auch nicht die normale Verwertung des Werkes beeinträchtigen. Dass ist der Fall, wenn die Freistellung unmittelbare Substitutionswirkung für die herkömmliche Werkverwertung der Urheber, also ihrer Primärverwertungsmöglichkeiten, entfaltet.³⁹³ In anderen Worten darf es nicht zu einer Verringerung der rechtmäßigen, primären Transaktionen der Urheber kommen.³⁹⁴ Ob die normale Werkverwertung im Sinne der Primärverwertung der Urheber beeinträchtigt ist, ist aufgrund der Unbestimmtheit des Tatbestandsmerkmals dabei eine normative Frage, sodass im Ergebnis eine umfassende Gesamtabwä-

391 Vgl. zur Voraussetzung *Loewenheim*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 44a Rn. 14; *Leenen*, in: Wandtke/Bullinger, Urheberrecht, InfoSoc-RL Art. 5 Rn. 162.

392 So zu § 50 UrhG BGH, Urt. v. 30.04.2020 - I ZR 228/15, NJW, 2554 Rn. 71 – Reformistischer Aufbruch II.

393 BGH, Urt. v. 30.04.2020 - I ZR 228/15, NJW, 2554 Rn. 72 – Reformistischer Aufbruch II; *Paul*, in: Borges/Hilber, Beck'scher Online-Kommentar IT-Recht, UrhG § 45 Rn. 10.

394 EuGH, Urt. v. 26.04.2017 - C-527/15, GRUR 2017, 610 Rn. 70 – Filmspeler.

gung durchzuführen ist.³⁹⁵ Hierbei kommt es zur Überschneidung mit der dritten Voraussetzung des Dreistufentests (ungebührliche Beeinträchtigung der berechtigten Interessen des Urhebers), womit es letztendlich auf eine umfassende Abwägung von Verwertungsinteressen der Urheber einerseits und Nutzerinteressen andererseits ankommt.

Damit ist die zweite und dritte Voraussetzung des Dreistufentests funktional identisch zur Interessenabwägung, die im zweiten Teil der vorliegenden Untersuchung im Rahmen der Ermittlung des regulatorischen Freistellungsbedürfnisses vorgenommen worden ist. Sie ergab, dass die Interessen der Trainingsdatensammler die betroffenen Urheberinteressen überwiegen.³⁹⁶ Dieses Abwägungsergebnis gilt demzufolge auch bei der Anwendung des Dreistufentests.

Alles in allem sprechen daher bessere Gründe dafür, dass die herkömmliche Werknutzung nicht in einer Art und Weise beeinträchtigt wird, von der eine unmittelbare Substitutionswirkung für betroffene Urheber ausgeht. Damit findet keine Beeinträchtigung der normalen Werkverwertung im Sinne des Dreistufentests statt. Gleichzeitig werden die berechtigten Interessen betroffener Urheber nicht ungebührlich verletzt. Im Ergebnis ist die Freistellung des Web Scrapings von urheberrechtlich geschützten Trainingsdaten durch die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG daher mit dem Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL vereinbar.³⁹⁷

395 Ebenso implizit OLG Hamburg, Urt. v. 28.04.2022 - 5 U 48/05, ZUM, 563 (574) – Metall auf Metall; ausdrücklich auch WIPO, Panel Report WT/DS160/R Rn. 6.180; vgl. auch *Runge*, GRUR Int. 2007, 130 (133). Ein entsprechendes Verständnis leitet sich auch in den Schlussanträgen des Generalanwalts Sánchez-Bordona v. 8.12.2016 – ECLI:EU:C:2016:938 Rn. 76-78 ab, der im Rahmen seiner Prüfung der Beeinträchtigung der normalen Werkverwertung eine umfassende Gesamtabwägung vornimmt.

396 Dazu siehe im Detail oben unter 2. Teil B. I.

397 Wie hier auch *Hofmann*, WRP 2024, 11 Rn. 23; *Leistner*, GRUR 2024, 1665 (1670); im Ergebnis auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 (13) Rn. 50 ff.; anderer Ansicht für manche Fallkonstellationen *Vesala*, IIC 2023, 351 (366 f.); außerdem *Jäger*, Artificial Creativity?, S. 371; *Schack*, NJW 2024, 113 Rn. 8, 24, 26; *Spindler*, ZGE 2018, 273 (287 f.) (aus einer de lege ferenda-Perspektive); *Wandtke*, MMR 2017, 367 (368) halten es nur bei Einführung einer gesetzlichen Vergütungspflicht für mit dem Dreistufentest vereinbar; *Schippan*, ZUM 2024, 670 (675) außerdem zweifelnd, wobei der Autor keine nähere Begründung für diese Zweifel liefert; *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 142 ff. außerdem andere Ansicht für das Training generativer KI-Systeme. Aus einer regulatorischen Perspektive de lege ferenda gegen eine Konformität der Frei-

e) Zwischenergebnis zur Anwendbarkeit von § 44b UrhG

Vervielfältigungen, die beim Web Scraping urheberrechtlich geschützter Trainingsdaten angefertigt werden, sind dem Grunde nach von der in § 44b UrhG geregelten Schrankenbestimmung für Text und Data Mining gedeckt.³⁹⁸ Hiergegen spricht auch nicht die Anwendung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL. Die im Rahmen seiner Anwendung vorzunehmende Gesamtabwägung zwischen den verwertungsbezogenen Interessen der Urheber und den Interessen der Werknutzer führt zu einem Überwiegen der Interessen auch kommerzieller Trainingsdatensammler.³⁹⁹

Damit erfüllt das geltende Urheberrecht grundsätzlich das regulatorische Bedürfnis nach einer Freistellung des Web Scrapings von Werken zum Training von KI durch eine urheberrechtliche Schrankenbestimmung.

f) Beschränkung auf rechtmäßig zugängliche Werke

Damit die Freistellungswirkung des § 44b Abs. 2 S. 1 UrhG eintreten kann, müssen die vom Web Scraping betroffenen Werke außerdem rechtmäßig zugänglich sein. Unklar ist aber, was genau unter dieser Voraussetzungen zu verstehen ist und welche Auswirkung eine Weitergabe von Werkzugängen oder eine Bereitstellung von auf Basis von Zugängen angefertigten

stellung mit dem Dreistufentest *Triaille/Mecùs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 102 f.

398 Wie hier auch *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn.11; *Hofmann*, WRP 2024, 11 Rn.15; *Leistner*, GRUR 2024, 1665 (1668); *Kowala*, IIC 2024, 1604 (1609); *Vesala*, IIC 2023, 351 (355 f.); ebenso zustimmend, aber ohne nähere Begründung: *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 1; *Bomhard*, InTeR 2023, 174 (176); *Chiou*, JIPITEC 2019, 398 Rn. 22; *de la Durantaye*, ZUM 2023, 645 (651); *Dregelies*, GRUR 2024, 1484 (1486); *Konertz/Schönhof*, WRP 2024, 289 Rn. 30; *Kögel*, DSRITB 2023, 285 (289); *Maamar*, ZUM 2023, 481 (483); *Raue*, ZUM 2021, 793 (796); *Schwarz/Söbbing*, RD 2023, 415 (417 f.); *Senftleben*, IIC 2023, 1535 (1544); *Käde*, Kreative Maschinen und Urheberrecht, S. 91; jedenfalls ausdrücklich in diesem Sinne auch *Kleinkopf*, Text- und Data-Mining, S. 238; im Ergebnis auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 38 ff., wobei die Entscheidung stark auf den entschiedenen Einzelfall zugeschnitten und (noch) nicht verallgemeinerbar ist; a. A. *Schippan*, ZUM 2024, 670 (675); für das Training generativer KI-Systeme *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 127; *Jäger*, Artificial Creativity?, S. 349.

399 Dazu im Detail siehe unmittelbar oben.

Vervielfältigungsvorlagen für das KI-Training auf die Rechtmäßigkeit der Zugänglichkeit von Werken im Sinne des § 44b Abs. 2 S. 1 UrhG hat.

aa) Begriff der rechtmäßigen Zugänglichkeit

Die in § 44b Abs. 2 S. 1 UrhG vorgesehene Rechtmäßigkeitsanforderung könnte sich zum einen auf die *öffentliche Zugänglichmachung* des vervielfältigten Werkes, zum anderen aber auch auf den *Abruf* des Werkes durch den Trainingsdatensammler beziehen. Da Trainingsdatensammler freien Zugang zum Werk haben, dasselbe aber mitunter nicht rechtmäßig im Internet zugänglich gemacht worden sein muss, hat das Verständnis des Tatbestandsmerkmals insbesondere Auswirkungen auf die Vervielfältigung von rechtswidrig öffentlich zugänglich gemachten Werken. Nach einer auf die Zugänglichmachung des Werkes bezogenen, im Übrigen aber zwischen beiden Standpunkten vermittelnden Ansicht ist ein Werk auch bei einer rechtswidrigen öffentlichen Wiedergabe so lange rechtmäßig zugänglich im Sinne des § 44b Abs. 2 S. 1 UrhG, wie der Inhalt nicht auf einer strukturell urheberrechtsverletzenden Webseite zugänglich gemacht wird.⁴⁰⁰ Hierunter sind solche Webseiten zu verstehen, auf denen im Wesentlichen urheberrechtsverletzende öffentliche Wiedergaben stattfinden. Insbesondere das Web Scraping von Piraterie-Webseiten würde danach nicht in den Anwendungsbereich der Schrankenbestimmung fallen.⁴⁰¹

Bei der Auslegung der Rechtmäßigkeitsanforderung aus § 44b Abs. 2 S. 1 UrhG ist auch die umgesetzte Richtlinienbestimmung, Art. 4 DSM-RL,⁴⁰² im Wege richtlinienkonformer Auslegung mit in den Blick zu nehmen.

(1) Wortlaut und systematischer Bezug zu Art. 3 DSM-RL

Für ein auf die *öffentliche Zugänglichmachung* des vervielfältigten Werkes bezogenes Verständnis spricht, dass der Richtliniengeber im Rahmen der allgemeinen Text und Data Mining-Schrankenbestimmung (Art. 4 DSM-RL) einerseits sowie der Schrankenbestimmung für Text und Data Mining zum Zwecke der wissenschaftlichen Forschung (Art. 3 DSM-RL) anderer-

400 Vesala, IIC 2023, 351 (359).

401 Vesala, IIC 2023, 351 (359).

402 BT-Drs. 19/27426, S. 87.

seits auf unterschiedliche Formulierungen zurückgegriffen hat.⁴⁰³ Während Art. 4 DSM-RL die Vervielfältigung von *rechtmäßig zugänglichen Werken* gestattet, sollen Forschungsorganisationen und Einrichtungen des Kulturerbes im Rahmen von Art. 3 DSM-RL Vervielfältigungen von Werken vornehmen dürfen, zu denen sie *rechtmäßigen Zugang* haben. Im Vergleich ist die in Art. 3 DSM-RL gewählte Tatbestandsformulierung dabei konkreter auf den *Zugang* der Werknutzer bezogen.⁴⁰⁴ Es wäre zudem jedenfalls teleologisch stringenter, die Schrankenbestimmung zum Zwecke der wissenschaftlichen Forschung weiter zu fassen als jene für das allgemeine Text und Data Mining. Denn sie wird dem besonderen wissenschaftlichen Fortschrittsinteresse der Allgemeinheit gerecht.⁴⁰⁵

Allerdings wurden die Rechtmäßigkeitsanforderungen in Art. 3 und 4 DSM-RL zwar in grammatikalischer Hinsicht unterschiedlich formuliert. In beiden Tatbestandsvarianten knüpft der Rechtmäßigkeitsbegriff aber an den *Zugang* zum Werk und nicht dessen öffentliche Wiedergabe an. Damit besteht semantisch kein Unterschied zwischen den Formulierungen. Beide Regelungsgehalte sind vielmehr identisch. Dementsprechend wird der scheinbare terminologische Unterschied zwischen Art. 3 und 4 DSM-RL auch nicht bei der deutschen Umsetzung in §§ 44b, 60d UrhG aufgegriffen. Aus diesen Gründen spricht die grammatikalische Divergenz zwischen Art. 3 und 4 DSM-RL nicht gegen das zugangsbezogene Verständnis der Rechtmäßigkeitsanforderung.⁴⁰⁶

Gegen ein enges Verständnis der Schrankenbestimmung für Text und Data Mining, nach der es auf die Rechtmäßigkeit der Zugänglichmachung des vervielfältigten Werkes ankäme, spricht außerdem, dass sich Art. 4 DSM-RL, der durch § 44b UrhG in nationales Recht umgesetzt wurde, ausdrücklich auf die rechtmäßige *Zugänglichkeit* des betreffenden Werkes, nicht auf die Rechtmäßigkeit der Quelle im Sinne der Rechtmäßigkeit der *Zugänglichmachung* bezieht. Hieraus folgt, dass es bereits nach einer wortlautgetreuen Anwendung der Norm nicht auf die urheberrechtliche Beurteilung der öffentlichen Wiedergabe des Werkes, sondern auf den Zugangsakt des Datensammlers ankommen muss.⁴⁰⁷

403 Margoni, Kluwer Copyright Blog v. 22.12.2023.

404 Margoni, Kluwer Copyright Blog v. 22.12.2023.

405 Margoni, Kluwer Copyright Blog v. 22.12.2023.

406 A. A. allerdings Margoni, Kluwer Copyright Blog v. 22.12.2023.

407 Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 8; Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 19 (anders noch in der Vorauflage); Hofmann, WRP 2024,

(2) Regelungszweck

Dafür, dass ein rechtmäßiger Zugang zum vervielfältigten Werk im Rahmen des § 44b Abs. 2 S. 1 UrhG ausreichend ist und es demnach nicht auf die Legitimität der Zugänglichmachung ankommt, spricht auch, dass ErwGr. 14 S. 4 DSM-RL ausdrücklich eine freie Zugänglichkeit der Werke im Internet als ausreichend erachtet.

Zweck der Schrankenbestimmung für Text und Data Mining ist im Weiteren, eine möglichst klare und rechtssichere Grundlage für das Sammeln großer Datenmengen bereitzustellen.⁴⁰⁸ Das unbestimmte Wesentlichkeitskriterium, anhand dessen teilweise differenziert werden soll, ob das betreffende Werk rechtswidrig auf einer strukturell urheberrechtsverletzenden Webseite wiedergegeben wurde und daher kein rechtmäßiger Zugang zum Inhalt im Sinne des § 44b Abs. 2 S. 1 UrhG bestand, wird den Anforderungen an einen rechtssicheren Schrankentatbestand für Datensammler jedoch nicht gerecht. Da das Web Scraping typischerweise automatisiert stattfindet, sind Trainingsdatensammler praktisch außerdem nicht in der Lage, die Rechtmäßigkeit der öffentlichen Wiedergabe einzelner Werke oder den Inhalt und die Zielrichtung vom Webcrawler erfasster Internetseiten zu beurteilen.⁴⁰⁹ Eine einschränkende Auslegung der Rechtmäßigkeitsvoraussetzung aus § 44b Abs. 2 S. 1 UrhG würde die Schrankenbestimmung damit praktisch weitestgehend unbrauchbar machen.⁴¹⁰ Das widerspräche dem Sinn und Zweck des § 44b UrhG.

(3) Ergebnis der Auslegung

Eine Einschränkung des Schrankentatbestands durch ein auf die öffentliche Zugänglichmachung des genutzten Werkes bezogenes Verständnis der Rechtmäßigkeitsanforderung ist aus den vorgenannten Gründen abzuleh-

11 Rn. 16; *Konertz/Schönhof*, WRP 2024, 289 Rn. 31; *Raue*, GRUR 2017, 11 (14); *Schack*, GRUR 2021, 904 (907); *v. Welser*, GRUR-Prax 2023, 516 Rn. 32; *Schack*, NJW 2024, 113 Rn. 11. Einschränkend lediglich *Dworschak*, in: *Fromm/Nordemann*, Urheberrecht, UrhG § 44b Rn. 12; *Margoni*, Kluwer Copyright Blog v. 22.12.2023.

408 ErwGr. 18 DSM-RL. So im vorliegenden Kontext auch *Bomhard*, InTeR 2023, 174 (176).

409 Im Grundsatz sieht das auch *Vesala* so, wobei der Autor allerdings annimmt, dass die vorgeschlagene kursorische Überprüfung der anvisierten Internetquellen zumutbar sei, siehe *Vesala*, IIC 2023, 351 (359); wie hier auch *Hofmann*, ZUM 2024, 166 (171).

410 So auch *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (26).

nen. Unabhängig davon, ob der Veröffentlichung der Inhalte vom Rechteinhaber zugestimmt wurde, ist damit ausreichend, dass der Datensammler rechtmäßigen Zugang zum Werk hat.⁴¹¹

(4) Beispiele für rechtmäßige Zugangskonstellationen

Rechtmäßiger Zugang zu Werken besteht insbesondere dann, wenn die Werke frei im Internet verfügbar sind, oder wenn Zugang auf Basis vertraglicher Vereinbarungen besteht.⁴¹² Die Anforderung wird auch dann gewahrt, wenn ein Nutzeraccount auf einer Webseite angelegt wurde und die vervielfältigten Werke nach ordnungsgemäßem Login frei zugänglich sind. Die Rechtmäßigkeit des Zugangs beurteilt sich insgesamt aus Perspektive der Datensammler. Letztendlich stellt die Voraussetzung klar, dass § 44b UrhG für Datensammler keinen Zugangsanspruch begründet.⁴¹³

(5) Technische Schutzmaßnahmen sowie Geoblocking-Technologien und ihre Auswirkungen auf die rechtmäßige Zugänglichkeit

Aus dem zugangsbezogenen Verständnis der Rechtmäßigkeitsanforderung folgt weiterhin, dass die Freistellung von Vervielfältigungshandlungen bei der Umgehung technischer Schutzmaßnahmen davon abhängt, ob die ergriffenen Schutzmaßnahmen erst die Vervielfältigung oder aber bereits den Werkzugang verhindern. Echte Kopierschutzmechanismen, also Maßnahmen zur Verhinderung von Vervielfältigungshandlungen, führen nicht zu

411 Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 8; Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 19 (anders noch in der Voraufgabe); Hofmann, ZUM 2024, 166 (171); Raue, GRUR 2017, II (14); Schack, GRUR 2021, 904 (907); Griffiths/Synodinou/Xalabarder, GRUR Int. 2023, 22 (26); v. Welser, GRUR-Prax 2023, 516 Rn. 32. Einschränkend lediglich Dworschak, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 12; Margoni, Kluwer Copyright Blog v. 22.12.2023.

412 ErwGr. 14 S. 4, 18 UAbs. 2 DSM-RL. Schippan spricht letzterer Alternative die praktische Bedeutung ab. Denn im Rahmen einer vertraglichen Zugangsgewährung würden gleichzeitig stets vertragliche Vorbehaltserklärungen (vgl. § 44b Abs. 3 UrhG) abgegeben. Damit käme es im Ergebnis praktisch immer zu einer Beendigung der Schrankenwirkung. Der rechtmäßige Zugang entfaltet in diesen Fällen daher keine Wirkung, siehe Schippan, ZUM 2024, 670 (675). Dies kann sicherlich, muss aber nicht immer der Fall sein.

413 Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 7; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 8.

einer Zugangshinderung im engeren Sinne und dementsprechend nicht zu einem Ausbleiben der Freistellungswirkung des § 44b UrhG.⁴¹⁴ Bei ihrer Umgehung können Unterlassungsansprüche aufgrund einer Verletzung des § 95a UrhG geltend gemacht werden, wobei § 95b UrhG zu beachten ist. Mechanismen wie beispielsweise Geo-Blocking-Maßnahmen haben hingegen zur Folge, dass das Werk für einen Trainingsdatensammler bereits nicht rechtmäßig zugänglich ist. Kommt es zu einer Umgehung der Geoblocking-Maßnahmen, sind die Vervielfältigungshandlungen daher nicht von § 44b Abs. 2 UrhG privilegiert. Dass eine Differenzierung zwischen zugangs- und vervielfältigungshindernden Schutzmechanismen dem Urheberrecht bereits bekannt ist, zeigt sich auch in § 95a Abs. 2 S. 2 UrhG. Hier werden beide Arten von Schutzmaßnahmen ausdrücklich aufgeführt.

(6) Abgleich mit den identifizierten Regelungsbedürfnissen und Schlussfolgerungen

Insgesamt kann das Web Scraping von Werken zur Verwendung als Trainingsdaten damit nur erfolgen, wenn die Werke für Datensammler frei und damit im Sinne des § 44b Abs. 2 UrhG rechtmäßig zugänglich sind. Zwar führt die Beschränkung der Freistellung auf rechtmäßig zugängliche Werke dazu, dass die Schrankenbestimmung im Hinblick auf ihre Freistellungsreichweite hinter den identifizierten Freistellungsbedürfnissen zurückbleibt. Allerdings ist die Rechtmäßigkeitsvoraussetzung des § 44b Abs. 2 UrhG weit. Nicht umfasst ist vor allem das parasitäre Eindringen in geschlossene Webseitenstrukturen, beispielsweise durch das unerlaubte Überwinden einer Login-Beschränkung. Demgegenüber hat ein Trainingsdatensammler selbst dann rechtmäßigen Zugang zu Werken, wenn sie hinter einer Bezahlschranke zugänglich gemacht werden und der Datensammler diese durch den Erwerb eines Zugangs ordnungsgemäß überwindet. Auch dann können seine Web Scraping-Aktivitäten demzufolge unter die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG gefasst werden.⁴¹⁵ Zugang zu den für das KI-Training benötigten Inhalten kann damit in der Regel ohne wesentliche Hindernisse erlangt werden. Infolgedessen ist die Rechtmäßigkeitsvoraussetzung aus einer praktischen Perspektive

414 Zu pauschal hingegen *Hofmann*, ZUM 2024, 166 (171); *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 137, die davon ausgehen, dass sämtliche technischen Schutzmaßnahmen die rechtmäßige Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG beenden.

415 Das verkennt hingegen *Schack*, NJW 2024, 113 Rn. 11.

keine wesentliche Hürde für das Web Scraping von urheberrechtlich geschützten Trainingsdaten. Trotz dessen, dass § 44b Abs. 2 S. 1 UrhG die Freistellungswirkung auf die Vervielfältigung rechtmäßig zugänglicher Werke beschränkt, wird die Schrankenbestimmung den identifizierten Freistellungsvoraussetzungen daher im Wesentlichen gerecht.

bb) Weitergabe des Zugangs oder Bereitstellung bereits angefertigter Vervielfältigungsvorlagen der Werke an Dritte

Auch die Entwicklung und Anwendung von KI kann arbeitsteilig von staten gehen. Entwickler sogenannter Basismodelle greifen für das Training ihrer KI beispielsweise auch auf von intermediären Trainingsdatensammlern zusammengestellte Trainingsdatensätze zurück.⁴¹⁶ Vor der Auslieferung können die Basismodelle weiterhin in einem Fine-Tuning mit anwendungsspezifischen Daten finalisiert werden. Diese stellt der nachgelagerte, dritte Abnehmer selbst dem Entwickler des Basismodells zur Verfügung.⁴¹⁷ Vor dem Hintergrund solcher Fallkonstellationen ist fraglich, inwieweit Beteiligte wie beispielsweise KI-Entwickler eines Basismodells rechtmäßigen Zugang zu Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben, wenn ihnen die Werke oder Zugänge zu diesen Werken von einem anderen Zugangsinhaber wie beispielsweise einem vorgelagerten intermediären Trainingsdatensammler oder einem nachgelagerten Abnehmer eines Basismodells für das Fine-Tuning vermittelt worden sind, ohne dass sie die Zugangsvoraussetzungen in ihrer eigenen Person erfüllen. Hierzu gehört beispielsweise, den Zugang zu einer Werkdatenbank erworben zu haben.

Allgemein kann dabei zwischen zwei Formen der Weitergabe unterschieden werden. Möglich ist einerseits, dass datenverarbeitenden KI-Entwicklern oder intermediären Trainingsdatensammlern, die selbst KI-Training

416 Beispielsweise das *The Pile*-Dataset, ein 800 GB großer, öffentlich verfügbarer Datensatz für das Training von Sprachverarbeitungsmodellen, siehe *Gao u. a.*, arXiv:2101.00027 2020. Das *The Pile*-Dataset wurde unter anderen von *Meta* (Mutterkonzern von *Facebook*) zum Training des Textgeneratoren-Basismodells *LLaMA* genutzt, siehe *Touvron u. a.*, arXiv:2302.13971 2023, S. 2.

417 So beispielsweise der Fall beim Textgenerator-Basismodell *GPT* von *OpenAI*, auf dem auch der bekannte Chatbot *ChatGPT* basiert, siehe <https://platform.openai.com/docs/guides/fine-tuning> (zuletzt abgerufen am 16.2.2024), oder *LLaMA 2*, bei dem mit Hilfe der *Microsoft Azure*-Umgebung Fine-Tuning durchgeführt werden kann, siehe <https://learn.microsoft.com/en-us/azure/ai-studio/how-to/fine-tune-model-llama> (zuletzt abgerufen am 16.2.2024).

durchführen oder eigenhändig Trainingsdatensätze erstellen und damit die erlaubnisbedürftigen Vervielfältigungshandlungen im Sinne des § 16 UrhG ausführen (im Folgenden „Datenverarbeiter“ genannt), nur der Zugang zu im Internet abrufbaren Werken vermittelt wird. Das ist beispielsweise durch eine Weitergabe von Login-Daten zu einer geschützten Datenbank möglich. Der Datenverarbeiter fertigt dann auf Basis dieses Zugangs eigenhändig Vervielfältigungen für das Training von KI an. Als Vervielfältigungsvorlage dienen dabei unmittelbar die Wiedergaben der Werke in der zugangsgeschützten Quelle. Zum anderen können dem Datenverarbeiter aber auch bereits vom originären Zugangsinhaber wie beispielsweise einem nachgelagerten Abnehmer eines Basismodells oder einem vorgelagerten intermediären Trainingsdatensammler (im Folgenden „Zugangsinhaber“ genannt) rechtmäßig vervielfältigte Werke als Vervielfältigungsvorlagen zur Verfügung gestellt werden. Dann fertigt der Datenverarbeiter neue Vervielfältigungsstücke der zuvor vom Zugangsinhaber abgespeicherten Werke an. Als Vervielfältigungsvorlage dienen in diesem Fall daher nicht die originären Werkwiedergaben im Internet, sondern sekundäre Vervielfältigungen aus der Sphäre des Zugangsinhabers. Dies ist insbesondere bei Kunden intermediärer Trainingsdatensammler der Fall, sofern die intermediären Trainingsdatensammler nicht nur Linklisten bereitstellen, sondern in einem ersten Schritt selbst die Daten abspeichern und vorverarbeiten, bevor die kuratierten und vorverarbeiteten Datensätze KI-Entwicklern angeboten werden. Klärungsbedürftig ist in beiden Fallkonstellationen, ob auch der nachgelagerte Datenverarbeiter vom rechtmäßigen Zugang des Zugangsinhabers im Sinne des § 44b Abs. 2 S. 1 UrhG profitiert.

Differenziert werden kann dabei, ob der Datenverarbeiter im eigenen Interesse handelt oder als Dienstleister zugunsten des Zugangsinhabers tätig wird. Ersteres ist insbesondere dann der Fall, wenn der Datenverarbeiter die generierten Inhalte in eigener Verantwortung zu Trainingsdatensätzen aufbereitet. Das umfasst auch die Tätigkeit intermediärer Trainingsdatensammler. Zugunsten des Zugangsinhabers wird hingegen tätig, wer in seinem Interesse KI-Training durchführt, beispielsweise, um in dessen Auftrag anwendungsfallsspezifische KI-Systeme zu entwickeln.

(1) Zugangsvermittlung im Interesse des Zugangsinhabers

Fraglich ist, ob ein im Interesse des Zugangsinhabers tätig werdender Datenverarbeiter Vervielfältigungen von für ihn rechtmäßig zugänglichen Werken im Sinne des § 44b Abs. 2 S. 1 UrhG anfertigt, wenn ihm der Zu-

gang zu den vervielfältigten Werken (beispielsweise Login-Daten zu einer Datenbank) vom Zugangsinhaber vermittelt worden ist.

(i) Datenverarbeiter als Vervielfältigungsdienstleister

Vorgelagert klärungsbedürftig ist zunächst aber, ob der Datenverarbeiter in der untersuchten Fallkonstellation überhaupt *eigene* Vervielfältigungshandlungen im Sinne von § 16 UrhG vornimmt. Nur in diesem Fall kommt es auf seinen rechtmäßigen Zugang zu den Werken nach der Zugangsweitergabe an. Eigene Vervielfältigungshandlungen liegen hingegen nicht vor, wenn er als Vervielfältigungsdienstleister⁴¹⁸ für den Zugangsinhaber tätig wird. Dann werden diesem die Vervielfältigungen unmittelbar selbst zugerechnet. Vervielfältigungsdienstleister sind dabei von Hilfspersonen im Sinne des § 53 Abs. 1 S. 2 UrhG oder im Kontext der Wissenschaftsschranken aus §§ 60c, d UrhG zu unterscheiden. Diese nehmen eine eigene Vervielfältigungshandlung vor, welche jedoch von den Schrankenbestimmungen mitprivilegiert wird.⁴¹⁹

(a) Voraussetzungen

Damit der Datenverarbeiter lediglich als Vervielfältigungsdienstleister auftritt, darf er nicht aktiv in die Vervielfältigungshandlung eingreifen.⁴²⁰ Ein aktiver Eingriff findet jedenfalls auch dann statt, wenn die angefertigten Vervielfältigungen nicht nur mit dem Auftraggeber, sondern auch mit einer unbestimmten Zahl Dritter geteilt werden.⁴²¹ Weiterhin muss die Tätigkeit des Datenverarbeiters auf die Organisation der Vervielfältigungen als Hilfsdienstleistung beschränkt bleiben. Hierfür ist zum einen notwendig, dass

418 Zum Begriff siehe EuGH, Urt. v. 29.11.2017 - C-265/16, GRUR 2018, 68 Rn. 35 – VCAST; EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 30 – Ocilion.

419 Zur Privatkopieschranke deutlich in BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 25 – Internet-Radiorecorder; Renner, in: Fromm/Nordemann, Urheberrecht, UrhG § 53 Rn. 23; vgl. auch Loewenheim/Stieper, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 53 Rn. 27, 30; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 53 Rn. 14; zu den Wissenschaftsschranken außerdem vgl. Stieper, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 60c Rn. 18; Stieper, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 60d Rn. 14.

420 Vgl. EuGH, Urt. v. 29.11.2017 - C-265/16, GRUR 2018, 68 Rn. 54 – VCAST.

421 Vgl. EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 45 f. – Ocilion.

die Vervielfältigungen auf die Initiative des Auftraggebers zurückgehen.⁴²² Der Datenverarbeiter darf weiterhin nicht den Zugang zu den Werken vermitteln, die Gegenstand der Vervielfältigungshandlungen sind.⁴²³ Darüber hinaus darf die Bereitstellung der Vervielfältigungen selbst nicht Hauptgegenstand des Angebots des Datenverarbeiters sein. Er muss sich vielmehr darauf beschränken, *Mittel* für Vervielfältigungshandlungen zur Verfügung zu stellen.⁴²⁴

(β) Aktiver Eingriff des Datenverarbeiters im vorliegenden Fall

In der untersuchten Fallgruppe prozessiert der Datenverarbeiter die Werke im Interesse eines auftraggebenden Zugangsinhabers. Demzufolge ist nicht davon auszugehen, dass die angefertigten Vervielfältigungsstücke mit einer unbestimmten Zahl anderer Interessenten geteilt werden. Außerdem verschafft der Zugangsinhaber und nicht der Datenverarbeiter den Werkzugang.

Fraglich ist allerdings, ob die Vervielfältigungen tatsächlich auf die Initiative des Zugangsinhabers zurückgehen. Dafür spricht zwar, dass er die Vervielfältigungshandlungen auf die vom Zugang umfassten Werke konkretisiert hat. Allerdings steht zwischen der Zugangsverschaffung und der Vervielfältigung noch ein Willensentschluss des Datenverarbeiters. Denn er muss den Zugang erst aktiv für die Anfertigung der Vervielfältigungsstücke nutzen. Der Zugangsverschaffung folgen, abhängig von der konkreten Fallgestaltung, aber jedenfalls nicht unbedingt automatisierte Vervielfältigungshandlungen des Datenverarbeiters nach. Demzufolge löst der Zugangsinhaber den Vervielfältigungsvorgang nicht unmittelbar selbst aus. Das spricht dafür, dass der Datenverarbeiter aktiv in die Vervielfältigungshandlungen eingreift.⁴²⁵

Darüber hinaus führt der Datenverarbeiter die Organisation und Durchführung der Vervielfältigungshandlungen durch, um auf Basis der vervielfältigten Werke KI-Training zu betreiben. Dafür müssen dem Zugangsinhaber die Vervielfältigungen nicht überlassen werden. Der Zugangsinhaber hat vielmehr nur ein Interesse am Erhalt oder der Anwendung eines

422 Vgl. EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 50 - Ocilion.

423 Vgl. EuGH, Urt. v. 29.11.2017 - C-265/16, GRUR 2018, 68 Rn. 37 f. – VCAST.

424 Vgl. EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 40, 44 – Ocilion.

425 Ein vergleichbares Kriterium zur Bestimmung der Herstellereigenschaft einer Vervielfältigung leitet der BGH in seiner Rechtsprechung auch aus der nationalen Urheberrechtsdogmatik ab, siehe BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 33 – Internet-Radiorecorder.

trainierten KI-Systems. Ziel der Tätigkeit des Datenverarbeiters ist daher nicht die Bereitstellung technischer Hilfsmittel für Vervielfältigungen des Zugangsinhabers, sondern anstelle des Zugangsinhabers die Vervielfältigungsleistung und das Training von KI durchzuführen. Die Vervielfältigungen und ein sich anschließendes KI-Training sind demzufolge Hauptgegenstand des Angebots des Datenverarbeiters. Insgesamt folgt daraus, dass der Datenverarbeiter im untersuchten Fall aktiv in die Vervielfältigungshandlungen eingreift.

(γ) Schlussfolgerungen

Aufgrund des aktiven Eingriffs in die Vervielfältigungshandlungen ist der Datenverarbeiter im vorliegenden Fall kein Vervielfältigungsdienstleister des Zugangsinhabers. Die auf Basis des vermittelten Zugangs vorgenommenen Vervielfältigungshandlungen zum Training der KI werden direkt dem Datenverarbeiter zugerechnet.

(ii) Eigener rechtmäßiger Zugang des Datenverarbeiters

Aufgrund der eigenen Vervielfältigungshandlungen des im Interesse des Zugangsinhabers handelnden Datenverarbeiters im Sinne des § 16 UrhG kommt es für ein Eingreifen der Schrankenbestimmung aus § 44b UrhG zu seinen Gunsten darauf an, ob er aufgrund der Zugangsvermittlung rechtmäßigen Zugang zu den betreffenden Werken im Sinne des § 44b Abs. 2 S. 1 UrhG gehabt hat. Inwieweit der Tatbestand ein solches Verständnis zulässt, hängt auch von der Auslegung des mit § 44b UrhG umgesetzten Art. 4 DSM-RL ab.

(α) Strenge Wortlautanwendung

Gegen einen rechtmäßigen Zugang des Datenverarbeiters zu den Werken spricht zunächst, dass er die vom Rechtsinhaber oder einem dritten Zugangsanbieter aufgestellten Zugangsvoraussetzungen nicht eigenhändig erfüllt. Insbesondere hat er, beispielsweise durch den Abschluss eines Abonnements, nicht selbst den Zugang zu einer Werkdatenbank erworben.

(β) Umkehrschluss aus § 53 Abs. 1 S. 2 UrhG

Gegen einen rechtmäßigen Zugang des Datenverarbeiters im vorliegenden Fall spricht weiterhin, dass § 53 Abs. 1 S. 2 UrhG den Einsatz von Hilfspersonen ausdrücklich mitumfasst. Da eine dem entsprechende Privilegierung von Hilfspersonen in § 44b UrhG nicht vorgesehen ist, könnte im Umkehrschluss hieraus gefolgert werden, dass § 44b UrhG den Einsatz von Hilfspersonen für das Text und Data Mining nicht ermöglichen wollte.

Voraussetzung eines Umkehrschlusses ist jedoch, dass das Gesetz die bezeichnete Rechtsfolge nur auf den konkret geregelten Fall beschränken wollte. Dies bestimmt sich nach dem Willen des Gesetzgebers und insbesondere dem Telos der Regelung.⁴²⁶ In Bezug auf die vorliegende Bestimmung dürfte der Einsatz von Hilfspersonen damit *allein* im Rahmen der Schrankenbestimmung für Privatkopien für freistellungswert gehalten werden.

Grund für die Erweiterung der Schrankenwirkung des § 53 UrhG auf Hilfspersonen war, dass vom Rechtsinhaber nicht effektiv kontrolliert werden kann, wer im privaten Umfeld tatsächlich die Vervielfältigungshandlung vornimmt.⁴²⁷ Ein solches Kontrolldefizit tritt allerdings immer bei Vervielfältigungshandlungen im nicht-öffentlichen Raum auf. Das kann nicht nur bei der Herstellung von Privatkopien, sondern beispielsweise auch bei Vervielfältigungshandlungen zu Lehr- oder Archivierungszwecken der Fall sein.⁴²⁸ Das Kontrolldefizit tritt bei der Anfertigung von Privatkopien lediglich besonders zutage. Aus diesem Grund ist die Freistellung von Vervielfältigungen durch Hilfspersonen nur im Rahmen von § 53 UrhG ausdrücklich normiert worden. Hieraus lässt sich aber nicht schlussfolgern, dass der Rückgriff auf Hilfspersonen zur Durchführung von durch Schrankenbestimmungen privilegierten Werknutzungen ausschließlich für die Anfertigung von Privatkopien gestattet sein soll. Ein Umkehrschluss aus § 53 Abs. 1 S. 2 UrhG ist daher nicht möglich.

426 Larenz/Canaris, Methodenlehre der Rechtswissenschaft, S. 209.

427 BT-Drs. 15/38, S. 20.

428 Für eine Freistellung der Hilfsdienstleistung hier vgl. Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 60a Rn. 9; Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60c Rn. 18; Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60d Rn. 14.

(γ) Ableitung eines allgemeinen Rechtsgedankens aus § 53 Abs. 1 S. 2 UrhG

Im Gegenteil könnte der in § 53 Abs. 1 S. 2 UrhG zum Ausdruck kommende Rechtsgedanke der Mitprivilegierung von Hilfspersonen als allgemeingültig angesehen werden. Damit wäre er auch auf § 44b UrhG übertragbar. Methodisch ist hierfür kein mehrfaches In-Erscheinung-Treten des Rechtsgedankens im UrhG notwendig.⁴²⁹ Es reicht aus, dass der einer Regelung zugrundeliegende Normsinn auch auf einen weiteren Kreis von Fällen übertragen werden kann.⁴³⁰ Wie bereits dargelegt wurde,⁴³¹ ist dies für das Kontrolldefizit gegenüber dem Einsatz von Hilfspersonen bei nicht-öffentlichen Vervielfältigungshandlungen der Fall.

Der Normsinn könnte darüber hinaus auch auf Vervielfältigungen von Werken für das Training von KI, die mit Hilfe von Webcrawlern angefertigt worden sind, zutreffen. Denn Maßnahmen wie IP-Adressbeschränkungen, die den Zugang zu einer Werkdatenbank auf ausgewählte Personen beschränken, können selbst mit eingeschränkten technischen Fähigkeiten umgangen werden. Damit lässt sich faktisch nur eine sehr eingeschränkte personelle Zugangs- und demzufolge auch Vervielfältigungskontrolle umsetzen. Wie bei der Schrankenbestimmung für Privatkopien spricht dieses Kontrolldefizit vor dem Einsatz von Hilfspersonen auch beim Training von KI dafür, einen solchen urheberrechtlich mit zu privilegieren. Der in § 53 Abs. 1 S. 2 UrhG zum Ausdruck kommende Normsinn kann damit auch auf die Schrankenbestimmung für Text und Data Mining übertragen werden.

Insgesamt sprechen die vorgenannten teleologischen Gründe dafür, aus § 53 Abs. 1 S. 2 UrhG einen allgemeinen Rechtsgedanken abzuleiten, der insbesondere auch auf die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG anwendbar ist. Folge dessen ist, dass der Einsatz von Hilfspersonen für Vervielfältigungen von Werken als Trainingsdaten im Interesse des Zugangsinhabers von § 44b UrhG umfasst ist. Dies kann bei der Anwendung des Schrankentatbestands erreicht werden, indem auch Datenverarbeiter rechtmäßigen Zugang zu Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben, denen der Werkzugang von einem Zugangsinhaber vermittelt worden ist, damit sie in seinem Interesse die Vervielfältigungen anfertigen.

429 Larenz/Canaris, Methodenlehre der Rechtswissenschaft, S. 207.

430 Larenz/Canaris, Methodenlehre der Rechtswissenschaft, S. 207.

431 Dazu siehe unmittelbar oben.

(δ) Teleologische Erwägungen anhand des umgesetzten Art. 4 DSM-RL

Gegen das aufgezeigte Tatbestandsverständnis könnte aus einer unionsrechtlichen jedoch Perspektive sprechen, dass ErwGr. 28 DSM-RL den Kulturerbe-Einrichtungen einen Rückgriff auf Hilfspersonen für die Zwecke der Sammlungserhaltung gestattet. Da sich in der Richtlinie keine entsprechende Aussage in Bezug auf das allgemeine Text und Data Mining findet, ließe sich hieraus ein Umkehrschluss ableiten, nach dem der Einsatz dritter Datenverarbeiter als Hilfspersonen im Interesse des Zugangsinhabers nicht vom Anwendungsbereich des § 44b UrhG umfasst wäre. Ob ein entsprechender Umkehrschluss gezogen werden kann, ist aber auch im Rahmen des ErwGr. 28 DSM-RL davon abhängig, ob die Richtlinie tatsächlich regeln wollte, dass der Einsatz dritter Hilfspersonen *ausschließlich* bei Vervielfältigungshandlungen zum Zwecke der Erhaltung von Sammlungen der Kulturerbe-Einrichtungen freigestellt sein sollte. Dagegen kann allerdings angeführt werden, dass ErwGr. 28 DSM-RL gemäß seines Wortlauts auf der Annahme beruht, dass Einrichtungen des Kulturerbes regelmäßig nicht über die technischen Mittel und Fähigkeiten verfügen, die für die umfangreichen Vervielfältigungshandlungen zur Sammlungserhaltung notwendig sind. Dasselbe trifft unter Umständen aber auch auf Zugangsinhaber zu, die mangels technischer Fähigkeiten und Ressourcen für umfangreiche Vervielfältigungshandlungen sowie eine automatisierte Analyse de facto nicht von der Schrankenbestimmung aus § 44b UrhG profitieren können. Grund für die besondere Herausstellung in ErwGr. 28 DSM-RL muss vielmehr gewesen sein, dass das Fähigkeiten- und Ressourcendefizit bei Einrichtungen des Kulturerbes besonders deutlich zutage tritt. Wie in Bezug auf § 53 Abs. 1 S. 2 UrhG ergeben sich demzufolge auch aus ErwGr. 28 DSM-RL keine Hinweise darauf, dass der dem Erwägungsgrund zugrundeliegende Rechtsgedanke allein für kulturelle Erhaltungszwecke Anwendung finden sollte. Vielmehr kann der in ErwGr. 28 DSM-RL zum Ausdruck kommende Rechtsgedanke auch hier verallgemeinert und ebenfalls auf die Schrankenbestimmung für Text und Data Mining und den Einsatz von Hilfspersonen für das Training von KI im Allgemeinen übertragen werden.

Für das Tatbestandsverständnis, nach dem die Datenverarbeitung durch Hilfspersonen im Wege eines weiten Verständnisses des Begriffs der rechtmäßigen Zugänglichkeit ermöglicht wird, spricht weiterhin, dass das von der Freistellung durch den in § 44b UrhG umgesetzten Art. 4 DSM-RL (unter anderem) betroffene Vervielfältigungsrecht der Urheber nach Art. 2 InfoSoc-RL dem Zweck dient, dem Urheber die Verwertung seines Wer-

kes zu sichern.⁴³² Umgekehrt schützt die Tatbestandsvoraussetzung der rechtmäßigen Zugänglichkeit den Urheber beziehungsweise Rechtsinhaber vor einer übergebührlichen Beeinträchtigung seiner Verwertungsinteressen. Hieraus folgt, dass eine Beendigung der Schrankenwirkung des Art. 4 DSM-RL zulasten eines dritten Datenverarbeiters bei der Weitergabe eines Zugangs durch den Schrankenprivilegierten Zugangsinhaber nur dann notwendig ist, wenn es durch die Weitergabe zu einer intensivierten Beeinträchtigung der wirtschaftlichen Verwertungsmöglichkeiten der Urheber kommt. Wird der Datenverarbeiter im Interesse des Zugangsinhabers tätig, heißt das allerdings, dass er die Vervielfältigungen der Werke *anstelle* des Zugangsinhabers anfertigt. Die Werke werden auf Grundlage des zur Verfügung stehenden Werkzugangs also nur einmal für das Text und Data Mining genutzt. Daraus folgt, dass die Zugangweitergabe an einen Datenverarbeiter in den hier betreffenden Fällen nicht zu einer intensivierten Beeinträchtigung der Verwertungsinteressen betroffener Urheber führt. Insgesamt spricht damit auch eine teleologische Auslegung des Art. 4 DSM-RL vor dem Hintergrund der verwertungsrechtlichen Funktion des Vervielfältigungsrechts dafür, dass auch Datenverarbeiter rechtmäßigen Zugang zu Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben, denen der Zugang als Hilfsperson des Zugangsinhabers vermittelt worden ist.

(€) Auslegungsergebnis

Insgesamt sprechen damit vor allem teleologische Gründe dafür, den Rechtmäßigkeitsbegriff so auszulegen, dass auch solche Datenverarbeiter rechtmäßigen Zugang zu Werken gemäß § 44b Abs. 2 S. 1 UrhG haben, denen der Zugang von einem Zugangsinhaber vermittelt worden ist, damit sie im Interesse des Zugangsinhabers Vervielfältigungen der Werke anfertigen. In diesen Fällen ist also nicht notwendig, dass die Werknutzer die Zugangsvoraussetzung eigenhändig erfüllen und sich beispielsweise den Zugriff auf eine Werkdatenbank durch Abschluss eines Nutzungsvertrags selbst verschafft haben.⁴³³

432 Nordemann, in: Fromm/Nordemann, Urheberrecht, Einl. zum UrhG Rn. 19; v. Ungern-Sternberg, GRUR 2012, 576 (579 f.); v. Ungern-Sternberg, GRUR 2012, 1198 (1200); Haberstumpf, ZUM 2020, 810 (816); Grisse/Kaiser, ZUM 2021, 401 (404). Außerdem ErwGr. 10 InfoSoc-RL.

433 Anders wohl Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 8.

(iii) Zugangsweitergabe trotz vertraglicher Weitergabebeschränkung

Fraglich ist allerdings, ob der Zugang des Datenverarbeiters zu den Werken auch dann im Sinne von § 44b Abs. 2 S. 1 UrhG rechtmäßig ist, wenn dieser zwar von einem Zugangsinhaber beispielsweise durch Weitergabe von Login-Daten verschafft worden ist, die Weitergabe des Zugangs dem Zugangsinhaber allerdings vertraglich untersagt worden ist.

Dagegen spricht, dass auch die vertragliche Beziehung zwischen dem zugangsgewährenden Rechtsinhaber oder sonstigen Zugangsanbieter und den Werknutzern wie beispielsweise Trainingsdatensammlern zur Bestimmung des rechtmäßigen Zugangs im Rahmen des § 44b Abs. 2 S. 1 UrhG herangezogen wird.⁴³⁴ Demzufolge wäre es konsequent, die vertragliche Ausgestaltung der Zugangsgewährung auch auf einer Sekundärebene zwischen Zugangsinhaber und Datenverarbeiter zu berücksichtigen.

Allerdings vervielfältigen die Datenverarbeiter in der vorliegenden Sachverhaltskonstellation die Werke im Interesse, also anstelle des Zugangsinhabers. Durch die Weitergabe der Zugänge kommt es demzufolge nicht zu einer Intensivierung der wirtschaftlichen Beeinträchtigung der betroffenen Rechtsinhaber beziehungsweise Anbieter von Werkzugängen wie beispielsweise Datenbankbetreibern. Dürften Zugangsinhaber aus einer urheberrechtlichen Perspektive keine Hilfspersonen für ihre Vervielfältigungen einschalten, würde dies weiterhin einen Wertungswiderspruch zu den §§ 53, 60c, 60d UrhG hervorrufen. Im Rahmen dieser Bestimmungen ist der Einsatz von Hilfspersonen zwar ebenfalls an Voraussetzungen gekoppelt, kann durch Rechtsinhaber aber nicht insgesamt untersagt werden.⁴³⁵

Aus Perspektive der Datenverarbeiter bestehen zudem strukturelle Ähnlichkeiten zwischen vertraglichen Beschränkungen der Zugangsweitergabe und der Rechtmäßigkeit der öffentlichen Zugänglichmachung der Werke. Weil letztere für Datenverarbeiter (und Zugangsinhaber) keine Bedeutung entfaltet, sondern allein ihr eigener Zugang zum Werk entscheidend ist, kann sich infolgedessen auch eine vertragliche Beschränkung der Zugangsweitergabe im dem Datenverarbeiter vorgelagerten Verhältnis nicht zu seinen Lasten auswirken. Im Gegenteil dient das Abstellen auf den eigenen Zugang und nicht die Rechtmäßigkeit der Zugänglichmachung im Rahmen des § 44b Abs. 2 S. 1 UrhG der Gewährleistung von Rechtssicherheit

434 Dazu siehe oben unter 3. Teil A. III. 2. f) aa) (4).

435 Vgl. zu den Schrankenbestimmungen m. w. N. zum Einsatz von Hilfspersonen unmittelbar oben.

für Trainingsdatensammler und andere Datenverarbeiter.⁴³⁶ Diese Rechtsicherheit würde für Datenverarbeiter konterkariert, wenn sich die Details des Vertragsverhältnisses zwischen dem Zugangsanbieter und dem Zugangsinhaber auf die Rechtmäßigkeit ihrer Datenverarbeitung im Rahmen des § 44b Abs. 2 S. 1 UrhG auswirken würden. Die Datenverarbeiter müssen vielmehr auf ihre erlangte freie Zugänglichkeit der Werke vertrauen dürfen.

Schließlich würde eine Berücksichtigung vertraglicher Beschränkungen der Zugangsweitergabe im Verhältnis zwischen Zugangsanbieter und Zugangsinhaber außer Acht lassen, dass der Nutzungsvorbehaltsmechanismus aus § 44b Abs. 3 UrhG de lege lata zentrales regulatorisches Instrument für die Steuerung der Nutzung von Werken zum Text und Data Mining ist. Es bestünde die Gefahr, dass durch „Zugangsabreden“ zwischen Zugangsanbietern und -inhabern die spezifischen, aus § 44b Abs. 3 UrhG folgenden Voraussetzungen unterlaufen würden.

Alles in allem sprechen damit sowohl systematische als auch teleologische Gründe dafür, dass eine vertragliche Untersagung der Zugangsweitergabe zwischen Zugangsanbieter und -inhaber keine Auswirkungen darauf hat, dass die betreffenden Werke für einen Datenverarbeiter, dem der Zugang zu den Werken vermittelt worden ist, im Sinne des § 44b Abs. 2 S. 1 UrhG rechtmäßig zugänglich sind. Unabhängig davon begeht der Zugangsinhaber in diesen Fällen aber eine sanktionsfähige vertragliche Pflichtverletzung.

(2) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzeugugängen für Datenverarbeitung im Interesse des Zugangsinhabers

Fraglich ist darüber hinaus, ob die Werke für einen Datenverarbeiter auch dann noch rechtmäßig zugänglich im Sinne des § 44b Abs. 2 S. 1 UrhG sind, wenn ihm nicht der Zugang zu den Werken (z. B. durch eine Weitergabe von Login-Daten) vermittelt wird, sondern wenn durch den Zugangsinhaber zuvor angefertigte Vervielfältigungsvorlagen zur Verfügung gestellt werden. Hierfür speichert der Zugangsinhaber die Werke zunächst ab, stellt also selbst Vervielfältigungsstücke her. Diese Werkstücke kann der Datenverarbeiter im Anschluss als Vervielfältigungsvorlage für eigene Vervielfältigungshandlungen nutzen, die er für die automatisierte Analyse der Werke benötigt.

436 ErwGr. 18 DSM-RL. So im vorliegenden Kontext auch *Bomhard*, InTeR 2023, 174 (176).

Um die rechtmäßige Zugänglichkeit der Werke für den Datenverarbeiter zu beurteilen, kommt es insbesondere darauf an, auf welches Zugangssubjekt bei der Bewertung abgestellt wird. Möglich wäre einmal, nur die Quelle in den Blick zu nehmen, die der Datenverarbeiter für die Anfertigung seiner Vervielfältigungsstücke verwendet. Das sind die Werkstücke, die bereits durch den Zugangsinhaber hergestellt worden sind. In der Folge käme es nur darauf an, ob der Datenverarbeiter rechtmäßigen Zugang zu den beim Zugangsinhaber verfügbaren Werkstücken hatte. Da der Zugangsinhaber dem Datenverarbeiter die Werkstücke zur Herstellung von Vervielfältigungen für die automatisierte Analyse aktiv zugänglich macht, ist das der Fall. Infolgedessen bestünde rechtmäßiger Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG. Andererseits könnte aber auch die gesamte Zugangskette mit in den Blick genommen werden. Mittelbar wäre damit auch der Zugang des Zugangsinhabers in die Betrachtung einzubeziehen. Für den Datenverarbeiter käme es, wie im Falle einer Zugangsvermittlung,⁴³⁷ demzufolge darauf an, ob der Zugangsinhaber rechtmäßigen Zugang zu den Werken im Sinne der Regelung gehabt hat und ob er hiervon als nachgelagerter Datenverarbeiter auch profitieren kann.

Für eine Berücksichtigung der gesamten Zugangskette, mit der auch der Zugang des Herstellers der Vervielfältigungsvorlagen in den Blick nimmt genommen wird, spricht, dass es auch im Rahmen des § 53 Abs. 1 S. 1 UrhG darauf ankommt, dass für die Vervielfältigungshandlungen keine offensichtlich rechtswidrigen Vorlagen genutzt werden. Hieraus ließe sich unter Umständen ein allgemeiner Rechtsgedanke ableiten, in dessen Folge auch im Rahmen des § 44b Abs. 2 S. 1 UrhG der Ursprung der Vervielfältigungsvorlage, also die gesamte Zugangskette, in den Blick genommen werden müsste.

Allerdings beruht die Rechtmäßigkeitsanforderung aus § 53 Abs. 1 S. 1 UrhG auf einem politischen Kompromiss im Vermittlungsausschuss zwischen Deutschem Bundestag und Bundesrat.⁴³⁸ Diese Regelungshistorie spricht dagegen, ihr eine allgemeingültige Bedeutung auch über den konkreten Regelungsgegenstand hinaus beizumessen. Außerdem konnte bereits gezeigt werden, dass aus § 53 Abs. 1 S. 2 UrhG ein allgemeiner Rechtsgedanke abgeleitet werden kann, wonach Vervielfältigungshandlungen von Hilfspersonen des originär Schrankenberechtigten ebenfalls von der Frei-

437 Dazu unmittelbar siehe oben unter 3. Teil A. III. 2. f) bb) (1).

438 BT-Drs. 15/1353.

stellungswirkung der Schrankenbestimmung miterfasst sind.⁴³⁹ Hierfür kann es keinen Unterschied machen, ob der Hilfsperson der Werkzugang vermittelt wird, oder ob ihm bereits Vervielfältigungen der Werke zum Zwecke der Hilfsdienstleistung verfügbar gemacht werden.

Gegen eine Berücksichtigung der Zugangskette spricht außerdem, dass die Beurteilung des rechtmäßigen Zugangs des Herstellers der Vervielfältigungsvorlage Ähnlichkeiten zu der Fragestellung aufweist, ob eine im Internet zur Verfügung gestellte Vervielfältigungsvorlage rechtmäßig öffentlich zugänglich gemacht worden ist. Die Rechtmäßigkeit der Zugänglichmachung ist für die Anwendung des § 44b Abs. 2 S. 1 UrhG aber ohne Bedeutung.⁴⁴⁰ Bei der Berücksichtigung der gesamten Zugangskette droht deswegen die Gefahr eines Wertungswiderspruchs.

Dafür, bei der Beurteilung der rechtmäßigen Zugänglichkeit allein auf die unmittelbare Quelle des Datenverarbeiter abzustellen, kann darüber hinaus angeführt werden, dass der Datenverarbeiter unter Umständen nicht weiß und wissen kann, aus welcher Primärquelle die Vervielfältigungsvorlagen herrühren und welche Zugangsvoraussetzungen hier hätten erfüllt sein müssen. Es erscheint vor diesem Hintergrund unbillig, ihm die außerhalb seines Einflussbereichs liegende Zugangskette zuzurechnen. Für den Datenverarbeiter bedeutete dies nicht zuletzt einen erheblichen Verlust an Rechtssicherheit. Dies zeigt sich insbesondere bei der Nutzung von frei verfügbar gemachten Trainingsdatensätzen. Zwar hätte der Datenverarbeiter selbst zu ihnen rechtmäßigen Zugang. Er könnte aber dennoch nicht beurteilen, ob die Schrankenwirkung des § 44b UrhG eingreift. Dieser Zustand widerspräche dem Ziel der Regelung, eine rechtssichere Grundlage für die Verarbeitung der Werke zu schaffen.⁴⁴¹

Alles in allem sprechen damit mehrere teleologische Gründe dafür, bei der Beurteilung der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG allein auf den unmittelbaren Zugang des Datenverarbeiters zu seiner Vervielfältigungsvorlage abzustellen. Die Zugangskette, welche der Vervielfältigungsvorlage zugrunde liegt, ist für die Beurteilung seines rechtmäßigen Zugangs damit nicht von Bedeutung.

Da dem Datenverarbeiter der Zugang zu seinen Vervielfältigungsvorlagen im konkreten Fall aktiv vom Zugangsinhaber bereitgestellt wird, damit er in seinem Interesse Analyseleistungen erbringt, hat er im Rahmen der vorliegenden Fallgruppe rechtmäßigen Zugang zu den Werken im Sinne

439 Oben unter 3. Teil A. III. 2. f) bb) (1).

440 Dazu siehe oben unter 3. Teil A. III. 2. f) aa).

441 ErwGr. 18 DSM-RL.

von § 44b Abs. 2 S. 1 UrhG. Auf das Verhältnis zwischen Zugangs- und Rechtsinhaber sowie den Ursprung der verwendeten Vervielfältigungsvorlagen kommt es nicht an.

- (3) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzeugängen für Datenverarbeitung im Eigeninteresse des Datenverarbeiters

Im Rahmen einer anderen Konstellation ist weiterhin denkbar, dass die vom Zugangsinhaber vervielfältigten Werke einem im eigenen Interesse handelnden Datenverarbeiter als Vervielfältigungsvorlage zur Verfügung gestellt werden. Hierunter fallen insbesondere intermediäre Trainingsdatensammler, die Zugangsinhabern beispielsweise Daten abkaufen, um diese für eine kommerzielle Verwertung in Trainingsdatensätze zu überführen. Denkbar ist aber auch, dass ein Zugangsinhaber die vervielfältigten Werke selbst als Vorlage für weitere Vervielfältigungshandlungen (gegebenenfalls rechtswidrig) öffentlich wiedergibt, damit andere die Inhalte zum eigenen KI-Training vervielfältigen können. Fraglich ist, ob die Datenverarbeiter auch in diesen Fällen rechtmäßigen Zugang den akquirierten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben, sodass ihre Vervielfältigungshandlungen in den Anwendungsbereich der Schrankenbestimmung für Text und Data Mining fallen können.

- (i) Rechtmäßiger Zugang der Datenverarbeiter

Da es für den rechtmäßigen Zugang im Sinne des § 44b Abs. 2 S. 1 UrhG nur auf den Zugang des Datenverarbeiters zu den unmittelbaren Vervielfältigungsvorlagen ankommt und ihm die Vorlagen auch in dieser Fallkonstellation aktiv bereitgestellt worden sind, hat der Datenverarbeiter auch hier rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG.⁴⁴² Demzufolge würden die Vervielfältigungshandlungen grundsätzlich in den Anwendungsbereich der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG fallen.

442 Dazu ausführlich siehe unmittelbar oben unter 3. Teil A. III. 2. f) bb) (2).

(ii) Begrenzung der Schrankenwirkung durch den Dreistufentest

Unklar ist aber, ob die Freistellung der Nutzungshandlung über § 44b UrhG in diesem Fall mit dem Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL vereinbar ist. Danach darf die Schrankenbestimmung nur in bestimmten Sonderfällen angewandt werden, in denen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt und die berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden.

Weil die Schrankenbestimmung nur Vervielfältigungen für das Text und Data Mining freistellt, betrifft sie nur einen bestimmten Sonderfall, sodass die erste Voraussetzungen des Dreistufentests im vorliegenden Fall gewahrt ist.⁴⁴³ Ob die normale Verwertung der Werke beeinträchtigt wird oder ob es zu einer ungebührlichen Verletzung der berechtigten Urheberinteressen kommt, richtet sich weiterhin nach einer umfassenden Gesamtabwägung der gegenläufigen Interessen insbesondere im Hinblick auf die Verwertungsmöglichkeiten betroffener Urheber.⁴⁴⁴

(a) Argumente für eine ungebührliche Beeinträchtigung der Urheberinteressen

Für eine ungebührliche Beeinträchtigung der Urheberinteressen und in der Folge für eine Verletzung des Dreistufentests spricht, dass der Datenverarbeiter im vorliegenden Fall nicht im Interesse des Zugangsinhabers tätig wird. Infolgedessen ist es möglich, dass neben dem Datenverarbeiter auch der Zugangsinhaber selbst Interesse an der Nutzung der betroffenen Werke zum Text und Data Mining hat. Aus diesem Grund kann es auf Grundlage eines Zugangs des Zugangsinhabers mitunter zu *zwei* wirtschaftlichen relevanten Nutzungen der Werke durch den Zugangsinhaber zum einen und den Datenverarbeiter zum anderen kommen. Damit ist im vorliegenden Fall im Vergleich zur Nutzung der Werke durch den Zugangsinhaber allein eine intensiviertere Beeinträchtigung der Verwertungsinteressen der Urheber möglich. Hierauf hat auch eine Vergütung, die der Datenverarbeiter dem Zugangsinhaber für die Bereitstellung der Vervielfältigungsvorlagen zahlt, keinen Einfluss. Denn diese Zahlung ist nicht an den betroffenen Urheber gerichtet.

443 Vgl. zur ersten Stufe des Dreistufentests ausführlicher oben unter 3. Teil A. III. 2. d).

444 Dazu bereits siehe oben unter 3. Teil A. III. 2. d).

Weiterhin kann für eine ungebührliche Beeinträchtigung der Urheberinteressen angeführt werden, dass die Bereitstellung von Vervielfältigungsvorlagen zugunsten eines nicht selbst zugangsberechtigten Datenverarbeiters möglicherweise die Entscheidung der betroffenen Urheber unterläuft, einem Datenverarbeiter keinen Zugang zu ihren Werken zu gewähren. Im Gegenteil müssten sie bei Zugrundelegung des obigen Tatbestandsverständnisses hinnehmen, dass ihre Werke für das KI-Training unkontrollierbar in die Welt gelangt sind, sobald sie die Werke in einer zugangsgeschützten Umgebung verfügbar gemacht haben.

Außerdem haben Urheber bei der Nutzung ihrer Werke zum Training von KI ein schutzwürdiges Amortisationsinteresse.⁴⁴⁵ Dieses Interesse kann auch durch eine vergütungspflichtige Zugangsbeschränkung geschützt werden. Um dem Amortisationsinteresse bestmögliche Geltung zu verleihen, sollte die Effektivität einer eingesetzten Zugangsbeschränkung weitestgehend aufrechterhalten werden. Vor diesem Hintergrund ist es interessengerecht, die untersuchte Fallkonstellation aus dem Anwendungsbereich der Schrankenbestimmung für Text und Data Mining auszuklammern. Um von der Freistellungswirkung der Schrankenbestimmung profitieren zu können, müsste der Datenverarbeiter infolgedessen einen eigenen Werkzugang erwerben.

(β) Argumente gegen eine ungebührliche Beeinträchtigung der Urheberinteressen

Gegen eine ungebührliche Beeinträchtigung der Urheberinteressen spricht allerdings, dass durch die Verwendung der Werke zum Training von KI auch in der hier betreffenden Fallkonstellation keine Kernverwertungsarten der Urheber berührt sind. Weiterhin gegen die Beeinträchtigung der normalen Werkverwertung der Urheber spricht, dass individuelle Lizenzverträge zwischen Urhebern und Datenverarbeitern aufgrund prohibitiv hoher Transaktionskosten im Regelfall nicht zustande kommen.⁴⁴⁶

Gegen eine ungebührliche Beeinträchtigung der Urheberinteressen spricht weiterhin, dass sich aus einer Abwägung von Nutzer- und Urheberinteressen ein umfassendes Freistellungsbedürfnis in Bezug auf das Web Scraping von Werken zum KI-Training ergeben hat.⁴⁴⁷ Vor diesem Hin-

445 Dazu siehe oben 2. Teil B. II. 1.

446 Siehe oben unter 2. Teil B. I. 2. und 3. Teil A. III. 2. d).

447 Dazu siehe oben 2. Teil B. I. 2.

tergrund ist das Interesse der Urheber an einem bestmöglichen Schutz der von ihnen eingeschalteten Zugangsbeschränkung trotz ihres grundsätzlich anerkennungswürdigen Amortisationsinteresses im Spezifischen nicht schutzbedürftig. Vielmehr sollten die Werke frei zum Training von KI genutzt werden dürfen.⁴⁴⁸ Dafür spricht auch, dass aus Perspektive des Datensammlers keine rechtswidrig erlangte Vervielfältigungsvorlage verwendet worden ist. Vielmehr wurden ihm die Werkstücke aktiv zugänglich gemacht. Vor diesem Hintergrund ist auch zu berücksichtigen, dass es Ziel des Normgebers ist, eine rechtssichere Grundlage für das Text und Data Mining zur Verfügung zu stellen.⁴⁴⁹ Haben Datenverarbeiter selbst keine Zugangsbeschränkungen umgangen, müssen sie daher darauf vertrauen dürfen, dass sie die Werke rechtmäßig auf Basis des § 44b UrhG verarbeiten dürfen, solange die übrigen Freistellungsvoraussetzungen erfüllt sind.

(γ) Schlussfolgerungen

Grundsätzlich sprechen insbesondere die Nichtbeeinträchtigung der Kernverwertungsmöglichkeiten der Urheber sowie das schutzwürdige Interesse der Datenverarbeiter an einer rechtssicheren Grundlage für Vervielfältigungen zum Training von KI gegen eine ungebührliche Beeinträchtigung der normalen Verwertung von Werken durch die Urheber. Vervielfältigungshandlungen von Datenverarbeitern, die sie auf Basis von durch Zugangsinhaber zur Verfügung gestellten Vervielfältigungsvorlagen anfertigen, verletzen damit grundsätzlich nicht den Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL. Damit sind sie im Allgemeinen von der Schrankenbestimmung für Text und Data Mining gemäß § 44b UrhG freigestellt.

(δ) Abweichende Beurteilung bei objektiver Erkennbarkeit der zugangsgeschützten Herkunft genutzter Vervielfältigungsvorlagen

Etwas anderes könnte allerdings dann gelten, wenn für den im eigenen Interesse handelnden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlagen einer zugangsgeschützten Quelle entstammen. Denn wenn sich der Datenverarbeiter auf seinen rechtmäßigen Zugang zu

448 Dazu siehe oben 2. Teil B. I. 2.

449 ErwGr. 18 DSM-RL.

den Vervielfältigungsvorlagen zurückzieht, obwohl erkennbar ist, dass für den originären Werkzugang Zugangsbeschränkungen überwunden werden müssen oder insbesondere ein Entgelt zu entrichten ist, wohnt seiner Nutzung ein Umgehungsmoment inne. Es ist mit der unmittelbaren Umgehung der zugangshindernden Maßnahmen vergleichbar, die zu einem Nichteinreifen der Schrankenbestimmung aus § 44b UrhG führt. Der Datenverarbeiter hätte beispielsweise selbst ein eigenes Entgelt für den Zugang zu den Werken entrichten müssen, um seine im Eigeninteresse erfolgende Nutzung der Werke angemessen auszugleichen. Außerdem hat der Datenverarbeiter in diesem Fall kein schützenswertes Interesse an einer rechtssicheren Freistellungsregelung durch die Schrankenbestimmung. Dies führt im Ergebnis zu einer ungebührlichen Beeinträchtigung der Urheberinteressen. Infolgedessen ist in diesen Fällen der Dreistufentest aus Art. 5 Abs. 5 Info-Soc-RL verletzt. Ist für den vervielfältigenden Datenverarbeiter objektiv erkennbar, dass die Vervielfältigungsvorlage einer zugangsgeschützten Quelle entstammt, sind seine Vervielfältigungshandlungen abweichend vom zuvor Gesagten demzufolge ausnahmsweise nicht von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 S. 1 UrhG freigestellt.⁴⁵⁰ Eine Rückausnahme hiervon muss nur dann gemacht werden, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, auf Basis des Werkzugangs angefertigte Werkstücke dritten Datenverarbeitern als Vorlagen für Vervielfältigungen zum Text und Data Mining bereitzustellen. Denn dann hätte die Ermöglichung weiterer Vervielfältigungshandlungen zum Text und Data Mining insbesondere in der Höhe des für den Zugangserwerb zu entrichtenden Entgelts berücksichtigt werden können. Zudem liegt kein Umgehungsmoment vor, welches ebenfalls maßgeblich zur Verletzung des Dreistufentests beiträgt.⁴⁵¹

Dass damit auch die Tätigkeit intermediärer Trainingsdatensammler im Wesentlichen vom Anwendungsbereich des § 44b Abs. 2 S. 1 UrhG erfasst ist, deckt sich auch mit den bereits herausgearbeiteten Regelungsbedürfnissen. Für ihre Tätigkeit ist aber insbesondere noch die Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG in den Blick zu nehmen.⁴⁵²

450 ähnlich auch *Dworschak*, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 15.

451 Das wird insbesondere bei Gedächtnisinstitutionen wie Bibliotheken relevant, siehe unten unter 3. Teil A. III. 4. a).

452 Dazu sogleich unter 3. Teil A. III. 2. h).

(4) Zugangsvermittlung für Datenverarbeitung im Interesse des Datenverarbeiters

Möglich ist auch, dass der Zugangsinhaber dem Datenverarbeiter seinen Zugang (beispielsweise Login-Daten) vermittelt, also weiterreicht, damit er auf Basis dieses Werkzugangs im eigenen Interesse Vervielfältigungstücke für das Training von KI oder die Herstellung von Trainingsdatensätzen anfertigen kann. Denkbar ist, dass der Zugangsinhaber hierfür eine Vergütung erhält. Fraglich ist auch in diesem Fall, ob der Datenverarbeiter aufgrund der Zugangsvermittlung rechtmäßigen Zugang zu den betroffenen Werken im Sinne des § 44b Abs. 2 S. 1 UrhG hat. Dann würden seine Vervielfältigungshandlungen in den Anwendungsbereich des Schrankentatbestands fallen.

Wie bereits gezeigt werden konnte, ist es für die Anwendung des Schrankentatbestands aus § 44b UrhG unschädlich, dass der Datenverarbeiter die Zugangsvoraussetzungen nicht eigenhändig erfüllt.⁴⁵³ Allerdings muss eine Anwendung der Schrankenbestimmung wegen einer Verletzung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL ausscheiden, wenn für den im eigenen Interesse tätig werdenden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlage einer zugangsgeschützten Quelle entstammt. Denn solchen Vervielfältigungshandlungen wohnt ein Umgehungsmoment inne. Sie beeinträchtigen die berechtigten Urheberinteressen daher übergebührlich.⁴⁵⁴ Wird dem Datenverarbeiter der Werkzugang beispielsweise in Form von Login-Daten vermittelt, ist für ihn allerdings objektiv erkennbar, dass die Vervielfältigungsvorlagen aus einer geschützten Quelle stammen, deren Zugangsvoraussetzungen er nicht selbst erfüllt. Damit beeinträchtigt eine Anwendung der Schrankenbestimmung aus § 44b UrhG auf diese Fallgruppe die berechtigten Urheberinteressen übergebührlich. Das führt zu einer Verletzung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL. Aus diesem Grund kann die Nutzung in der betreffenden Fallgruppe nicht von der Schrankenbestimmung aus § 44b UrhG freigestellt werden. Eine Ausnahme hiervon muss nur dann gemacht werden, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, den Werkzugang an dritte Datenverarbeiter zum Text und Data Mining weiterzugeben.⁴⁵⁵

453 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (1).

454 Dazu ausführlich siehe unmittelbar oben unter 3. Teil A. III. 2. f) bb) (3).

cc) Fernwirkung einer fehlenden Zugangsberechtigung insbesondere bei der öffentlichen Wiedergabe von Trainingsdatensätzen

In einer umgekehrten Fallkonstellation ist schließlich auch klärungsbedürftig, wie sich eine Bereitstellung von Vervielfältigungsvorlagen an Datenverarbeiter auf die rechtmäßige Zugänglichkeit der Werke für die Datenverarbeiter im Sinne des § 44b Abs. 2 S. 1 UrhG auswirkt, wenn der Vermittler selbst keinen rechtmäßigen Zugang zu den Werken gehabt hat. Das umfasst insbesondere auch Fälle, in denen die Werke als Teil von Trainingsdatensätzen angeboten werden, obwohl der Anbieter des Datensatzes keinen rechtmäßigen Zugang zu den Werken nutzen konnte.

Wie bereits gezeigt wurde, kommt es für die Bestimmung der Rechtmäßigkeit des Zugangs des nachgelagerten Datenverarbeiters allein auf seine Beziehung zu den von ihm genutzten Vervielfältigungsvorlagen an.⁴⁵⁶ Sofern der Zugang für den Datenverarbeiter frei möglich ist, beispielsweise, weil die Werke frei im Internet wiedergegeben oder ihm aktiv überlassen werden, hat der Datenverarbeiter rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG. Auf die Urheberrechtskonformität der Zugänglichmachung, also der öffentlichen Wiedergabe der Vervielfältigungsvorlagen, kommt es nicht an.⁴⁵⁷ Ebenso wenig ist grundsätzlich die Rechtmäßigkeit des Zugangs des die Vervielfältigungsvorlage zur Verfügung Stellenden von Bedeutung.⁴⁵⁸ Das Bedürfnis nach einer rechtssicheren Grundlage für das Text und Data Mining, welches in ErwGr. 18 DSM-RL zum Ausdruck kommt, überwiegt daher grundsätzlich die urheberrechtsverletzende Herkunft der Text und Data Mining-Grundlage.

Allerdings steht auch in diesem Fall der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL einer Anwendung des Schrankenatbestands entgegen, wenn für den vervielfältigenden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlage einer zugangsgeschützten Quelle entstammt.⁴⁵⁹ Wann das der Fall ist, ist unter Berücksichtigung aller Umstände des Einzelfalls zu bestimmen. Fraglich ist dabei insbesondere, ob von der Größe des bereitgestellten Datensatzes in Verbindung mit der Art der darin enthaltenen Werke auf eine zugangsgeschützte Herkunft der Inhalte geschlossen werden muss. Denkbar wäre

455 Vgl. dazu ausführlich oben unter 3. Teil A. III. 2. f) bb) (3).

456 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (2).

457 Dazu siehe oben unter 3. Teil A. III. 2. f) aa).

458 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (2).

459 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (3).

das beispielsweise beim öffentlich verfügbaren Trainingsdatensatz *Books3*. Er besteht aus 197.000 Büchern.⁴⁶⁰ Zugunsten einer objektiven Erkennbarkeit kann hier angeführt werden, dass eine unrechtmäßige Herkunft der Inhalte aufgrund der großen Menge an verfügbar gemachten Schriftwerke naheliegen dürfte. Gegen einen solchen Schluss spricht jedoch, dass es auch umfangreiche Trainingsdatensätze gibt, die sich aus urheberrechtlich freien Inhalten zusammensetzen. Hierzu gehört beispielsweise der *Project Gutenberg*-Datensatz. Er besteht ausschließlich aus gemeinfrei gewordenen Schriftwerken.⁴⁶¹ Allein aus der Größe des bereitgestellten Datensatzes und der Art der enthaltenen Werke kann daher nicht ohne weiteres auf eine zugangsgeschützte Herkunft der Inhalte geschlossen werden. Als Ausgangspunkt der Betrachtung muss vielmehr das konkret vervielfältigte Werk dienen. Die zugangsgeschützte Quelle eines Werkstücks ist beispielsweise dann objektiv erkennbar, wenn es sich um ein audiovisuelles Werk handelt, das als Eigenproduktion eines Streaming-Anbieters erkennbar ist. Diese Inhalte werden vom Rechtsinhaber stets nur auf dem zugangsgeschützten Streaming-Portal verfügbar gemacht. Solche Werke könnten aus einer praktischen Perspektive beispielsweise durch einen automatisierten Abgleich mit der Wortbildmarke des Streaming-Anbieters herausgefiltert werden. Bei der Bestimmung der objektiven Erkennbarkeit muss im Übrigen aber stets auch berücksichtigt werden, wie erkennbar die Nutzung des Werkes für den Datenverarbeiter ist. Die zugangsgeschützte Herkunft einer Vervielfältigungsvorlage ist beispielsweise dann nicht objektiv erkennbar, wenn das jeweilige Werk in einem ansonsten unspezifischen, urheberrechtskonform erscheinenden Trainingsdatensatz eingebettet ist. Von Bedeutung ist auch, inwieweit die Datensammlung automatisiert und generalisiert erfolgt, beispielsweise durch ein breit angelegtes Web Scraping von Inhalten für das Training von KI. Für objektive Erkennbarkeit spricht hingegen, wenn das betreffende Werk gezielt vervielfältigt wird oder der vervielfältigte Datensatz insgesamt aus Werkstücken besteht, die einen zugangsgeschützten Ursprung nahelegen.

460 Siehe dazu https://huggingface.co/datasets/the_pile_books3 (zuletzt angerufen am 22.3.2024); vgl. außerdem *Gao u. a.*, arXiv: 2101.00027 2020, S. 3 f.

461 Dazu <https://www.gutenberg.org> (zuletzt abgerufen am 22.3.2024); https://data.ncl.ac.uk/articles/dataset/Gutenberg_Dataset/24574753 (zuletzt abgerufen am 22.3.2024); *Gao u. a.*, arXiv: 2101.00027 2020, S. 5.

g) Kreis der Schrankenbegünstigten

§ 44b UrhG spezifiziert den Kreis der Schrankenbegünstigten im Gegensatz zu § 60d UrhG nicht. Aus diesem Grund ist der Anwendungsbereich der Schrankenbestimmung auch dann eröffnet, wenn die Trainingsdatensammler gleichzeitig in den persönlichen Anwendungsbereich der speziellen Schrankenbestimmung für Text und Data Mining zu wissenschaftlichen Zwecken, § 60d UrhG, fallen.⁴⁶²

aa) Kommerziell tätige Trainingsdatensammler

Darüber hinaus profitieren von der Schrankenbestimmung sowohl kommerziell auch als nicht-kommerziell tätige Trainingsdatensammler. Etwas anderes ergibt sich auch nicht aus der Anwendung des Dreistufentests.⁴⁶³ Das Freistellungsbedürfnis, welches sowohl für kommerziell als auch für nicht-kommerziell tätige Trainingsdatensammler besteht, wird in personeller Hinsicht durch § 44b UrhG demzufolge erfüllt.

bb) Intermediäre Trainingsdatensammler

Fraglich ist allerdings, ob der persönliche Anwendungsbereich der Schrankenbestimmung aus § 44b UrhG auch intermediäre Trainingsdatensammler umfasst. Sie speichern ebenfalls Werke mit Hilfe von Webcrawlern zur Verwendung als KI-Trainingsdaten ab. Damit ist auch ihre Tätigkeit von dem herausgearbeiteten Freistellungsinteresse abgedeckt. Im Gegensatz zum KI-Entwickler führen sie aber selbst kein Training von KI-Systemen durch. Infolgedessen findet das Text und Data Mining im Sinne des § 44b Abs.1 UrhG nicht bei ihnen statt – jedenfalls so lange die intermediären Trainingsdatensammler nicht in besonderen Fallkonstellationen eine vorbereitende Analyse der Werke durchführen, die unter § 44b Abs.1 UrhG gefasst werden kann.⁴⁶⁴

Für eine Freistellung der intermediären Trainingsdatensammler über § 44b UrhG ist somit maßgeblich, ob sie bei der Speicherung von Wer-

462 Vgl. Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 4.

463 Dazu und zu einer abweichenden Auffassung siehe oben unter Teil. A. I. 2. b) cc).

464 So insbesondere in LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

ken für ein späteres KI-Training durch Dritte Vervielfältigungen „für das Text und Data Mining“ beziehungsweise „für die Zwecke des Text und Data Mining“ vornehmen, wie es § 44b Abs. 2 S. 1 UrhG respektive Art. 4 Abs. 1 DSM-RL tatbestandlich voraussetzen. Das hängt davon ab, ob die Schrankenbestimmung mit diesen Formulierungen einen Bezug der Vervielfältigungshandlungen zu einem bereits konkretisierten Text und Data Mining-Projekt notwendig macht. Wäre das der Fall, könnten Vervielfältigungen intermediärer Trainingsdatensammler, die ohne Bezug zu einem spezifischen Text und Data Mining-Projekt stattfinden, nicht unter den Schrankentatbestand aus § 44b Abs. 2 S. 1 UrhG subsumiert werden. Bei der Auslegung der Schrankenbestimmung ist insbesondere auch der unionsrechtliche Hintergrund der Regelung aus Art. 4 DSM-RL zu berücksichtigen.

(1) Wortlaut der Regelung

Als Ausgangspunkt der Auslegung ist auf den natürlichen Wortsinn des Tatbestandsmerkmals abzustellen.⁴⁶⁵ Im nationalen Umsetzungsrecht wird vorausgesetzt, dass die Vervielfältigungen *für* das Text und Data Mining angefertigt werden müssen. Mit der Präposition „für“ wird ihrem natürlichen Wortsinn nach der Zweck einer Handlung beschrieben.⁴⁶⁶ Die gleiche Bedeutung kann weiterhin der Formulierung aus Art. 4 Abs. 1 DSM-RL entnommen werden. Diese bezieht sich ausdrücklich auf die „Zwecke“ der Vervielfältigungshandlungen.

Bezugspunkt der Zweckbindung ist „das Text und Data Mining“. Diese Formulierung ist unbestimmt. Die Zweckbindung wird damit pauschal mit der Analysetechnik als solcher und nicht mit *einem konkreten* Text und Data Mining-Vorhaben verknüpft.⁴⁶⁷ Demzufolge sprechen der Wortlaut des § 44b Abs. 2 S. 1 UrhG und des umgesetzten Art. 4 Abs. 1 DSM-RL dafür, dass es ausreichend ist, wenn die in Rede stehenden Vervielfältigungen angefertigt werden, damit zu einem späteren Zeitpunkt die Analysetechnik des Text und Data Mining als solche durchgeführt wird. Hierfür bedarf

465 Zur Auslegungsregel für alle *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 163 f. Im europäischen Recht ebenso unbestritten, siehe EuGH, Urt. v. 05.07.2012 - C-49/11, EuZW 2012, 638 Rn. 32 – Content Services; EuGH, Urt. v. 10.03.2005 - C-336/03, MMR 2005, 364 Rn. 21 – easycar.

466 *Dudenredaktion* (*Bibliographisches Institut*), Duden: das Bedeutungswörterbuch, Schlagwort: „für“.

467 Vgl. zur identischen Formulierung im Rahmen der Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG: *Wulf/Löbeth*, GRUR 2024, 737 (740).

es diesem Wortlaut entsprechend noch keiner Konkretisierung eines einzelnen Text und Data Mining-Projekts. Der Wortlaut der Bestimmungen spricht demzufolge dafür, dass auch intermediäre Trainingsdatensammler Vervielfältigungen für das Text und Data Mining im Sinne von § 44b Abs. 2 S. 1 UrhG anfertigen können. Denn auch ihre Vervielfältigungen sind allgemein für das KI-Training, also für die Analysetechnik des Text und Data Mining bestimmt, auch wenn noch kein spezifisches Trainingsprojekt konkretisiert ist.

(2) Systematische Anknüpfungspunkte

Systematische Bezüge der Schrankenbestimmung aus § 44b UrhG respektive Art. 4 DSM-RL zu anderen urheberrechtlichen Regelungen könnten hingegen für eine enge Auslegung des Tatbestands sprechen. Nach dieser müsste für eine Privilegierung der Vervielfältigungshandlungen bereits ein konkretes Text und Data Mining-Projekt bestehen, dem die Vervielfältigungen gewidmet werden.

Hierfür spricht zum einen, dass im Rahmen von § 60d UrhG a. F., der vorherigen Fassung der Schrankenbestimmung für Text und Data Mining zum Zwecke wissenschaftlicher Forschung, vereinzelt der Bezug der Vervielfältigungshandlungen zu einem konkreten Text und Data Mining-Projekt vorausgesetzt wurde.⁴⁶⁸

Wenn für die Tatbestandserfüllung kein Bezug der Vervielfältigungshandlungen zu einem konkreten Text und Data Mining-Projekt notwendig wäre, könnte außerdem die Gefahr eines Wertungswiderspruchs zwischen Art. 3 und Art. 4 DSM-RL bestehen. Denn Forschungseinrichtungen müssen bei der Aufbewahrung der Werke gemäß Art. 3 Abs. 2 DSM-RL respektive § 60d Abs. 5 UrhG n. F. angemessene Sicherheitsvorkehrungen vorsehen. Bei intermediären Trainingsdatensammlern, die Werke vervielfältigen, ohne ein bereits konkretisiertes Text und Data Mining-Projekt ermöglichen zu wollen, ist jedoch ebenfalls davon auszugehen, dass die Werke zunächst für einen unbestimmten Zeitraum gespeichert werden. Auch hier findet demzufolge eine vorgelagerte Aufbewahrung der Werke statt. Hierfür müssten sie gemäß § 44b UrhG respektive Art. 4 DSM-RL jedoch keine Sicherheitsvorkehrungen einsetzen.

468 Vgl. dazu *Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60d Rn. 13.

Allerdings ist im Rahmen der Freistellungsregelung aus § 44b UrhG respektive Art. 4 DSM-RL eine Löschungspflicht zu beachten.⁴⁶⁹ Diese besteht im Rahmen des Art. 3 DSM-RL beziehungsweise § 60d UrhG nicht. Das rechtfertigt bereits die Verpflichtung zur Einsetzung angemessener Sicherheitsmaßnahmen. Darüber hinaus richtet sich die Schrankenbestimmung für Text und Data Mining zu allgemeinen Zwecken aus § 44b UrhG respektive Art. 4 DSM-RL insbesondere an gewinnwirtschaftlich tätige Trainingsdatensammler. Da die geschäftliche Betätigung eng mit den vervielfältigten Werken verbunden ist, bestehen bereits natürliche Anreize für eine Einsetzung angemessener Sicherheitsmaßnahmen zur Sicherung des wirtschaftlichen Guts. Es bedarf vor diesem Hintergrund keiner besonderen Verpflichtung zur Einhaltung solcher Sicherheitsstandards. Ein Wertungswiderspruch zwischen Art. 3 und Art. 4 DSM-RL bei einem weiten Tatbestandsverständnis der Zweckbestimmung besteht deswegen nicht.

Darüber hinaus geht § 60d UrhG a. F. auf eine ältere unionsrechtliche Grundlage zurück, die durch Art. 3 DSM-RL als neue Spezialregelung hinsichtlich der Vervielfältigung von Werken für wissenschaftlichen Zwecken dienendes Text und Data Mining überlagert worden ist. Ein systematischer Vergleich zwischen § 60d UrhG a. F. und § 44b UrhG ist demzufolge nicht möglich.

Aus den systematischen Bezügen der Schrankenbestimmung für allgemeines Text und Data Mining aus § 44b UrhG respektive Art. 4 DSM-RL zu anderen urheberrechtlichen Regelungen könnten damit keine belastbaren Schlüsse für eine Auslegung des Tatbestands zur Beantwortung der hier aufgeworfenen Fragestellung abgeleitet werden.

(3) Telos der Schrankenbestimmung

Weitere Anhaltspunkte für die Auslegung der Schrankenbestimmung könnten sich allerdings aus dem Regelungszweck der Schrankenbestimmung ergeben.

Für ein enges Verständnis der von § 44b Abs. 2 S. 1 UrhG vorausgesetzten Zweckbindung der Vervielfältigungshandlungen, nach der für eine Freistellung der Nutzungshandlungen über § 44b UrhG bereits ein konkretes Text und Data Mining-Projekt notwendig ist, spricht, dass im Falle einer Freistellung intermediärer Trainingsdatensammler diese und die KI-Entwickler, also zwei Beteiligte, Vervielfältigungen zum Training von KI anfertigen

469 Dazu im Detail siehe unten unter 3. Teil A. III. 2. h).

dürften. Damit käme es zu einer intensivierten Beeinträchtigung der urheberrechtlichen Verwertungsinteressen.

Allerdings würden bei einem weiten Verständnis der von § 44b Abs. 2 S. 1 UrhG vorausgesetzten Zweckbindung trotz alledem die Vervielfältigungen – auch diejenige des intermediären Trainingsdatensammlers – ausschließlich zum Zwecke des Text und Data Mining vorgenommen. Auch die Vervielfältigungen der intermediären Trainingsdatensammler sind demzufolge noch von der gesetzgeberischen Regelungsentention gedeckt, Text und Data Mining auf Grundlage von Werken über § 44b UrhG zu ermöglichen. Außerdem werden auch Vervielfältigungen intermediärer Trainingsdatensammler zum Zwecke des KI-Trainings angefertigt, sodass im Ergebnis dieselbe urheberrechtliche Nutzungsart wie auch bei Vervielfältigungen durch KI-Entwickler betroffen ist.⁴⁷⁰ Hierfür ist ohne Belang, ob schon ein konkretes KI-Trainingsprojekt besteht, oder nicht.

Für ein weites Verständnis der von § 44b Abs. 2 S. 1 UrhG vorausgesetzten Zweckbindung spricht darüber hinaus, dass die Schrankenbestimmung gemäß ErwGr. 18 DSM-RL die Entwicklung neuer Technologien ermöglichen soll. Das können insbesondere auch Technologien sein, die auf KI-Systemen beruhen. Die Erstellung von hierfür benötigten Trainingsdatensätzen ist jedoch ressourcen- und zeitintensiv.⁴⁷¹ Sie kann nicht von jedem KI-Entwickler geleistet werden. Intermediäre Trainingsdatensammler, die über die Ressourcen zur Erstellung von Trainingsdatensätzen verfügen, sind für die Entwicklung neuer Technologien im Bereich der KI daher wesentlich. Ein arbeitsteiliges Vorgehen entspricht auch der modernen Wirtschaftsstruktur. Umgekehrt würde eine Zweckbindung der Vervielfältigungen an ein bereits konkretisiertes Text und Data Mining-Projekt kleine KI-Entwickler benachteiligen, die nicht über die Ressourcen zur vollständig eigenhändigen Erstellung der notwendigen Trainingsdatensätze verfügen.

Überdies liegt dem Urheberrecht ein allgemeiner Rechtsgedanke zugrunde, nach dem Vervielfältigungshandlungen von Hilfspersonen originär Schrankenbegünstigter ebenfalls von der Freistellungswirkung einer Schrankenregelung mitumfasst sind.⁴⁷² Auch wenn die Tätigkeit intermediärer Trainingsdatensammler noch keinem konkretisierten KI-Trainingsprojekt und keinem individualisierten KI-Entwickler zugeordnet werden kann, handelt es sich dennoch um eine vorgelagerte Hilfstätigkeit für das

470 Zur selbstständigen Nutzungsart vgl. unten unter 3. Teil A. III. 2. i) hh) (2) (i).

471 Mit einem anschaulichen Beispiel *Paafß/Hecker*, Künstliche Intelligenz, S.15 f.; *Tenyks*, The Unseen Cost of “Low Quality” Large Datasets.

472 Dazu ausführlich siehe oben unter 3. Teil A. III. 2. f) bb) (ii) (y).

Speichern von Werken durch KI-Entwickler zur Durchführung konkreter Trainingsprojekte, das auch nach der engen Auslegung des Tatbestands von § 44b Abs. 2 UrhG freigestellt ist. Als eine solche vorgelagerte Hilfstätigkeit sollten demnach auch die Vervielfältigungen intermediärer Trainingsdatensammler vom Schrankentatbestand miterfasst sein.

Vervielfältigungen nur dann über § 44b Abs. 2 UrhG freizustellen, wenn sie für ein bereits konkretisiertes, spezifisches Text und Data Mining-Projekt bestimmt sind, würde schließlich dazu führen, dass die Erstellung von Trainingsdatensätzen nur *on demand* möglich wäre. Insbesondere umfangreiche Trainingsdatensätze, die zum etablierten Standard für die Entwicklung ganzer Systemarten wie beispielsweise Textgenerierungs-KI gehören, könnten nicht im Vorhinein und damit ohne die Dringlichkeit eines spezifischen KI-Entwicklungsprojekts erstellt werden. Es gäbe infolgedessen keine qualitativ bewährte Datenbasis, auf Grundlage derer KI-Systeme diverser Arten entwickelt werden könnten. Das würde die Entwicklung von KI erheblich verlangsamen sowie verteuern und widerspräche damit direkt dem Regelungsziel aus ErwGr. 18 DSM-RL, Innovationen zu fördern.

In diesem Zusammenhang ist auch zu berücksichtigen, dass die Schrankenbestimmung für Text und Data Mining zu wissenschaftlichen Zwecken in § 60d Abs. 1 S. 1 UrhG – abzüglich des semantisch nicht relevanten Artikels „das“ – dieselbe Tatbestandsvoraussetzung wie § 44b Abs. 1 S. 1 UrhG enthält. Ein enges Verständnis, nach dem für jede Vervielfältigungshandlung der Bezug zu einem konkretisierten Text und Data Mining-Projekt erforderlich ist, müsste daher auch im Kontext wissenschaftlicher Datenaggregation der Regelung aus § 60d UrhG zugrunde gelegt werden. Insbesondere für wissenschaftliche Forschung auf Grundlage großer Datensätze bedarf es aber einer Auslagerung der Datensatzerstellung und -verwaltung an hierfür eigens geschaffene Stellen. Denn eben jene Tätigkeiten sind zu ressourcenintensiv, als dass sie von einzelnen Forschern für ihre jeweiligen Forschungsprojekte durchgeführt werden könnten. Insbesondere forschende Personen wie beispielsweise Doktoranden, die nur einen eng begrenzten Zeitraum für ihre Forschungsleistung zur Verfügung haben, wären in ihrer Tätigkeit erheblich beeinträchtigt, wenn sie, anstatt datengetriebene Forschung durchzuführen, ihre Zeit für die Erstellung großer Datensätze aufwenden müssten. Vor diesem Hintergrund ist in der Wissenschaftslandschaft diverser Datenwissenschaften auch praktisch üblich, dass die Datensatzerstellung und -pflege von konkreten Forschungsprojekten ausgelagert erfolgt. Damit wird auch ein besserer Zugang diverser forschender Personen zum Datensatz und damit eine effiziente Ressourcennutzung ermög-

licht. Letztere ist insbesondere auch für öffentliche Forschungseinrichtungen von Bedeutung. Da die Datensammlung im Regelfall außerdem vor einem konkretisierten Forschungsprojekt stattfindet, ist hierfür auch nicht die Aufbewahrungsmöglichkeit aus § 60d Abs. 5 UrhG entscheidend. Aus alledem folgt, dass ein enges Verständnis der Tatbestandsvoraussetzung „für (das) Text und Data Mining“ die datengetriebene Wissenschaft unzumutbar in ihrer Forschungstätigkeit beeinträchtigen würde.

Alles in allem sprechen damit durchgreifende teleologische Gründe insbesondere unter Berücksichtigung des in ErwGr. 18 DSM-RL niedergelegten Zwecks der umgesetzten Richtlinienbestimmung dafür, dass bei der Herstellung von Vervielfältigungen kein spezifisches Text und Data Mining- beziehungsweise KI-Trainings-Projekt gefördert werden muss, damit die Vervielfältigungshandlungen unter § 44b Abs. 2 S. 1 UrhG subsumiert werden können. Ausreichend ist vielmehr die Zweckbindung der Vervielfältigungshandlung an die Analysetechnik des Text und Data Minings, beispielsweise also die Verwendung der Werke zum KI-Training, im Zeitpunkt der Herstellung des Vervielfältigungsstücks. Die teleologischen Erwägungen sprechen demzufolge auch dafür, dass intermediäre Trainingsdatensammler Vervielfältigungen „für das Text und Data Mining“ im Sinne des § 44b Abs. 2 S. 1 UrhG vornehmen können.

Sobald die Zweckbindung der Vervielfältigungsstücke an das Text und Data Mining aufgehoben wird, ist § 44b Abs. 2 S. 2 UrhG im Weiteren allerdings so auszulegen, dass die Vervielfältigungsstücke ab diesem Zeitpunkt nicht mehr für das Text und Data Mining *erforderlich* sind. Demzufolge greift im Zeitpunkt der Zweckumwidmung die Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG ein.

(4) Zwischenergebnis

Insgesamt führen insbesondere der Zweck der Schrankenbestimmung aus § 44b UrhG respektive Art. 4 DSM-RL sowie die Wortlaute sowohl der unionsrechtlichen Schrankenbestimmung als auch der nationalen Umsetzungsnorm dazu, dass die Schrankenwirkung begründende Tatbestandsmerkmal der Vervielfältigungen „für das Text und Data Mining“ im Sinne des § 44b Abs. 2 S. 1 UrhG so ausgelegt werden könnte, dass hierfür kein Bezug zu einem bereits konkretisierten Text und Data Mining-Projekt

erforderlich ist.⁴⁷³ Infolgedessen würden auch intermediäre Trainingsdatensammler in den personellen Anwendungsbereich der Schrankenbestimmung fallen, solange die von ihnen angefertigten Vervielfältigungsstücke im Zeitpunkt ihrer Herstellung allgemein der Ermöglichung des Text und Data Minings und damit insbesondere des KI-Trainings gewidmet sind.⁴⁷⁴

(5) Dreistufentest

Fraglich ist allerdings, ob das oben genannte Tatbestandsverständnis und der daraus resultierende weite personelle Anwendungsbereich der Schrankenbestimmung auch mit dem Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL vereinbar wären. Wäre das nicht der Fall, würde dies einer entsprechenden Auslegung des Tatbestandsmerkmals „für das Text und Data Mining“ aus § 44b Abs. 2 S. 1 UrhG entgegenstehen.

Damit die Anwendung der Schrankenbestimmung auch zugunsten intermediärer Trainingsdatensammler erfolgen kann, darf es sich hierbei nur um einen bestimmten Sonderfall (erste Stufe) handeln, im Rahmen dessen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt (zweite Stufe) und die berechtigten Interessen der Rechtsinhaber nicht ungebührlich verletzt werden (dritte Stufe). Klärungsbedürftig ist im vorliegenden Fall allein, ob die zweite sowie dritte Stufe der Schutzvorschrift bei der Anwendung der Schrankenbestimmung auf intermediäre Trainingsdatensammler gewahrt wären. Diese münden methodisch in einer umfassenden Gesamtabwägung der betroffenen Urheber- und Nutzerinteressen.⁴⁷⁵

(i) Erwägungen zugunsten der betroffenen Urheber

Gegen eine Vereinbarkeit der Freistellung mit dem Dreistufentest spricht zunächst, dass die Vervielfältigungen intermediärer Trainingsdatensammler unter Umständen der Vorbereitung rechtswidriger Nutzungen dienen.

473 In diese Richtung wohl auch *Leistner*, GRUR 2024, 1665 (1667), der eine Privilegierung zur Erstellung von Basisdatensammlungen begrüßt. Andere Ansicht zu § 60d UrhG a. F.: *Stieper*, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 60d Rn. 13.

474 A. A. wohl *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 7; wie hier wohl implizit *Dworschak*, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 21.

475 Dazu methodisch ausführlicher siehe oben unter 3. Teil A. III. 2. d).

Denn die Schrankenbestimmung umfasst weder die Verbreitung noch die öffentliche Wiedergabe der Werke zum Zwecke des Text und Data Minings.

Darüber hinaus intensiviert die Tätigkeit intermediärer Trainingsdatensammler die Nutzung betroffener Werke als KI-Trainingsdaten. Hieraus könnte auch eine intensivierete wirtschaftliche Beeinträchtigung der Urheber folgen. Zudem sind am Sachverhalt beim Hinzutreten eines intermediären Trainingsdatensammlers zwei (zumeist) wirtschaftlich tätige Akteure – der vorgelagert tätige intermediäre Trainingsdatensammler einerseits sowie der KI-Entwickler als dessen Abnehmer andererseits – beteiligt. Aus der Verwendung eines Werks als Trainingsdatum wird in diesen Fallkonstellationen deswegen mehr Wert geschöpft.

(ii) Erwägungen zugunsten intermediärer Trainingsdatensammler

Intermediären Trainingsdatensammlern muss allerdings zugutegehalten werden, dass ein zwingender Schluss von ihren Vorbereitungshandlungen, also den in Rede stehenden Vervielfältigungen, auf eine nachfolgende, rechtswidrige Nutzung nicht gezogen werden kann. Zwar wirkt die Ausgestaltung des § 44b UrhG in Bezug auf ihr Geschäftsmodell prohibitiv.⁴⁷⁶ Das bedeutet aber noch nicht unweigerlich, dass ihre Betätigungen in Urheberrechtsverletzungen münden. Bereits im Rahmen dieser Arbeit konnte ein Raum für urheberrechtskonforme Geschäftsmodelle aufgezeigt werden.⁴⁷⁷ Auch erste Gerichtsverfahren zur Anwendung der Schrankenbestimmung aus § 44b UrhG auf die Erstellung von Trainingsdatensätzen liefern Ansatzpunkte hierfür.⁴⁷⁸ Letztendlich liegt es in der Verantwortung des intermediären Trainingsdatensammlers, auf Grundlage des urheberrechtlichen Möglichkeitsraums rechtskonforme Geschäftsmodelle zu entwickeln. Es sind keine Gründe ersichtlich, warum ihm dies mittels einer pauschalen Schuldvermutung von vornherein versagt werden sollte.

In diesem Zusammenhang kann auch festgestellt werden, dass die Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a Nr. 1 UrhG die Rechtmäßigkeit der Vervielfältigungshandlung eines Vermittlers von der Rechtmäßigkeit der Werknutzung desjenigen Dritten, der den Vermittlungsdienst nutzt, entkoppelt. Hieran wird deutlich,

476 Siehe dazu unten unter 3. Teil A. III. 2. g) bb) (7).

477 Siehe dazu unten unter 3. Teil A. III. 2. g) bb) (7).

478 Siehe LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 - Laion.

dass technische Vorbereitungshandlungen nach dem UrhG nicht unweigerlich das Schicksal der Rechtmäßigkeit nachgelagerter Nutzungshandlungen teilen. Dies muss auch in Ein-Personen-Konstellationen und damit auch für die Tätigkeit intermediärer Trainingsdatensammler gelten.

Gegen eine ungebührliche Verletzung der Urheberinteressen spricht weiterhin, dass das UrhG den wirtschaftlichen Wert des Werkes als Trainingsdatum dem urheberrechtlichen Ausschließlichkeitsrecht durch die Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) entzogen hat. Auch die Tätigkeit intermediärer Trainingsdatensammler beschränkt sich darauf, diesen Wert zu realisieren. Die Wertschöpfung erstreckt sich insbesondere nicht auf andere Verwertungsmöglichkeiten des Werkes. Vor diesem Hintergrund kann es keinen Unterschied machen, ob der Wert allein durch KI-Entwickler oder eben auch durch intermediäre Trainingsdatensammler realisiert wird. Zudem werden die Werke auch nur „für das KI-Training“ angefertigt. Damit ordnet sich die Nutzung intermediärer Trainingsdatensammler als dieselbe urheberrechtliche Nutzungsart ein wie die Nutzung durch KI-Entwickler. Weiterhin sind auch von der Tätigkeit intermediärer Trainingsdatensammler nicht die Kernverwertungsarten der betroffenen Urheber berührt.

Ohnehin können KI-Entwickler selbst in personeller Hinsicht uneingeschränkt von der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) profitieren und eigenhändig Vervielfältigungen von Werken für das Training von KI herstellen. Die Tätigkeit intermediärer Trainingsdatensammler beschränkt sich darauf, diese bereits freigestellten Nutzungen zu ermöglichen. Das kann auch als vorgelagerte Hilfstätigkeit eingeordnet werden, wobei der Einsatz von Hilfspersonen im Urheberrecht grundsätzlich mitprivilegiert ist.⁴⁷⁹

Weiterhin kommt es im Wirtschaftszweig der KI-Entwicklung – wie in vielen anderen Lebensbereichen auch – zur arbeitsteiligen Ausgestaltung von Prozessen. Intermediäre Trainingsdatensammler sind Teil dieser arbeitsteilig organisierten KI-Entwicklung. Sie ermöglichen beispielsweise die Betätigung kleinerer, ressourcenärmerer KI-Entwickler. Sie gehören damit als integraler Bestandteil zum Ökosystem der KI-Entwicklung. Dies bewirkt, dass sie auch vom Innovationsförderungsgedanken aus ErwGr. 18 DSM-RL mitumfasst sind.

Letztlich profitieren von einer Erfassung intermediärer Trainingsdatensammler zudem auch die Urheber mittelbar. Denn eine konsolidierte

479 Siehe dazu ausführlich unter 3. Teil A. III. 2. f) bb) (1) (ii) (γ).

Marktstruktur erleichtert es ihnen, den de lege lata vorgesehenen Opt-out-Mechanismus aus § 44b Abs. 3 UrhG zu bedienen und die Einhaltung ihrer Nutzungsvorbehaltserklärungen zu überwachen. Intermediäre Trainingsdatensammler, deren Augenmerk auf der Erstellung von Trainingsdatensätzen liegt, könnten zudem mehr Know-How aufbauen und Ressourcen aufwenden, um Nutzungsvorbehalte zu beachten und auch darüber hinaus urheberrechtskonforme Trainingsdatenakquise zu betreiben. Hierzu gehört auch, qualitativ hochwertige Trainingsdatensätze zu erstellen, deren Nutzung nicht zuletzt das Risiko für Urheberrechtsverletzungen im Output eines KI-Systems verringert. Auch eine mögliche gesetzliche Vergütungspflicht könnte de lege ferenda leichter gegenüber wenigen größeren Marktteilnehmern durchgesetzt werden.

(6) Schlussfolgerungen

Insgesamt streiten bessere Gründe gegen eine unzumutbare Beeinträchtigung der Urheber. Der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL steht einer Anwendung der Schrankenbestimmung aus § 44b UrhG auf intermediäre Trainingsdatensammler daher nicht entgegen. Im Ergebnis können auch intermediäre Trainingsdatensammler daher von der Schrankenwirkung des § 44b UrhG profitieren, solange die von ihnen angefertigten Vervielfältigungsstücke im Zeitpunkt ihrer Herstellung der Ermöglichung des Text und Data Minings und damit insbesondere des KI-Trainings gewidmet sind.⁴⁸⁰

(7) Auswirkungen der Schrankenbestimmung auf die Geschäftsmodelle intermediärer Trainingsdatensammler

Eine vom personellen Anwendungsbereich der Schrankenbestimmung unabhängig zu beantwortende Fragestellung ist, welche urheberrechtskonformen Geschäftsmodelle § 44b UrhG den intermediären Trainingsdatensammlern ermöglicht. Diese der Speicherung von Werken als Trainingsdaten nachgelagerte Problemstellung ist nicht Teil des Forschungsgegenstands der vorliegenden Untersuchung. Anzumerken ist allerdings, dass § 44b UrhG zwar die Vervielfältigung, nicht aber die Verbreitung oder öffent-

480 A. A. wohl Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 7; wie hier wohl implizit Dworschak, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 21.

liche Wiedergabe der vervielfältigten Werke gestattet. Das schränkt die Möglichkeiten für eine urheberrechtskonforme Betätigung intermediärer Trainingsdatensammler ein. Dennoch verbleiben hier gewisse Spielräume. Denkbar und praktisch verbreitet ist beispielsweise, dass die öffentlich wiedergegebenen Trainingsdatensätze neben Annotationen nur Links zu den erfassten, im Internet abrufbaren Werken enthalten.⁴⁸¹ Diese Linklisten können durch gewinnwirtschaftlich handelnde intermediäre Trainingsdatensammler urheberrechtlich frei wiedergegeben werden, insofern nicht auf rechtswidrig öffentlich zugänglich gemachte Werke verlinkt wird.⁴⁸² Die vom intermediären Trainingsdatensammler angefertigten Vervielfältigungsstücke werden in diesem Fall nur zur Herstellung der Annotationen benötigt und können danach wieder gelöscht werden.

Auch das Verbreitungsrecht aus § 17 UrhG wird nicht durch jede Weitergabe hergestellter Trainingsdatensätze berührt. Möglich ist beispielsweise eine konzerninterne Bereitstellung von Trainingsdatensätzen durch konzernzugehörige Dienstleister, die als Intermediäre Trainingsdatensätze für Konzernzwecke zusammenstellen.⁴⁸³

In Betracht käme darüber hinaus auch, dass ein Teil des Trainingsprozesses an den intermediären Trainingsdatensammler ausgelagert wird, sodass es nicht zur Verbreitung eines von ihm erstellten, urheberrechtlich relevanten Trainingsdatensatzes im Sinne des § 17 UrhG kommt. Hierfür könnte der auftraggebende KI-Entwickler gegebenenfalls auch Rechenkapazität bereitstellen.

481 Ein solcher Sachverhalt lag dem Hamburger Urteil im Fall „Laion“ zugrunde, siehe LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

482 Vgl. zu den Voraussetzungen an eine öffentliche Wiedergabehandlung beim Linking EuGH, Urt. v. 08.06.2016 - C-160/15, GRUR 2016, 1152 Rn.39 ff. – GS Media; es werden jedoch Überlegungen angestellt, auf die automatisierte Erstellung von Linklisten gewisse Privilegierungen anzuwenden, die für Suchmaschinenbetreiber entwickelt worden sind, siehe Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 97 Rn. 167 ff.

483 Zu konzerninternen Weitergaben und dem Verbreitungsrecht aus § 17 UrhG siehe BGH, Urt. v. 06.05.1981 - I ZR 92/78, GRUR 1982, 100 (101f.) – Schallplattenexport; Dustmann, in: Fromm/Nordemann, Urheberrecht, UrhG § 17 Rn.19; Raue, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 17 Rn. 35; Loewenheim, in: Schricker/Loewenheim, Urheberrecht, UrhG § 17 Rn. 41, jeweils alle m. w. N.

h) Löschungspflicht der Trainingsdatensammler

§ 44b Abs. 2 S. 2 UrhG setzt weiterhin voraus, dass die angefertigten Vervielfältigungsstücke gelöscht werden müssen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.

Vor dem Hintergrund dieser Löschungspflicht ist insbesondere klärungsbedürftig, ob auch intermediäre Trainingsdatensammler praktisch von der Schrankenbestimmung des § 44b UrhG profitieren können. Denn sie sammeln die Daten zunächst nur und bereiten sie zu Trainingsdatenkorpora auf. Erst später werden die Daten einem dritten KI-Entwickler zur Verfügung gestellt, wobei Zeitpunkt der Datenweitergabe und Identität des Abnehmers beim Web Scraping noch nicht bekannt sein müssen. Inwieweit intermediäre Trainingsdatensammler von der Schrankenbestimmung profitieren können, hängt davon ab, wann die gesetzlich geregelte Löschungspflicht eingreift.

aa) Zeitpunkt des Eingreifens der Löschungspflicht

Um den Zeitpunkt zu bestimmen, ab dem die Löschungspflicht für die vervielfältigten Werke in Kraft tritt, kommt es auf den in § 44b Abs. 2 S. 2 UrhG angelegten Bezugspunkt der Verpflichtung zur Löschung der Werke an. Denkbar ist zum einen, § 44b Abs. 2 S. 2 UrhG so auszulegen, dass er sich generell auf das Text und Data Mining als Form der Analysetechnik bezieht. Dies legt auch die Formulierung in Art. 4 DSM-RL nahe, die sich nicht auf *ein konkretes* Text und Data Mining-Vorhaben, sondern nur pauschal auf Text und Data Mining als solches bezieht.⁴⁸⁴ Folge dieser Auslegung wäre, dass die Löschungspflicht so lange nicht entstehen würde, wie die Daten grundsätzlich noch für etwaige Mining-Aktivitäten beziehungsweise KI-Training erforderlich sind. Jedenfalls unter der Voraussetzung, dass grundsätzlich noch KI-Systeme trainiert werden, sind Trainingsdatensätze noch so lange für ein potenzielles KI-Training notwendig, wie sie auf dem Trainingsdatenmarkt angeboten werden. Für intermediäre Trainingsdatensammler bedeutete das, dass sie Trainingsdatensätze erstellen und für KI-Entwickler auf unbestimmte Zeit bereithalten könnten. Die durch § 60d Abs. 5 UrhG beziehungsweise Art. 3 Abs. 2 DSM-RL gewährte Aufbewahrungserlaubnis zum Zwecke (nicht-kommerzieller) wissenschaftlicher For-

484 Wulf/Löbeth, GRUR 2024, 737 (740).

schung, die in § 44b UrhG beziehungsweise Art. 4 DSM-RL hingegen nicht vorgesehen ist, legt aber nahe, dass im Umkehrschluss hierzu ein strengeres Verständnis der Löschungspflicht angezeigt ist. Danach bezieht sich § 44b Abs. 2 S. 2 UrhG auf einen konkreten Analysevorgang. Die gespeicherten Werke müssen demzufolge bereits gelöscht werden, sobald sie für *diesen* Analysevorgang nicht mehr benötigt werden. Das ist dann der Fall, wenn die konkrete Datenanalyse beendet ist.⁴⁸⁵ Für das KI-Training bedeutet das, dass die Trainingsdaten gemäß § 44b Abs. 2 S. 2 UrhG unmittelbar nach dem Abschluss eines Trainingsvorgangs gelöscht werden müssen. Nach einer zweckorientierten Auslegung, die die Entwicklung neuer Technologien und Innovationen im Allgemeinen fördert⁴⁸⁶ und damit auch Nachbesserungen an den Systemen ermöglichen sollte, ist das jedenfalls dann der Fall, wenn das KI-System in die Anwendungsphase übergeht.⁴⁸⁷ Insoweit die Werke noch während der Anwendungsphase des Systems verwendet werden müssen, um am selben KI-System durch fortlaufendes Training Nachbesserungen vorzunehmen, greift die Löschungspflicht erst nach Abschluss dieser Nachbesserungen ein.⁴⁸⁸ Da Qualitätssicherungs- und Transparenzinitiativen nicht mehr zur jeweiligen automatisierten Analyse, also dem Training des konkreten KI-Systems, gehören, können die Werke auf Basis von § 44b UrhG hierfür allerdings nicht aufbewahrt werden.

Für intermediäre Trainingsdatensammler bedeutet das, dass sie ihren Trainingsdatensatz gemäß § 44b Abs. 2 S. 2 UrhG vernichten müssen, sobald mit ihm das Training eines KI-Systems abgeschlossen worden ist. Da

485 Vgl. zu diesem Ansatz in *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44b Rn. 15.

486 ErwGr. 18 DSM-RL.

487 Für ein solch weites Verständnis aufgrund derselben zweckorientierten Erwägungen auch *Novelli u. a.*, *Computer Law & Security Review* 2024, (10); wohl außerdem *Bullinger*, in: *Wandtke/Bullinger*, Urheberrecht, UrhG § 44b Rn. 9, denn der Zweck des Minings, also des KI-Trainings, ist erreicht, wenn das KI-System in die Anwendungsphase übergeht; insoweit übereinstimmend, als dass eine zweckorientierte, weitere Auslegung des Tatbestandsmerkmals erfolgen muss, auch *Bomhard*, in: *Ahlberg/Götting/Lauber-Rönsberg*, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 22, im Detail aber abweichend; ähnlich ohne direkten KI-Bezug auch *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44b Rn. 15, der im Ergebnis von einem projektbezogenen Verständnis ausgeht und innerhalb eines solchen Projekts auch mehrere Analysevorgänge als rechtmäßig erachtet.

488 Folgerichtig *Bomhard*, in: *Ahlberg/Götting/Lauber-Rönsberg*, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 22b; *Konertz/Schönhof*, WRP 2024, 289 Rn. 32; a. A. *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44b Rn. 15; *Wulf/Löbeth*, GRUR 2024, 737 (740).

dies dem Sinn des Datensammelns zuwiderläuft, wirkt die Löschungspflicht für intermediäre Trainingsdatensammler prohibitiv.⁴⁸⁹ Ähnliche Effekte bestehen für Trainingsdatensätze selbst aufbauende KI-Entwickler, die die unter erheblichem Ressourceneinsatz vorbereiteten Trainingsdatensätze gemäß § 44b Abs. 2 S. 2 UrhG ebenfalls nur einmalig verwenden dürfen.

bb) Regulatorische Defizite des geltenden Urheberrechts

Die Pflicht zur Löschung von mittels Webcrawlern generierter, urheberrechtlich geschützter Trainingsdaten ist, wie im zweiten Teil dieser Untersuchung gezeigt wurde, nicht erforderlich, um einen angemessenen Ausgleich von Urheber- und Nutzerinteressen herbeizuführen. Werden die als Trainingsdaten genutzten Vervielfältigungsstücke nicht gelöscht, wird mit den Daten darüber hinaus weiteres KI-Training ermöglicht. Die Löschungspflicht beeinträchtigt außerdem Qualitäts- und Nachweisanforderungen bei der Entwicklung von KI.⁴⁹⁰ Alles in allem wird die in § 44b Abs. 1 S. 2 UrhG vorgesehene Löschungspflicht den identifizierten Regelungsbedürfnissen damit nicht gerecht.⁴⁹¹ Sie beeinträchtigt die Interessen der Trainingsdatensammler und KI-Entwickler unangemessen. Das gilt insbesondere für intermediäre Trainingsdatensammler, deren Geschäftsmodell durch die Löschungspflicht erheblich gestört wird. De lege lata könnte dem sich aus der Löschungspflicht ergebenden regulatorischen Defizit mit Hilfe eines Erwerbs von ECL für das Training von KI abgeholfen werden.⁴⁹²

i) Opt-out-Mechanismus für Rechtsinhaber

§ 44b Abs. 3 UrhG sieht für Urheber die Möglichkeit vor, die Freistellungswirkung der Schrankenbestimmung durch Erklärung eines Nutzungsvorbe-

489 So auch Kögel, InTeR 2023, 179 (181).

490 Margoni/Kretschmer, GRUR Int. 2022, 685 (693); Vgl. auch Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 23, die das Problem aber durch eine de facto Aufhebung der Löschungspflicht lösen wollen, die die Wortlautgrenze der Regelung sprengt und daher nicht mit den gängigen Auslegungsmethoden begründet werden kann. Mit rechtspolitischer Kritik noch in Bomhard, InTeR 2023, 174 (177).

491 Dass die Löschungspflicht für den Umgang mit Trainingsdaten für KI problematisch sei, betont auch de la Durantaye, ZUM 2023, 645 (659). Für eine Streichung plädiert auch Schack, NJW 2024, 113 Rn. 10.

492 Dazu im Detail siehe Pukas, GRUR 2023, 614.

halts zu beenden. Bei online zugänglichen Werken muss die Ausschlussklärung gemäß § 44b Abs. 3 S. 2 UrhG hierfür in maschinenlesbarer Form erfolgen.⁴⁹³

Wie im zweiten Teil der vorliegenden Untersuchung gezeigt werden konnte, ist aus einer regulatorischen Perspektive kein Opt-out-Mechanismus für Urheber vorzusehen. Ein angemessener Ausgleich zwischen Urheber- und Nutzerinteressen kann vielmehr auch ohne Ausschlussmöglichkeit, sondern durch eine gesetzliche Vergütungspflicht sichergestellt werden. Der in § 44b Abs. 3 UrhG geregelte Opt-out-Mechanismus wird den identifizierten Freistellungsbedürfnissen demzufolge nicht gerecht.⁴⁹⁴ Wie stark der Widerspruch zwischen den herausgearbeiteten Regelungsbedürfnissen und den geltenden urheberrechtlichen Rahmenbedingungen ausgeprägt ist, hängt im Folgenden von den konkreten Voraussetzungen des Nutzungsvorbehalts und seiner Wirkung ab.

aa) Maschinenlesbarkeit des Nutzungsvorbehalts

Damit ein Nutzungsvorbehalt bei online zugänglichen Werken wie den vom Web Scraping urheberrechtlich geschützter Trainingsdaten betroffenen wirksam ist, muss er gemäß § 44b Abs. 3 S. 2 UrhG in maschinenlesbarer Form erfolgen. Im UrhG wird allerdings nicht näher konkretisiert, was hierunter zu verstehen ist. Der Rechtsbegriff ist daher auslegungsbedürftig.

(1) Notwendigkeit einer digitalen Hinterlegung der Informationen

Fraglich ist dabei zunächst, ob die mit dem Nutzungsvorbehalt übermittelten Informationen digital in einer Art und Weise hinterlegt sein müssen, die ihre unmittelbare Verarbeitung durch einen Computer ermöglicht, damit sie im Sinne des § 44b Abs. 3 S. 2 UrhG maschinenlesbar sind. Ist das nicht der Fall, genügt eine mittelbare maschinelle Verarbeitbarkeit, die erst durch eine Digitalisierung des Erklärungsinhalts (z. B. Einscannen einer handschriftlichen Notiz) erreicht wird.

493 Zu den Anforderungen an den Nutzungsvorbehalt siehe sogleich unter 3. Teil A. III. 2. i) aa).

494 So konkret zu Art. 4 DSM-RL auch *Chiou*, JIPITEC 2019, 398 Rn. 23 f.

Durch erhebliche Fortschritte im Bereich maschinellen Lernens⁴⁹⁵ können nahezu alle auch nicht elektronisch dargestellten Informationen wie handgeschriebene Texte digitalisiert und damit mittelbar maschinell verarbeitet werden.⁴⁹⁶ Aufgrund leistungsfähiger Texterkennungssoftware gilt das insbesondere für analoge Textdaten. Bilderkennungssystemen versetzen ihre Nutzer zudem in die Lage, Bildinhalte semantisch zu erfassen. Würde eine mittelbare maschinelle Verarbeitbarkeit der Informationen ausreichen, um Maschinenlesbarkeit im Sinne des § 44b Abs. 3 UrhG herbeizuführen, bedeutete dies, dass faktisch nahezu alle Vorbehaltserklärungen unabhängig von ihrer Verkörperungsform maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG wären.⁴⁹⁷ In diesem Falle würde das Tatbestandsmerkmal der Maschinenlesbarkeit weitestgehend leerlaufen. Bereits denklogisch muss es aber einen Unterschied zwischen maschinenlesbaren und nicht maschinenlesbaren Erklärungen geben.⁴⁹⁸ Darüber hinaus kann vermutet werden, dass der Normgeber dem Tatbestandsmerkmal mit seiner Aufnahme in den Gesetzestext eine eigenständige Bedeutung verliehen hat.

Beides spricht dafür, dass ein solch weites Verständnis des Tatbestandsmerkmals der Maschinenlesbarkeit nicht intendiert ist. Insgesamt führen damit teleologische Erwägungen dazu, dass der Inhalt einer Nutzungsvorbehaltserklärung digital in einer Art und Weise hinterlegt werden muss, die eine unmittelbare Verarbeitung durch einen Computer ermöglicht, damit er als maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG gilt.⁴⁹⁹ Analoge Erklärungen wie handgeschriebene Texte, die digitalisiert werden können, erfüllen die Voraussetzung der Maschinenlesbarkeit daher nicht.

(2) Bestehen eines Formalisierungs- und Strukturierungserfordernisses

Klärungsbedürftig ist darüber hinaus aber auch, ob die Nutzungsvorbehaltserklärung in einem formalisierten und strukturierten Format, also in einer Programmier- oder vergleichbar formalisierten Sprache abgegeben werden muss, damit die Voraussetzung der Maschinenlesbarkeit im Sinne des § 44b Abs. 3 S. 1 UrhG erfüllt ist. Sofern das nicht notwendig ist, könnte

495 Vgl. *Maslej u. a.*, The AI Index 2023 Annual Report by Stanford University, S. 54 ff.

496 Vgl. zur Sprachverarbeitung *Menzel*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 601.

497 *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; *Bomhard*, InTeR 2023, 174 (178).

498 Ebenso *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332.

499 A. A. *Wulf/Löbeth*, GRUR 2024, 737 (739 f.), wohl bedingt durch ein grammatikalisches Missverständnis des § 44b Abs. 3 UrhG zugrundeliegenden Art. 4 Abs. 3 DSM-RL und entgegen dem insoweit eindeutigen Wortlaut des § 44b Abs. 3 UrhG.

eine maschinenlesbare Vorbehaltserklärung auch wirksam in natürlicher Sprache abgegeben werden, sofern die Erklärung unmittelbar durch einen Computer verarbeitet werden kann. Hierfür genügt es, wenn sie in elektronischer Form, also insbesondere als elektronischer Text, verkörpert ist.

Auch diese Frage muss im Wege der Auslegung des Tatbestandsmerkmals der Maschinenlesbarkeit aus § 44b Abs. 3 S. 2 UrhG beantwortet werden.⁵⁰⁰ Im Rahmen dieser Auslegung ist die Bedeutung des Tatbestandsmerkmals dabei insbesondere im Lichte des Art. 4 Abs. 3 DSM-RL und des hier verwendeten Begriffs der Maschinenlesbarkeit zu bestimmen. Seiner Umsetzung dient § 44b Abs. 3 S. 2 UrhG.⁵⁰¹ Weiterhin zu berücksichtigen ist, dass es für die Erklärung von Nutzungsvorbehalten in natürlicher Sprache keine einheitliche Formulierung gibt. Anders als beispielsweise § 312j Abs. 3 S. 2 BGB gibt das UrhG dem Erklärenden keinen Erklärungswortlaut vor. Aufgrund ihrer Varianz bieten natürlichsprachliche Vorbehaltserklärungen damit keine vollständige Erkennungssicherheit.⁵⁰² Folge dessen ist, dass die Auslegung des Maschinenlesbarkeitsbegriffs auch maßgeblich davon mitbeeinflusst wird, wieviel Erkennungssicherheit mit den Nutzungsvorbehaltserklärungen erreicht werden muss.

(i) Natürlicher Sprachgebrauch des Begriffs „maschinenlesbar“ bzw. „machine-readable“

Ausgangspunkt der Auslegung des Rechtsbegriffs der Maschinenlesbarkeit ist der allgemeine Sprachgebrauch. Da die Computerwissenschaft und -anwendung erheblich durch die englische Sprache geprägt ist, muss hierbei auch das Begriffsäquivalent „machine-readable means“ mit in den Blick genommen werden. Es wird in der englischen Sprachfassung des durch § 44b Abs. 3 UrhG umgesetzten Art. 4 Abs. 3 DSM-RL genutzt. Maschinenlesbar beziehungsweise machine-readable sind Daten nach einem teilweise zugrunde gelegten Begriffsverständnis dann, wenn sie in einer Form gespeichert sind, die von einem Computer verarbeitet werden kann.⁵⁰³ Volltextsuchen und Suchmaschinen zeigen bereits praktisch, dass eine Verarbeitung von digital hinterlegten, natürlich-sprachlichen Textinhalten durch Computer ohne weiteres möglich ist. Dem entsprechend könnten auch in

500 Vgl. hierzu ausführlich auch *Hamann*, ZGE 2024, 113.

501 BT-Drs. 19/27426, S. 88.

502 *Maamar*, ZUM 2023, 481 (484).

503 *Waite*, Oxford English Dictionary, Begriff: machine-readable.

natürlicher Sprache abgegebene Nutzungsvorbehaltserklärungen unter den Begriff der Maschinenlesbarkeit (beziehungsweise der machine-readability) gefasst werden.

Das aufgezeigte Verständnis wird dem Begriff allerdings nicht einhellig zugrunde gelegt. Im natürlichen Sprachgebrauch ist vielmehr auch eine Interpretation verbreitet, nach der die Daten in strukturierter Form gespeichert sein müssen, sodass sie durch ein einheitliche und gleichbleibende Verarbeitungslogik anwendendes Computerprogramm verarbeitet werden können.⁵⁰⁴ Auch wenn sie in einem digitalen Format vorgehalten werden, können hierunter mangels einer Strukturierung der Daten keine Erklärungen in natürlicher Sprache subsumiert werden. Beide Verständnismöglichkeiten zeigen, dass sich im natürlichen Sprachgebrauch kein einheitliches Verständnis des Begriffs der Maschinenlesbarkeit beziehungsweise der machine-readability herausgebildet hat. Die dargelegten Interpretationsmöglichkeiten entscheiden sich insbesondere im Hinblick auf die Notwendigkeit einer Strukturierung der Daten. Gerade hiernach richtet sich aber, ob eine Vorbehaltserklärung in natürlicher Sprache, also unstrukturiert abgegeben werden kann, oder nicht. Aus dem natürlichen Sprachgebrauch können daher keine Schlussfolgerungen für die Auslegung des Rechtsbegriffs der Maschinenlesbarkeit abgeleitet werden.

(ii) Begriffsverwendung im nationalen Regelungskontext

Der Begriff der Maschinenlesbarkeit wird allerdings in § 12 Abs. 1 S. 2 EGovG legaldefiniert. Danach sind Informationen maschinenlesbar, wenn sie durch Software automatisiert ausgelesen und verarbeitet werden können. Eine vergleichbare Legaldefinition ist auch in § 3 Nr. 5 DNG normiert. Wie die Möglichkeit zur Volltextsuche und das Suchmaschinenindexing, bei dem nicht strukturierte und formalisierte Webseiteninhalte systematisch und automatisiert ausgewertet werden, praktisch zeigen, bedarf es für eine Verarbeitung digitaler Informationen durch (automatisierte) Software keiner Strukturierung der betroffenen Inhalte. Informationen müssten für eine Maschinenlesbarkeit in diesem Sinne demzufolge digital hinterlegt, nicht aber strukturiert oder formalisiert abrufbar sein. Ein Vergleich zu den gleichlautenden Tatbestandsvoraussetzungen in EGovG und DNG spricht

504 Glossar des National Institute of Standards and Technology (U.S. Department of Commerce), Begriff: machine-readable, siehe https://csrc.nist.gov/glossary/term/Machine_Readable (zuletzt abgerufen am 14.4.2024).

deswegen dafür, das Tatbestandsmerkmal der Maschinenlesbarkeit auch in § 44b Abs. 3 S. 2 UrhG so auszulegen, dass zwar eine digitale Hinterlegung, aber keine Strukturierung oder Formalisierung der Vorbehaltserklärungen notwendig ist.⁵⁰⁵

(iii) Einheitliches europäisches Begriffsverständnis

Im Übrigen geht § 44b Abs. 3 UrhG aber auf Art. 4 Abs. 3 DSM-RL zurück und muss daher insbesondere richtlinienkonform ausgelegt werden. Auch in Art. 4 Abs. 3 DSM-RL ist der Begriff der Maschinenlesbarkeit zwar nicht ausdrücklich definiert. Es könnte sich aber bereits ein einheitliches europäisches Begriffsverständnis herausgebildet haben,⁵⁰⁶ dass auch dem Rechtsbegriff der Maschinenlesbarkeit aus Art. 4 Abs. 3 DSM-RL und damit richtlinienkonform auch § 44b Abs. 2 S. 2 UrhG zugrunde gelegt werden kann. Um dies zu ermitteln, muss herausgearbeitet werden, in welchen Gemeinschaftsrechtsakten der Begriff der Maschinenlesbarkeit verwendet wird und ob den Rechtsbegriffen in den Regelungswerken ein einheitliches Verständnis zugrunde liegt.

(a) Maschinenlesbarkeit in der PSI-Richtlinie

Maschinenlesbarkeit ist zum einen in Art. 2 Nr. 13 RL (EU) 2019/1024 (PSI-RL) definiert. Er liegt dem bereits an anderer Stelle herangezogenen⁵⁰⁷ § 3 Nr. 5 DNG zugrunde. Gemäß Art. 2 Nr. 13 PSI-RL ist ein Dateiformat maschinenlesbar, wenn es so strukturiert ist, dass Softwareanwendungen konkrete Daten, einschließlich einzelner Sachverhaltsdarstellungen und deren interner Struktur, leicht identifizieren, erkennen und extrahieren können. Unabhängig davon, welche Voraussetzungen erfüllt werden müssen, damit eine leichte Identifizierung und Extrahierung ermöglicht wird, setzt Art. 2 Nr. 13 PSI-RL eine Datenstrukturierung voraus. Eine solche Strukturierung wird durch Erklärungen in natürlicher Sprache nicht gewährleistet.

505 A. A. *Hamann*, ZGE 2024, 113 (131).

506 Vgl. für eine einheitliche Auslegung von Rechtsbegriffe innerhalb des Gemeinschaftsrechts EuGH, Urt. v. 27.01.2005 - C-188/03, NZA 2005, 213 Rn. 29 – Junk; EuGH, Urt. v. 27.02.2003 - C-373/00, EuZW 2003, 315 Rn. 35.

507 Dazu siehe unmittelbar oben in diesem Abschnitt.

(β) Maschinenlesbarkeit in der Richtlinie zur Nachhaltigkeitsberichterstattung von Unternehmen

Der Begriff der Maschinenlesbarkeit wird darüber hinaus in ErwGr. 55 RL (EU) 2022/2464 (CSRD) (Regelungsziel aufgegriffen durch Art. 29d CSRD) verwendet. Die Richtlinie regelt die Nachhaltigkeitsberichterstattung von Unternehmen. Die Dokumentation der Unternehmen soll danach maschinenlesbar sein. Zur näheren Ausgestaltung wird auf Art. 3 VO (EU) 2019/815 (Delegierte Verordnung) verwiesen. Aus diesem ergibt sich wiederum, dass die Nachhaltigkeitsberichte im xhtml-Format veröffentlicht werden müssen. Das xhtml-Format sieht dabei eine gewisse Formalisierung durch die xhtml-Syntax vor. Eine Strukturierung der Erklärungsinhalte findet hierdurch allerdings nicht statt. Der Erklärungsinhalt wird, eingebettet in die xhtml-Syntax, in natürlicher Sprache wiedergegeben. Der Begriff der Maschinenlesbarkeit, welcher in ErwGr. 55 CSRD verwendet wird, setzt demnach zwar eine formalisierte Einbettung des Erklärungsinhalts voraus. Darüber hinaus lässt er aber ausreichen, wenn der eigentliche Inhalt der Erklärung in natürlicher Sprache vorgehalten. Eine Strukturierung der Erklärung, also eine Übersetzung „in Code“, ist nicht notwendig.

(γ) Maschinenlesbarkeit in der KI-Verordnung

Auch in der KI-VO wird das Tatbestandsmerkmal der Maschinenlesbarkeit verwendet. ErwGr. 133 adressiert die maschinenlesbare Kennzeichnung von KI-generierten Inhalten. Als Beispiel für eine maschinenlesbare Kennzeichnungstechnik werden unter anderem Wasserzeichen genannt. Hierbei handelt es sich um eine Markierungstechnik, bei der selbstgewählte, identifizierbare Zeichen direkt in den Inhalt, beispielsweise ein Lichtbild, eingearbeitet werden. Im Gegensatz zur Einbettung von Informationen in Metadaten werden die Zeichen dabei unmittelbar mit dem Inhalt verflochten. Die Wasserzeichen können, müssen aber nicht mit dem bloßen Auge wahrnehmbar sein. Sie folgen keiner strukturierten oder formalisierten Logik. Sie sind nicht codiert. Vielmehr werden oft verwendertypische Zeichen wie Bildmarken oder zufällige Rauschmuster eingesetzt.

Darüber hinaus verpflichtet Art. 47 KI-VO Anbieter von Hochrisiko-KI-Systemen, maschinenlesbare EU-Konformitätserklärung auszustellen. Diese Erklärungen können physisch oder maschinell unterzeichnet werden (Abs. 1) und müssen in einer Sprache verfügbar sein, die für die jeweils zuständige, nationale Aufsichtsbehörde leicht verständlich ist (Abs. 2). Aus der

gesetzlichen Sprachanforderung sowie der Möglichkeit, die Erklärungen auch physisch zu unterzeichnen, folgt, dass die EU-Konformitätserklärungen nach Art. 47 KI-VO in natürlicher Sprache abgegeben werden müssen. Maschinenlesbarkeit im Sinne von Art. 47 KI-VO setzt demzufolge ebenso wenig eine Strukturierung wie eine Formalisierung der Erklärungsinhalte voraus. Vielmehr ist eine Erklärung in natürlicher Sprache ausreichend. Insgesamt setzt der Begriff der Maschinenlesbarkeit innerhalb der KI-VO keine Codierung der jeweiligen Inhalte voraus.

(δ) Maschinenlesbarkeit in der Datenschutz-Grundverordnung

Das Tatbestandsmerkmal der Maschinenlesbarkeit wird außerdem in der DS-GVO aufgegriffen. Gemäß Art. 12 Abs. 7 DS-GVO können bei der Datenerhebung zu übermittelnde Informationen auch in Form standardisierter Bildsymbole bereitgestellt werden. In elektronischer Form müssen diese Bildsymbole gemäß Art. 12 Abs. 7 S. 2 DS-GVO maschinenlesbar sein. Dem Regelungszweck der Herstellung von Barrierefreiheit zugunsten sehbehinderter Menschen⁵⁰⁸ entsprechend ist hierunter zu verstehen, dass die Symbole computergestützt vorlesbar sein müssen.⁵⁰⁹ Hierfür bedarf es zwar einer gewissen Formalisierung, aber keiner Strukturierung der vorzulesenden Textinhalte. Vielmehr genügt auch eine elektronisch auslesbare Erklärung in natürlicher Sprache.

Nach Art. 20 DS-GVO haben Personen außerdem das Recht, die sie betreffenden personenbezogenen Daten in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten. Die Norm erfasst die Maschinenlesbarkeit und die Strukturierung der Daten als eigenständige Merkmale. Umgekehrt folgt hieraus, dass die Voraussetzung der Maschinenlesbarkeit aus Art. 20 DS-GVO selbst ebenfalls keine Strukturierung der Daten notwendig macht.

508 *Quaas*, in: Wolff/Brink/v. Ungern-Sternberg, Beck'scher Online-Kommentar Datenschutzrecht, DS-GVO Art. 12 Rn. 54; *Heckmann/Paschke*, in: Ehmman/Selmayr, Datenschutz-Grundverordnung, DS-GVO Art. 12 Rn. 55.

509 Vgl. *Heckmann/Paschke*, in: Ehmman/Selmayr, Datenschutz-Grundverordnung, DS-GVO Art. 12 Rn. 55.

(€) Schlussfolgerung zur Existenz eines einheitlichen europäischen Begriffsverständnisses

Insgesamt wird der Rechtsbegriff der Maschinenlesbarkeit innerhalb der europäischen Rechtsordnung an verschiedenen Stellen aufgegriffen. Teilweise setzt er dabei eine Strukturierung und Formalisierung der übermittelten Informationen, teilweise aber auch weder eine Strukturierung noch eine Formalisierung der Inhalte voraus. Mitunter wird vom Normgeber sogar eine Erklärung in natürlicher Sprache implizit vorausgesetzt. Hieraus folgt, dass sich kein einheitliches europäisches Verständnis des Rechtsbegriffs der Maschinenlesbarkeit herausgebildet hat, welches Art. 4 Abs. 3 DSM-RL und im Wege richtlinienkonformer Auslegung auch § 44b Abs. 3 UrhG zugrunde gelegt werden kann.⁵¹⁰ Insbesondere kann damit auch keine Parallele zur Legaldefinition der Maschinenlesbarkeit aus Art. 2 Nr. 13 PSI-RL gezogen werden.⁵¹¹ Eine Parallele gerade zu dieser Regelung zu ziehen, wäre willkürlich. Außerdem dient sie der Erfüllung eines anderen Zweckes als Art. 4 DSM-RL.⁵¹² Mangels eines einheitlichen europäischen Begriffsverständnisses ist die Bedeutung des Rechtsbegriffs der Maschinenlesbarkeit in Art. 4 Abs. 3 DSM-RL vielmehr rechtsaktsautonom zu bestimmen. Hierfür kann sich insbesondere auch am Telos der Regelung orientiert werden.

(iv) Telos der Regelung

Zweck der Voraussetzung der Maschinenlesbarkeit aus Art. 4 Abs. 3 DSM-RL respektive § 44b Abs. 3 S. 2 UrhG ist, eine automatisierte Vervielfältigung von Werken zu ermöglichen.⁵¹³ Dafür müssen die Nutzungsvorbehalte in einer Art und Weise erklärt werden, die es automatisierten Webcrawlern ermöglicht, den Erklärungsinhalt zu erfassen und zu verarbeiten. Dass ein automatisiertes Auffinden und infolgedessen eine Verarbeitung von Textinhalten in natürlicher Sprache technisch möglich ist, zeigen bereits

510 A. A. Griffiths/Synodinou/Xalabarder, GRUR Int. 2023, 22 (29); Bomhard, David/Siglmüller, RD 2024, 45 (50) Fn. 45.

511 Wie hier auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 69; a. A. Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 31.

512 Vgl. dazu ausführlicher LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 69.

513 Auch Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 10; BT-Drs. 19/27426.

der Betrieb von Suchmaschinen und die Suchfunktion in Dokumenten. Auch HTML-Code kann mit Hilfe simpler Skripte automatisiert ausgewertet und nach bestimmten Sprachformulierungen durchsucht werden. Solche Suchalgorithmen können auch für die Verarbeitung von Inhalten genutzt werden, die mit Hilfe automatisierter Webcrawler für das Training von KI vervielfältigt werden sollen. Eine automatisierte Verarbeitung ist somit möglich, solange die Daten digital hinterlegt und insofern durch Maschinen verarbeitbar sind. In natürlicher Sprache abgegebene Nutzungsvorbehaltserklärungen stehen damit einer automatisierten Vervielfältigung aus einer technischen Perspektive nicht per se entgegen.⁵¹⁴ Dass die Informationen strukturiert und formalisiert vorgehalten werden, ist für eine automatisierte Verarbeitung aus einer technischen Perspektive demzufolge nicht notwendig.⁵¹⁵ Die Frage nach der Angemessenheit der konkreten Nutzungsvorbehaltserklärung im weiteren Sinne ist dabei von der Frage der Maschinenlesbarkeit der Erklärung zu trennen.⁵¹⁶

Nach ErwGr. 18 UAbs. 2 S. 2 DSM-RL sollen darüber hinaus in Allgemeine Geschäftsbedingungen einer Webseite eingebettete Nutzungsvorbehaltserklärungen die Voraussetzungen des Art. 4 Abs. 3 DSM-RL erfüllen können. Zwar bezieht sich der Erwägungsgrund ausweislich seiner Formulierung nur auf in den Allgemeinen Geschäftsbedingungen einer Webseite eingebettete Vorbehaltserklärungen, *sofern* diese maschinenlesbar sind. Daraus kann aber nicht geschlossen werden, dass im Umkehrschluss aus ErwGr. 18 UAbs. 2 S. 2 DSM-RL Vorbehaltserklärungen in natürlicher Sprache grundsätzlich nicht maschinenlesbar seien.⁵¹⁷ Denn es folgt aus der allgemeinen Lebenserfahrung, dass Erklärungen innerhalb von Allgemeinen Geschäftsbedingungen einer Webseite so wie die Vertragsbestimmungen selbst in natürlicher Sprache abgegeben werden. Das muss auch dem Richtliniengeber bei der Formulierung des ErwGr. 18 DSM-RL bewusst gewesen sein. Dass nach ErwGr. 18 UAbs. 2 S. 2 DSM-RL Vorbehaltserklärungen in Allgemeinen Geschäftsbedingungen einer Webseite grundsätzlich geeignet sein *können*, die Voraussetzungen aus Art. 4 Abs. 3 DSM-RL zu erfüllen, spricht demzufolge dafür, dass die Abgabe der Vorbehaltserklärung in natürlicher Sprache der Maschinenlesbarkeit im Sinne des Art. 4 Abs. 3

514 Auch Leistner verweist auf die Leistungsfähigkeit moderner Sprachverarbeitungssysteme, siehe *Leistner*, GRUR 2024, 1665 (1673).

515 A. A. *Maamar*, ZUM 2023, 481 (484); *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; wohl auch *Hamann*, ZGE 2024, 113 (131).

516 Zur Angemessenheit sogleich unter 3. Teil A. III. 2. i) bb).

517 So aber *Hamann*, ZGE 2024, 113 (147).

DSM-RL jedenfalls nicht entgegenstehen kann. Ansonsten wäre ErwGr. 18 UAbs. 2 S. 2 DSM-RL gegenstandslos. Umgekehrt kann damit auch aus ErwGr. 18 UAbs. 2 S. 2 DSM-RL abgeleitet werden, dass eine digitale Hinterlegung des Nutzungsvorbehalts ausreichend sein muss. Eine Strukturierung oder Formalisierung der Informationen wäre demnach nicht erforderlich.

Für ein Formalisierungs- und Strukturierungserfordernis der Vorbehaltserklärungen könnte jedoch sprechen, dass eine formalisierte und strukturierte Informationsübermittlung, beispielsweise in Form von HTML-Metadaten,⁵¹⁸ zur größtmöglichen Erkennungs- und damit Rechtssicherheit für den Erklärungsempfänger führt. Allerdings bezweckt Art. 4 Abs. 3 DSM-RL respektive § 44b Abs. 3 UrhG nicht, die größtmögliche Rechtssicherheit für Trainingsdatensammler herbeizuführen. Um Abgrenzungsschwierigkeiten im Anwendungsbereich der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen, Art. 5 Abs. 1 InfoSoc-RL respektive § 44a UrhG zu beenden,⁵¹⁹ sollte Rechtssicherheit lediglich durch das Ob der Freistellung von Vervielfältigungshandlungen zum Text und Data Mining, nicht aber den Opt-out-Mechanismus im Speziellen herbeigeführt werden.

In diesem Zusammenhang spricht außerdem gegen die Notwendigkeit einer Formalisierung oder Strukturierung der Vorbehaltserklärungen, dass ohnehin nicht jede Software alle formalisierten Erklärungsmuster auslesen und verarbeiten kann. Ein Trainingsdatensammler, der HTML-Metadaten auswertet, nimmt beispielsweise keine Vorbehaltserklärung zur Kenntnis, die mittels des Robots Exclusion Standards abgegeben worden ist. Ohne gesetzlich normierten Erklärungsstandard besteht selbst bei einer Strukturierung und Formalisierung der Erklärungsmuster damit immer das Risiko, dass Nutzungsvorbehaltserklärungen nicht ausgelesen und verarbeitet werden können.⁵²⁰ Unabhängig davon, ob eine Strukturierung und Formalisierung der Vorbehaltserklärungen erfolgt, oder ob der Nutzungsvorbehalt in natürlicher Sprache erklärt wird, gibt es demzufolge keine vollständige Erkennungssicherheit für Trainingsdatensammler. Ein Grundrisiko ist damit immer vom Datensammler zu tragen. Eine Formalisierung oder Strukturierung führt nicht zwangsläufig zur Eliminierung dieses Risikos.

Darüber hinaus müssen Anbieter von KI-Modellen mit allgemeinem Verwendungszweck nach Art. 53 Abs. 1 lit. c KI-VO eine Strategie auf den

518 Mit einem Überblick über diese Technologie *Hamann*, ZGE 2024, 113 (157 ff.).

519 Vgl. ErwGr. 18 UAbs. 1 DSM-RL.

520 *Wulf/Löbeth*, GRUR 2024, 737 (738).

Weg bringen, auf Grundlage derer Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG ausgelesen und beachtet werden sollen. Hierfür muss gemäß der gesetzlichen Anforderung auch auf „modernste Technik“ zurückgegriffen werden. Diese wird aber nur dann benötigt, wenn in natürlicher Sprache erklärte Vorbehaltserklärungen digital erfasst und semantisch verarbeitet werden sollen.⁵²¹ Auch das spricht im Ergebnis gegen eine Formalisierungs- und Strukturierungsnotwendigkeit der Vorbehaltserklärungen.

Zweck des Opt-out-Mechanismus aus Art. 4 Abs. 3 DSM-RL ist außerdem nicht, größtmögliche Sicherheit für Trainingsdatensammler herzustellen. Der Opt-out-Mechanismus dient vorrangig der Wahrung der Interessen der Rechtsinhaber. Dem Interesse der Rechtsinhaber wird umgekehrt aber eine niederschwellige Möglichkeit zur Erklärung des Nutzungsvorbehalts gerecht. Eine solche wird durch die Erklärbarkeit von Nutzungsvorhalten in natürlicher Sprache gewährleistet.

Gegen das Erfordernis der Formalisierung und Strukturierung von Vorbehaltserklärungen spricht zudem, dass dies zu einem Ausschluss solcher Rechtsinhaber führen würde, die nicht in der Lage sind, die Strukturierungs- und Formalisierungsmechanismen zu bedienen. Das ist insbesondere dann der Fall, wenn die Rechtsinhaber nicht über die notwendigen technischen Fähigkeiten verfügen. Da die Schrankenbestimmung für Text und Data Mining einen Opt-out-Mechanismus vorsieht, sollte jedoch allen Rechtsinhabern unabhängig von ihrer technischen Vorbildung ermöglicht werden, diesen Mechanismus auch auszuüben. Das gilt im Weiteren auch für solche Rechtsinhaber, die auf die technische Infrastruktur eines Dritten, beispielsweise eines Content-Management-Systemanbieters, zurückgreifen. Ihre Möglichkeit zur Erklärung des Nutzungsvorbehalts darf nicht davon abhängen, inwieweit der Dienstleister dem Rechtsinhaber Einfluss auf die technische Infrastruktur gewährt, also ob er ihm beispielsweise die Möglichkeit eröffnet, HTML-Metadaten zu bearbeiten.

Für eine Erklärbarkeit des Nutzungsvorbehalts in natürlicher Sprache spricht schließlich auch, dass auch der nationale Umsetzungsgesetzgeber davon ausging, dass der Nutzungsvorbehalt im Impressum oder den Nutzungsbedingungen einer Webseite erklärt werden kann.⁵²² Auch dafür ist notwendig, dass die Vorbehaltserklärung in natürlicher Sprache abgegeben werden können muss.

521 LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 67.

522 BT-Drs. 19/27426, S. 89.

(v) Systematik innerhalb des nationalen Regelungsgefüges

Weiterhin möglich wäre, einen systematischen Schluss aus § 49 UrhG zu ziehen. Denn die Schrankenbestimmung ermöglicht gemäß § 49 Abs. 1 S. 1 UrhG die Nutzung einzelner Rundfunkkommentare sowie Artikel aus Zeitungen und Informationsblättern zu privilegierten Zwecken, wenn die geschützten Inhalte nicht mit einem Vorbehalt der Rechte versehen sind. Allerdings ist der in § 49 UrhG angelegte Vorbehaltsmechanismus zwar funktional mit dem Nutzungsvorbehalt aus § 44b Abs. 3 UrhG vergleichbar. Nichtsdestotrotz sind beide Tatbestände nicht identisch formuliert. Insbesondere sieht § 49 UrhG keine Voraussetzung der Maschinenlesbarkeit vor. Die Schrankenbestimmung adressiert zudem vorwiegend analoge Nutzungen, während § 44b Abs. 3 S. 2 UrhG auf Online-Nutzungen von Werken zugeschnitten ist. Beide Sachverhaltskonstellationen unterscheiden sich demzufolge voneinander. Ein systematischer Schluss aus § 49 UrhG kann deswegen nicht gezogen werden.

(vi) Schlussfolgerung

Insgesamt sprechen damit die Legaldefinitionen des Rechtsbegriffs der Maschinenlesbarkeit im EGovG und DNG, vor allem aber durchgreifende teleologische Gründe auch auf Ebene des umgesetzten Unionsrechts dafür, das Tatbestandsmerkmal der Maschinenlesbarkeit im Sinne des § 44b Abs. 3 S. 2 UrhG so auszulegen, dass keine Formalisierung oder Strukturierung des Erklärungsinhalts notwendig sind. Ein einheitliches europäisches Begriffsverständnis steht dem nicht entgegen. Ausreichend ist eine digitale Hinterlegung der Vorbehaltserklärung, mittels derer die Erklärung aus einer technischen Perspektive unmittelbar durch den Computer verarbeitet werden kann. Um maschinenlesbar und damit wirksam zu sein, müssen Nutzungsvorbehalte im Sinne des § 44b Abs. 3 S. 2 UrhG daher nicht „in codierter Form“ erklärt werden. Die Vorbehaltserklärung kann vielmehr auch in natürlicher Sprache abgegeben werden.⁵²³

523 Wie hier auch *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 16; *Steinrötter/Schauer*, in: Barudi, Das neue Urheberrecht, § 14 Text und Data Mining Rn. 14; *Akinci/Heidrich*, IPRB 2023, 270 (272); *Leistner*, GRUR 2024, 1665 (1673); wohl ebenso *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 9; *Schippan*, ZUM 2024, 670 (675), wobei es hier etwas unklar bleibt; BT-Drs. 19/27426, S. 89; dazu tendierend auch LG Hamburg, Urt. v. 27.09.2024 - 310

bb) Angemessenheit des Nutzungsvorbehalts

Gemäß Art. 4 Abs. 3 DSM-RL müssen die Werke zudem in angemessener Weise mit dem Nutzungsvorbehalt versehen werden.

(1) Verhältnis zwischen Angemessenheits- und Maschinenlesbarkeitsvoraussetzung

Im Richtlinientext wird beispielhaft konkretisiert („*etwa*“), dass die Angemessenheitsvoraussetzung bei online veröffentlichten Inhalten zur Folge hat, dass die Nutzungsvorbehalte in maschinenlesbarer Form abgegeben werden müssen. Die Voraussetzung der Maschinenlesbarkeit, welche in § 44b Abs. 3 S. 2 UrhG umgesetzt worden ist, leitet sich im Richtlinientext damit aus der Grundvoraussetzung der Angemessenheit ab. Da in der Richtlinie zwar eine beispiel-, aber keine regelgebende Formulierung („*in der Regel angemessen, wenn*“; „*insbesondere dann angemessen, wenn*“; „*Angemessenheit liegt in der Regel vor, wenn*“) verwendet wird, kommt der Voraussetzung der Maschinenlesbarkeit damit keine Indizwirkung für die Erfüllung der Angemessenheitsvoraussetzung als Regelbeispiel zu. Hieraus folgt, dass Vorbehaltserklärungen gegen die Nutzung von online veröffentlichten Werken gemäß Art. 4 Abs. 3 DSM-RL aufgrund des Angemessenheitserfordernisses zwar maschinenlesbar sein müssen. Aus der Maschinenlesbarkeit der Vorbehaltserklärung kann umgekehrt allerdings noch nicht geschlussfolgert werden, dass die Erklärung aufgrund ihrer Maschinenlesbarkeit bereits angemessen im Sinne des Art. 4 Abs. 3 DSM-RL ist. Die Angemessenheit der Vorbehaltserklärung bleibt vielmehr als eigene Tatbestandsvoraussetzung bestehen.

O 227/23, GRUR-RS 2024, 25458 Rn. 66; a. A. hingegen *Hamann*, ZGE 2024, 113 (144); im Ergebnis auch *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; *Gausling*, CR 2021, 609 Rn. 11; *Müller-ter Jung/Rexin*, CR 2023, 169 (174); sich hieran anschließend *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 31, solange es keine einheitliche Standardformulierung in natürlicher Sprache gibt; auch schon in *Bomhard*, InTeR 2023, 174 (278); *Maamar*, ZUM 2023, 481 (484); *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (30).

(2) Richtlinienkonforme Auslegung des § 44b UrhG zur nationalen Anknüpfung der Angemessenheitsvoraussetzung

Trotz des Wortlauts von Art. 4 Abs. 3 DSM-RL ist die Angemessenheitsvoraussetzung nicht in § 44b Abs. 3 UrhG umgesetzt worden. Sie wird lediglich in der Gesetzesbegründung aufgegriffen.⁵²⁴ Es bedarf deswegen einer diese Diskrepanz korrigierenden, richtlinienkonformen Auslegung des nationalen Schrankentatbestands.⁵²⁵ Fraglich ist jedoch, an welches Tatbestandsmerkmal in § 44b UrhG die richtlinienkonforme Auslegung angeknüpft werden kann. Denkbar ist hierfür zum einen, die Auslegung im unbestimmten Rechtsbegriff der Maschinenlesbarkeit aus § 44b Abs. 3 S. 2 UrhG zu verankern. Hiergegen spricht aber, dass jede Nutzungsvorbehaltserklärung nach Art. 4 Abs. 3 DSM-RL angemessen sein muss, unabhängig davon, für welche Art von Inhalten der Vorbehalt erklärt wird. Die Voraussetzung der Maschinenlesbarkeit muss hingegen nur bei online veröffentlichten Inhalten erfüllt werden. Eine richtlinienkonforme Anknüpfung an dieses Tatbestandsmerkmal ist damit nicht möglich.⁵²⁶

Die Angemessenheitsvoraussetzung ist deswegen im Wege einer richtlinienkonformen Auslegung des Tatbestandsmerkmals „vorbehalten“ in § 44b Abs. 3 UrhG hineinzulesen. Damit haben sich Rechtsinhaber Nutzungen ihrer Werke zum Text und Data Mining nur dann gemäß § 44b Abs. 3 S. 1 Hs. 2 UrhG überhaupt *vorbehalten*, wenn sie ihre Erklärung auch in einer angemessenen Art und Weise abgegeben haben.

(3) Konkrete Anforderungen an die Vorbehaltserklärungen

Die Vorbehaltserklärung muss so ausgestaltet sein, dass sie mit üblichen Mitteln und zumutbarem Aufwand erkannt werden kann. Zur Beurteilung dessen kommt es unter anderem auf die Verkehrssitte und bei online zugänglichen Werken auf den Stand der Technik an.⁵²⁷

524 Siehe in BT-Drs. 19/27426, S. 89.

525 Insoweit auch *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332. Auch *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 15; *Raue*, ZUM 2021, 793 (796 f.) beziehen sich auf die Angemessenheitsvoraussetzung, ohne jedoch eine dogmatische Verankerung im nationalen Recht vorzunehmen.

526 A. A. *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332.

527 Vgl. *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; für eine Berücksichtigung der Verkehrssitte auch *Raue*, ZUM 2021, 793 (797).

(i) Ort der Vorbehaltserklärung bei online veröffentlichten Inhalten

Bei der Erklärung von Vorbehalten gegen die Nutzung online verfügbarer Inhalte in natürlicher Sprache ist in diesem Zusammenhang fraglich, an welchem Ort der Nutzungsvorbehalt erklärt werden muss, damit er mit zumutbarem Aufwand erkannt werden kann. Hierfür ist maßgeblich, an welchen Orten der durchschnittliche Werknutzer mit einer Vorbehaltserklärung rechnet. Diese Verkehrserwartung wird durch die Nennung des Impressums sowie der Allgemeinen Geschäfts- beziehungsweise Nutzungsbedingungen einer Webseite in den Gesetzesbegründungen geprägt.⁵²⁸ Grund hierfür ist, dass die sich aus den Gesetzesmaterialien ergebenden Anforderungen erheblich sowohl im fach- und populärwissenschaftlichen als auch im anwendungsbezogenen Diskurs aufgegriffen wurden.⁵²⁹ Dies hat zum Herausbilden einer entsprechenden Nutzererwartung geführt. Infolgedessen ist es für Trainingsdatensammler zumutbar, diese Orte einer Webseite nach Nutzungsvorbehaltserklärungen gegen die Verwendung von Werken zum Text und Data Mining abzusuchen.

Werden auf einer Webseite mehrere Werke wiedergegeben, so würden Urheber und Rechtsinhaber insbesondere bei der Verwendung umfangreicher Werkdatenbanken darüber hinaus erheblich belastet, wenn sie für jede ihrer Werkwiedergaben eine individuelle Vorbehaltserklärung abgeben müssten. Gleichzeitig besteht kein schutzwürdiges Interesse der Trainingsdatensammler an einer Vielzahl individueller Erklärungen, sofern die abgegebene Vorbehaltserklärung auch in generalisierter, also mehrere Inhalte gleichzeitig erfassender Form zweifelsfrei den jeweils betroffenen Inhalten zugeordnet werden kann. Die Angemessenheitsvoraussetzung bringt die Interessen von Urhebern und Rechtsinhabern sowie Trainingsdatensammlern in einen Ausgleich, sodass beides zugunsten der Urheber und Rechtsinhaber zu berücksichtigen ist. Infolgedessen kann eine Vorbehaltserklärung in den Nutzungsbedingungen oder dem Impressum einer Webseite auch für sämtliche der auf einer Webseite abrufbaren Werkwiedergaben oder eine

528 ErwGr. 18 UAbs. 2 DSM-RL; BT-Drs. 19/27426, S. 89.

529 Siehe beispielsweise <https://bff.de/news/die-bff-justiziarin-dorothe-lanc-informiert-zum-thema>; <https://www.zdf.de/nachrichten/wirtschaft/urheberrecht-kuenstliche-intelligenz-ki-internet-100.html>; <https://www.boersenverein.de/beratung-service/recht/kuenstliche-intelligenz>; <https://bvpa.org/text-und-data-mining-durch-nutzung-svorbehalt-untersagen>; <https://www.golem.de/news/chatgpt-stable-diffusion-dall-e-die-crux-mit-dem-opt-out-aus-text-und-data-mining-2303-172254-2.html> (alle zuletzt abgerufen am 4.7.2025).

ausgewählte Gruppe hiervon erklärt werden.⁵³⁰ Demnach ist ausreichend, dass sich die abgegebene Vorbehaltserklärung zweifelsfrei auf eine oder mehrere Werkwiedergaben bezieht.⁵³¹ Ebenso zumutbar ist der Aufwand für eine Überprüfung der jeweiligen Subpage, also Einzelseite, auf der ein konkretes, online zugängliches Werk verfügbar gemacht worden ist.⁵³² Grund hierfür ist, dass der Ort des Nutzungsvorbehalts durch den Fundort des Werkes bereits hinreichend konkretisiert ist.

Von Werknutzern, die für das Text und Data Mining gezielt und in wesentlichem Umfang in einem spezifischen Dateiformat verkörperte Werke vervielfältigen, kann außerdem erwartet werden, dass sie sich in Bezug auf das konkrete Dateiformat vertiefte technische Kenntnisse aneignen. Denn an die Vervielfältigung schließt sich eine Verarbeitung der Werke an, die einen Umgang mit den Besonderheiten des jeweiligen Dateiformats notwendig macht. Infolgedessen ist es für Werknutzer, die gezielt und in wesentlichem Umfang in einem spezifischen Dateiformat verkörperte Werke vervielfältigen, zumutbar, zusätzlich auch die Metadaten der Werke im Hinblick auf hierin erklärte Nutzungsvorbehalte auszulesen.⁵³³ Im Ergebnis ist damit grundsätzlich sowohl die Abgabe einer „asset-“ im Sinne einer dateibezogenen als auch einer „location-based“ Vorbehaltserklärung möglich.⁵³⁴

Eine automatisierte Vervielfältigung und Verarbeitung von Werken wäre hingegen nicht gewährleistet, wenn eine umfassende, über die jeweilige Quelle des Inhalts hinausgehende Suche nach Vorbehaltserklärungen des betroffenen Rechtsinhabers beispielsweise auf einer persönlichen Webseite oder auf Nachrichtenportalen durchgeführt werden müsste.⁵³⁵ Hinzu kommt, dass dem Datensammler ein wesentlicher Teil der vom Text und Data Mining betroffenen Rechtsinhaber online verfügbarer Werke mangels hinreichender Urhebernennungen nicht bekannt ist. Daher folgt aus der

530 LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 63; im Anschluss hieran auch *Leistner*, GRUR 2024, 1665 (1667, 1673); a. A. *Hamann*, ZGE 2024, 113 (148).

531 LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 62.

532 *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; vgl. außerdem *Raue*, ZUM 2021, 793 (797); im Anschluss hieran *Hamann*, ZGE 2024, 113 (135); *Kleinkopf*, Text- und Data-Mining, S. 249.

533 Vgl. ErwGr. 18 UAbs. 2 DSM-RL.

534 Vgl. zur Terminologie *Bäcker*, Der Text und Data Mining Vorbehalt: Technische Umsetzung der Maschinenlesbarkeit; außerdem in *Stieper/Denga*, GRUR 2024, 1473 (1476) aufgegriffen.

535 Ähnlich auch *Leistner*, GRUR 2024, 1665 (1673).

Angemessenheitsvoraussetzung, dass nur im Umfeld der als Vervielfältigungsvorlage genutzten Quelle nach Vorbehaltserklärungen gesucht werden muss. Für online veröffentlichte Werke bedeutet dies, dass sich eine Suche nach Vorbehaltserklärungen auf die Webseite beschränken kann, auf der die jeweilige Vervielfältigungsvorlage des Werkes öffentlich wiedergegeben wird.⁵³⁶ Nicht angemessen sind demzufolge Pauschalerklärungen, beispielsweise auf der Webseite einer Verwertungsgesellschaft „für ihr gesamtes Repertoire“.⁵³⁷

(ii) Sprache der Vorbehaltserklärung

Vor dem Hintergrund der Angemessenheitsvoraussetzung ist darüber hinaus fraglich, in welcher Sprache Vorbehaltserklärungen abgegeben werden dürfen, die in natürlicher Sprache verfasst sind. Auch hierfür kommt es insbesondere auf die Verkehrsgewohnheiten an. Da es sich um die am weitesten verbreitete Verkehrssprache handelt, ist das Suchen nach Nutzungsvorbehalten in englischer Sprache zumutbar.⁵³⁸ Wegen der leichten Identifizierbarkeit müssen Werknutzer außerdem mit Nutzungsvorbehalten in der Hauptsprache einer jeweiligen Webseite rechnen, auf der die betreffenden Werke veröffentlicht worden sind. Da die betroffenen Rechtsinhaber für Werknutzer hingegen oftmals nicht bekannt sind,⁵³⁹ ist es für sie nicht zumutbar, die Sprache der jeweiligen Rechtsinhaber mit zu berücksichtigen.⁵⁴⁰

(iii) Mindestmaß an Struktur und Klarheit

Neben alledem müssen die Erklärungen ein Mindestmaß an Struktur und Klarheit aufweisen. Es soll ein digitales Verständnis der Vorbehaltserklärungen ermöglichen. Berücksichtigt werden müssen demnach nur übliche, klar verständliche Formulierungen und Wortbausteine sowie Keywords wie

536 Heine, GRUR-Prax 2024, 87 Rn. 17.

537 So zum Beispiel aber die Verwertungsgesellschaften GEMA (<https://www.gema.de/documents/d/guest/tm-vorbehaltserklaerung-2023-12-pdf>) sowie VG Bild-Kunst (<https://www.bildkunst.de/service/reservation-of-rights-for-text-and-data-mining>).

538 Bomhard, DSRITB 2023, 255 (266); Bomhard, InTeR 2023, 174 (178).

539 Vgl. dazu oben unter 2. Teil B. I. 2.

540 A. A. wohl Bomhard, DSRITB 2023, 255 (266); auch in Bomhard, InTeR 2023, 174 (178): „sämtliche in Betracht kommende Sprachen“.

beispielsweise „TDM reserved“, „Text und Data Mining vorbehalten“, „Art. 4 Abs. 3 DSM-RL“, „§ 44b Abs. 3 UrhG“ sowie „Nutzungsvorbehalt“.⁵⁴¹ Diese Aufzählung ist allerdings nicht abschließend.

cc) Ausdrücklichkeit des Nutzungsvorbehalts

Darüber hinaus müssen die Nutzungsvorbehaltserklärungen gemäß Art. 4 Abs. 3 DSM-RL ausdrücklich erklärt werden. Die Voraussetzung wurde ebenfalls nicht mit § 44b UrhG in nationales Recht umgesetzt. Auch zur Behebung dieses Umsetzungsdefizits bedarf es deswegen einer richtlinienkonformen Auslegung des in § 44b Abs. 3 S. 1 UrhG verwendeten Rechtsbegriffs des sich *Vorbehaltens*.⁵⁴² Mithin wurde ein Nutzungsvorbehalt nur dann gemäß § 44b Abs. 3 UrhG wirksam abgegeben, wenn der Vorbehalt auch ausdrücklich erklärt worden ist. Das ist der Fall, wenn der Inhalt der Erklärung expressis verbis transportiert wird.⁵⁴³ Entscheidend ist dabei, ob die Handlung des Vorbehaltserklärenden unmittelbar auf die Abgabe des Erklärungsinhalts, also des Vorbehalts, gerichtet ist. Nicht ausreichend sind demnach Handlungen, mit denen der Vorbehaltserklärende vorrangig einen anderen, selbstständigen Zweck verfolgt und dabei den Vorbehaltswillen indirekt miterklärt.⁵⁴⁴ Eine ausdrückliche Erklärung kann nach diesen Maßstäben auch in formalisierter und standardisierter Form abgegeben werden, sofern der verwendete Befehl gerade dazu dient, den Vorbehaltswillen des Erklärenden auszudrücken. Eine Erklärung in Codeform steht dem Ausdrücklichkeitserfordernis daher nicht entgegen.

dd) Notwendigkeit einer Beschränkung der Vorbehaltserklärung auf Nutzung zum Zwecke des Text und Data Minings

Fraglich ist weiterhin aber, ob aus dem Ausdrücklichkeitskriterium des Art. 4 Abs. 3 DSM-RL, welches richtlinienkonform in § 44b Abs. 3 S. 1

541 So auch *Bomhard*, InTeR 2023, 174 (178).

542 Vgl. dazu unmittelbar oben unter 3. Teil A. III. 2. i) bb); wie hier ebenfalls LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 62.

543 Vgl. *Armbrüster*, in: Münchener Kommentar zum BGB, BGB Vor § 116 Rn. 6.

544 Vgl. v. *Savigny*, System des heutigen römischen Rechts, S. 245; auch *Armbrüster*, in: Münchener Kommentar zum BGB, BGB Vor § 116.

UrhG hineingelesen werden muss,⁵⁴⁵ abgeleitet werden kann, dass nur solche Vorbehaltserklärungen ausdrücklich und damit wirksam sind, die „konkret-individuell“⁵⁴⁶ auf ein Verbot der Nutzung des Werkes für das Text und Data Mining beschränkt sind. Dies hätte zur Folge, dass durch zu weit gefasste Vorbehaltserklärungen nicht die Rechtsfolge des § 44b Abs. 3 UrhG ausgelöst werden könnte.

Dafür, dass Vorbehaltserklärungen für ein Wirksamwerden auf Nutzungen zum Zwecke des Text und Data Mining beschränkt werden müssen spricht, dass gemäß ErwGr. 18 UAbs. 2 S. 3 DSM-RL andere Nutzungen als die zum Text und Data Mining Vorgenommenen nicht von der Vorbehaltserklärung betroffen sein sollen.⁵⁴⁷ Allerdings wird aus ErwGr. 18 DSM-RL nicht deutlich, ob dies als Voraussetzungen für die Wirksamkeit des Nutzungsvorbehalts zu verstehen ist. Denn der Erwägungsgrund könnte auch als rechtsfolgenbezogene Klarstellung verstanden werden, nach der die quasi-dingliche Wirkung des in Art. 4 Abs. 3 DSM-RL verankerten Nutzungsvorbehalts allein die Freistellung der Nutzungshandlung über die Schrankenbestimmung für Text und Data Mining beendet. Für letzteres spricht, dass gemäß ErwGr. 45 InfoSoc-RL die unionsrechtlichen Schrankenbestimmungen einer vertraglichen Abweichung im Grundsatz nicht entgegenstehen.⁵⁴⁸ Es wäre vor diesem Hintergrund wertungswidersprüchlich, dürfte ein Rechtsinhaber beispielsweise die Privatkopiefreiheit gegenüber einem Werknutzer wirksam vertraglich einschränken, in Verbindung mit einer Erklärung des Nutzungsvorbehalts gemäß Art. 4 Abs. 3 DSM-RL eine solche Einschränkung aber faktisch nicht mehr vornehmen, ohne die Unwirksamkeit der Vorbehaltserklärung in Bezug auf das Text und Data Mining in Kauf nehmen zu müssen. Eine Auslegung des ErwGr. 18 UAbs. 2 S. 3 DSM-RL, nach der der Erwägungsgrund eine Wirksamkeitsvoraussetzung für Nutzungsvorbehaltserklärungen beinhaltet, würde demzufolge zu einer allgemeinen Schlechterstellung derjenigen Rechtsinhaber führen, deren Werke potenziell von Text und Data Mining-Aktivitäten betroffen sind. Eine solche Schlechterstellung wird von Art. 4 Abs. 3 DSM-RL aber nicht bezweckt. Im Ergebnis spricht dies dafür, dass ErwGr. 18 UAbs. 2 S. 3 DSM-

545 Dazu siehe oben unter 3. Teil A. III. 2. i) cc).

546 *Hamann*, ZGE 2024, 113 (135).

547 *Hamann*, ZGE 2024, 113 (134 f.).

548 Zur vertraglichen Abweichung von Schrankenbestimmungen, die nach hier vertreter Auffassung im Grundsatz möglich ist, umfassend auch mit weiteren Nachweisen in: *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 263 f.

RL keine Wirksamkeitsvoraussetzung, sondern eine rechtsfolgenbezogene Klarstellung in Bezug auf die quasi-dingliche Wirkung des Nutzungsvorbehalts enthält.⁵⁴⁹ Der Richtlinienggeber wollte demnach nur einem Schluss vom Kleinen aufs Große aus Art. 4 Abs. 3 DSM-RL zulasten der Anwendung anderer Schrankenbestimmungen vorbeugen. Das bedeutet, dass keine Notwendigkeit besteht, eine Vorbehaltserklärung auf Nutzungen zum Zwecke des Text und Data Mining zu beschränken. Solange die Vorbehaltserklärung nicht implizit, sondern ausdrücklich abgegeben worden ist, kann sie sich neben dem Text und Data Mining demzufolge auch andere Nutzungszwecke einbeziehen.⁵⁵⁰

ee) Weitere formalisierte Einzelfälle

Untersucht werden kann in diesem Zusammenhang im Weiteren, ob die im Verkehr üblichen, strukturiert-formalisierten Mechanismen zur Erklärung von Nutzungsvorbehalten den aus § 44b Abs. 3 S. 2 UrhG folgenden Anforderungen genügen. Nur wenn die zuvor spezifizierten Voraussetzungen erfüllt sind, wird mit einer Verwendung einer dieser Mechanismen die Freistellungswirkung des § 44b Abs. 2 S. UrhG beendet.

(1) Robots Exclusion Standard (robots.txt)

Mithilfe des Robots Exclusion Protocols können vom Webseitenbetreiber Zugangsregeln für automatisierte Webcrawler aufgestellt werden. Diese Webcrawler können grundsätzlich auch Werke für das KI-Training generieren. Möglich ist, alle Webcrawler allgemein (als „*“ symbolisiert) oder ausgewählte Webcrawler namentlich anzusprechen. Das Auslesen der Webseite oder spezifizierter Seitenbereiche kann gestattet oder untersagt werden.

549 A. A. Hamann, ZGE 2024, 113 (134 f.).

550 A. A. Hamann, ZGE 2024, 113 (134 f.).

```
robots.txt — Bearbeitet

User-Agent: Googlebot
Allow: /info
Disallow: /personalia

User-Agent: MSNBot
Disallow: /personalia

User-Agent: BacklinkCrawler
Disallow: /personalia

User-Agent: *
Disallow: /ueberuns
```

Um mit Hilfe des Robots Exclusion Protocols einen im Sinne des § 44b Abs. 3 UrhG wirksamen Nutzungsvorbehalt zu erklären, müsste die Verwendung der robots.txt-Datei den Vorbehaltswillen angemessen und ausdrücklich sowie gemäß § 44b Abs. 3 S. 2 UrhG maschinenlesbar übermitteln.

Für eine Angemessenheit des Erklärungsmusters spricht zunächst, dass die Verwendung des Robots Exclusion Protocols von der Internet Engineering Task Force im September 2022 als Internetstandard vorgeschlagen worden ist.⁵⁵¹ Das Robots Exclusion Protocol gehört damit zum Stand der Technik, der für die Übermittlung von Vorbehaltserklärungen eingesetzt wird. Von Datensammlern kann deswegen erwartet werden, dass sie mit dem Robots Exclusion Protocol übermittelte Vorbehaltserklärungen mit üblichen Mitteln und zumutbarem Aufwand erkennen können.

Darüber hinaus wenden sich Vorbehaltserklärende mit einem Einsatz des Robots Exclusion Protocols nicht spezifisch gegen die Verwendung ihrer Werke für das Text und Data Mining. Vielmehr gilt der Vorbehalt zweckunabhängig gegen jede automatisierte Vervielfältigung durch Webcrawler. Eine solche überschießende Vorbehaltserklärung ist für die Wirksamkeit im Sinne des § 44b Abs. 3 UrhG jedoch unschädlich.⁵⁵²

Fraglich ist jedoch, ob die Vorbehaltserklärungen mit Hilfe des Robots Exclusion Protocols auch ausdrücklich abgegeben werden. Dies ist nach richtlinienkonformer Auslegung des § 44b Abs. 3 S. 1 UrhG notwendig.⁵⁵³ Im Gegensatz dazu erfüllen konkludente Vorbehaltserklärungen nicht die gesetzlichen Anforderungen.⁵⁵⁴ Konkludent sind Vorbehaltserklärungen, die vordergründig einem anderen Zweck dienen, implizit aber auch den

551 <https://datatracker.ietf.org/doc/html/rfc9309> (zuletzt abgerufen am 11.7.2024).

552 Dazu siehe unmittelbar oben unter 3. Teil A. III. 2. i) dd).

553 Dazu ebenfalls siehe oben unter 3. Teil A. III. 2. i) cc).

554 Raue, ZUM 2021, 793 (796).

Erklärungsinhalt übermitteln.⁵⁵⁵ Auch wenn das Robots Exclusion Protocol nutzungszweckunabhängig jede automatisierte Vervielfältigung der betroffenen Werke durch Webcrawler untersagt, dient es nichtsdestotrotz keinem anderen Zweck als der Übermittlung der (umfassenden) Vorbehaltserklärung. Die mit Hilfe des Robots Exclusion Protocols übermittelte Erklärung ist trotz ihrer Formalisierung und Strukturierung sowie ungeachtet der überschießenden Vorbehaltswirkung demzufolge ausdrücklich. Im Ergebnis erfüllt sie damit die Voraussetzungen, die an eine Vorbehaltserklärung im Sinne des § 44b Abs. 3 S. 1 UrhG gestellt werden. Da das Robots Exclusion Protocol auch in einer Art und Weise digital hinterlegt ist, die eine unmittelbare Verarbeitung durch Maschinen ermöglicht, sind die mit Hilfe dieses Protokolls abgegebenen Vorbehaltserklärungen auch maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG. Insgesamt sind Nutzungsvorbehalte, die mit Hilfe des Robots Exclusion Protocols erklärt werden, demzufolge wirksam.⁵⁵⁶ Sie lösen die in § 44b Abs. 3 UrhG angelegte Beendigungswirkung aus.

(2) Spawning-Protokoll (ai.txt)

Neben dem Robots Exclusion Protocol wird auch das Spawning-Protokoll zur Abgabe von Vorbehaltserklärungen genutzt. Das Spawning-Protokoll ermöglicht Webseitenbetreibern, sich spezifisch gegen Datensammler zu wenden, die die auf der Webseite wiedergegebenen Werke für das Training von KI vervielfältigen wollen. Die Syntax entspricht dabei grundsätzlich derjenigen des Robots Exclusion Protocols, bei dem die Vorbehaltserklärung über eine im Wurzelverzeichnis der Webseite abgelegte Textdatei übermittelt werden.⁵⁵⁷ Allerdings wird die ai.txt-Datei nicht von den Datensammlern selbst, sondern vom Intermediär *Spawning* ausgelesen. Dieser

555 Vgl. v. Savigny, System des heutigen römischen Rechts, S. 245; auch *Armbrüster*, in: Münchener Kommentar zum BGB, BGB Vor § 116.

556 Wie hier auch *Baumann*, NJW 2023, 3673 Rn. 19; *de la Durantaye*, ZUM 2023, 645 (651); *Gausling*, CR 2021, 609 Rn. 5; *Heine*, GRUR-Prax 2024, 87 Rn. 16; *Hofmann*, WRP 2024, 11 Rn. 18; *Konertz/Schönhof*, WRP 2024, 289 Rn. 33; *Maamar*, ZUM 2023, 481 (484); *Müller-ter Jung/Rexin*, CR 2023, 169 Rn. 29; *Schack*, NJW 2024, 113 Rn. 20; *Schwarz/Söbbing*, RD 2023, 415 Rn. 33; a. A. hingegen *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 33; *Hamann*, ZGE 2024, 113 (153).

557 *Hamann*, ZGE 2024, 113 (155). Außerdem ausdrücklich referiert in: <https://spawning.substack.com/p/aitxt-a-new-way-for-websites-to-set> (zuletzt aufgerufen am 2.5.2024).

gibt die Vorbehaltserklärungen über eine Schnittstelle gebündelt an die Trainingsdatensammlern weiter.⁵⁵⁸

Da auch die mit Hilfe des Spawning-Protokolls abgegebenen Vorbehaltserklärungen zwar formalisiert und strukturiert, aber ausschließlich zum Zwecke der Vorbehaltsübermittlung erklärt worden sind, erfüllen sie trotz ihrer Formalisierung und Strukturierung die Ausdrücklichkeitsanforderungen, die sich im Wege einer richtlinienkonformen Auslegung des § 44b Abs. 3 UrhG ergeben.

Über die bereitgestellte Anwendungsschnittstelle können Datensammler die mit Hilfe des Spawning-Protokolls hinterlegten Vorbehalte darüber hinaus automatisiert und ohne Informationsverlust abrufen, sodass der Vorbehaltszweck auch mit Hilfe des Spawning-Protokolls sicher erreicht wird. Das Spawning-Protokoll gehört damit ebenfalls zum Stand der Technik.⁵⁵⁹ Da die Nutzung der Anwendungsschnittstelle im Übrigen niederschwellig und kostenlos möglich ist, ist auch die Berücksichtigung des Spawning-Protokolls für Datensammler zumutbar. Im Ergebnis sind deswegen auch Vorbehaltserklärungen mit Hilfe des Spawning-Protokolls angemessen. Infolgedessen qualifizieren sich die auf Basis des Spawning-Protokolls abgegebenen Erklärungen auch als Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG. Aufgrund ihrer unmittelbaren digitalen Hinterlegung, die sogar eine Standardisierung und Formalisierung der Erklärungsinhalte mit sich bringt, sind auch die Anforderungen aus der Maschinenlesbarkeitsvoraussetzung aus § 44b Abs. 3 S. 2 UrhG erfüllt. Alles in allem sind mit Hilfe des Spawning-Protokolls erklärte Nutzungsvorbehalte insgesamt wirksam im Sinne des § 44b Abs. 3 UrhG. Sie beenden daher die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining.⁵⁶⁰

(3) World Wide Web Consortium (W3C) TDM Reservation Protocol

Möglich ist darüber hinaus auch, Vorbehaltserklärungen auf Basis des vom *World Wide Web Consortium* entwickelten TDM Reservation Protocol abzugeben. Hierin sind verschiedene Kommunikationsmethoden zusammen-

558 Vgl. <https://spawning.substack.com/p/aitxt-a-new-way-for-websites-to-set> (zuletzt abgerufen am 30.4.2024).

559 Vgl. zum Begriff des Stands der Technik: *Martini*, in: Paal/Pauly, Kommentar zu Datenschutz-Grundverordnung und Bundesdatenschutzgesetz, DS-GVO Art. 25 Rn. 39a; *Mayen*, in: Scheurle/Mayen, Telekommunikationsgesetz, TKG § 113d Rn. 6.

560 In diesem Sinne wohl auch *Hamann*, ZGE 2024, 113 (155).

gefasst.⁵⁶¹ Da auch das W3C TDM Reservation Protocol ausschließlich den Zweck der Übermittlung von Vorbehaltserklärungen erfüllt und der Vorbehaltswille mit ihm ebenfalls ohne Informationsverlust sowie unentgeltlich und automatisiert übermittelt werden kann, erfüllen auch Erklärungen auf Basis des W3C TDM Reservation Protocols die Voraussetzungen der Ausdrücklichkeit und Angemessenheit aus § 44b Abs. 3 UrhG. Eine digitale Hinterlegung der Vorbehaltserklärungen beispielsweise als HTML-Metadaten-Tags ermöglicht zudem das unmittelbare computergestützte Auslesen der Vorbehalte. Demzufolge sind die Vorbehaltserklärungen auch maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG. Infolgedessen erfüllen auch Nutzungsvorbehalte auf Basis des W3C TDM Reservation Protocols die Voraussetzungen des § 44b Abs. 3 UrhG. Auch durch sie wird die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining damit beendet.⁵⁶²

(4) Technische Schutzmaßnahmen gemäß § 95a UrhG

Unklar ist demgegenüber, ob Nutzungsvorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG durch den Einsatz technischer Schutzmaßnahmen gemäß § 95a UrhG wirksam miterklärt werden.

Nutzungsvorbehalte dienen dem Zweck, den Vorbehaltswillen der Rechtsinhaber kommunikativ zum Ausdruck zu bringen. Im Unterschied hierzu werden technische Schutzmaßnahmen, die über § 95a UrhG vor Umgehung geschützt sind, eingesetzt, um auf einer faktischen Ebene den Zugriff auf oder die Vervielfältigung von Werken zu verhindern. Sie wirken also tatsächlich gegen die Nutzungshandlung.⁵⁶³ Technische Schutzmaßnahmen können den Vorbehaltswille neben der faktischen Nutzungsbeschränkung demzufolge nur mittelbar neben dem Hauptzweck der Schutzmaßnahmen miterreichen. Läge im Einsatz technischer Schutzmaßnahmen eine Vorbehaltserklärung, würde diese damit nur konkludent abgegeben.⁵⁶⁴ Schon deswegen wäre das Ausdrücklichkeitserfordernis nicht erfüllt, welches richtlinienkonform in das Tatbestandsmerkmal des Vorbehaltens aus

561 Siehe zum Inhalt des TDM Reservation Protocol <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202> (zuletzt abgerufen am 2.5.2024).

562 So auch *Hamann*, ZGE 2024, 113 (160): „gerade visionär“; *Schippan*, ZUM 2024, 670 (676); außerdem wohl ebenso *de la Durantaye*, ZUM 2023, 645 (651); *Hofmann*, WRP 2024, 11 Rn. 18; *Konertz/Schönhof*, WRP 2024, 289 Rn. 33; *Maamar*, ZUM 2023, 481 (484); *Schwarz/Söbbing*, RD 2023, 415 Rn. 15.

563 Vgl. auch *Conrad/Schubert*, GRUR 2018, 350 (355).

564 Ebenso *Hamann*, ZGE 2024, 113 (163).

§ 44b Abs. 3 S. 1 UrhG hineingelesen werden muss. Demzufolge werden beim Einsatz technischer Schutzmaßnahmen im Sinne des § 95a Abs. 2 S. 1 UrhG keine Nutzungsvorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG miterklärt.⁵⁶⁵

Darüber hinaus beenden zugangsbeschränkende Maßnahmen bereits die rechtmäßige Zugänglichkeit der betreffenden Werke im Sinne des § 44b Abs. 2 S. 1 UrhG. In diesen Fällen kann die Schrankenbestimmung für Text und Data Mining daher schon deswegen nicht angewendet werden.

ff) Werk- oder wiedergabebezogene Wirkung des Nutzungsvorbehalts

Gemäß § 44b Abs. 3 S. 1 UrhG beendet eine wirksame Erklärung des Nutzungsvorbehalts die Freistellungswirkung, die von der Schrankenbestimmung für Text und Data Mining, § 44b UrhG, ausgeht. Im Detail regelt die Schrankenbestimmung aber nicht die Wirkung und Reichweite des Nutzungsvorbehalts. Das betrifft unter anderem die Fragestellung, ob die Vorbehaltswirkung nur die Nutzung einer konkreten Werkwiedergabe beziehungsweise eines spezifischen Vervielfältigungsstücks als Vervielfältigungsvorlage erfasst, in deren beziehungsweise dessen Zusammenhang der Nutzungsvorbehalt erklärt worden ist (wiedergabebezogene Wirkung). Nach einem weiteren Verständnis könnte die Vorbehaltswirkung aber auch werkbezogen eintreten, also über die konkrete Werkwiedergabe oder das spezifische Vervielfältigungsstück hinaus die Freistellung der Nutzung des betreffenden Werkes insgesamt beendet (werkbezogene Wirkung).

Für eine wiedergabebezogene Wirkung des Nutzungsvorbehalts spricht, dass Nutzungsvorbehaltserklärungen aufgrund der Angemessenheitsvoraussetzung mit zumutbarem Aufwand erkannt werden können müssen. Nur dann sind sie wirksam und vom Datensammler zu berücksichtigen. Würden Nutzungsvorbehaltserklärungen werkbezogen wirken, müsste ein Datensammler bei der Vervielfältigung eines Werkes sämtliche der jemals verfügbar gemachten Repräsentationen eines Werkes, also Vervielfältigungsstücke oder öffentliche Wiedergaben, daraufhin absuchen, ob an anderer Stelle und zu irgendeinem Zeitpunkt bereits eine Vorbehaltserklärung für das Werk abgegeben worden ist. Gerade bei online verfügbaren Werken wäre dieser Aufwand kaum zu erbringen. Vorbehaltserklärungen an anderen Repräsentationen des Werkes aufzufinden ist für Datensammler

565 Wie hier auch *Hamann*, ZGE 2024, 113 (163); a. A. *Meys*, GRUR Int. 2020, 457 (467).

infolgedessen nicht zumutbar. Im Sinne des § 44b Abs. 3 S. 1 UrhG sind aber nur in zumutbarer Weise erkennbare Nutzungsvorbehaltserklärungen überhaupt wirksam.⁵⁶⁶ Die wiedergabebezogene Wirkung des Nutzungsvorbehalts folgt damit als Kehrseite aus der in § 44b Abs. 3 S. 1 UrhG verankerten Angemessenheitsvoraussetzung.

Darüber hinaus könnten Datensammler im Falle einer werkbezogenen Wirkung des Nutzungsvorbehalts nie sicher sein, ob nicht doch ein nicht identifizierter Nutzungsvorbehalt gegen die Verwendung des Werkes erklärt worden ist. Die hieraus resultierende Unsicherheit würde zu prohibitiven Effekten und damit zu einer Einschränkung der effektiven Wirksamkeit der Schrankenbestimmung führen. Dies wäre auch mit dem Grundsatz der praktischen Wirksamkeit des Unionsrechts nicht vereinbar.

Im Ergebnis sprechen beide Argumente für eine wiedergabebezogene Wirkung des Nutzungsvorbehalts. Die von der Vorbehaltserklärung ausgehende Beendigungswirkung beschränkt sich damit stets auf die Nutzung der jeweiligen öffentlichen Wiedergabe beziehungsweise des spezifischen Vervielfältigungsstücks als Vervielfältigungsvorlage für die Zwecke des Text und Data Mining, an der beziehungsweise dem der Nutzungsvorbehalts erklärt worden ist. Die Nutzung anderer Werkrepräsentationen ist hiervon nicht betroffen.⁵⁶⁷

gg) Zeitliche Wirkung des Nutzungsvorbehalts

§ 44b Abs. 3 UrhG regelt weiterhin nicht ausdrücklich, welche zeitliche Wirkung einem Nutzungsvorbehalt zukommt. Einerseits könnte er Wirkung lediglich für die Zukunft entfalten (*ex nunc*), also sich nur auf solche Vervielfältigungshandlungen beziehen, die ab dem Zeitpunkt der Vorbehaltserklärung vorgenommen werden. Auf Grundlage eines strengeren Verständnisses, nach dem ein Nutzungsvorbehalt auch rückwirkend (*ex tunc*) eingreift, müssten auch vor der Vorbehaltserklärung für das Text und Data Mining hergestellte Vervielfältigungsstücke mit Wirksamwerden des Vorbehalts beseitigt werden. Für eine *ex nunc* Wirkung der Vorbehaltserklärung spricht, dass der Begriff des *sich Vorbehaltens* dem natürlichen Sprachverständnis nach voraussetzt, dass sich die vorbehaltende Person

566 Siehe oben unter 3. Teil A. III. 2. i) bb).

567 So auch *Leistner*, GRUR 2024, 1665 (1673); wohl außerdem implizit *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 140 f.

eine Möglichkeit oder Entscheidung für die Zukunft offenhalten will.⁵⁶⁸ Das Vorbehalten bezieht sich damit auf ein zukünftiges Tun oder Unterlassen. Bereits abgeschlossene Vervielfältigungshandlungen werden deswegen schon begrifflich nicht erfasst. Es gibt einen terminologischen Unterschied zwischen dem *Löschen* einer Vervielfältigung und dem *Vorbehalten* einer Vervielfältigungshandlung.⁵⁶⁹

Für ein auf die Zukunft bezogenes Verständnis der Vorbehaltswirkung spricht zudem, dass der Nutzungsvorbehalt nach Art. 4 Abs. 3 DSM-RL die „Anwendung der Ausnahme und Beschränkung (für Text und Data Mining)“ regelt. Das betroffene Vervielfältigungsrecht bezieht sich gemäß Art. 2 InfoSoc-RL auf die Vervielfältigung als *Handlung*, also die *Herstellung* des Vervielfältigungsstücks.⁵⁷⁰ Dies führt dazu, dass es einer Anwendung der Schrankenbestimmung nur so lange bedarf, wie die Herstellung des Vervielfältigungsstücks als *Handlung* des Werknutzers andauert. Ist das Vervielfältigungsstück bereits hergestellt worden, kommt es nicht mehr auf die „Anwendung der Ausnahme und Beschränkung“ an. Die hergestellten Vervielfältigungsstücke können vielmehr weiterhin bestehen. Die in Art. 4 Abs. 3 DSM-RL terminologisch angelegte Wirkung des Nutzungsvorbehalts deutet demzufolge ebenfalls auf eine ex nunc-Wirkung der Vorbehaltserklärung hin. Von einer solchen Rechtswirkung ist darüber hinaus auch der Umsetzungsgesetzgeber ausgegangen.⁵⁷¹

Im Ergebnis sprechen durchgreifende Gründe dafür, dass der Nutzungsvorbehalt aus § 44b Abs. 3 UrhG Wirkung ex nunc entfaltet.⁵⁷² Vervielfältigungsstücke, die vor Wirksamwerden einer Vorbehaltserklärung für das Text und Data Mining und insbesondere für das KI-Training hergestellt worden sind, werden vom Nutzungsvorbehalt demzufolge nicht erfasst. Eine Löschung der Werke aus bereits angefertigten Trainingsdatensets muss demzufolge nicht erfolgen. Weiterhin müssen Werke nicht aus trainierten KI-Systemen „herausgerechnet“ werden. Hiervon zu unterscheiden

568 *Dudenredaktion (Bibliographisches Institut)*, Duden: das Bedeutungswörterbuch, Begriff: vorbehalten.

569 *Heesen/Jüngels*, RuZ 2021, 45 (55).

570 Dies wird auch deutlich in der deutschen Umsetzung, § 16 Abs. 1 UrhG.

571 BT-Drs. 19/27426, S. 89.

572 Allgemeine Meinung, siehe unter anderem: *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 27; *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 9; *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 10; *Steinrötter/Schauer*, in: Barudi, Das neue Urheberrecht, § 14 Text und Data Mining Rn. 14; *Heesen/Jüngels*, RuZ 2021, 45 (55); BT-Drs. 19/27426, S. 89.

ist allerdings die Zulässigkeit der Herstellung weiterer, mitunter für das KI-Training notwendiger Vervielfältigungen. Eine solche ist in Folge des Nutzungsvorbehalts nicht mehr über § 44b UrhG freigestellt.

hh) Der Vorbehaltserklärende

Werknutzungen werden nur dann durch § 44b Abs. 2 S. 1 UrhG erlaubnisfrei gestellt, wenn sich der Rechtsinhaber die Nutzungen nicht vorbehalten hat (§ 44b Abs. 3 S. 1 UrhG). § 44b UrhG weist die Vorbehaltsbefugnis damit „dem Rechtsinhaber“ zu. Gleichzeitig wird im Regelungstext nicht spezifiziert, welche Anforderungen an die Vorbehaltsbefugnis im Sinne des § 44b Abs. 3 UrhG im Einzelnen zu stellen sind.

Auch wenn sowohl das UrhG als auch die unionsrechtlichen Bestimmungen des Urheberrechts begrifflich zwischen *Urhebern* und *Rechtsinhabern* differenzieren, folgt aus der gesetzlich gewährten Vorbehaltsbefugnis der Rechtsinhaber, dass erst recht auch Urheber Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG abgeben können.

(1) Einfache und ausschließliche Rechtsinhaber

Mit Blick auf die Vorbehaltsbefugnis der Rechtsinhaber ist darüber hinaus jedoch unklar, ob § 44b Abs. 3 UrhG nur *ausschließlichen* Rechtsinhabern die Möglichkeit eröffnet, einen Nutzungsvorbehalt zu erklären, oder ob auch Inhaber eines *einfachen* Nutzungsrechts nach § 44b Abs. 3 UrhG vorbehaltsbefugt sind.

Für eine Vorbehaltsbefugnis auch einfacher Rechtsinhaber spricht, dass auch sie, unabhängig von Umfang und Qualität ihrer Rechtsposition, *Rechtsinhaber* im Sinne des § 44b Abs. 3 S. 1 UrhG sind. Damit fallen sie nach dem Wortlaut des § 44b Abs. 3 S. 1 UrhG in den persönlichen Anwendungsbereich der Regelung. Darüber hinaus haben auch einfache Rechtsinhaber mit dem Erwerb des Nutzungsrechts eine Investition getätigt, an der ein Schutzinteresse besteht. Dieses ließe sich durch die Vorbehaltserklärung realisieren.

Allerdings bezieht sich der Begriff des „Rechtsinhabers“ innerhalb des UrhG auch ohne spezifizierende Zusätze („ausschließlich“ oder „einfach“) entweder auf ausschließliche (beispielsweise in § 10 Abs. 3 UrhG) oder auf einfache (beispielsweise in § 60d Abs. 6 UrhG oder § 95a Abs. 1 UrhG) Rechtsinhaber. Die konkrete Bedeutung ist vom Kontext der jeweiligen

Regelung abhängig. Es besteht damit kein einheitliches Verständnis des Rechtsinhaberbegriffs, das auch § 44b Abs. 3 S.1 UrhG zugrunde gelegt werden könne. Ein Schluss auf die notwendige Qualität der Rechtsposition ist aus dem Wortlaut des § 44b Abs. 3 S.1 UrhG demzufolge nicht möglich.

Allerdings versetzt ein einfaches Nutzungsrecht einen Rechtsinhaber nicht in die Lage, anderen Rechtsinhabern die Nutzung desselben Werkes zu verbieten (§ 31 Abs. 2 UrhG). Es wäre vor diesem Hintergrund wertungswidersprüchlich, wäre der einfache Rechtsinhaber im Anwendungsbereich der Schrankenbestimmung für Text und Data Mining über § 44b Abs. 3 UrhG befugt, anderen, von der Schrankenbestimmung privilegierten Nutzern die rechtmäßige Nutzung des Werkes zum Zwecke des Text und Data Mining zu verbieten. Die Nichtexklusivität des einfachen Nutzungsrechts hat außerdem zur Folge, dass sich auch das Interesse des einfachen Rechtsinhabers an einem Schutz seiner Investition in die Ermöglichung der Werknutzung nicht darauf erstreckt, dass die Werknutzung unter Ausschluss anderer, also exklusiv erfolgt. Ein Schutz seiner Investition über § 44b Abs. 3 UrhG ist demzufolge nicht nötig.

Insgesamt sprechen damit bessere Gründe dafür, dass neben Urhebern nur ausschließliche Rechtsinhaber Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG abgeben können. Inhaber einfacher Nutzungsrechte sind hingegen nicht vorbehaltbefugt.⁵⁷³

Ungeachtet dessen zielt die Abgabe der Nutzungsvorbehaltserklärung auf die Herbeiführung der von § 44b Abs. 3 UrhG vorgeschriebenen Rechtsfolge ab. Damit wird zwar keine vom Vorbehaltserklärenden selbstgewählte Rechtsfolge spezifiziert. Es fehlt also an den Voraussetzungen einer Willenserklärung im bürgerlichrechtlichen Sinne.⁵⁷⁴ Dennoch ist die Erklärung des Nutzungsvorbehalts eine geschäftsähnliche Handlung.⁵⁷⁵ Damit finden auch die Regelungen zur Stellvertretung bei der Abgabe von Willenserklärungen, §§ 164 ff. BGB, entsprechende Anwendung.⁵⁷⁶ Der Urheber oder ausschließliche Rechtsinhaber kann demzufolge auch einfachen Rechtsin-

573 A. A. LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 58; im Anschluss hieran *Leistner*, GRUR 2024, 1665 (1667, 1671); weiterhin auch *Brockmeyer*, Text und Data Mining, S. 79, nach dem alle Rechtsinhaber berechtigt sind; offengelassen von *Vesala*, IIC 2023, 351 (357); *Hamann*, ZGE 2024, 113 (138 ff.).

574 Zu den Voraussetzungen einer Willenserklärung vgl. *Köhler*, BGB Allgemeiner Teil § 6 Rn. 1f.

575 *Hamann*, ZGE 2024, 113 (138); allgemein zur geschäftsähnlichen Handlung und ihren Voraussetzungen siehe *Köhler*, BGB Allgemeiner Teil § 5 Rn. 7.

576 *Mansel*, in: Jauernig, Bürgerliches Gesetzbuch, BGB § 164 Rn. 2.

habern Vertretungsmacht für die Erklärung eines Nutzungsvorbehalts in Bezug auf die jeweilige Werkwiedergabe erteilen oder eine Vorbehaltserklärung nachträglich genehmigen (§ 184 BGB analog). Hierfür spricht auch der in Art. 4 Abs. 3 EUV verankerte Grundsatz der effektiven Anwendung des § 44b Abs. 3 UrhG zugrundeliegenden Unionsrechts.⁵⁷⁷ Eine solche Genehmigung kann auch konkludent abgegeben werden, beispielsweise durch die (gerichtliche) Geltendmachung der Rechtsverletzung im betreffenden Fall.⁵⁷⁸

(2) Inhaltliche, räumliche und zeitliche Beschränkungen des Nutzungsrechts

Fraglich ist des Weiteren aber auch, wie das Ausschließlichkeitsrecht eines Rechtsinhabers konkret ausgestaltet sein muss, damit er vorbehaltsbefugt ist. Problematisch wird dies zwar nicht bei Inhabern unbeschränkt ausschließlicher Nutzungsrechte. Deren umfassende Rechtsposition beinhaltet auch die Befugnis, einen Nutzungsvorbehalt nach § 44b Abs. 3 UrhG zu erklären. Im Gegensatz hierzu ist allerdings unklar, ob und wann Inhaber von inhaltlich, räumlich oder zeitlich beschränkten Nutzungsrechten, wie sie durch § 31 Abs. 1 S. 1 UrhG ermöglicht werden, zur Abgabe von Nutzungsvorbehaltserklärungen befugt sind.

(i) Inhaltliche Beschränkungen und die Vorbehaltsbefugnis

Inhaltliche Beschränkungen der eingeräumten Rechtsposition können insoweit stattfinden, als dass die Aufspaltung selbstständig verkehrsfähige Rechte herausbildet. Der Nutzungsvorbehalt entfaltet seine Wirkung vor diesem Hintergrund gegen die Erlaubnisfreiheit des Rechts der Vervielfältigung von Werken zum Text und Data Mining. Dies umfasst auch Vervielfältigungen zum Training von KI.⁵⁷⁹ Da ein Markt für Trainingsdatensätze vorhanden ist und das Training selbst auch von anderen Sachverhaltskonstellationen wie der Archivierung oder der allgemeinen wissenschaftlichen Verwendung von Werken abgegrenzt werden kann, ist diese Vervielfältigung von Werken für das Training von KI eine eigenständige

577 *Leistner*, GRUR 2024, 1665 (1671).

578 *Leistner*, GRUR 2024, 1665 (1671).

579 Dazu siehe oben unter 3. Teil A. III. 2. a) bis e).

Nutzungsart.⁵⁸⁰ Die Vorbehaltswirkung des § 44b Abs. 3 UrhG erstreckt sich damit auf diese selbstständige Nutzungsart. Nutzungsvorbehalt und Nutzungsart sind über § 44b Abs. 3 UrhG also miteinander verbunden. Aus diesem Grund sind nur solche Inhaber inhaltlich beschränkter Nutzungsrechte vorbehaltsbefugt, die das ausschließliche Nutzungsrecht zur Vervielfältigung von Werken für das Training von KI innehaben, insoweit die Vervielfältigung nicht von § 44b Abs. 2 UrhG freigestellt ist.⁵⁸¹ Umgekehrt hat dies zur Folge, dass beispielsweise solche Rechtsinhaber keinen wirksamen Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklären können, die das ausschließliche, aber beschränkte Recht innehaben, ein Werk zu verfilmen oder auszustellen.

Ob den Rechtsinhabern das Nutzungsrecht für nicht von § 44b UrhG freigestellte Vervielfältigungshandlungen zum KI-Training eingeräumt wurde, ist im Weiteren von der Ausgestaltung und (insbesondere) Auslegung der jeweiligen Nutzungsverträge abhängig. Insbesondere bei pauschalen Rechtseinräumungen kommt es dabei vor allem auf den Zweck der Rechtsübertragung an (§ 31 Abs. 5 UrhG). Von Bedeutung kann auch eine Einräumung von Rechten an unbekanntem Nutzungsarten gemäß § 31a UrhG sein. Unbekannt ist eine Nutzungsart, wenn sie dem durchschnittlichen Urheber zum Zeitpunkt des Abschlusses des Lizenzvertrags nicht als praktisch durchführbare eigenständige Nutzungsart geläufig ist.⁵⁸² Hierzu genügt schon, dass zwar die technische Möglichkeit, nicht aber die wirtschaftliche Bedeutung und urheberrechtliche Verwertbarkeit der Nutzungsart für den durchschnittlichen Urheber bekannt ist.⁵⁸³ Unschädlich ist ebenso Kenntnis in (technischen) Fachkreisen.⁵⁸⁴ Dass Werke in großem Umfang verwertet werden können, um insbesondere generative KI zu trainieren, und in der Verwertung von Werken als Trainingsdaten daher auch eine erhebliche wirtschaftliche Auswertungsmöglichkeit liegt, ist dem durchschnittlichen Urheber wohl seit der von einer gesellschaftlichen Debatte begleiteten Veröffentlichung des KI-gestützten Chatbots

580 *Pukas*, GRUR 2023, 614 (617); im Anschluss hieran *Schack*, NJW 2024, 113 Rn. 16; *Hamann*, ZGE 2024, 113 (138). Ebenso *Kögel*, InTeR 2023, 179 (181).

581 *Heine*, GRUR-Prax 2024, 87 Rn. 19.

582 *Nordemann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 31a Rn. 44.

583 BGH, Urt. v. 11.10.1990 - I ZR 59/89, GRUR 1991, 133 (136) – Videozweitauswertung; BGH, Urt. v. 26.01.1995 - I ZR 63/93, GRUR 1995, 212 (213) – Videozweitauswertung III.

584 *Nordemann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 31a Rn. 44; *Strobl*, ZUM 2023, 492 (498).

ChatGPT für breite Verbraucherkreise am 30. November 2022 bekannt. Vor diesem Veröffentlichungsdatum ist die Nutzung (also Vervielfältigung) von Werken für das Training von KI im Umkehrschluss eine unbekanntete Nutzungsart im Sinne des § 31a UrhG.⁵⁸⁵ Rechtsinhabern, denen vor diesem Datum Rechte an unbekannteten Nutzungsarten im Sinne des § 31a UrhG eingeräumt worden sind, kommt deswegen auch das Nutzungsrecht zur Vervielfältigung von Werken für das Training von KI zu, insoweit die Vervielfältigungen nicht von § 44b Abs. 2 UrhG freigestellt sind. Sie sind daher vorbehaltbefugt, solange kein Widerruf der Rechtseinräumung im Sinne des § 31a Abs. 1 S. 3 UrhG erfolgt ist.

Verwertungsgesellschaftsrechtliche Wahrnehmungsverträge, von denen pauschal das Vervielfältigungsrecht umfasst wird, sind so auszulegen, dass hierunter auch das Recht zur Vervielfältigung von Werken für das Training von KI fällt.⁵⁸⁶ Weiterhin ausreichend ist eine Einräumung des Vervielfältigungsrechts „für die Entwicklung neuer Software“ oder die „Auswertung zugrundeliegender Informationen wie Muster, Trends oder Korrelationen“. Denn beides umfasst unmittelbar auch die Nutzung der Werke zum Training neuer KI-Systeme.

Nicht vorbehaltbefugten Rechtsinhabern kann trotz alledem auch Vertretungsmacht für die Erklärung des Nutzungsvorbehalts in Bezug auf ihre jeweilige Werkwiedergabe eingeräumt werden. Dies richtet sich nach den Regelungen der Stellvertretung, §§ 164 ff. BGB, die analog angewendet werden können.⁵⁸⁷

(ii) Zeitliche sowie räumliche Beschränkungen und die Vorbehaltsbefugnis

Der Nutzungsvorbehalt gegen die Freistellung von Vervielfältigungshandlungen für das Text und Data Mining wirkt darüber hinaus sowohl zeitlich als auch räumlich uneingeschränkt. Das wirft die Frage auf, ob Inhaber räumlich oder zeitlich beschränkter Nutzungsrechte ebenfalls vorbehaltbefugt sind.

Gegen eine Vorbehaltsbefugnis des Inhabers eines zeitlich oder räumlich beschränkten Nutzungsrechts spricht, dass es widersprüchlich erscheint, ihm ein Vorbehaltsrecht zuzubilligen, welches umfassender ist als das ihm eingeräumte Verwertungsrecht. In Bezug auf zeitlich beschränkte Rechts-

585 So auch Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 31a Rn. 49.

586 Pukas, GRUR 2023, 614 (618).

587 Dazu vgl. bereits unmittelbar oben.

einräumungen muss allerdings berücksichtigt werden, dass der Nutzungsvorbehalt nur wiedergabebezogene Wirkung entfaltet.⁵⁸⁸ Damit wird die Freistellungswirkung der Schrankenbestimmung stets nur in Bezug auf die Nutzung einer konkreten öffentlichen Wiedergabe oder eines spezifischen Vervielfältigungsstücks als Vervielfältigungsvorlage beendet. Der Nutzungsvorbehalt beendet die Freistellungswirkung zwar zeitlich unbeschränkt. Aufgrund der wiedergabebezogenen Wirkung des Nutzungsvorbehalts kann er von einem neuen Rechtsinhaber oder dem Urheber nach Erlöschen des ersten, zeitlich beschränkten Nutzungsrechts in Bezug auf seine eigene öffentliche Wiedergabe des Werkes oder neu angefertigten Vervielfältigungsstücke aber erneut ausgeübt werden. Aufgrund der wiedergabebezogenen Wirkung des Nutzungsvorbehalts kommt es demzufolge nicht zu einem Widerspruch zwischen der unbeschränkten Vorbehaltsfolge und einer zeitlich beschränkten Nutzungsrechtseinräumung.

Die wiedergabebezogene Wirkung des Nutzungsvorbehalts führt weiterhin dazu, dass die absolut wirkende Vorbehaltserklärung auch bei einer räumlichen Beschränkung des Nutzungsrechts de facto nur Sperrwirkung in Bezug auf die Nutzung des jeweiligen Vervielfältigungsstücks oder der jeweiligen öffentlichen Wiedergabe entfaltet, die von der räumlich beschränkten Nutzungsrechtseinräumung erfasst ist. Ein Nutzungsvorbehalt, der beispielsweise an der deutschen Sprachfassung eines Spielfilms erklärt worden ist, entfaltet seine Vorbehaltswirkung nur gegen die Nutzung dieser Sprachfassung. Auch wenn dem Vorbehalt de jure eine absolute Wirkung zukommt, berührt die Erklärung damit nicht die Vervielfältigung von Werkexemplaren anderer Sprachfassungen. Die wiedergabebezogene Wirkung des Nutzungsvorbehalts führt daher auch bei einer räumlichen Beschränkung des Nutzungsrechts dazu, dass es nicht zu einem Widerspruch zwischen der unbeschränkten Wirkung des Nutzungsvorbehalts und der räumlichen Beschränkung des Nutzungsrechts kommt. Nicht zuletzt vermittelt auch ein ausschließliches, aber räumlich oder zeitlich beschränktes Nutzungsrecht den Rechtsinhabern eine Rechtsposition, Kraft derer sie in die Lage versetzt sind, anderen die Nutzung des betreffenden Werkes zu untersagen. Damit ist auch die Vorbehaltsbefugnis des Inhabers eines räumlich oder zeitlich beschränkten, ausschließlichen Nutzungsrechts interessengerecht.

Alles in allem ergeben sich keine Wertungswidersprüche zwischen der räumlichen oder zeitlichen Beschränkung eines ausschließlichen Nutzungs-

588 Dazu siehe oben unter 3. Teil A. III. 2. i) ff).

rechts und der Möglichkeit zur Abgabe eines de jure unbeschränkten Nutzungsvorbehalts, wie § 44b Abs. 3 UrhG ihn ermöglicht. Urheber oder ausschließliche Rechtsinhaber werden von der Vorbehaltserklärung des Inhabers eines anderen ausschließlichen, aber räumlich oder zeitlich beschränkten Nutzungsrechts nicht beeinträchtigt. Maßgeblich hierfür ist die wiedergabebezogene Wirkung der Vorbehaltserklärung. Insgesamt sprechen die vorgenannten Gründe dafür, dass auch Inhaber eines ausschließlichen, aber räumlich oder zeitlich beschränkten Nutzungsrechts uneingeschränkt vorbehaltbefugt sind. Auch sie können damit Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG abgeben.

(3) Vertragliche Schutzpflicht beim Auseinanderfallen von Wiedergabeherrschaft und Vorbehaltbefugnis

In der praktischen Umsetzung der durch § 44b Abs. 3 UrhG gewährten Vorbehaltsmöglichkeit kann es mitunter zu einem Auseinanderfallen der Herrschaftsgewalt über die Wiedergabe des Werkes und der Befugnis zur Erklärung des Nutzungsvorbehalts im Sinne des § 44b Abs. 3 UrhG kommen. Dies ist insbesondere bei der öffentlichen Wiedergabe von Werken durch einfache Rechtsinhaber der Fall. Sie sind kraft ihres Nutzungsrechts zwar wiedergabe-, nicht aber vorbehaltbefugt. Der in diesen Fällen betroffene Urheber oder ausschließliche Rechtsinhaber ist de facto nicht in der Lage, in der Sphäre des nutzenden Rechtsinhabers die Vorbehaltserklärung umzusetzen.⁵⁸⁹ Auch die Möglichkeit, den Nutzungsvorbehalt als Stellvertreter des Vorbehaltbefugten analog §§ 164 ff. BGB erklären zu können,⁵⁹⁰ hilft hier nur bedingt weiter. Denn aus dem rechtlichen Können folgt nicht unbedingt, dass der nutzende Rechtsinhaber diese Befugnis auch tatsächlich ausübt. Prohibitive Effekte können zum Beispiel bestehen, wenn Kosten mit der Umsetzung der Vorbehaltserklärungen verbunden sind und der nutzende Rechtsinhaber im Übrigen kein eigenes Interesse daran hat, die Verwendung der Werke zum KI-Training zu unterbinden. Aus dem Abhängigkeitsverhältnis zwischen wiedergebendem Nutzer und vorbehaltbefugtem Urheber beziehungsweise ausschließlichem Rechtsinhaber leitet sich jedoch eine vertragliche Schutzpflicht (§ 241 Abs. 2 BGB) in Form einer Mitwirkungspflicht des Nutzers ab. Sie beinhaltet, dass auf Verlangen des Vorbehaltbefugten diejenigen erforderlichen Maßnahmen

589 Heine, GRUR-Prax 2024, 87 Rn. 20.

590 Dazu siehe oben unter 3. Teil A. III. 2. i) hh) (1).

zu ergreifen sind, damit der Urheber oder ausschließliche Rechtsinhaber die Vorbehaltserklärung in angemessener Weise abgeben kann.⁵⁹¹ Hiervon betroffen sind unter anderem auch Verwertungsgesellschaften, die Nutzern einfache Lizenzen zur Nutzung des Werkes einräumen, ohne aber in die Wiedergabehandlung der Lizenznehmer eingebunden zu sein.

Die Mitwirkungspflicht, die sich grundsätzlich bereits im Wege ergänzender Vertragsauslegung aus dem Lizenzvertrag zwischen den Parteien ableitet, kann auch durch eine ausdrückliche Bestimmung in den Lizenzbedingungen abgesichert werden. Das ist beispielsweise in Abschnitt I. Nr. 3 Abs. 3 GEMA Tarif VR-OD15 der Fall.⁵⁹²

ii) Fernwirkung des Nutzungsvorbehalts bei Bereitstellung von Vervielfältigungsvorlagen oder einer Zugangsvermittlung an Dritte

Online verfügbare Werke können insbesondere durch Inhaber von Datenbankzugängen, über die die Werke abrufbar sind, Dritten Datenverarbeitern für Text und Data Mining-Aktivitäten und damit auch für das Training von KI weitervermittelt werden. Das gilt auch dann, wenn in diesen Datenbanken Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG erklärt worden sind. Vor dem Hintergrund dieses Dreiecksverhältnisses zwischen Vorbehaltserklärendem, Zugangsinhaber und Datenverarbeiter ist klärungsbedürftig, ob und inwiefern für den Zugangsinhaber wahrnehmbare Nutzungsvorbehalte, beispielsweise in den Nutzungsbedingungen der Online-Datenbank, zulasten des außenstehenden Datenverarbeiters weiterwirken, auch wenn dieser unter Umständen keine Kenntnis von der Vorbehaltserklärung hat.⁵⁹³

Die Weitervermittlung ist auf zwei Arten umsetzbar. Zum einen kann einem Datenverarbeiter der Zugang zu im Internet verfügbaren Werken direkt weitergegeben werden. Das ist beispielsweise bei der Überlassung von Login-Daten zu einer geschützten Datenbank der Fall, aus der der Datenverarbeiter die zum Training benötigten Werke selbstständig erstmals vervielfältigt. Zum anderen können dem Datenverarbeiter auch Werkstücke als Vervielfältigungsvorlagen bereitgestellt werden, die zuvor vom Zu-

591 Zur Angemessenheit des Nutzungsvorbehalts siehe oben unter 3. Teil A. III. 2. i) bb).

592 Siehe den Tarif unter <https://ogy.de/Orf7> (zuletzt abgerufen am 30.7.2024).

593 Zu einer vergleichbaren Konstellation bei der Weitergabe des Zugangs zu Werken im Kontext der Tatbestandsvoraussetzung der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG bereits oben unter 3. Teil A. III. 2. f) bb).

gangsinhabern angefertigt worden sind. In diesem Fall fertigt der Datenverarbeiter in seinem eigenen Herrschaftsbereich Vervielfältigungsstücke auf Basis der zunächst vom Zugangsinhaber hergestellten Vervielfältigungsvorlagen an. Dies ist insbesondere bei Endkunden intermediärer Trainingsdatensammler der Fall, insofern die Intermediären nicht nur Linklisten bereitstellen, sondern die Werke selbst abspeichern und vorverarbeiten, bevor sie auf deren Grundlage fertige Trainingsdatensätze auf dem Markt anbieten.

(1) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen

Ob der durch den Zugangsinhaber wahrnehmbare Nutzungsvorbehalt auch zulasten des nachgelagerten Datenverarbeiters wirkt, der vom Zugangsinhaber angefertigte Werkstücke als Vervielfältigungsvorlage nutzt, lässt sich mit Hilfe der wiedergabebezogenen Wirkung des Nutzungsvorbehalts beantworten.⁵⁹⁴ Denn der Zugangsinhaber bietet dem Datenverarbeiter seine eigenen, zuvor auf Grundlage des Zugangs angefertigten Werkstücke als neue Vervielfältigungsvorlagen an. Demzufolge kann eine mit den Ausgangswiedergaben, nicht aber mit den neuen Vervielfältigungsvorlagen verknüpfte Vorbehaltserklärung keine Beendigungswirkung für die Nutzung dieser neuen Vervielfältigungsvorlagen entfalten. Die Vorbehaltswirkung bleibt vielmehr auf die Nutzung der Ausgangswiedergabe beschränkt. Die Nutzungshandlungen des nachgelagerten Datenverarbeiters sind dementsprechend von § 44b Abs. 2 UrhG gedeckt und erlaubnisfrei gestellt, auch wenn an der ursprünglichen Werkwiedergabe ein Nutzungsvorbehalt erklärt worden ist. Unabhängig hiervon sind die Vervielfältigungshandlungen des Zugangsinhabers urheberrechtswidrig, wenn er bei ihnen wirksame Nutzungsvorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG nicht beachtet. Zudem könnte durch den Zugangsinhaber eine rechtswidrige Verbreitung (§ 17 UrhG) oder öffentliche Wiedergabe (§ 15 Abs. 2 UrhG) der vervielfältigten Werke stattfinden. Das ist im Detail aber einzelfallabhängig.

Fraglich ist jedoch, ob der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL einer Anwendung der Schrankenbestimmung zugunsten des Datenverarbeiters in der untersuchten Fallgruppe dann entgegensteht, wenn die Bereitstellung der Vervielfältigungsvorlagen der Werke im Interesse des Datenverarbeiters, also nicht im Rahmen einer Hilfstätigkeit für den Zugangsinhaber erfolgt, und der Datenverarbei-

⁵⁹⁴ Zur wiedergabebezogenen Wirkung siehe oben unter 3. Teil A. III. 2. i) ff).

ter objektiv erkennen kann, dass die ursprünglich genutzten Inhalte mit einem Nutzungsvorbehalt versehen waren.⁵⁹⁵ Gegen eine Unvereinbarkeit mit dem Dreistufentest spricht aber, dass die Bereitstellung der Vervielfältigungsvorlagen der Werke im vorliegenden Fall nicht generell zu einer urheberrechtskonformen Nutzung der Werke auf Basis von § 44b UrhG führt, wie es bei der Bereitstellung von Vervielfältigungsvorlagen *zugangsbeschränkter* Werke an Dritte der Fall ist. Vielmehr verletzen Zugangsinhaber, die die Werke trotz des Nutzungsvorbehalts zunächst vervielfältigen, bereits bei der Herstellung der Vervielfältigungsvorlagen mangels Anwendung der Schrankenbestimmung aus § 44b Abs. 2 UrhG die Rechtsposition betroffener Urheber beziehungsweise Rechtsinhaber. Es kommt damit nicht zu einer vollständigen Umgehung der Wirkung des Nutzungsvorbehalts. Die Urheberrechtsverletzung wird nur beim Zugangsinhaber lokalisiert. Der vorliegenden Sachverhaltskonstellation wohnt demzufolge kein Umgehungsmoment inne. Damit besteht ein Unterschied zur Bereitstellung von Vervielfältigungsvorlagen *zugangsbeschränkter* Werke an Dritte, die de jure zu einer Umgehung der eingesetzten Zugangsbeschränkung führt.⁵⁹⁶ Infolgedessen kommt es nicht zu einer ungebührlichen Verletzung der Urheberinteressen. Eine Einschränkung der Anwendung von § 44b Abs. 2 UrhG für den Fall, dass der Datenverarbeiter im eigenen Interesse tätig wird und objektiv erkennen kann, dass die ursprünglichen Werkwiedergaben mit einem Nutzungsvorbehalt versehen waren, findet über den Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL demzufolge nicht statt.

(2) Zugangsvermittlung

Auch die Auswirkung einer Vorbehaltserklärung auf Vervielfältigungshandlungen, die auf Grundlage eines dem Datenverarbeiter vom Zugangsinhaber weitervermittelten Werkzugangs vorgenommen worden sind, richtet sich nach der wiedergabebezogenen Wirkung des Nutzungsvorbehalts.⁵⁹⁷ Aufgrund dessen, dass dem Datenverarbeiter lediglich der Werkzugang vermittelt, also beispielsweise Login-Daten weitergegeben worden sind, nutzt er als Vervielfältigungsvorlage weiterhin die ursprüngliche Wiedergabe beziehungsweise das ursprüngliche Werkstück, an dem der Nutzungs-

595 Vgl. zu einem Parallelproblem bei der Weiterwirkung von Zugangsbeschränkungen im Rahmen des § 44b Abs. 2 S. 1 UrhG oben unter 3. Teil A. III. 2. f) bb).

596 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (3).

597 Zur wiedergabebezogenen Wirkung siehe oben unter 3. Teil A. III. 2. i) ff).

vorbehalt erklärt worden ist. Damit bleibt der wiedergabebezogen wirkende Nutzungsvorbehalt ausschlaggebend für die Freistellungswirkung des § 44b Abs. 2 UrhG. Die Zugangsweitergabe hat auf die Wirkung des Nutzungsvorbehalts demzufolge keinen Einfluss. Auch wenn Datenverarbeitern der Werkzugang von einem Zugangsinhaber vermittelt wird, müssen sie beim Vervielfältigen der Inhalte Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG auslesen. Dies setzt unter Umständen voraus, dass sich die Datenverarbeiter Zugriff auf die Nutzungsbedingungen einer Werkdatenbank verschaffen, bevor sie Vervielfältigungen für die Datenanalyse anfertigen.

jj) Zwischenergebnis zum Nutzungsvorbehalt

Alles in allem müssen Nutzungsvorbehaltserklärungen ausdrücklich und angemessen sein, um die in § 44b Abs. 3 UrhG angelegte Beendigungswirkung auszulösen. Beides wird zwar nicht ausdrücklich in § 44b Abs. 3 UrhG vorausgesetzt. Dennoch sind die Anforderungen im Wege einer richtlinienkonformen Auslegung des „sich Vorbehalten“ im Sinne des § 44b Abs. 3 S. 1 UrhG mitzubersichtigen. Trotz der Ausdrücklichkeitsanforderung dürfen die Nutzungsvorbehalte allerdings mit anderen, nicht auf § 44b Abs. 3 UrhG bezogenen Vorbehaltserklärungen verbunden werden. Vorbehaltserklärungen gegen die Nutzung online zugänglicher Werke müssen darüber hinaus gemäß § 44b Abs. 3 S. 2 UrhG maschinenlesbar sein. Dies setzt zwar voraus, dass die Vorbehaltserklärungen digital hinterlegt sind, sodass sie unmittelbar durch einen Computer verarbeitet werden können. Im Übrigen ist aber weder eine Formalisierung noch eine Strukturierung der Vorbehaltserklärungen notwendig. Damit können Nutzungsvorbehalte bei online verfügbaren Werken grundsätzlich auch in natürlicher Sprache abgegeben werden. Dies wird aber durch die richtlinienkonform zu berücksichtigende Angemessenheitsvoraussetzung eingegrenzt. Hiernach müssen die Erklärungen insbesondere gewisse Klarheitsanforderungen erfüllen. Es bestehen weiterhin gewisse Einschränkungen hinsichtlich der verwendbaren Sprache und des Ortes, an dem die digitalen Vorbehaltserklärungen hinterlegt werden dürfen. Insgesamt wird durch die Erklärbarkeit des Nutzungsvorbehalts in natürlicher Sprache eine niedrige Hürde für Rechteinhaber zur Abgabe der Vorbehaltserklärung aufgestellt. Praktisch können die Anforderungen unter anderem auch durch die Verwendung des Robots Exclusion Standards, des Spawning-Protokolls und des W3C TDM Reservation Protocols erfüllt werden. Wegen der niederschweligen Wirk-

samkeitsvoraussetzungen und der leichten Erklärbarkeit des Vorbehalts ist aus einer praktischen Perspektive zu befürchten, dass viele Rechtsinhaber die Freistellungswirkung der Schrankenbestimmung aus § 44b UrhG für die Nutzung ihrer Werke beenden werden. Folge dessen ist eine erhebliche Unterminierung des im zweiten Teil der vorliegenden Untersuchung identifizierten Freistellungsbedürfnisses für die Speicherung von urheberrechtlich geschützten Trainingsdaten aus Internetquellen. Nur die wiedergabebezogene Wirkung des Nutzungsvorbehalts trägt dem Freistellungsbedürfnis insoweit Rechnung, als das eine weitreichende, werkbezogene Ausschlusswirkung der Vorbehaltserklärungen vermieden wird.

j) Fazit zur Schrankenbestimmung für Text und Data Mining

Alles in allem kann das KI-Training nach einer richtlinienkonformen Auslegung des Tatbestands unter den Text und Data Mining-Begriff des § 44b Abs. 1 UrhG gefasst werden. Damit wird das Web Scraping von urheberrechtlich geschützten Trainingsdaten grundsätzlich von § 44b Abs. 2 UrhG freigestellt. Das gilt sowohl für kommerziell als auch für nicht-kommerziell tätige Trainingsdatensammler. Allerdings bleibt die Schrankenbestimmung aufgrund der in § 44b Abs. 2 S. 2 UrhG geregelten Löschungspflicht und dem in § 44b Abs. 3 UrhG vorgesehenen Opt-out-Mechanismus dennoch hinter den im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnissen zurück.

Bei der Erfassung und Speicherung von Werken zum Training von KI haben Datenverarbeiter auch dann rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG, wenn ihnen Vervielfältigungsvorlagen von einem Zugangsinhaber zur Verfügung gestellt worden sind, ohne dass die Datenverarbeiter selbst die Zugangsvoraussetzungen erfüllen. Aufgrund des Dreistufentests muss hiervon aber eine Ausnahme gemacht werden, wenn für den im eigenen Interesse handelnden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlagen auf Grundlage von Werkstücken aus einer zugangsgeschützten Quelle hergestellt wurden. Das gilt jedoch nicht, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, auf Basis des Werkzugangs angefertigte Werkstücke dritten Datenverarbeitern als Vorlagen für Vervielfältigungen zum Text und Data Mining bereitzustellen. Nach diesen Maßstäben ist zwar auch eine direkte Zugangsweitergabe eines vermittelnden Zugangsinhabers,

beispielsweise durch die Weitergabe von Login-Daten an einen Datenverarbeiter möglich. Hier führt eine Anwendung des Dreistufentests aber dazu, dass eine Anwendung der Schrankenbestimmung aus § 44b UrhG zugunsten des Datenverarbeiters im Regelfall dann ausscheidet, wenn er nicht als Hilfsperson des Zugangsinhabers, sondern im eigenen Interesse tätig wird.

Vervielfältigungen müssen gemäß § 44b Abs. 2 S. 1 UrhG gelöscht werden, wenn sie für das Text und Data Mining nicht mehr erforderlich sind. Dies bezieht sich auf ein Training eines konkreten KI-Systems. Die Vervielfältigungen der Werke dürfen aber aufbewahrt werden, solange noch Nachbesserungen am jeweiligen KI-System vorgenommen werden, für die die Vervielfältigungen notwendig sind.

Vervielfältigungen von im Internet abrufbaren Werken können jedoch nicht durch die Schrankenbestimmung für Text und Data Mining freigestellt werden, wenn ein maschinenlesbarer Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist. Hierfür genügt grundsätzlich eine Vorbehaltserklärung in natürlicher Sprache. Es müssen aber die Angemessenheits- und Ausdrücklichkeitsvoraussetzungen erfüllt sein, die sich aus einer richtlinienkonformen Auslegung des § 44b Abs. 3 UrhG ergeben. Die Möglichkeit, einen Nutzungsvorbehalt zu erklären, konterkariert erheblich die im zweiten Teil der vorliegenden Untersuchung identifizierten Freistellungsbedürfnisse.

3. Andere Schrankenbestimmungen

Klärungsbedürftig bleibt damit, ob sich Trainingsdatensammler zur Verwirklichung der identifizierten Freistellungsinteressen auf andere urheberrechtliche Schrankenbestimmungen berufen können.

a) Schrankenbestimmung für Zitate (§ 51 UrhG)

Da trainierte KI-Systeme bei ihrer Anwendung mittelbar auf die Trainingsdaten Bezug nehmen, könnte zum einen die Schrankenbestimmung für Zitate (§ 51 UrhG) angewendet werden. Nach § 51 Abs. 1 S. 1 UrhG ist hierfür notwendig, dass von den Trainingsdatensammlern ein Zitatziel verfolgt wird. Beim Web Scraping von Werken für das KI-Training könnte das nur der Fall sein, um eine geistige Auseinandersetzung zwischen den für das Training vervielfältigten Werken und den Aussagen des Trainings-

datensammlers oder KI-Nutzers zu ermöglichen.⁵⁹⁸ Die Erstellung eines Trainingsdatenkorpus beinhaltet jedoch regelmäßig keinen Aussagegehalt, sodass eine geistige Auseinandersetzung mit den vervielfältigten Werken bereits deswegen nicht stattfindet.⁵⁹⁹ Auch bei der Datenannotation wird ein für das Training genutztes Werk lediglich beschrieben. Einen eigenständigen Aussagegehalt enthält die Annotation hingegen nicht. Selbst wenn die KI-Nutzer mit der auf das Training folgenden Anwendung der KI einen Aussagegehalt transportieren, mangelt es ihnen außerdem typischerweise an der Kenntnis über die konkret verwendeten Trainingsdaten. Eine geistige Auseinandersetzung mit den Trainingswerken ist dann ebenfalls nicht möglich. Insgesamt wird mit dem Web Scraping von Werken zum Training von KI keine geistige Auseinandersetzung angestrebt, sodass die Datensammlung mangels Verfolgung eines Zitatzwecks nicht über § 51 UrhG freigestellt werden kann.⁶⁰⁰

b) Schrankenbestimmung für die Benutzung eines Datenbankwerks (§ 55a UrhG)

Zumindest auf das Web Scraping von Datenbankwerken könnte darüber hinaus § 55a UrhG anwendbar sein. Da die dauerhafte Vervielfältigung, so wie sie bei der Erfassung und Speicherung von Werken für das Training von KI stattfindet, das für die übliche Benutzung der Datenbankwerke erforderliche Maß im Regelfall aber überschreitet,⁶⁰¹ kann § 55a UrhG im Ergebnis nicht das Web Scraping von Datenbankwerken für das Training von KI freistellen.⁶⁰² Wenn es sich bei den zum Training vervielfältigten Datenbankwerken hingegen um öffentlich geteilte Trainingsdatensätze handelt, liegt im Upload der Datenbankwerke bereits eine schlichte Einwilligung bezogen auf ihre Weiterverwendung für das Training von KI. In diesen Fällen ist eine Freistellung demnach auch ohne Eingreifen der Schrankenbestimmung gewährleistet.⁶⁰³

598 Vgl. zur Voraussetzung EuGH, Urt. v. 29.07.2019 - C-516/17, GRUR 2019, 940 Rn. 78 – Reformistischer Aufbruch.

599 So auch *Spindler*, GRUR 2016, 1112 (1116); *Spindler*, ZGE 2018, 273 (278).

600 Etwas weniger streng, größtenteils aber ablehnend *Kleinkopf*, Text- und Data-Mining, S. 183 f.

601 *Lüft/Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 55a Rn. 75; sehr kritisch auch *Spindler*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 55a.

602 *Kleinkopf*, Text- und Data-Mining, S. 187.

c) Schrankenbestimmung für die Nutzung als unwesentliches Beiwerk (§ 57 UrhG)

Eine Freistellung des Web Scrapings von Werken für das Training von KI ließe sich im Einzelfall auch über die Schrankenbestimmung für unwesentliches Beiwerk, § 57 UrhG, erreichen.

aa) Irrtümlich nicht erfasste Nutzungsvorbehalte (§ 44b Abs. 3 UrhG)

Denkbar ist zum einen, dass Trainingsdatensammler trotz einer Implementierung technischer Sicherungsmechanismen zur Berücksichtigung von Nutzungsvorbehalten unbeabsichtigterweise Inhalte vervielfältigen, für deren Nutzung Vorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG abgegeben worden sind. Diese unbeabsichtigten Vervielfältigungen könnten in den Anwendungsbereich der Schrankenbestimmung für die Nutzung von Werken als unwesentliches Beiwerk (§ 57 UrhG) fallen.⁶⁰⁴

Entscheidend ist dabei zum einen, welches Objekt in diesen Fällen Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG ist. Dafür, dass sich der anvisierte Trainingsdatenkörper bestehend aus den über § 44b UrhG rechtmäßig nutzbaren Werken in seiner Gesamtheit als Hauptgegenstand der Nutzungshandlung qualifiziert, spricht, dass Trainingsdaten vom relevanten Verkehrskreis schon deswegen als Datensätze und nicht als einzelne Datenpunkte betrachtet werden, weil das Training von KI in der Regel einer Vielzahl an Datenpunkten bedarf und daher stets nur der Datensatz als Ganzes technische und wirtschaftliche Bedeutung entfaltet.⁶⁰⁵ Eine Gruppierung von Vervielfältigungsobjekten kann im Rahmen des § 57 UrhG darüber hinaus trotz dessen, dass § 57 UrhG lediglich im Singular von *dem Gegenstand* der Vervielfältigung spricht, als Hauptgegenstand im Sinne des § 57 UrhG angesehen werden.⁶⁰⁶ Denn Zweck des § 57 UrhG ist, eine Werknutzung freizustellen, die die urheberrechtlichen Verwertungsinteressen nur unerheblich beeinträchtigt. Dafür ist nur von untergeordneter Bedeutung, im Verhältnis zu welchem Objekt oder welchen Objekten

603 Anders *Kleinkopf*, Text- und Data-Mining, S. 187, die in diesen Fällen § 55a UrhG zur Anwendung bringen will.

604 Dazu im Kontext erweiterter kollektiver Lizenzen ähnlich auch in *Pukas*, GRUR 2023, 614 (619 f.).

605 Vgl. zu den quantitativen Anforderungen für die Trainingsdatensätze im Überblick oben I. Teil B. II. 3.

606 *Jacobs*, in: FS Ahrens, S. 225 (227).

die Werknutzung als unwesentliches Beiwerk erscheint. Dafür, dass sich dieser *anvisierte* Trainingsdatenkorpus bestehend aus den über § 44b UrhG rechtmäßig nutzbaren Werken als Hauptgegenstand der Nutzungshandlung im Sinne des § 57 UrhG qualifiziert, spricht weiterhin, dass Trainingsdatensammler aufgrund der Einsetzung technischer Sicherungsmechanismen zur Berücksichtigung von Nutzungsvorbehalten im vorliegenden Fall darauf abzielen, nur solche Inhalte zu erfassen, die nicht mit einem Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG versehen worden sind. Deswegen ist eine Abgrenzung zu den fälschlicherweise trotz eines erklärten Nutzungsvorbehalts vervielfältigten Werken möglich. Beide Werkgruppen sind in der vorliegenden Fallkonstellation demzufolge nicht gleichrangig. Insgesamt sprechen damit mehrere Gründe dafür, dass der Trainingsdatenkorpus bestehend aus den Inhalten, die nicht mit einem Nutzungsvorbehalt versehen worden sind, Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG ist.⁶⁰⁷

Einzelne Werke, die unbeabsichtigt trotz eines wirksamen Nutzungsvorbehalts vervielfältigt worden sind, könnten im Verhältnis zu diesem Hauptgegenstand als unwesentliches Beiwerk erscheinen.⁶⁰⁸ Voraussetzung dessen ist, dass sie nur beiläufig einbezogen und vom Durchschnittsbetrachter nicht wahrgenommen werden, sodass sie beliebig weggelassen oder ausgetauscht werden können und die Gesamtwirkung des Hauptgegenstands der Vervielfältigung nicht beeinflussen.⁶⁰⁹ Dies ist unter Berücksichtigung aller Umstände des Einzelfalles zu beurteilen.⁶¹⁰ In Relation zum gesamten Trainingsdatenkorpus aus den aufgrund von § 44b UrhG rechtmäßig erfassten Werken machen einzelne, unbeabsichtigterweise trotz Nutzungsvorbehaltserklärung vervielfältigte Werke einen nur untergeordneten Teil des Datenkorpus aus. Die einzelnen Werke aus dem Trainingsdatensatz herauszunehmen, würde seine Gesamtwirkung nicht beeinflussen. Die betroffenen Werke sind daher beliebig austauschbar. Ob darüber hinaus eine nur beiläufige Einbeziehung der Werke stattfindet, die von § 57 UrhG freigestellt ist, kann bei der Berücksichtigung sämtlicher Umstände des Einzelfalles davon abhängig gemacht werden, ob der Trainingsdatensammler wirksame Maßnahmen zur Beachtung erklärter Nutzungsvorbehalte gemäß § 44b

607 Ähnlich zum Widerspruch bei ECL-Lizenzlösungen schon in *Pukas*, GRUR 2023, 614 (619 f.).

608 Vgl. *Pukas*, GRUR 2023, 614 (619 f.).

609 BGH, Urt. v. 17.11.2014 - I ZR 177/13, GRUR 2015, 667 Rn. 27, 31 – Möbelkatalog; *Vogel*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 57 Rn. 10 m. w. N. Kritisch zur Austauschbarkeit *Jacobs*, in: FS Ahrens, S. 225 (229 f.).

610 BGH, Urt. v. 17.11.2014 - I ZR 177/13, GRUR 2015, 667 Rn. 21 – Möbelkatalog.

Abs. 3 UrhG ergriffen und damit bestmögliche Anstrengungen unternommen hat, dass es nicht zur Vervielfältigung der betreffenden Werke kommt. Dann ist die fehlerhafte, unbeabsichtigte Vervielfältigung einzelner, über § 44b Abs. 3 UrhG zur Nutzung durch den Urheber vorbehaltener Werke als unwesentliches Beiwerk des im Übrigen urheberrechtskonform akquirierten Trainingsdatensatzes von § 57 UrhG gedeckt. Die Vervielfältigung der Werke ist in diesem Fall demzufolge urheberrechtlich frei möglich. Mit dieser einzelfallbezogenen Anwendung von § 57 UrhG werden auch nicht strukturell die Voraussetzungen des § 44b UrhG unterlaufen.

Die Anwendung des § 57 UrhG auf diesen Einzelfall, ändert aber nichts daran, dass die im Sinne des § 44b Abs. 3 UrhG erklärten Nutzungsvorbehalte bestmöglich berücksichtigt werden müssen. Selbst die einzelfallbezogene Anwendung der Schrankenbestimmung für die Nutzung von Werken als unwesentliches Beiwerk führt daher nicht zu einer allgemeinen Freistellung des Web Scrapings von im Internet abrufbaren Werken für das Training von KI. Auch die Anwendung von § 57 UrhG kann die im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnisse daher nicht erfüllen.

bb) Trainingsdatensätze aus primär nicht urheberrechtlich geschützten Inhalten

Möglich ist darüber hinaus auch, dass Trainingsdatensammler, deren Ziel die Erfassung und Verarbeitung urheberrechtlich nicht geschützter Inhalte ist, für die Vervielfältigung einzelner Werke, die trotz dessen irrtümlich gespeichert werden, von der Schrankenbestimmung aus § 57 UrhG profitieren könnten. Denkbar ist das beispielsweise, wenn der Trainingsdatensammler amtliche und damit gemäß § 5 UrhG nicht geschützte Werke vervielfältigt, in einem dieser Werke aber urheberrechtlich geschützte Inhalte eingebettet sind, beispielsweise ein Kartenausschnitt.⁶¹¹ Damit § 57 UrhG für die Vervielfältigung dieser Werke eingreifen kann, müssen die unbeabsichtigt vervielfältigten Werke auch hier gemäß § 57 UrhG neben dem eigentlichen Gegenstand der Vervielfältigung, dem Trainingsdatensatz aus urheberrechtsfreien Inhalten, als unwesentliches Beiwerk erscheinen.

611 Vgl. zu einer ähnlichen Sachverhaltskonstellation BGH, Urt. v. 21.01.2021 - I ZR 59/19, ZUM 2021, 527 – Kastellaun.

Für eine Einordnung des vom Trainingsdatensammler zielgerichtet akquirierten Trainingsdatensatzes aus urheberrechtsfreien Inhalten als Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG ist zunächst unschädlich, dass der Trainingsdatensatz als Objektgruppe aus mehreren Datenpunkten besteht.⁶¹² Weiterhin dient § 57 UrhG dem Zweck, der Allgemeinheit wirtschaftlich unerhebliche Werknutzungen freizugeben. Dabei kann es nicht darauf ankommen, im Verhältnis zu welchem Objekt die wirtschaftlich unerhebliche Werknutzung stattfindet. Aus diesen Gründen ist ebenfalls nicht notwendig, dass der Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG selbst urheberrechtlich geschützt ist.⁶¹³ Darüber hinaus besteht aufgrund der Zweckrichtung der Datensammlung keine Gleichwertigkeit zwischen den unbeabsichtigt vervielfältigten Werken als Nebenfolge der Datensammlung und den zielgerichtet gespeicherten urheberrechtsfreien Inhalten.⁶¹⁴ Letztere verkörpern im Verhältnis zu den vervielfältigten Werken demzufolge einen eigenständigen Gegenstand. Auch im vorliegenden Fall kann der Trainingsdatensatz aus urheberrechtsfreien Inhalten in seiner Gesamtheit damit als Hauptgegenstand der Vervielfältigungshandlung im Sinne des § 57 UrhG angesehen werden.

Für die Anwendung des § 57 UrhG kommt es infolgedessen darauf an, ob die versehentlich erfassten Werke im Verhältnis zum Trainingsdatensatz aus urheberrechtsfreien Inhalten als unwesentliches Beiwerk im Sinne des § 57 UrhG erscheinen.⁶¹⁵ Auch im vorliegenden Fall machen die vervielfältigten Werke nur einen unwesentlichen Teil des gesamten Trainingsdatensatzes aus. Sie könnten demzufolge weggelassen oder ausgetauscht werden, ohne dass die Qualität und damit die Gesamtwirkung des Datensatzes beeinflusst wird. In die anschließend vorzunehmende Beurteilung der Beiläufigkeit der Vervielfältigung, die in einer Gesamtabwägung mündet, kann eingestellt werden, dass die Vervielfältigungsstücke nicht zielgerichtet hergestellt werden. Ob dies tatsächlich der Fall ist, lässt sich unter anderem anhand der für den Webcrawler definierten Regeln bewerten. Von Bedeutung ist auch, auf welche Quellen der Trainingsdatensammler das Web Scraping ausgerichtet hat. Im Detail ist das einzelfallabhängig. Sind die Vervielfältigungen

612 Dazu siehe unmittelbar oben unter 3. Teil A. III. 3. c) aa).

613 Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 57 Rn. 1; Jacobs, in: FS Ahrens, S. 225 (227); vgl. z. B. auch OLG Köln, Urt. v. 23.08.2013 - 6 U 17/13, GRUR-RR 2014, 58 (59), das als Hauptgegenstand einen „Verkaufskatalog“ und eine „Internetseite“ ansieht.

614 Zu einer vergleichbaren Fallkonstellation etwas ausführlicher siehe unmittelbar oben unter 3. Teil A. III. 3. c) aa).

615 Zu den Voraussetzungen siehe unmittelbar oben unter 3. Teil A. III. 3. c) aa).

beiläufig, ist davon auszugehen, dass sie unwesentliches Beiwerk des Trainingsdatensatzes aus primär urheberrechtsfreien Inhalten im Sinne des § 57 UrhG sind. Infolgedessen werden die Vervielfältigungshandlungen in diesem Fall durch die Schrankenbestimmung aus § 57 UrhG erlaubnisfrei gestellt. Auch das führt jedoch nicht zu einer umfassenden Freistellung des Web Scrapings von Werken als KI-Trainingsdaten. Durch die Anwendung des § 57 UrhG in diesem Einzelfall kann das im zweiten Teil der vorliegenden Untersuchung identifizierte Freistellungsbedürfnis daher ebenfalls nicht vollständig erfüllt werden.

d) Schrankenbestimmung für Karikatur, Parodie und Pastiche (§ 51a UrhG)

Im Einzelfall könnte die Vervielfältigung von Werken als Trainingsdaten auch durch die Schrankenbestimmung für Karikatur, Parodie und Pastiche aus § 51a UrhG erlaubt sein.⁶¹⁶ Denkbar ist, dass bereits die Vervielfältigung konkret ausgewählter Werke, die zum Training einer künstlerisch eingesetzten KI genutzt werden sollen, als vorgelagerte Nutzung von der Schrankenbestimmung privilegiert wird, wenn mit der KI anschließend Karikaturen, Parodien oder Pastiches erzeugt werden sollen.⁶¹⁷ Denn dann könnte bereits die (Auswahl und) Vervielfältigung der als Trainingsdaten genutzten Werke eine künstlerische Betätigung darstellen, die für einen effektiven Grundrechtsschutz und eine wirksame Anwendung des § 51a UrhG von der Schrankennutzung mit zu umfassen ist.⁶¹⁸ Notwendig wäre hierfür aber, dass bereits zum Zeitpunkt der Speicherung der Trainingsdaten hinreichend konkretisiert ist, dass der Output des trainierten, künstlerisch genutzten KI-Systems in den Anwendungsbereich der Schrankenbestimmung für Karikatur, Parodie und Pastiche aus § 51a UrhG fallen wird. Zudem darf nicht *auch*, sondern *nur* die Erzeugung von Karikaturen, Parodien und Pastiches bezweckt sein.⁶¹⁹ Auch die Abwägung im Rahmen des Dreistufen-

616 Insgesamt dazu im Detail *Lennartz/Kraetzig*, GRUR 2024, 1258, wobei die AutorInnen von einem weiten Pastichebegriff ausgehen, der jede abgeleitete künstlerische Betätigung erfasst.

617 *Lennartz/Kraetzig*, GRUR 2024, 1258 (1261, 1262).

618 *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260 f.).

619 Großzügiger hingegen *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 158 f., nach denen es ausreicht, wenn die Erfüllung der Voraussetzung des § 51a UrhG „Hauptzweck“ des KI-Systems ist.

tests, die bei der Anwendung des § 51a UrhG eine wesentliche Bedeutung entfaltet,⁶²⁰ müsste für den Output damit vorweggenommen werden. Ohne Kenntnis vom konkreten Output des KI-Systems ist das jedoch kaum prognostizierbar. Insgesamt ist eine Freistellung der Vervielfältigungen über § 51a UrhG damit nur in begrenzten Einzelfällen denkbar.

e) Intermediäre Trainingsdatensammler als Bibliotheken (§ 60d UrhG)

Von den regulatorischen Freistellungsbedürfnissen für das Web Scraping urheberrechtlich geschützter Trainingsdaten ist auch die Tätigkeit intermediärer Trainingsdatensammler umfasst. Sie nehmen eine Vermittlerrolle zwischen Urhebern und KI-Entwicklern ein. Vor diesem Hintergrund könnte ihre Datensammlung nicht nur durch § 44b UrhG, sondern auch durch die erweiterte Schrankenbestimmung aus § 60d UrhG freigestellt sein. Diese privilegiert unter anderem Forschungseinrichtungen sowie Vermittler wie Archive oder Bibliotheken. Da es für die Erzielung einer generellen Freistungswirkung zugunsten intermediärer Trainingsdatensammler nicht auf eine Forschungstätigkeit ankommen kann, müssen sie unter den Begriff der Kulturerbe-Einrichtung im Sinne des § 60d Abs. 3 Nr. 1 UrhG gefasst werden können, um in den persönlichen Anwendungsbereich der erweiterten Schrankenbestimmung zu fallen. In Betracht kommt lediglich, dass sie sich als Bibliotheken im Sinne der Regelung qualifizieren.

Der Bibliotheksbegriff ist weder in § 60d UrhG noch in Art. 2 Nr. 3, Art. 3 DSM-RL legaldefiniert, dessen Umsetzung § 60d UrhG dient. Auch in der InfoSoc-RL (Art. 5 Abs. 2 lit. c) sowie der verwaiste Werke-RL⁶²¹ (Art. 1) wird er ohne eine nähere Begriffsbestimmung aufgegriffen. Seine Bedeutung muss deswegen mit Hilfe der üblichen Auslegungsmethoden bestimmt werden. Die Basis der Auslegung bildet dabei auch im europäischen Recht der natürliche Wortsinn des verwendeten Rechtsbegriffs.⁶²² Im allgemeinen deutschen Sprachgebrauch wird unter einer Bibliothek ein Aufbewahrungsort für eine systematisch geordnete Sammlung von Büchern verstanden.⁶²³

620 Stieper, GRUR 2020, 699 (703).

621 Richtlinie 2012/28/EU des Europäischen Parlaments und des Rates vom 25. Oktober 2012 über bestimmte zulässige Formen der Nutzung verwaister Werke.

622 EuGH, Urt. v. 05.07.2012 - C-49/11, EuZW 2012, 638 Rn. 32 – Content Services; EuGH, Urt. v. 10.03.2005 - C-336/03, MMR 2005, 364 Rn. 21 – easyCar.

Selbst wenn davon abgesehen wird, dass intermediäre Trainingsdatensammler nicht nur Schriftwerke, also „Bücher“, vervielfältigen, ist darüber hinaus fraglich, ob sie eine Systematisierung ihrer Inhalte vornehmen. Dagegen spricht, dass die Werke innerhalb der Trainingsdatensätze nicht in Kataloge aufgenommen werden. Hierfür ursächlich sind unter Umständen ein Fehlen der für eine Systematisierung notwendigen Metadaten sowie die Menge der verarbeiteten Inhalte. Eine Systematisierung ist zudem auch nicht für das Training notwendig. Zwar kann auf Plattformen intermediärer Trainingsdatensammler mitunter eine Katalogisierung, also systematische Anordnung verschiedener Trainingsdatensätze stattfinden.⁶²⁴ Vom Bibliotheksbegriff wird allerdings eine systematische Anordnung der gesammelten Inhalte selbst vorausgesetzt. Das sind im vorliegenden Fall die für das Training bestimmten Werkstücke.

Mangels Systematisierung der vervielfältigten Werke fallen intermediäre Trainingsdatensammler demzufolge nicht unter den in § 60d Abs. 3 Nr. 1 UrhG respektive Art. 2 Nr. 3 DSM-RL normierten Bibliotheksbegriff. Es kommt damit bereits nicht darauf an, ob als Bibliothek eine räumliche Einrichtung erforderlich ist.⁶²⁵ Intermediäre Trainingsdatensammler können damit jedenfalls so lange nicht von der erweiterten Schrankenwirkung des § 60d UrhG profitieren, wie sie nicht aus anderen Gründen in den persönlichen Anwendungsbereich der Schrankenbestimmung fallen.⁶²⁶ Die regulatorischen Defizite im Hinblick auf die Löschungspflicht (vgl. § 44b Abs. 2 S. 2 UrhG)⁶²⁷ sowie den Nutzungsvorbehalt (vgl. § 44b Abs. 3 UrhG)⁶²⁸ können durch die Anwendung der Schrankenbestimmung aus § 60d UrhG damit nicht zugunsten intermediärer Trainingsdatensammler geheilt werden.

623 Dudenredaktion (Bibliographisches Institut), Duden: das Fremdwörterbuch, Eintrag: „Bibliothek“.

624 Siehe zum Beispiel die Plattform „huggingface.co“, auf der unter anderem eine große Zahl an Trainingsdatensätzen angeboten wird.

625 Vgl. dazu Henke, E-Books im Urheberrecht, S. 118 ff.

626 Zu einem entsprechenden Fall, bei dem eine wissenschaftliche Analysen vorbereitende Datensammlung als ausreichend erachtet wurde, siehe LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

627 Siehe dazu oben unter 3. Teil A. III. 2. h).

628 Siehe dazu oben unter 3. Teil A. III. 2. i).

4. Rolle der Gedächtnisinstitutionen für das Training von KI

Kulturerbe-Einrichtungen im Sinne des § 60d Abs. 3 Nr. 1 UrhG, zu denen insbesondere Bibliotheken und Archive gehören, könnten eine wesentliche Rolle bei der Speicherung von KI-Trainingsdaten einnehmen. Denn sie halten einen erheblichen Teil des Kulturerbes und damit auch eine große Menge an Werken vor. Diese können zum Training von KI-Systemen genutzt werden. Für Trainingsdatensammler ist die Verwendung dieser Werkbestände mitunter vorteilhaft, weil eine große Zahl an Werken spezifischer Gattungen aus einer Quelle akquiriert werden kann. Die Werke und ihre Werkrepräsentationen erfüllen zudem gewisse Qualitätsanforderungen. Nutzer, die Trainingsdaten für die Entwicklung von KI-Systemen benötigen, könnten die Sammlungen der Gedächtnisinstitutionen vor diesem Hintergrund auf Basis von §§ 44b, 60d UrhG umfassend zur Erstellung von Trainingsdatenkorpora vervielfältigen. Dies umfasst auch digitale Sammlungsbestandteile wie elektronische Zeitschriftenbibliotheken oder anderen Datenbanken, zu denen jedenfalls größere Gedächtnisinstitutionen üblicherweise Zugänge bereithalten.

a) Rechtmäßige Zugänglichkeit der Werke für Nutzer der Gedächtniseinrichtungen

Damit Nutzer solcher Gedächtnisinstitutionen von der Freistellungswirkung der Schrankenbestimmungen aus den §§ 44b, 60d UrhG profitieren können, ist zunächst notwendig, dass sie gemäß § 44b Abs. 2 S. 1 UrhG rechtmäßigen Zugang zu den Werken haben, die von den Gedächtnisinstitutionen bereitgestellt werden. Zu unterscheiden ist dabei einmal zwischen Werken, die sich wie beispielsweise Buchbestände oder eigene Digitalisate physisch oder elektronisch im unmittelbaren Herrschaftsbereich der Gedächtnisinstitutionen befinden, und solchen Werken, die über Zugänge abrufbar sind, die eine Gedächtnisinstitution bei Dritten einlizenziert hat. Hierzu zählen beispielsweise Inhalte einer elektronischen Zeitschriftenbibliothek.

Zu Werken, die sich physisch oder elektronisch im unmittelbaren Herrschaftsbereich der Gedächtnisinstitutionen befinden, haben Nutzer aufgrund dessen rechtmäßigen Zugang im Sinne des § 44b Abs. 2 S. 1 UrhG, dass sie mit Abschluss ihres Nutzungsvertrags ordnungsgemäß die Zugangsbeschränkungen der Gedächtnisinstitution (z. B. Login-Beschrän-

kung mit Nutzernummer und Passwort) überwinden.⁶²⁹ Hier bestehen keine Unterschiede zu Nutzern von Online-Datenbanken, die mit Abschluss eines Abonnements rechtmäßigen Zugang zu den Inhalten der Datenbanken im Sinne des § 44b Abs. 2 S. 1 UrhG erhalten.

Etwas anderes könnte allerdings für Werke gelten, die über Zugänge abrufbar sind, die die Gedächtnisinstitutionen bei Dritten einlizenziert haben. In diesem Fall vermitteln die Gedächtnisinstitutionen den Nutzern den Zugang weiter. Die Nutzer handeln dabei nicht als Hilfspersonen der Gedächtnisinstitution, sondern im eigenen Interesse. Zwar ist für die Anwendung des Schrankenbestands aus § 44b UrhG grundsätzlich unschädlich, dass die Nutzer die vom Dritten aufgestellten Zugangsvoraussetzungen nicht selbst erfüllen, also insbesondere nicht selbst Zugangsdaten erworben haben.⁶³⁰ Allerdings können die Nutzer im vorliegenden Fall zumeist objektiv erkennen, dass die Werke einer zugangsgeschützten Quelle entstammen, die nicht zum unmittelbaren Herrschaftsbereich der Gedächtnisinstitution, also der eigenen Sammlung im engeren Sinne, gehört. Vervielfältigungen der durch diese Drittzugänge akquirierten Werke wohnt damit ein Umgehungsmoment inne. Denn der Nutzer spart sich eigene Aufwendungen, die für den Erwerb des Zugangs zu den Drittinhalten notwendig gewesen wären. Diese Umgehung beeinträchtigt die berechtigten Urheberinteressen übergebühlich. Darüber hinaus stellen Dritte (Urheber beziehungsweise Rechtsinhaber) durch die Zugangskontrolle den Schutz ihres Amortisationsinteresses sicher. Die Wahrung dieses Interesses würde durch eine Zugangsweitergabe an Nutzer von Gedächtnisinstitutionen zur Speicherung von Trainingsdaten vereitelt. Infolgedessen führt die Anwendung der Schrankenbestimmung aus § 44b UrhG in diesem Fall zu einer Verletzung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL.⁶³¹ Infolgedessen fällt die Vervielfältigung zugangsgeschützter Drittinhalte, die nicht im engeren Sinne zur Sammlung der Gedächtnisinstitutionen gehören und zu denen diese lediglich den Zugang an Nutzer weitervermitteln, um die Werke etwa für Trainingsdatensätze zu nutzen, nicht unter die Schrankenregelung des § 44b UrhG. Etwas anderes könnte aber dann gelten, wenn es für den Urheber oder sekundären Rechtsinhaber beim Zugangserwerb der Gedächtnisinstitution erkennbar gewesen ist, dass es sich um eine Gedächtnisinstitution handelt. Denn dann hätte er, bedingt durch die Funktion dieser Einrichtungen, wissen müssen,

629 Vgl. zur Voraussetzung siehe oben unter 3. Teil A. III. 2. f) aa).

630 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (1).

631 Hierzu ausführlich unter 3. Teil A. III. 2. f) bb) (3).

dass eine Zugangsvermittlung an Einrichtungsnutzer naheliegt. Dies hätte auch bei der Bestimmung der Gegenleistung für die Zugangsgewährung berücksichtigt werden können. Ein Schutz der Amortisationsinteressen hätte demzufolge sichergestellt werden können. Dem Sachverhalt haftet in diesem Fall außerdem kein Umgehungsmoment an.

Darüber hinaus findet die Zugangsvermittlung an Nutzer der Gedächtnisinstitutionen nur innerhalb der Infrastruktur der Einrichtungen statt. Eine unkontrollierte Weitergabe des Zugangs an beliebige Dritte, die faktisch zu einem Leerlaufen der Zugangsbeschränkung führen könnte, ist nicht zu befürchten.

Alles in allem führen die vorgenannten Gründe dazu, dass der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL nicht verletzt ist, wenn für den Urheber oder sekundären Rechtsinhaber im Rahmen der Zugangsverschaffung erkennbar gewesen ist, dass es sich um eine Gedächtnisinstitution handelt.⁶³² Alles in allem haben Nutzer von Gedächtnisinstitutionen zu von der Einrichtung bereitgestellten Werken rechtmäßigen Zugang im Sinne von § 44b Abs. 2 S. 1 UrhG, insofern die betreffenden Werke nicht in Datenbanken Dritter wie beispielsweise Verlagsdatenbanken wiedergegeben werden oder die Gedächtnisinstitutionen ihren Nutzern den Zugang zu Werken über Drittdatenbanken weitervermitteln, wobei beim Zugangserwerb durch die Gedächtnisinstitution für den jeweiligen Datenbankbetreiber erkennbar gewesen sein muss, dass es sich um eine Gedächtnisinstitution handelt.

b) Freistellung über § 60d UrhG zugunsten der Nutzer von Gedächtnisinstitutionen

Klärungsbedürftig ist weiterhin auch, ob Nutzer von Gedächtnisinstitutionen wie die Einrichtungen selbst von der erweiterten Freistellungswirkung des § 60d Abs. 3 S. 1 UrhG profitieren können. Hiervon könnten insbesondere kommerziell Forschende profitieren. Denn sie sind zwar nicht vom persönlichen Anwendungsbereich der Schrankenbestimmung aus § 60d

632 Ein etwas weiteres Verständnis legt unter Umständen der Regierungsentwurf zugrunde. Danach haben die Nutzer von Gedächtniseinrichtungen stets rechtmäßigen Zugang zu solchen Inhalten, eine teilweise Begrenzung über den Dreistufentest findet nicht statt, siehe BT-Drs. 19/27426, S. 97. Es bleibt allerdings unklar, ob in der Gesetzesbegründung tatsächlich Fälle der Zugangsvermittlung adressiert werden, dazu vgl. unten unter 3. Teil A. III. 5. b).

UrhG erfasst. Dennoch führt ihre wissenschaftliche Forschungstätigkeit dazu, dass die erweiterte Freistellungswirkung des § 60d UrhG sachlichen anwendbar wäre.

Damit Nutzer wie kommerziell Forschende von der erweiterten Freistellungswirkung profitieren können, müssen sie als Teil der Gedächtnisinstitutionen verstanden werden. Aus einer wortlautbezogenen Perspektive spricht zunächst dafür, dass der persönliche Anwendungsbereich nach § 60 Abs. 3 Nr. 1 UrhG nur solche Bibliotheken und Museen umfasst, die öffentlich zugänglich sind. Um allein angestelltes Forschungspersonal von Museen und Bibliotheken zu privilegieren, bedürfte es keiner Beschränkung der Privilegierungswirkung auf *öffentlich zugängliche* Einrichtungen. Vielmehr kommt der Einschränkung dann ein eigenständiger Bedeutungsgehalt zu, wenn auch die Nutzer der Einrichtungen in den persönlichen Anwendungsbereich der Schrankenbestimmung mit einbezogen würden. Gegen ein solches Tatbestandsverständnis spricht aus einer systematischen Perspektive allerdings, dass eben diese Nutzer als Einzelforscher nach § 60d Abs. 3 Nr. 2 UrhG vom Schrankentatbestand privilegiert werden, solange sie keine kommerziellen Zwecke verfolgen. Die Nutzer als Teil der Gedächtnisinstitutionen aufzufassen und ihnen damit über § 60d Abs. 3 Nr. 1 UrhG die Schrankenwirkung zu eröffnen, unterläuft die besondere Voraussetzung des § 60d Abs. 3 Nr. 2 UrhG. Nicht zuletzt bestehen Parallelen zur in § 60e Abs. 4 S. 2 UrhG geregelten Anschlusservielfältigung, die über Terminals in Bibliotheken ermöglicht wird. Auch diese Regelung umfasst ungeachtet ihres Wortlauts, der auf den ersten Blick eine Mitprivilegierung der Bibliotheksnutzer nahelegt, nur die öffentliche Wiedergabe der Bibliotheken an den Terminals. Die Nutzervielfältigungen richten sich hingegen nach gesondert anzuwendenden Schrankenbestimmungen, insbesondere § 53 UrhG.⁶³³ Eine Mitprivilegierung der Nutzer erfolgt demzufolge nicht.

Für eine Mitprivilegierung der Nutzer von Gedächtnisinstitutionen spricht aus einer geschichtshistorischen Perspektive jedoch, dass der nationale Umsetzungsgesetzgeber davon ausging, dass „individuelle Nutzer“, die der Gedächtnisinstitution „angehören“, in den Anwendungsbereich der Schrankenbestimmung fallen sollen.⁶³⁴ Es ist allerdings bereits unklar, welche Personengruppe der Umsetzungsgesetzgeber hiermit adressiert hat.

633 BGH, Urt. v. 16.04.2015 - I ZR 69/11, GRUR 2015, 1101 Rn. 41 – Elektronische Leseplätze II.

634 BT-Drs. 19/27426, S. 97.

Denn der Nutzerbegriff könnte sich zum einen auf die Nutzer von Gedächtnisinstitutionen im allgemeinsprachlichen Sinn, also insbesondere Nutzer von Bibliotheken, beziehen. Nach einem anderen, engeren Verständnis könnten hingegen individualisierte *Werknutzer* innerhalb der privilegierten Organisation adressiert worden sein. Bei Zugrundelegung des zweiten Begriffsverständnisses hätte der Umsetzungsgesetzgeber nur klargestellt, dass die rechtstechnische Privilegierung von Organisationsstrukturen zu einer Privilegierung der individuellen Personen führt, die in der Organisationsstruktur zusammengefasst werden. Demnach könnte beispielsweise die Adressierung von Hochschulen im Schrankentatbestand zu einer Privilegierung der individuellen, an den Hochschulen forschenden Mitglieder wie Professoren oder wissenschaftlichen Mitarbeitern führen. Für das zweite Begriffsverständnis spricht, dass der nationale Umsetzungsgesetzgeber mit seiner Erläuterung zu § 60d Abs. 3 UrhG an ErwGr. 14 DSM-RL anknüpft.⁶³⁵ Danach sollen auch Forschungsorganisationen oder Einrichtungen kulturellen Erbes „angehörige Personen“ von der erweiterten Schrankenbestimmung erfasst sein. Der Nutzerbegriff, welcher seinem allgemeinen Begriffsverständnis nach eine weite Auslegung des Tatbestands nahelegt,⁶³⁶ wird in ErwGr. 14 DSM-RL nicht verwendet.

Vielmehr könnte die in ErwGr. 14 DSM-RL gewählte Formulierung nach einer richtlinienkonformen Auslegung gegen eine Erstreckung der Schrankenwirkung auf Nutzer von Gedächtnisinstitutionen über § 60d Abs. 3 Nr. 1 UrhG sprechen. Denn es besteht ein qualitativer Unterschied zwischen einer Institution *angehörender* Personen, also eigenem Forschungspersonal, und *Nutznießern* einer Institution. Letztere profitieren als Außenstehende von der (Forschungs-)Leistung der Gedächtnisinstitution. Hierunter fallen beispielsweise Nutzer einer Bibliothek. Sie machen sich die Sammlungsleistung und eine gegebenenfalls stattfindende wissenschaftliche Aufbereitung der Sammlung durch die Bibliothek zunutze, ohne der Einrichtung unmittelbar zugehörig zu sein.

Gegen ein Verständnis des § 60d Abs. 3 Nr. 1 UrhG, nach dem auch Nutzer der Gedächtnisinstitutionen von der erweiterten Schrankenwirkung des § 60d UrhG profitieren, spricht darüber hinaus auch, dass Zweck der Erweiterung der Schrankenwirkung auf Gedächtnisinstitutionen in § 60d Abs. 3 Nr. 1 UrhG die Ermöglichung von Forschung ist, die die Einrichtung

635 BT-Drs. 19/27426, S. 96.

636 Dazu siehe unmittelbar oben in diesem Abschnitt.

gen selbst im Zusammenhang mit ihrer Haupttätigkeit betreiben.⁶³⁷ Eine Erweiterung der Schrankenwirkung auf individuelle Nutzer von Gedächtnisinstitutionen ist in diesem Regelungszweck nicht angelegt.

Alles in allem sprechen damit sowohl systematische als auch geschichtshistorische sowie teleologische Erwägungen dafür, § 60d Abs. 3 Nr. 1 UrhG so auszulegen, dass sich die Schrankenwirkung nicht auf Nutzer von Gedächtnisinstitutionen erstreckt. Nach § 60d Abs. 3 Nr. 1 UrhG privilegiert sind demnach nur die Gedächtnisinstitutionen selbst. Die urheberrechtsrelevanten und freigestellten Nutzungshandlungen werden dabei von den Gedächtnisinstitutionen unmittelbar angehörendem Forschungspersonal vorgenommen. Nutzer von Gedächtnisinstitutionen, denen über die Sammlungen und Datenbankzugänge der Institutionen rechtmäßiger Zugang zu Werken vermittelt wird,⁶³⁸ können Vervielfältigungen zum Training von KI daher nur auf Basis der allgemeinen Schrankenbestimmung für Text und Data Mining, § 44b UrhG, vornehmen.⁶³⁹

5. Zwischenergebnis zum Eingreifen einer urheberrechtlichen Schrankenbestimmung

Alles in allem kann der eigentliche KI-Trainingsvorgang unter die in § 44b Abs. 1 UrhG normierte Legaldefinition des Text und Data Mining subsumiert werden. Damit sind die beim Web Scraping vorgenommenen Vervielfältigungshandlungen grundsätzlich von der Schrankenbestimmung des § 44b Abs. 2 S. 1 UrhG gedeckt. Löschungspflicht (§ 44b Abs. 2 S. 2 UrhG) und Opt-out-Mechanismus für Urheber (§ 44b Abs. 3 UrhG) führen allerdings dazu, dass die Reichweite der Schrankenbestimmung für Text und Data Mining dennoch hinter dem im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnis zurückbleibt. Es kann auch nicht durch die Anwendung anderer Schrankenbestimmungen, insbesondere § 44a UrhG und § 57 UrhG, befriedigt werden, auch wenn es im Einzelfall insbesondere zu einer Freistellung der Vervielfältigungen von Werken zum Training von KI über § 57 UrhG kommen kann. Das ist der Fall, wenn die Werke unbeabsichtigt bei der Speicherung urheberrechtlicher Trainingsdaten miterfasst, oder wenn trotz wirksamer technischer Sicherungsmaßnahmen zur Berücksichtigung von Nutzungsvorbehaltserklärungen im Sinne des

637 ErwGr. 8 DSM-RL.

638 Dazu siehe unmittelbar oben unter 3. Teil A. III. 5. a).

639 Zu den Voraussetzungen siehe oben unter 3. Teil A. III. 2.

§ 44b Abs. 3 UrhG Vorbehaltserklärungen im Einzelfall nicht erkannt und berücksichtigt wurden. Von der erweiterten Schrankenbestimmung aus § 60d UrhG können außerdem keine Nutzer von Gedächtnisinstitutionen wie Bibliotheken oder Archiven profitieren. Wollen sie über die Gedächtnisinstitutionen erlangte Werke zum Training von KI vervielfältigen, sind sie demzufolge insbesondere auf § 44b UrhG verwiesen. Von der erweiterten Freistellungswirkung des § 60d UrhG profitieren darüber hinaus auch intermediäre Trainingsdatensammler nicht grundsätzlich. Sie sind keine Bibliotheken und damit Kulturerbe-Einrichtungen im Sinne des § 60d Abs. 3 Nr. 1 UrhG. Demzufolge eröffnen sie nicht den persönlichen Anwendungsbereich der Schrankenbestimmung. Sie müssen daher ebenfalls insbesondere auf § 44b UrhG zurückgreifen.

IV. Gesetzliche Vergütung für Urheber

Das im zweiten Teil dieser Untersuchung identifizierte Regelungsbedürfnis sieht neben der Freistellung des Web Scrapings von Werken zum Training von KI vor, dass den Urhebern für die anfallenden Vervielfältigungen ihrer Werke eine angemessene Vergütung gezahlt werden sollte.⁶⁴⁰ Die Schrankenbestimmung für Text und Data Mining (§ 44b UrhG), die grundsätzlich auf Vervielfältigungen von Werken zum Training von KI angewendet werden kann, gewährt den betroffenen Urhebern eine solche Vergütung allerdings nicht. Gleiches gilt gemäß § 60h Abs. 2 Nr. 3 UrhG für die spezielle Schrankenbestimmung des Text und Data Minings zum Zwecke wissenschaftlicher Forschung (§ 60d UrhG).

V. Zwischenfazit zum Web Scraping von Werken als Trainingsdaten für KI

Insgesamt hält das Urheberrecht mit der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) eine Schrankenbestimmung bereit, von deren Freistellungswirkung das Web Scraping urheberrechtlich geschützter Werke als Trainingsdaten umfasst ist. Mit den im zweiten Teil der vorliegenden Untersuchung identifizierten Freistellungserfordernissen übereinstimmend ist § 44b UrhG auch zugunsten kommerziell tätiger Trainingsdatensammler anwendbar. Allerdings führen die Löschungspflicht der

640 Dazu ausführlich siehe oben unter 2. Teil B. II. 1.

Trainingsdatensammler (§ 44b Abs. 2 S. 2 UrhG), der selbst durch eine Vorbehaltserklärung in natürlicher Sprache auslösbare Opt-out-Mechanismus (§ 44b Abs. 3 UrhG) und die Vergütungsfreiheit der Vervielfältigungshandlungen dazu, dass die Schrankenbestimmung die identifizierten Regelungsbedürfnisse für ein interessengerechtes urheberrechtliches Regelungsregime dennoch nicht vollständig erfüllt. De lege ferenda sind deswegen Anpassungen an der Schrankenbestimmung für Text und Data Mining insbesondere direkt in § 44b UrhG notwendig.

B. Durchführung des Trainings auf Grundlage urheberrechtlich geschützter Trainingsdaten

Nach der Speicherung der für das Training bestimmten Werke mit Hilfe von Webcrawlern wird auf Basis der erfassten Werke das KI-Training durchgeführt. Für die Durchführung des gesamten Trainingsprozess konnte dabei ein regulatorisches Frestellungsinteresse identifiziert werden. Die Freistellung sollte im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte erfolgen.

Damit das geltende Urheberrecht das im zweiten Teil der Untersuchung identifizierte Frestellungsbedürfnis erfüllt, ist es erforderlich, dass alle Teilaspekte des Trainingsvorgangs⁶⁴¹ durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte aus den §§ 15 ff. UrhG vom Ausschließlichkeitsrecht ausgenommen sind.

I. Vorverarbeitung der Trainingsdaten (Normalisierung und Feature Encoding)

Unter dem Ausdruck der Vorverarbeitung von Trainingsdaten werden mehrere Verarbeitungsschritte zusammengefasst. Die erfassten Trainingsdaten müssen zunächst normalisiert werden. Als letztes wird das Feature Encoding der Inhalte durchgeführt.⁶⁴² Erst dann eignen sich die urheberrecht-

641 Dazu siehe unter 2. Teil C. am Anfang.

642 Zur Bedeutung beider technischen Begriffe siehe oben unter 1. Teil B. II. 2.

lich geschützten Trainingsdaten als Berechnungsgrundlage für den Algorithmus.

1. Normalisierung der Trainingsdaten

Da bei der Normalisierung digitale Kopien der als Trainingsdaten erfassten Werke angefertigt werden, ist ein Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG denkbar. Teil des Normalisierungsverfahrens, also der Transponierung der Daten auf eine vergleichbare Skala, ist darüber hinaus regelmäßig eine Veränderung der geschützten Inhalte.⁶⁴³ Berücksichtigt werden muss damit außerdem, inwieweit solche Veränderungen Einfluss auf den möglichen Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG und eine mögliche urheberrechtliche Freistellung der Normalisierungshandlungen haben.

a) Normalisierung als Vervielfältigung im Sinne des § 16 UrhG: Anwendung eines normativierten Vervielfältigungsbegriffs

Ob bei der Normalisierung von Werken für das Training von KI Vervielfältigungsstücke im urheberrechtlichen Sinne angefertigt werden, hängt davon ab, was unter einem Vervielfältigungsstück im Sinne des § 16 UrhG zu verstehen ist. Denn auch wenn der Vervielfältigung im urheberrechtlichen Sinne ein tatsächlicher Speichervorgang zugrunde liegt, handelt es dennoch um einen normativen Rechtsbegriff, der infolgedessen auch der Auslegung zugänglich ist.⁶⁴⁴ Daraus folgt, dass nicht jede tatsächliche Speicherung auch eine Vervielfältigung im Rechtssinne sein muss.⁶⁴⁵ Eine normative Aufladung des Vervielfältigungsbegriffs ergibt sich auch aus der Rechtspre-

643 Zu möglichen Veränderungsschritten im Detail unten unter 3. Teil B. I. 1. a) bb) (2).

644 *Hugenholtz*, EIPR 2000, 482 (485); sich hieran anschließend *Schönberger*, ZGE 2018, 35 (53); a. A. wohl *Heerma*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; *Chiou*, JIPITEC 2019, 398 Rn. 6; außerdem *Jager*, Artificial Creativity?, S. 332.

645 So auch *Hugenholtz*, EIPR 2000, 482 (485); sich hieran anschließend *Schönberger*, ZGE 2018, 35 (53); de lege ferenda auch *Becker*, ZGE 2016, 239 (267); *Hofmann*, ZGE 2016, 482 (502); a. A. BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 25 – Internet-Radiorecorder; *Sucker*, Der digitale Werkgenuss im Urheberrecht, S. 89; außerdem wohl *Heerma*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; *Chiou*, JIPITEC 2019, 398 Rn. 6; außerdem *Jager*, Artificial Creativity?,

chung des EuGH, nach der das Kopieren eines Schutzgegenstands nicht in das Vervielfältigungsrecht des Rechtsinhabers eingreift, wenn der Inhalt nach der tatsächlichen Speicherung so verändert wird, dass der Schutzgegenstand nicht mehr wiedererkennbar ist.⁶⁴⁶

aa) Der Begriff des Vervielfältigungsstücks nach dem UrhG

Was unter einem Vervielfältigungsstück zu verstehen ist, ist im UrhG nicht ausdrücklich definiert. Dennoch hat sich ein allgemeines Begriffsverständnis herausgebildet. Unter einer Vervielfältigung im Sinne des UrhG wird dabei jede in beliebiger Weise vorgenommene körperliche Festlegung eines Werkes verstanden, die unmittelbar oder mittelbar zur menschlichen Wahrnehmung der geistigen Schöpfung geeignet ist.⁶⁴⁷ Dabei wird ausschließlich eine technische Perspektive eingenommen.⁶⁴⁸ Bei der Normalisierung urheberrechtlich geschützter Trainingsdaten zur Entwicklung von KI-Systemen werden auf eine einheitliche Skala transformierte, digitale Kopien von Werken, also technisch gesehen Repräsentationen der Werke angefertigt, sodass eine tatbestandsmäßige Festlegung der Werke erfolgt. Sie ist jedenfalls mittelbar auch zur menschlichen Wahrnehmung der geistigen Schöpfung geeignet. Nach dem tradierten, national geprägten Begriffsverständnis würden bei der Normalisierung von zum Training erfassten Werken daher Vervielfältigungsstücke im Sinne des § 16 UrhG hergestellt.

bb) Funktionsbezogene Auslegung des Vervielfältigungsbegriffs

Das Vervielfältigungsrecht ist durch Art. 2 InfoSoc-RL allerdings unionsrechtlich harmonisiert. Welche Bedeutung der Begriff des Vervielfältigungsstücks (§ 16 Abs. 1 UrhG) beziehungsweise der Vervielfältigung (Art. 2 Info-

S. 333, die allesamt ein streng technisches Verständnis des Vervielfältigungsbegriffs zugrunde legen.

646 EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 31 – Pelham.

647 BGH, Urt. v. 04.10.1990 - I ZR 139/89, GRUR 1991, 449 (453) – Betriebssystem; BGH, Urt. v. 06.10.2016, GRUR 2017, 266 Rn. 37 – World of Warcraft I; BT-Drs. IV/270, S. 47.

648 BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 25 – Internet-Radiorecorder; Heerma, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; Sucker, Der digitale Werkgenuss im Urheberrecht, S. 89; Chiou, JIPITEC 2019, 398 Rn. 6; Jager, Artificial Creativity?, S. 333.

Soc-RL) hat, muss sich daher nach einer richtlinienkonformen Auslegung des Vervielfältigungsbegriffs richten. Das betrifft auch die Frage, ob Werkfestlegungen zum Zwecke rein technischer Verwendung, wie sie bei der Normalisierung urheberrechtlich geschützter Trainingsdaten angefertigt werden, in den Anwendungsbereich des Vervielfältigungsrechts fallen.

Auch in der InfoSoc-RL wird der Vervielfältigungsbegriff nicht ausdrücklich definiert. Das führt dazu, dass eine richtlinienkonforme Auslegung daher insbesondere anhand des Sinn und Zwecks des Vervielfältigungsrechts aus Art. 2 InfoSoc-RL erfolgen muss.⁶⁴⁹ Diese teleologische Perspektive kann unter anderem auch zu einer funktionsbezogenen, einschränkenden Auslegung des Vervielfältigungsrechts führen.⁶⁵⁰ Die dem Sachverhalt zugrundeliegende technische Ausgangslage ist hierfür nicht allein maßgeblich. Die Sozialbedeutung der Nutzungshandlung kann in diesem Zusammenhang jedoch nur dann zur Konkretisierung des Vervielfältigungsbegriffs genutzt werden, insofern sie sich mit Hilfe einer der Auslegungsmethoden an den Tatbestand aus Art. 2 InfoSoc-RL anknüpfen lässt.⁶⁵¹

649 v. *Ungern-Sternberg*, GRUR 2012, 1198 (1200); v. *Ungern-Sternberg*, GRUR 2015, 533 (534); insbesondere im Kontext des urheberrechtlichen Vervielfältigungsrechts nun auch an erster Stelle betont in EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 26 – *Ocilion*; außerdem bereits in EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 28 – *Pelham*, wobei die hier zum Tonträgerherstellerrecht getätigten Ausführungen ebenfalls auf das in Art. 2 InfoSoc-RL harmonisierte Urheberrecht übertragen werden können. In diesem Sinne auch *Grisse/Kaiser*, ZUM 2021, 401 (404); *Haberstumpf*, ZUM 2020, 810 (817). Eine teleologisch geprägte Auslegung des Vervielfältigungsrechts deutete sich bereits in der EuGH-Entscheidung *Allposters* an. Entscheidend ist für den EuGH hier, dass die Übertragung eines Kunstwerkes von einem Poster auf eine Leinwand die Reproduktion näher an das Original heranrückt, siehe EuGH, Urt. v. 22.01.2015 - C-419/13, GRUR 2015, 256 Rn. 42 – *Allposters*. Damit geht, was für den EuGH entscheidend gewesen sein muss, eine Vergrößerung der vom Werkstück ausgehenden Substitutionsgefahr, also eine Intensivierung der Beeinträchtigung der Verwertungsinteressen betroffener Urheber, einher.

650 v. *Ungern-Sternberg*, GRUR 2012, 1198 (1204 f.); im Anschluss hieran auch *Antoine*, ZGE 2023, 387 (399); im Kontext digitalen Werkgenusses ähnlich auch *Raue*, ZGE 2017, 514 (530); vgl. zu entsprechenden Überlegungen de lege ferenda *Becker*, ZGE 2016, 239 (267); *Hofmann*, ZGE 2016, 482 (502); ähnlich außerdem *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260), die die Abkehr vom technisch-zentrierten zu einem auf die Sozialbedeutung bezogenen Verständnis des Vervielfältigungsrechts fordern.

651 A. A. wohl *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260), die losgelöst von den anerkannten Auslegungsmethoden eine Konkretisierung anhand der Sozialbedeutung der Nutzungshandlung vornehmen wollen.

(1) Herleitung einer Schutzbereichsbegrenzung aus der verwertungsrechtlichen Funktion des Vervielfältigungsrechts

Grundsätzlich strebt die InfoSoc-RL an, ein hohes Schutzniveau für Urheber zu gewährleisten.⁶⁵² Allerdings ist Sinn und Zweck der urheberrechtlichen Verwertungsrechte und damit des Vervielfältigungsrechts im Speziellen, dem Urheber die Verwertung seines Werkes, also die Monetarisierung dessen Nutzung, zu sichern.⁶⁵³ Daraus folgt, dass nur solche Werkverwendungen auch erlaubnispflichtige Werknutzungen und damit Vervielfältigungen im urheberrechtlichen Sinne sein können, die die Verwertungsmöglichkeit der Urheber zumindest potenziell beeinträchtigen.⁶⁵⁴ Hierfür genügt, dass aus der Werkverwendung ein unmittelbares oder mittelbares Vermarktungspotenzial entsteht und damit eine Substitutionsgefahr für die Verwertung des Werkes durch den Urheber hervorgerufen wird.⁶⁵⁵ Eine solche kann grundsätzlich auch durch veränderte Erzeugnisse begründet werden, sodass auch diese im Allgemeinen unter das Vervielfältigungsrecht fallen.⁶⁵⁶ Geht von der Werkverwendung hingegen keine Substitutionsge-

652 ErwGr. 4, 9 InfoSoc-RL.

653 Nordemann, in: Fromm/Nordemann, Urheberrecht, Einl. zum UrhG Rn. 19; v. Ungern-Sternberg, GRUR 2012, 576 (579 f.); v. Ungern-Sternberg, GRUR 2012, 1198 (1200); Haberstumpf, ZUM 2020, 810 (816); Grisse/Kaiser, ZUM 2021, 401 (404). Außerdem ErwGr. 10 InfoSoc-RL.

654 öst. OGH, Beschl. v. 23.02.2010 - 4 Ob 208/09 f, ZUM 2010, 629 (630); v. Ungern-Sternberg, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23; v. Ungern-Sternberg, GRUR 2012, 576 (580); Schönberger, ZGE 2018, 35 (53); Antoine, Veränderte Werknutzungen, S. 179 f.; in diese Richtung auch Grisse/Kaiser, ZUM 2021, 401 (404); grundsätzlich wie hier auch Antoine, ZGE 2023, 387 (398), wobei sie eine Einschränkung der funktionsbezogenen Auslegung auf die „Nutzungsfälle an den Rändern des urheberrechtlichen Ausschließlichkeitsrechts“ befürwortet, dazu sogleich. Zum Verhältnis zu den urheberpersönlichkeitsrechtlichen Bestimmungen siehe Antoine, ZGE 2023, 387 (401 ff.).

655 v. Ungern-Sternberg, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23. Die Intensivierung der Substitutionsgefahr durch eine weitere Annäherung des Werkstücks an das Original scheint auch für den EuGH in der *Allposters*-Entscheidung maßgebend dafür gewesen zu sein, beim Medientransfer eines Bildes von Plakat auf Leinwand eine neue Vervielfältigung anzunehmen, vgl. EuGH, Urt. v. 22.01.2015 - C-419/13, GRUR 2015, 256 Rn. 42 – *Allposters*.

656 Hiervon ging der EuGH implizit bereits im Jahr 2012 aus, als er die Herstellung veränderter Erzeugnisse beiläufig unter das Vervielfältigungsrecht aus Art. 2 InfoSoc-RL fasste und auf Grundlage dessen ein Eingreifen der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus Art. 5 Abs. 1 InfoSoc-RL untersuchte, vgl. EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 53 – *Infopaq II*.

fahr für die Verwertung des Werkes durch den Urheber hervor, besteht auch kein Grund für eine Beteiligung des Urhebers an von Dritten erwirtschafteten Einnahmen. Anderes folgt auch nicht aus dem Grundsatz einer möglichst umfassenden Urheberbeteiligung. Denn eine Beteiligung der Urheber ohne Beeinträchtigung ihrer Verwertungsinteressen stellt keinen angemessenen Ausgleich, sondern eine ungerechtfertigte Überkompensation dar. Sie ist nicht durch das Urheberrecht bezweckt.⁶⁵⁷

Keine Substitutionsgefahr wird durch beiläufige oder zufällige, also nicht zielgerichtete und insgesamt untergeordnete Werkverwendungen begründet.⁶⁵⁸ Andererseits genügt aber, dass die Kopie des Werkes eine Werkrezeption ermöglichen soll. Das ist beispielsweise im Arbeitsspeicher von Computern der Fall. Eine Substitutionsgefahr für die Verwertung des genutzten Werks durch den Urheber wird im Kontext der Anfertigung veränderter Werkkopien weiterhin dann nicht ausgelöst, wenn die eigenschöpferischen Züge eines Werkes nach der Bearbeitung in einer Werkkopie nicht mehr wiedererkennbar sind.⁶⁵⁹ Dass die Werkverwendung von

657 Anders Jager, *Artificial Creativity?*, S. 333, der sein rein-technisches Verständnis gerade auf den Grundsatz der möglichst umfassenden Urheberbeteiligung stützt.

658 v. Ungern-Sternberg, GRUR 2012, 576 (580).

659 Vgl. zum Tonträgerherstellerrecht EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 38 - Pelham. Da auch der Harmonisierung des originären Urheberrechts ein verwertungsschützender Charakter zugutekommt und sowohl das Tonträgerherstellerrecht als auch das Urheberrecht gemeinsamen auf Art. 2 InfoSoc-RL beruhen, ist eine Übertragung des Wiedererkennbarkeitskriteriums aus der Rechtsprechung des EuGH auf das Urheberrecht möglich. So auch Nordemann, in: Fromm/Nordemann, *Urheberrecht*, UrhG § 23 Rn. 41; Lauber-Rönsberg/Ahlberg, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar *Urheberrecht*, UrhG § 23 Rn. 35; Loewenheim/Leistner, in: Loewenheim, *Handbuch des Urheberrechts*, § 8 Schutzzumfang Rn. 13; Nordemann/Czychowski, in: Hasselblatt, *Münchener Anwalts-Handbuch Gewerblicher Rechtsschutz*, § 40 *Urheberrecht und verwandte Schutzrechte* Rn. 560; Paul, in: Borges/Hilber, Beck'scher Online-Kommentar *IT-Recht*, UrhG § 23 Rn. 6; Raue, in: Dreier/Schulze, *Urheberrechtsgesetz*, UrhG § 23 Rn. 30; BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 49 - Porsche 911; BGH, Urt. v. 15.12.2022 - I ZR 173/21, GRUR 2023, 571 Rn. 29 - Vitrienenleuchte; Grisse/Kaiser, ZUM 2021, 401 (404); Grünberger, ZUM 2023, 309 (318); v. Ungern-Sternberg, GRUR 2022, 1777 (1779); Haberstumpf, ZUM 2020, 810 (416); Lauber-Rönsberg, ZUM 2020, 733 (736); Wagner, MMR 2020, 728 (730); kritisch allerdings Ohly, ZUM 2021, 745 (746); Schack, ZGE 2023, 263 (272 ff.); Stieper, GRUR 2020, 699 (706); ablehnend allerdings Antoine, *Verändernde Werknutzungen*, S. 181 ff. Antoine, die ihre Skepsis vor allem auf dem investitionsschützenden Charakter des Tonträgerherstellerrecht begründet, berücksichtigt dabei allerdings nicht, dass auch dem harmonisierten Urheberrecht ein Amortisationsgedanke zugrunde liegt (ErwGr. 10 InfoSoc-RL, er adressiert sowohl das Urheberrecht als auch

der Kunstfreiheit umfasst ist, ist vor dem Hintergrund der teleologischen Auslegung demnach nicht notwendig.⁶⁶⁰ Darüber hinaus betreffen die urheberrechtlichen Schrankenbestimmungen Sachverhaltskonstellationen, in denen die Verwertungsinteressen der Urheber jedenfalls potenziell beeinträchtigt sind, also zumindest eine mittelbare Verwertungsmöglichkeit für Urheber besteht.⁶⁶¹ Daher besteht auch kein Konflikt zwischen der funktionsbezogenen Auslegung des Vervielfältigungsrechts, die sich danach richtet, ob das Verwertungsinteresse eines Urhebers *überhaupt* beeinträchtigt ist, und dem Ausgleich gegenläufiger Interessen, den das Urheberrecht mit Hilfe der Schrankenbestimmungen vornimmt.⁶⁶² Die funktionsbezogene Auslegung auf den Randbereich des Urheberrechts zu beschränken, ist daher nicht notwendig.⁶⁶³

die verwandten, in der InfoSoc-RL geregelten Schutzrechte). Die für den EuGH maßgebliche Erwägung, dass Rechtsinhaber in die Lage versetzt werden müssen, ihre in die Schaffung des Schutzgegenstands fließenden ökonomischen Ressourcen durch die Verwertung des Werkes ausgleichen zu können, um auch das Schaffen aus einer wirtschaftlichen Perspektive zu ermöglichen, greift deswegen ebenso für den originären urheberrechtlichen Schutz. Vor diesem Hintergrund ist eine Anwendung des Wiedererkennbarkeitskriteriums auf das Urheberrecht sogar angezeigt. Eine Umsetzung der Wiedererkennbarkeitsschwelle im Vervielfältigungsrecht ist dabei dogmatisch konsequenter als eine Berücksichtigung in § 23 UrhG, in diese Richtung auch *Grünberger*, ZUM 2023, 309 (320 ff.). Etwas widersprüchlich hier der BGH, der die Wiedererkennbarkeitsformel zwar i. R. d. § 23 UrhG prüft (BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 47 ff. – Porsche 911.), letztlich aber zutreffend konstatiert, dass es bei mangelnder Wiedererkennbarkeit bereits an einer Vervielfältigungshandlung im Sinne des § 16 UrhG fehlt, sodass es auf das Vorliegen einer freien Benutzung nach § 24 UrhG a. F. schon nicht mehr ankommt, siehe BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 63 – Porsche 911.

660 *Grisse/Kaiser*, ZUM 2021, 401 (414); *Haberstumpf*, ZUM 2020, 810 (817).

661 Das gilt selbst für die Schrankenbestimmung zur Nutzung von Werken als unwesentliches Beiwerk (§ 57 UrhG). Denn auch sie geht auf eine unionsrechtliche Grundlage zurück, siehe *Vogel*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 57 Rn. 5. Damit kann auch diese Schrankenbestimmung nur dann eingreifen, wenn im Ergebnis der dargelegten funktionsbezogenen Auslegung überhaupt eine Vervielfältigung angefertigt worden ist. Nur dann liegt eine urheberrechtsrelevante und erlaubnisbedürftige Werknutzung vor. Das setzt jedenfalls eine potenzielle Beeinträchtigung der urheberrechtlichen Verwertungsinteressen voraussetzt. Dieses kann beispielsweise auch bei der hintergründigen Nutzung eines Werkes gegeben sein. Es ist daher von der Frage zu trennen, ob die Werknutzung für den Nutzer unwesentlich im Sinne der Bestimmung ist.

662 Anders *Antoine*, ZGE 2023, 387 (400), die eine Kollision mit den urheberrechtlichen Schrankenbestimmungen annimmt und den in den Schrankenbestimmungen verhandelten Fällen des Interessenausgleichs eine Begrenzung für die funktionsbezogene Auslegung des Vervielfältigungsrechts entnimmt.

(2) Wiedererkennbarkeit eigenschöpferischer Züge nach der Normalisierung

Bei der Normalisierung von urheberrechtlich geschützten Trainingsdaten, die mit Veränderungen an den Werken einhergeht, ist insbesondere klärungsbedürftig, ob die eigenschöpferischen Züge eines Werkes nach der Herstellung normalisierter Werkkopien noch wiedererkennbar sind. Das hängt davon ab, welche konkreten Verarbeitungsschritte Teil des Normalisierungsprozesses sind. Dies ist wiederum von der Art der verwendeten Inhalte sowie der Art des zu trainierenden KI-Systems abhängig.

Praktisch verbreitet ist beispielsweise eine Expansion von Textinhalten. Diese umfasst unter anderem eine Auflösung von Abkürzungen.⁶⁶⁴ Der Text bleibt dabei aber noch im Wesentlichen erkennbar. Weitere mögliche Normalisierungsschritte sind unter anderem eine Entfernung der Interpunktion oder eine Reduktion der Wörter auf ihren Wortstamm.⁶⁶⁵ Am Ende kommt es regelmäßig zu einer Tokenisierung der Inhalte, bei der der Text in einzelne Token aufgeteilt wird.⁶⁶⁶ Sie können aus ganzen Sätzen, Wörtern oder Silben bestehen. Ob die schöpferischen Züge eines Sprachwerks in normalisierter Form noch im urheberrechtlichen Sinne wiedererkennbar sind, hängt vor diesem Hintergrund von den im Einzelfall angewendeten Normalisierungsschritten sowie den schutzbegründenden Elementen des jeweiligen Werks ab. Für die Zwecke dieser Untersuchung ist daher davon auszugehen, dass die schutzbegründenden Elemente eines Schriftwerks auch nach seiner Normalisierung noch wiedererkennbar sind.

Auch für die Normalisierung von visuellen Werken kann auf unterschiedliche Normalisierungsverfahren zurückgegriffen werden. Teilweise bleiben dabei wesentliche schöpferische Merkmale der Inhalte wie Motiv, Bildausschnitt, Perspektive, Bildschärfe oder Kontrastgebung für einen Betrachter identifizierbar.⁶⁶⁷ Im Detail kommt es auch hierbei auf die im Einzelfall angewendeten Normalisierungsverfahren und die schutzbegrün-

663 So aber *Antoine*, ZGE 2023, 387 (400); *Antoine*, Verändernde Werknutzungen, S. 176 f.

664 *Yse*, Text Normalization for Natural Language Processing (NLP).

665 *Yse*, Text Normalization for Natural Language Processing (NLP).

666 Mit Beispielen *Menzli*, Tokenization in NLP; *Yse*, Text Normalization for Natural Language Processing (NLP).

667 Zu den schöpferischen Elementen eines Lichtbildwerkes siehe *Loewenheim/Leistner*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 2 Rn. 213. Zur zwei beispielhaften Normalisierungsreihen von Bildinhalten siehe *NVIDIA Corporation*, Normalize Operator — NVIDIA DALI 1.29.0 documentation.

denden Elemente der normalisierten Werke an. Daher ist betreffend der Normalisierung von visuellen Werken für die Zwecke dieser Untersuchung ebenfalls davon auszugehen, dass schutzbegründende Elemente der Werke im urheberrechtlichen Sinne wiedererkennbar bleiben.

An dieser Stelle kann demzufolge dahinstehen, wo genau die Wiedererkennbarkeitsgrenze des Schutzbereichs zu ziehen ist und wie sich dies auf einzelne Normalisierungsverfahren auswirkt. Vielmehr führen Veränderungen bei der Normalisierung von Werken nicht grundsätzlich zur mangelnden Wiedererkennbarkeit eigenschöpferischer Elemente in den urheberrechtlich geschützten Trainingsdaten. Im Einzelfall kann das aber durchaus anders zu beurteilen sein.

(3) Weitere funktionsorientierte Erwägungen zur Normalisierung von Werken

Weiterhin können die generierten Trainingsdaten auch in normalisierter Form noch auf dem Trainingsdatenmarkt verwertet werden. Damit kommt den Werkrepräsentationen auch nach der Normalisierung noch eine wirtschaftliche Bedeutung zu. Infolgedessen beeinträchtigt die Herstellung normalisierter Kopien von Werken auch aus dieser Perspektive die Verwertungsinteressen der Urheber.

cc) Schlussfolgerungen

Alles in allem können auch normalisierte Kopien von Werken Vervielfältigungsstücke im urheberrechtlichen Sinne sein.⁶⁶⁸ Solange die schutzbegründenden Elemente im normalisierten Erzeugnis noch wiedererkennbar sind, hat hierauf grundsätzlich auch die Veränderung der Werke bei der Normalisierung keinen Einfluss. Die Normalisierung urheberrechtlich geschützter Trainingsdaten greift damit – je nach angewendetem Normalisierungsverfahren – in das Vervielfältigungsrecht betroffener Urheber im Sinne des § 16 UrhG ein.⁶⁶⁹ Vervielfältigungen, die im Zuge der Normalisierung möglicherweise hergestellt werden, sind demzufolge erlaubnis-

668 Wie hier auch Käde, *Kreative Maschinen und Urheberrecht*, S. 70; Pesch/Böhme, GRUR 2023, 997 (1004); wohl außerdem Konertz/Schönhof, WRP 2024, 289 Rn. 6.

669 Weitergehend Dornis/Stober, *Urheberrecht und Training generativer KI-Modelle*, S. 70; Dregelies, GRUR 2024, 1484 (1484), die pauschal eine Vervielfältigung annehmen.

pflichtig, wenn nicht eine urheberrechtliche Schrankenbestimmung die Werknutzung gesetzlich gestattet. Bereits hieraus folgt, dass das UrhG das Freistellungsbedürfnis für die als Teil des Trainingsprozesses durchgeführte Normalisierung der urheberrechtlich geschützten Trainingsdaten nicht erfüllt. Denn nach der durchgeführten regulatorischen Bedarfsanalyse ist eine Freistellung des gesamten Trainingsprozesses im Wege einer unmittelbar begrenzenden Ausgestaltung der Verwertungsrechte interessengerecht.

Nichtsdestotrotz könnte die Freistellung von im Rahmen der Normalisierung vorgenommenen Vervielfältigungshandlungen durch die Anwendung einer urheberrechtlichen Schrankenbestimmung erreicht werden. Damit ließe sich zumindest das Ob des regulatorischen Freistellungsbedürfnisses erfüllen. Ob eine Schrankenbestimmung die Normalisierung der Trainingsdaten freistellt, hängt aber auch davon ab, ob sie die Herstellung veränderter Vervielfältigungsstücke umfasst.

b) Anwendung urheberrechtlicher Schrankenbestimmungen zur Erlaubnis der Vervielfältigungshandlungen

Erforderlich ist also, dass die im Rahmen der Normalisierung von urheberrechtlich geschützten Trainingsdaten möglicherweise vorgenommenen Vervielfältigungshandlungen durch die Anwendung einer urheberrechtlichen Schrankenbestimmung erlaubnisfrei gestellt sind.

aa) Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)

Vervielfältigungen von Werken im Rahmen der Normalisierung urheberrechtlich geschützter Trainingsdaten könnten von der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG gedeckt sein. Wie auch bei der Speicherung der Trainingsdaten mit Hilfe von Webcrawlern scheidet die Anwendung von § 44a UrhG allerdings zum einen an der eigenen wirtschaftlichen Bedeutung der Vervielfältigungen, die neben dem Einsatz für das Training einer konkreten KI auch im normalisierten Zustand auf dem Trainingsdatenmarkt veräußert werden können. Darüber hinaus ist auch davon auszugehen, dass dauerhafte Ver-

vielfältigungen der Inhalte stattfinden.⁶⁷⁰ Eine Freistellung von im Rahmen der Normalisierung stattfindenden Vervielfältigungshandlungen über die Schrankenbestimmung aus § 44a UrhG findet demzufolge nicht statt.

bb) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)

Die Vervielfältigungen von Werken, die gegebenenfalls im Rahmen der Normalisierung urheberrechtlich geschützter Trainingsdaten angefertigt werden, könnten wie die Speicherung der Trainingsdaten jedoch von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 UrhG freigestellt sein. Denn das KI-Training auf Basis von Werken fällt unter den Begriff des Text und Data Mining im Sinne des § 44b Abs. 1 UrhG.⁶⁷¹

(1) Grundsätzliche Anwendbarkeit von § 44b UrhG

Von der Schrankenbestimmung aus § 44b Abs. 2 UrhG gedeckt sind alle Vervielfältigungshandlungen, die *für* das KI-Training durchgeführt werden, hierfür also notwendig sind. Da das KI-Training auf Basis normalisierter Daten stattfindet, ist der vorgelagerte Normalisierungsschritt für das Training von KI notwendig. Infolgedessen werden die bei der Normalisierung urheberrechtlich geschützter Trainingsdaten gegebenenfalls stattfindenden Vervielfältigungshandlungen ebenfalls grundsätzlich vom Schrankentatbestand aus § 44b Abs. 2 UrhG umfasst.⁶⁷²

(2) Auswirkungen von Veränderungen der Werke auf die Anwendbarkeit der Schrankenbestimmung

Fraglich ist allerdings, ob die Schrankenbestimmung aus § 44b Abs. 2 UrhG auch die Herstellung *veränderter* Vervielfältigungsstücke der Werke

670 Zu dieser Voraussetzung ausführlich oben unter 3. Teil A. III. 1.

671 Dazu bereits ausführlich siehe oben unter 3. Teil A. III. 2. i) ee) (3) aa) bis dd); wie hier auch *Jager*, *Artificial Creativity?*, S. 351, 157 ff.

672 So für im Kontext des allgemeinen Text und Data Mining auch ErwGr. 8 S. 6 DSM-RL; zudem auch *Bullinger*, in: *Wandtke/Bullinger*, *Urheberrecht*, UrhG § 60d Rn. 11; *Bomhard*, *InTeR* 2023, 174 (176); *Kögel*, *InTeR* 2023, 179 (181); *Raue*, *ZUM* 2021, 793 (795, 797); für die alte deutsche Schrankenbestimmung für Text und Data Mining, § 60d UrhG a. F., außerdem *Geiger/Frosio/Bulayenko*, *IIC* 2018, 814 (831). Ebenso in ErwGr. 8 S. 6 DSM-RL.

freistellt, wie sie im Rahmen des Normalisierungsverfahrens angefertigt werden.

Grundsätzlich bestimmt § 62 Abs. 1 S. 1 UrhG, dass die aus den Schrankenbestimmungen folgenden Nutzungsberechtigungen nicht zu einer Erlaubnis von Änderungen an den betreffenden Werken führen. Allerdings sind Änderungen abweichend hiervon gemäß § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG zulässig, wenn der Urheber seine Einwilligung nach Treu und Glauben nicht versagen kann. Dieser unbestimmte Rechtsbegriff führt über den Rechtsgrundverweis aus § 62 Abs. 1 S. 2 UrhG auch im Anwendungsbereich des Änderungsverbots zu einer umfassenden Abwägung von Urheber- und Nutzerinteressen.⁶⁷³ Verändernde Werknutzungen sind danach zulässig, wenn sie zur Erfüllung des jeweiligen Schrankenzwecks erforderlich und zumutbar sind.⁶⁷⁴ Darüber hinaus müssen verändernde Werknutzungen über §§ 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG im Wege richtlinienkonformer Auslegung des Tatbestandsmerkmals der Erlaubnis nach „Treu und Glauben“ aber auch dann zulässig sein, wenn die mit einer Schrankenregelung umgesetzte Richtlinienbestimmung des Unionsurheberrechts eine Veränderung des Werks gestattet.

Dafür, dass Art. 4 DSM-RL, der mit § 44b UrhG umgesetzt worden ist,⁶⁷⁵ verändernde Vervielfältigungen von Werken zum Zwecke des Text und Data Minings freistellt, spricht zum einen, dass das Umarbeitungsrecht an Computerprogrammen (Art. 4 Abs. 1 lit. b Computerprogram-RL) vom Schrankentatbestand mitumfasst ist. Nach dieser Regelungslogik müssten auch verändernde Vervielfältigungen sonstiger Werke von der Schrankenbestimmung freigestellt sein. Vor allem zielt die DSM-RL mit Art. 4 weiterhin ausdrücklich darauf ab, unter anderem Vervielfältigungen von Werken freizustellen, die während der Normalisierung der gesammelten Daten für das Text und Data Mining angefertigt werden.⁶⁷⁶ Wie dargelegt wurde, führt die Normalisierung in vielen Fällen aber zu einer Veränderung der Werke. Demzufolge entspricht es dem niedergelegten Zweck des Art. 4

673 Peukert, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 62 Rn. 9; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 62 Rn. 6; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 62 Rn. 10; a. A. Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 62 Rn. 7.

674 OLG Hamburg, Urt. v. 05.06.1969 - 3 U 21/69, GRUR 1970, 38 (39) – Heintje; Engels, in: Götting/Lauber-Rönsberg/Ahlberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 62 Rn. 10 f.; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 62 Rn. 6, 12; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 62 Rn. 10.

675 BT-Drs. 19/27426, S. 87.

676 ErwGr. 8 DSM-RL.

DSM-RL, auch Veränderungen der Werke jedenfalls im Rahmen des Normalisierungsverfahrens urheberrechtlich freizustellen. Insgesamt ist deswegen davon auszugehen, dass die Schrankenbestimmung aus Art. 4 DSM-RL auch die Herstellung veränderter Vervielfältigungen jedenfalls für die Normalisierung von Werken umfasst.⁶⁷⁷ Dies ist richtlinienkonform über das Tatbestandsmerkmal der Erlaubnis nach „Treu und Glauben“ in § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG auch bei der Bestimmung der Reichweite des Änderungsverbots zu berücksichtigen. Änderungen an Werken, die im Rahmen des Normalisierungsverfahrens stattfinden und zur Herstellung veränderter Vervielfältigungsstücke führen, sind demzufolge nicht vom Änderungsverbot aus § 62 Abs. 1 S. 1 UrhG betroffen. Sie werden vielmehr durch § 44b Abs. 2 UrhG freigestellt.

c) Zwischenergebnis zur Normalisierung urheberrechtlich geschützter Trainingsdaten

Alles in allem können bei der Normalisierung urheberrechtlich geschützter Trainingsdaten, die regelmäßig vor dem eigentlichen Trainingsvorgang durchgeführt wird, Vervielfältigungshandlungen gemäß § 16 UrhG stattfinden. Diese Vervielfältigungen sind jedoch von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG umfasst. Das gilt auch für die Herstellung von im Rahmen des Normalisierungsverfahrens veränderten Vervielfältigungsstücken. Einer individuellen Erlaubnis durch die betroffenen Urheber bedarf es demzufolge jedenfalls dann nicht, wenn kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist.

Für den gesamten Trainingsprozess, zu dem auch der Normalisierungsschritt gehört, konnte aus einer regulatorischen Perspektive ein Freistellungsbedürfnis identifiziert werden. Danach sollte das Training von KI auf Basis von Werken durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte aus dem urheberrechtlichen Schutzbereich ausgeklammert werden. Vor diesem Hintergrund folgt aus dem möglichen Eingriff in das Vervielfältigungsrecht der von der Normalisierung betroffenen Urheber, dass das geltende Urheberrecht die im zweiten Teil der Untersuchung herausgearbeiteten Freistellungsbedürfnisse für die Durchführung des KI-Trainings nicht vollständig erfüllt. Aufgrund des Eingreifens der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG wird

677 A. A. Novelli u. a., Computer Law & Security Review 2024, S. 10.

das geltende Urheberrecht in Bezug auf die Normalisierungsprozesse aber jedenfalls dem Ob des regulatorischen Freistellungsbedürfnisses gerecht.

2. Feature Encoding

Im Anschluss an die Normalisierung wird das Feature Encoding der als Trainingsdaten gespeicherten Werke durchgeführt. Dabei werden die Werke in numerische Repräsentationen, sogenannte Tensoren, umgewandelt. Nur sie können im KNN, einem statistischen Modell, weiterverarbeitet werden.

Auch die Umwandlung der urheberrechtlich geschützten Trainingsdaten in Tensoren könnte einen Eingriff in das Vervielfältigungsrecht aus § 16 UrhG begründen. Wie bei der Normalisierung urheberrechtlich geschützter Trainingsdaten würden dies dazu führen, dass das im zweiten Teil der Untersuchung identifizierte Freistellungsbedürfnis vom geltenden Urheberrecht nicht verwirklicht wird.⁶⁷⁸

a) Umwandlung in Tensoren als Vervielfältigung im Sinne des § 16 UrhG

Fraglich ist daher zum einen, ob mit dem Feature Encoding eine weitere Vervielfältigungshandlung verbunden ist, die gemäß § 16 UrhG dem Urheber vorbehalten wird.

Unter einer Vervielfältigung im Sinne des UrhG wird im Allgemeinen jede in beliebiger Weise vorgenommene körperliche Festlegung eines Werkes verstanden, die unmittelbar oder mittelbar zur menschlichen Wahrnehmung der geistigen Schöpfung geeignet ist.⁶⁷⁹ Allerdings sind nach einer richtlinienkonformen Auslegung des Vervielfältigungsbegriffs nur solche Werkverwendungen auch erlaubnispflichtige Vervielfältigungen, die die Verwertungsmöglichkeit der Urheber zumindest potenziell beeinträchtigen.⁶⁸⁰ Das ist insbesondere dann nicht der Fall, wenn die eigenschöpfe-

678 Zum regulatorischen Freistellungsbedürfnis siehe 2. Teil C.

679 BGH, Urt. v. 04.10.1990 - I ZR 139/89, GRUR 1991, 449 (453) – Betriebssystem; BGH, Urt. v. 06.10.2016, GRUR 2017, 266 Rn. 37 – World of Warcraft I; BT-Drs. IV/270, S. 47.

680 öst. OGH, Beschl. v. 23.02.2010 - 4 Ob 208/09 f, ZUM 2010, 629 (630); v. Ungern-Sternberg, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23; v. Ungern-Sternberg, GRUR 2012, 576 (580); Schönberger, ZGE 2018, 35 (53); Antoine, Verändernde Werknutzungen, S. 179 f.; in diese Richtung auch Grisse/Kaiser, ZUM 2021, 401 (404); grundsätzlich wie hier auch Antoine, ZGE 2023, 387 (398), wobei sie

rischen Züge eines Werkes im umgewandelten Erzeugnis nicht wiedererkennbar sind.⁶⁸¹ Die Verwertungsmöglichkeiten der Urheber werden darüber hinaus auch dann nicht beeinträchtigt, wenn die Werkverwendungen beiläufig oder zufällig und insgesamt untergeordnet sind.⁶⁸²

aa) Wiedererkennbarkeitsgrenze

Beim Feature Encoding werden die erfassten Werke in einem numerischen Repräsentationsformat abgespeichert, sodass digitale Festlegungen der Werke entstehen. Fraglich ist aber, ob in diesen numerischen Festlegungen noch die eigenschöpferischen Züge der umgewandelten Werke wiedererkennbar sind. Dafür kommt es darauf an, wo genau die Wiedererkennbarkeitsgrenze verläuft, wie der Begriff also zu verstehen ist.

(1) Faktische oder normative Wiedererkennbarkeit

Zum einen ist möglich, die aus der Wiedererkennbarkeit folgende Schutzbereichsbegrenzung rein-faktisch zu verstehen. Dann wäre darauf abzustellen, ob übernommene Inhalte von Betrachtern rein tatsächlich noch als Werke der Urheber identifiziert werden können.⁶⁸³ Dabei ist stets auf die konkreten schutzbegründenden Merkmale eines Werkes und nicht auf andere Faktoren wie insbesondere Urhebernennungen, identifizierende Informationsgehalte oder andere nicht schutzfähige Elemente abzustellen.⁶⁸⁴ Ist

eine Einschränkung der funktionsbezogenen Auslegung auf die „Nutzungsfälle an den Rändern des urheberrechtlichen Ausschließlichkeitsrechts“ befürwortet, dazu sogleich. Zum Verhältnis zu den urheberpersönlichkeitsrechtlichen Bestimmungen siehe *Antoine*, ZGE 2023, 387 (401 ff.). Ausführlich dazu und zur funktionsbezogenen Auslegung des Vervielfältigungsrechts allgemein siehe oben unter 3. Teil B. I. 1. a) aa) und bb).

681 Vgl. EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 38 – Pelham; BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 49, 56 – Porsche 911.

682 v. *Ungern-Sternberg*, GRUR 2012, 576 (580).

683 In diesem Sinne LG Hamburg, Urt. v. 09.12.2020 - 308 O 431/17, ZUM-RD, 164 (174) – Hey, Pippi Langstrumpf; *Grisse/Kaiser*, ZUM 2021, 401 (413); *Lauber-Rönsberg*, ZUM 2020, 733 (736); wohl ebenfalls *Homar*, ZUM 2019, 731 (736); *Wagner*, MMR 2020, 728 (730); in Bezug auf die Wiedererkennbarkeitsgrenze auch *Antoine*, Verändernde Werknutzungen, S. 187 ff., wobei sie im Allgemeinen ebenfalls eine funktional-beschränkende Auslegung des Vervielfältigungsrechts fordert, siehe S. 193.

684 BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 49 – Porsche 911; deutlich auch *Grisse/Kaiser*, ZUM 2021, 401 (411).

ein Werk tatsächlich nicht identifizierbar, sind die übernommenen eigenschöpferischen Elemente also vollständig unkenntlich gemacht worden, scheidet eine Wiedererkennbarkeit im urheberrechtlichen Sinne aus.

Zum anderen könnte die Wiedererkennbarkeitsgrenze aber auch normativ verstanden werden. In der Folge bestünden verschiedene Möglichkeiten, die Schutzbereichsbegrenzung auszugestalten. Denkbar ist zum einen, über den Begriff der Wiedererkennbarkeit weitere teleologische Erwägungen zu berücksichtigen, die in einer Abwägung der im konkreten Fall widerstrebenden Grundrechte münden.⁶⁸⁵ Darüber hinaus könnte auch darauf abgestellt werden, ob dem Ursprungswerk in der veränderten Gestaltung noch eine wesentliche Bedeutung zukommt.⁶⁸⁶ Schließlich ließen sich auch alle schöpferischen Züge in einer Gesamtschau berücksichtigen. Dabei käme es darauf an, ob der Gesamteindruck der veränderten Gestaltung noch mit dem Ursprungswerk übereinstimmt.⁶⁸⁷

Unabhängig davon, ob es zu einem – jedenfalls denkbaren – normativen Verständnis des Wiedererkennbarkeitsbegriffs kommt, muss das natürliche Sprachverständnis des Begriffs aber Grundlage der Auslegung bleiben. Es bildet die Basis, aber auch die Grenze der Rechtsauslegung.⁶⁸⁸ Um in den Schutzbereich des Vervielfältigungsrechts zu fallen, muss das Werk im veränderten Erzeugnis daher stets tatsächlich identifizierbar sein. Hierfür spricht auch, dass aus einer funktionsbezogenen Perspektive auf das Vervielfältigungsrecht⁶⁸⁹ keine Substitutionsgefahr für die Verwertung des veränderten Werks durch den Urheber besteht, wenn es im Erzeugnis tatsächlich nicht identifizierbar ist. Die mit dem Vervielfältigungsrecht geschützten Verwertungsinteressen des Urhebers sind in diesem Fall daher nicht betroffen. Eine Normativierung des Wiedererkennbarkeitskriteriums kann demzufolge nur bei bestehender Identifizierbarkeit schutzbereichsbegrenzend, nicht aber in einer Art und Weise stattfinden, die die Wiedererkennbarkeit des Werkes allein auf Basis normativer Kriterien begründet.⁶⁹⁰ Voraussetzung jeder Wiedererkennbarkeit ist demzufolge zunächst, dass

685 So insbesondere *Stieper*, ZUM 2019, 713 (719); *Schulze*, GRUR 2020, 128 (130).

686 Vgl. *Finke*, Werk und Veränderung, S. 204.

687 So vgl. BGH, Urt. v. 15.12.2022 - I ZR 173/21, GRUR 2023, 571 Rn. 28, 32 – Vitrienenleuchte; BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 56 – Porsche 911.

688 *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 143.

689 Dazu ausführlich siehe oben unter 3. Teil B. I. 1. a) bb).

690 So auch *Grisse/Kaiser*, ZUM 2021, 401 (405, 413); *Haberstumpf*, ZUM 2020, 810 (817); in diese Richtung auch *Stieper*, ZUM 2019, 713 (719).

die schutzbegründenden Elemente des Werkes in der veränderten Gestaltung noch objektiv erkennbar und nicht vollständig unkenntlich gemacht worden sind.⁶⁹¹ Vollständig unkenntlich gemacht worden sind die eigenschöpferischen Züge des Ursprungswerkes, wenn die veränderte Gestaltung dem Ursprungswerk bei seiner Rezeption nicht mehr zugeordnet werden kann.⁶⁹² Nur wenn im tatsächlichen Sinne Identifizierbarkeit im veränderten Erzeugnis besteht, kann in einem zweiten Schritt untersucht werden, ob die Wiedererkennbarkeit im Sinne der EuGH-Rechtsprechung aufgrund normativ-einschränkender Kriterien dennoch abzulehnen ist.⁶⁹³ Die Herstellung des Erzeugnisses wäre nach richtlinienkonformer Auslegung damit keine Vervielfältigungshandlung gemäß § 16 UrhG.

(2) Beurteilungsperspektive für die faktische Wiedererkennbarkeit schutzbegründender Elemente

Für die Bestimmung der faktischen Wiedererkennbarkeit ist dem natürlichen Wortsinn nach die menschliche Wahrnehmung entscheidend.⁶⁹⁴ Unklar ist aber, welche Betrachtungsperspektive bei Bestimmung der tatsächlichen Identifizierbarkeit genau eingenommen wird.

Einerseits könnte auf die Wahrnehmungsfähigkeit eines durchschnittlichen Rezipienten abgestellt werden.⁶⁹⁵ Damit müsste die tatsächliche Wiedererkennbarkeit schöpferischer Elemente aus der Perspektive eines durchschnittlich geschulten und aufmerksamen Betrachters mitunter abgelehnt werden, obwohl bei einem gezielten und aufmerksamen Vergleich ein überdurchschnittlich vorgeprägter Betrachter wie beispielsweise ein Fachrichter

691 *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 42; *Grisse/Kaiser*, ZUM 2021, 401 (405); *Haberstumpf*, ZUM 2020, 810 (817); *Stieper*, ZUM 2019, 713 (719); vgl. außerdem BGH, Urt. v. 15.12.2022 - I ZR 173/21, GRUR 2023, 571 Rn. 29 – Vitrienenleuchte, woraus deutlich wird, dass der BGH jedenfalls davon ausgeht, dass vor einer Gesamtschau übernommene eigenschöpferische Züge identifiziert werden müssen.

692 *Grisse/Kaiser*, ZUM 2021, 401 (410).

693 Das Zweistufenverhältnis betont auch *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 42, 45.

694 *Grisse/Kaiser*, ZUM 2021, 401 (414).

695 BGH, Urt. v. 30.04.2020 - I ZR 115/16, GRUR 2020, 843 Rn. 29 – Metall auf Metall IV; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 43; *Apel*, MMR 2019, S.601 (602); *Haberstumpf*, ZUM 2020, 810 (817); *Finke*, Werk und Veränderung, S. 203; vgl. mit der gleichen Beurteilungsperspektive zur Rechtslage vor Umsetzung der InfoSoc-RL LG Hamburg, Urt. v. 09.12.2020 - 308 O 431/17, ZUM-RD, 164 Rn. 37 – Hey Pippi Langstrumpf.

noch eigenschöpferische Elemente des Ursprungswerkes im veränderten Erzeugnis identifizieren könnte. Nach einem anderen, den Schutzbereich des Urheberrechts weiter fassenden Verständnis könnte es aber auch allein auf eine objektive Identifizierbarkeit ankommen, selbst wenn nur überdurchschnittlich geprägtes Fachpublikum zur Identifizierung in der Lage wäre. Gegen ein so weites Verständnis der tatsächlichen Wiedererkennbarkeitsgrenze spricht allerdings, dass die Perspektive des durchschnittlichen Betrachters auch bei der Beurteilung der für den urheberrechtlichen Schutz notwendigen Individualität nach § 2 UrhG herangezogen wird.⁶⁹⁶ Es ist damit konsequent, wenn eben jene Beurteilungsperspektive auch für die Wiedererkennbarkeit eingenommen wird und damit auch aus dem urheberrechtlichen Schutzbereich herausführt. Die Wiedererkennbarkeitsgrenze liefe bei einem weiten Verständnis, nach dem sogar die Identifizierbarkeit durch überdurchschnittlich geschultes Fachpublikum wie Fachgerichte ausreichend wäre, in den meisten Fällen außerdem faktisch leer. Insbesondere im prozessualen Kontext, der denklogisch eine Identifizierung durch den Urheber oder Rechtsinhaber voraussetzt, erschöpfte sie sich vielmehr in der Feststellung einer Selbstverständlichkeit.⁶⁹⁷ Um zu erfassen, ob die Verwertungsinteressen der Urheber durch die Verwendung eines Werkes beeinträchtigt sind, kommt es schließlich maßgeblich auf die Perspektive der Konsumenten an. Denn sie entscheiden darüber, ob das Erzeugnis eines Dritten Substitutionsgefahr für die Verwertung des Ursprungswerkes durch den Urheber auslöst, oder nicht. Die Wiedererkennbarkeitsgrenze leitet sich indes aus einer an dem Schutz der Verwertungsinteressen orientierten Auslegung des Vervielfältigungsrechts ab. Demzufolge ist die Konsumentensicht auch für die Bestimmung der Wiedererkennbarkeit heranzuziehen.⁶⁹⁸

Alles in allem sprechen damit durchgreifende Erwägungen dafür, die Wiedererkennbarkeit auf Grundlage der Wahrnehmungsfähigkeit eines durchschnittlichen Rezipienten zu beurteilen. Auch bei der Beurteilung der

696 OLG Düsseldorf, Urt. v. 21.01.1997 - 20 U 13/96, ZUM 1998, 65 (67); OLG Karlsruhe, Urt. v. 10.11.1999 - 6 U 220/97, ZUM 2000, 327 (329); OLG Köln, Urt. v. 12.06.2009 - 6 U 215/08, GRUR-RR 2010, 182 (184) – Pferdeskulptur; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 2 Rn. 8; Loewenheim/Leistner, in: Schricker/Loewenheim, Urheberrecht, UrhG § 2 Rn. 162; vgl. Raue, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 2 Rn. 54.

697 Antoine, Verändernde Werknutzungen, S. 185.

698 So auch BGH, Urt. v. 30.04.2020 - I ZR 115/16, GRUR 2020, 843 Rn. 29 – Metall auf Metall IV.

tatsächlichen Wiedererkennbarkeit im Sinne einer Identifizierbarkeit der eigenschöpferischen Züge eines Werkes wird damit eine in gewissem Grad normative Betrachtungsperspektive zugrunde gelegt. Wie auch bei Beurteilungen aus Sicht des Durchschnittsverbrauchers, die im Lauterkeitsrecht verbreitet sind, können Gerichte die Perspektive des durchschnittlichen Rezipienten kraft ihres Erfahrungswissens selbst einnehmen.⁶⁹⁹ Verkehrsbefragungen in Rezipientenkreisen können dementsprechend zwar eine gewisse Indizwirkung entfalten. Erforderlich sind sie aber nicht.⁷⁰⁰

Eine Wiedererkennung ist weiterhin ohne Kenntnis über das im Ausgangspunkt wiederzuerkennende Werk bereits begrifflich nicht möglich. Deswegen ist bei der Beurteilung der faktischen Wiedererkennbarkeit davon auszugehen, dass die maßgeblichen Durchschnittsrezipienten auch Kenntnis vom jeweiligen Ursprungswerk haben.⁷⁰¹

(3) Technische Rekonstruierbarkeit

Um einem Missbrauch der Schutzbereichsbegrenzung über die Wiedererkennbarkeitsvoraussetzung vorzubeugen, müssen auch solche Festlegungen eines Werkes vom urheberrechtlichen Vervielfältigungsrecht mitumfasst sein, im Rahmen derer die eigenschöpferischen Elemente des Werkes zwar unkenntlich gemacht wurden, diese allerdings mit Hilfe technischer Mittel wieder rekonstruiert werden können.⁷⁰² Sofern für die Rekonstruktion allerdings nicht-übliche, über die durchschnittlichen technischen Fähigkeiten der Rezipienten hinausgehende Mittel eingesetzt werden müssen, ist eine Rekonstruktion der Werke jedenfalls von einem durchschnittlichen Rezipienten tatsächlich nicht zu erwarten. Es ist damit auch kein Multiplikations- bezie-

699 So auch *Leistner*, GRUR 2019, 1008 (1010); vgl. zum Durchschnittsverbraucher im Lauterkeitsrecht EuGH, Urt. v. 18.10.2012 - C-428/11, GRUR 2012, 1269 Rn. 53 - Purely Creative u. a.; außerdem ErwGr. 18 RL 2005/29/EG.

700 Gleiches gilt für Sachverständigengutachten, siehe *Leistner*, GRUR 2019, 1008 (1010); vgl. zum Lauterkeitsrecht *Dreyer*, in: Harte-Bavendamm/Henning-Bodewig, Gesetz gegen den unlauteren Wettbewerb (UWG): mit Preisangabenverordnung und Geschäftsgeheimnisgesetz: Kommentar, UWG § 5 Rn. 157; a. A. für die Wiedererkennbarkeit aber *Grisse/Kaiser*, ZUM 2021, 401 (414).

701 So auch *Apel*, MMR 2019, S.601 (602).

702 Insoweit übereinstimmend *Grisse/Kaiser*, ZUM 2021, 401 (414); *Pesch/Böhme*, GRUR 2023, 997 (1005); *Käde*, Kreative Maschinen und Urheberrecht, S. 69 f.; *Jäger*, Artificial Creativity?, S. 332; im allgemeineren Kontext auch *Dustmann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 16 Rn. 16.

hungsweise Substitutionseffekt auf dem relevanten Markt zu erwarten.⁷⁰³ Damit fehlt es in diesen Fällen an einer Beeinträchtigung der Verwertungsinteressen betroffener Urheber. Die technisch umgewandelten, unkenntlich gemachten Werkrepräsentationen sind in diesem Fall infolgedessen keine Vervielfältigungsstücke der Werke im urheberrechtlichen Sinne.

Zur Bestimmung des nötigen Rekonstruktionsaufwands ist nicht auf subjektive Merkmale wie konkrete technische Fähigkeiten eines Nutzers oder gegen eine Nutzergruppe gerichtete technische Zugangsbeschränkungen, sondern auf den objektiv, also für jedermann erforderlichen Aufwand für eine Wiederherstellung der Wahrnehmbarkeit der eigenschöpferischen Elemente eines Werkes abzustellen.⁷⁰⁴ Im Ergebnis sind die eigenschöpferischen Züge in technischen Kopien dann wiedererkennbar, wenn die Rekonstruktion vom durchschnittlichen Rezipienten ohne besondere technische Fähigkeiten oder mit üblicherweise bekannten technischen Hilfsmitteln und Programmen vorgenommen werden kann.⁷⁰⁵ Das ist zum Beispiel bei der Konvertierung eines Werkes in eine RAR-Datei der Fall. Sie ermöglicht zwar unmittelbar keine Rezeption der schöpferischen Eigenschaften eines Werkes. Von durchschnittlichen Computernutzern kann das Werk mit leicht öffentlich verfügbaren Programmen aber ohne Weiteres wieder extrahiert werden. Hierdurch werden die schöpferischen Elemente des Werkes wieder wahrnehmbar. Gleiches gilt für eine Zerlegung in eine Vielzahl kleinster Werksegmente, die sich unter Zuhilfenahme öffentlich und kostenlos verfügbarer Computerprogramme für einen technisch auch nur durchschnittlich versierten Werkrezipienten ohne besondere Schwierigkeiten wieder zu einer vollständigen, wahrnehmbaren Festlegung des Werkes zusammenfügen lassen.⁷⁰⁶

703 A. A. zu einem etwas anders gelagerten Problem *Jager*, *Artificial Creativity?*, S. 333, wobei die Erwägungen auf einer sehr theoretischen Betrachtungsweise beruhen; außerdem *Dornis*, CR 2024, 765 Rn. 26 f., der die teleologische Perspektive auf die Reichweite des Schutzbereichs jedoch unberücksichtigt lässt.

704 Lediglich von subjektiven Schranken geht allerdings *Sesing-Wagenpfeil* aus, siehe *Sesing-Wagenpfeil*, ZGE 2024, 212 (243 ff.). Die dort zugrunde gelegte Beurteilungsgrundlage trägt deswegen nicht vollständig.

705 *Grise/Kaiser*, ZUM 2021, 401 (414); a. A. aber *Dustmann*, in: *Fromm/Nordemann*, *Urheberrecht*, UrhG § 16; *Pesch/Böhme*, GRUR 2023, 997 (1005); *Käde*, *Kreative Maschinen und Urheberrecht*, S. 69 f.; *Sesing-Wagenpfeil*, ZGE 2024, 212 (243 ff.); in einem etwas anderen Kontext auch *Jager*, *Artificial Creativity?*, S. 332, nach denen es nicht auf den Aufwand für die Rekonstruktion der Inhalte ankommt, was aus den o. g. Gründen allerdings zu weitreichend ist.

(4) Wiedererkennbarkeitsgrenze bei ausschließlich technisch bedingten Erzeugnissen ohne eigene Werkqualität

Die beim Feature Encoding von Trainingsdaten erzeugten Werkrepräsentationen werden als Teil eines technischen Verfahrens angefertigt. Klärungsbedürftig ist in diesem Zusammenhang, ob die Wiedererkennbarkeitsgrenze das Vervielfältigungsrecht nach einer richtlinienkonformen, funktionsbezogenen Auslegung auch dann beschränkt, wenn das veränderte Erzeugnis ausschließlich technisch bedingt ist. Ist das nicht der Fall, so wären ausschließlich technisch bedingte Werkrepräsentationen wie die Produkte des Feature Encodings auch dann Vervielfältigungsstücke im Sinne des § 16 UrhG, wenn die Werke in der Repräsentation nicht wiedererkennbar sind.

Dafür, die Wiedererkennbarkeitsgrenze des urheberrechtlichen Schutzgegenstands nicht auf ausschließlich technisch bedingte Produkte anzuwenden, spricht, dass sie von der überwiegenden rechtswissenschaftlichen Literatur in § 23 Abs. 1 S. 2 UrhG verortet wird.⁷⁰⁷ Nach dem Wortlaut des § 23 Abs. 1 S. 2 UrhG muss zwischen dem Ausgangs- und einem neugeschaffenen *Werk* ein hinreichender Abstand (unter anderem im Sinne fehlender Wiedererkennbarkeit) bestehen. Neue Werke mit eigenen schöpferischen Merkmalen werden bei ausschließlich technisch bedingten Prozessen hingegen nicht angefertigt.

Allerdings *muss* die die richtlinienkonforme Auslegung im nationalen Recht nicht an § 23 Abs. 1 S. 2 UrhG angeknüpft werden. Denn sie geht auf den Sinn und Zweck von Art. 2 InfoSoc-RL zurück. Die funktional-einschränkende Auslegung des Vervielfältigungsrechts und damit auch die Wiedererkennbarkeitsgrenze kann demzufolge unmittelbar in einer richtlinienkonformen Auslegung des § 16 UrhG berücksichtigt werden.⁷⁰⁸ Infolgedessen bedarf es auch keiner Überwindung des Wortlauthindernisses aus § 23 Abs. 1 S. 2 UrhG, nach dem ein neues Werk geschaffen worden sein muss.⁷⁰⁹ Dies macht deutlich, dass die Frage nach der Werkqualität des

706 Vgl. zu einem ähnlichen Fall, aber im Kontext der öffentlichen Wiedergabe nach Art. 3 InfoSoc-RL: EuGH, Urt. v. 17.06.2021 - C-597/19, GRUR 2021, 1067 Rn. 59 – Mircom/Telenet.

707 So bisher die weit überwiegende Ansicht, beispielsweise *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 42; *Lauber-Rönsberg/Ahlberg*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 23 Rn. 3 f.; *Grünberger*, ZUM 2023, 309 (317 ff.). Vgl. auch BT-Drs. 19/27426, S. 79.

708 Siehe oben unter 3. Teil B. I. 1. a) bb).

709 In diese Richtung auch BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 52, 55, 58 – Porsche 911; zustimmend *Grünberger*, ZUM 2023, 309 (318).

veränderten Erzeugnisses keine Frage der richtlinienkonformen Einschränkung des Vervielfältigungsrechts, sondern eine Frage der nationalen Anknüpfung im UrhG ist. Für die funktionsbezogene Auslegung ist nur maßgeblich, ob ein abgeleitetes Erzeugnis die Verwertungsinteressen betroffener Urheber beeinträchtigt.⁷¹⁰ Die Werkqualität des generierten Erzeugnisses ist hierfür nicht von Bedeutung.

Insgesamt sprechen damit bessere Gründe dafür, dass die § 16 UrhG nach richtlinienkonformer Auslegung immanente Schutzbereichsgrenze der Wiedererkennbarkeit nicht nur bei der Schöpfung abgeleiteter Werke überschritten werden kann.⁷¹¹ Damit können auch ausschließlich technisch bedingte Verfahrenserzeugnisse Werke in nicht wiedererkennbarer und demzufolge urheberrechtlich nicht relevanter Art und Weise enthalten. Werkepräsentationen, die beim Feature Encoding urheberrechtlich geschützter Trainingsdaten hergestellt werden, qualifizieren sich nach der funktionsbezogenen Auslegung des Vervielfältigungsrechts damit ebenfalls dann nicht als Vervielfältigungsstück im Sinne des § 16 UrhG, wenn die eigenschöpferischen Merkmale der umgewandelten Werke nach Abschluss des technischen Verfahrens nicht mehr wiedererkennbar sind.

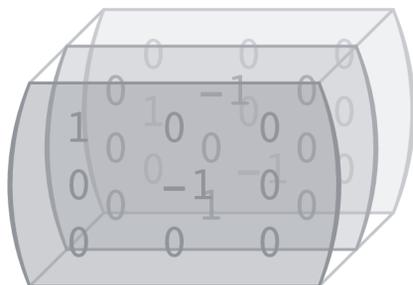
bb) Anwendung des Maßstabs der Wiedererkennbarkeit auf den konkreten Fall

Beim Feature Encoding werden die digital erfassten und normalisierten Werke in numerische Repräsentationen umgewandelt. Die Werke werden nach Abschluss des Umwandlungsprozesses also nur noch als Anordnung von Zahlen anhand von Achsen dargestellt.

710 Dazu im Detail oben unter 3. Teil B. I. 1. a) bb).

711 *Bullinger*, in: *Wandtke/Bullinger*, Urheberrecht, UrhG § 23 Rn. 28; *Grünberger*, ZUM 2023, 309 (318); *Siglmüller/Gassner*, RD 2023, 124 Rn. 15; *Maamar*, ZUM 2023, 481 (489 f.); wohl auch *Pesch/Böhme*, GRUR 2023, 997 (1006); *Schwarz/Söbbing*, RD 2023, 415 Rn. 13; a. A. *Lauber-Rönsberg/Ahlberg*, in: *Ahlberg/Götting/Lauber-Rönsberg*, Beck'scher Online-Kommentar Urheberrecht, UrhG § 23 Rn. 36; *Wandtke/Hauck*, GRUR Prax 2020, 542 (543), die allesamt Werkqualität voraussetzen. Wohl ebenso, aber noch zu § 24 UrhG a. F.: *Ory/Sorge*, NJW 2019, 710 (713).

Diese können beispielsweise wie folgt aussehen:



Für einen Rezipienten besteht bei diesen rein-numerischen Darstellungen keine Möglichkeit mehr, den eigentlichen Inhalt oder die ursprüngliche Form eines Werkes, also die jeweiligen schutzbegründenden Elemente, wahrzunehmen und zu verarbeiten. Die schutzbegründenden Merkmale eines Werkes aus diesen Zahlenmustern zu erkennen, ist für einen durchschnittlichen Rezipienten nicht mehr möglich. Die durchschnittlichen Rezipienten sind auch nicht in der Lage, die numerischen Darstellungen noch den Ursprungswerken zuzuordnen. Selbst wenn Tensor und zugehöriges Werk nebeneinander abgebildet sind, kann ein durchschnittlicher Rezipient keine inhaltliche Verwandtschaft beider Darstellungen erkennen.

Zwar können die Tensoren mit Hilfe technischer Mittel in wahrnehmbare Darstellungsformate retransformiert werden. Eine Rückumwandlung der Tensor-Repräsentation setzt aber vertiefte technische Kenntnisse und die Nutzung von Software voraus, die außerhalb des vom durchschnittlichen Rezipienten üblicherweise Beherrschten liegt.⁷¹² Im Gegensatz zu anderen technischen Maßnahmen wie dem Entpacken von RAR-Dateien ist die Retransformation von Tensoren für durchschnittlich technisch gebildete Werkrezipienten daher nicht möglich. Auch wenn sie technisch möglich ist, kann sie von einem durchschnittlichen Rezipienten demzufolge nicht erwartet werden.

Insgesamt folgt damit auf Grundlage der herausgearbeiteten Maßstäbe, dass die umgewandelten Werke in den rein-numerischen Tensoren tatsächlich nicht mehr identifizierbar sind. Dies führt zu einem Ausbleiben der Wiedererkennbarkeit im urheberrechtlichen Sinn. Auf ein normatives Ver-

712 Das verkennt Käde, *Kreative Maschinen und Urheberrecht*, S. 69 f.

ständnis des Wiedererkennbarkeitsbegriffs kommt es an dieser Stelle demzufolge nicht an.

Nach richtlinienkonformer Auslegung des Vervielfältigungsbegriffs findet beim Feature Encoding urheberrechtlich geschützter Trainingsdaten demzufolge keine Vervielfältigung der Inhalte gemäß § 16 Abs. 1 UrhG statt.⁷¹³ Die Festlegung der Werke in Form numerischer Tensoren genügt nicht, um das Werk auch nur mittelbar menschlich wahrnehmbar zu machen.⁷¹⁴

b) Schlussfolgerung

Im Rahmen des Feature Encodings von urheberrechtlich geschützten Trainingsdaten werden urheberrechtlich nicht relevante Werkrepräsentationen angefertigt. Dies gilt auch für weitere Kopien der Tensoren, die während des KI-Trainings angefertigt werden. Es bestehen systemische Ähnlichkeiten zu anderen Arten sekundärer Werkwiedergaben wie beispielsweise Inhaltsbeschreibungen, die ebenfalls kein Vervielfältigungsstück im Sinne des § 16 UrhG sind.

Der Vorverarbeitungsschritt des Feature Encodings urheberrechtlich geschützter Trainingsdaten fällt demzufolge bereits aufgrund einer unmittelbar begrenzenden, teleologisch geprägten Auslegung des Vervielfältigungsrechts nicht in den Schutzbereich des Ausschließlichkeitsrechts der Urheber. Damit werden die im zweiten Teil der vorliegenden Untersuchung für den Trainingsprozess identifizierten Regelungsbedürfnisse insoweit vollständig erfüllt.

Letztendlich folgt hieraus aber nicht, dass die weitergehende Verarbeitung des Tensors keine urheberrechtlich relevante Nutzungshandlung mehr begründen könnte. Denn sobald die Retransformation eines Tensor in ein menschlich wahrnehmbares Format stattfindet, wird das Werk wieder tatsächlich identifizierbar. Dies kann auch zur Wiedererkennbarkeit im urheberrechtlichen Sinn führen. Aufgrund dessen ist möglich, dass im Moment der Retransformation ein Vervielfältigungsstück im Sinne des § 16 UrhG

713 A. A. Käde, *Kreative Maschinen und Urheberrecht*, S. 69 f.; Spindler, GRUR 2016, 1112 (1113); im Anschluss hieran auch *Wiss. Dienst des Deutschen Bundestages*, KI und ML, eine urheberrechtliche Betrachtung (BT WD 10 – 3000 – 67/18), S. 7; wohl auch Dornis/Stober, *Urheberrecht und Training generativer KI-Modelle*, S. 70; Dregelies, GRUR 2024, 1484 (1484 f.).

714 A. A. Käde, *Kreative Maschinen und Urheberrecht*, S. 69 f.

hergestellt wird. Diese Umwandlung und die Weiterverarbeitung des umgewandelten Erzeugnisses bedürften dann der Erlaubnis durch den Urheber oder eine Schrankenbestimmung.

3. Zwischenfazit zur Vorverarbeitung urheberrechtlicher geschützter Trainingsdaten

Alles in allem können die Freistellungsbedürfnisse, die für die Vorverarbeitung urheberrechtlich geschützter Trainingsdaten als Teil des Trainingsprozesses ermittelt wurden, durch das geltende Urheberrecht nur teilweise erfüllt werden. Im Hinblick auf die Normalisierung von Trainingsdaten hängt dies damit zusammen, dass die bei der Normalisierung angefertigten Kopien der Werke Vervielfältigungsstücke des § 16 UrhG sein können. Damit bleibt das geltende Urheberrecht hinter den für die Normalisierung als Teil des Trainingsprozesses identifizierten Freistellungsbedürfnissen zurück. Allerdings gilt das nur für das Wie, nicht aber das Ob der Freistellung. Denn über eine Anwendung der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG kann dennoch eine Freistellung der Normalisierung erreicht werden. Voraussetzung hierfür ist jedoch, dass der Rechtsinhaber keinen Nutzungsvorbehalt gemäß § 44b Abs. 3 UrhG erklärt hat.

Im Gegensatz hierzu werden die hinsichtlich des Trainingsprozesses identifizierten Regelungsbedürfnisse beim anschließenden Feature Encoding der urheberrechtlich geschützten Trainingsdaten vollständig erfüllt. Die Herstellung der Tensoren, also der numerischen Repräsentationen der Werke, greift nicht in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG ein.

II. Trainingsanalyse der Werke selbst

Auch die eigentliche Trainingsanalyse der Werke selbst muss durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte urheberrechtlich freigestellt sein. Die Trainingsanalyse der Werke ist mit dem (automatisierten) Werkgenuss von Werken vergleichbar.⁷¹⁵ Der eigentliche Werkgenuss wird von den urheberrechtlichen Verwertungsrechten nicht

⁷¹⁵ Zur ausführlichen Begründung siehe oben unter 2. Teil C. II. 1.

erfasst.⁷¹⁶ Das gilt auch für den automatisierten Werkgenuss, also die automatisierte, informationsbezogene Auswertung von Werken, das Data Mining.⁷¹⁷ Dies spricht dafür, dass auch das mit dem automatisierten Werkgenuss wesensverwandte, eigentliche Training einer KI auf Basis von Werken nicht Teil des urheberrechtlichen Ausschließlichkeitsrechts ist. Wenn die in einem Werk enthaltenen Informationen nicht Teil des urheberrechtlichen Schutzbereichs sind und ihre Extraktion damit frei möglich ist, kann es außerdem keinen Unterschied machen, wie viele dieser Informationen aus dem Werk gewonnen worden sind oder aus wie vielen Werken eine Informationsextraktion stattfindet. Gegen die Gleichstellung des KI-Trainings mit dem Werkgenuss kann daher auch nicht eingewendet werden, dass das KI-Training quantitativ als auch qualitativ über den menschlichen Werkgenuss hinausgeht.⁷¹⁸ Im Ergebnis ist die eigentliche Analyse der Werke beim Training von KI damit – wie auch der automatisierte Werkgenuss – nicht Teil der urheberrechtlichen Verwertungsrechte.⁷¹⁹ Die Trainingsanalyse selbst wird demzufolge durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte aus dem Schutzbereich des Urheberrechts ausgeklammert. Das Freistellungsbedürfnis für das Training von KI wird hinsichtlich des eigentlichen Analysevorgangs damit erfüllt. Das „*right to read*“ ist insofern nicht nur das „*right to mine*“ im datenwis-

716 *Heerma*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 15 Rn. 10 m. w. N. außerdem bereits BT-Drs. IV/270, S. 28. Nach *Sucker* gilt das jedenfalls für den analogen Werkgenuss, während der digitale Werkgenuss mittelbar durch die zwangsläufig stattfindenden Vervielfältigungshandlungen vom urheberrechtlichen Ausschließlichkeitsrecht erfasst wird, siehe *Sucker*, Der digitale Werkgenuss im Urheberrecht, S. 47 ff., 89. Diese – zutreffende – folgenorientierte Betrachtungsweise hat aber keinen Einfluss darauf, dass der eigentliche menschliche Analysevorgang, also der Werkgenuss im engeren Sinne, auch im digitalen Kontext nicht dem Ausschließlichkeitsrecht der Urheber zugewiesen ist.

717 Das zeigt schon die Regelung des § 44b Abs. 2 UrhG. Siehe außerdem *Stieper*, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 60d Rn. 3; *Raue*, ZUM 2019, 684 (685); *de la Durantaye*, Allgemeine Bildungs- und Wissenschaftsschranke, S. 240 f.; *Kleinkopf*, Text- und Data-Mining, S. 181; *Cocoru/Boehm*, An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, S. 8; *European Copyright Society*, General Opinion on the EU Copyright Reform Package, S. 5; BT-Drs. 18/12329, S. 40; *Hilty/Moscon*, Modernisation of the EU Copyright Rules, S. 127 f.

718 So aber *Schack*, NJW 2024, 113 Rn. 9.

719 Vgl. auch *Raue*, ZUM 2019, 684 (684 f.); *Wulf/Löbeth*, GRUR 2024, 737 (738); ebenso *Kögel*, DSRITB 2023, 285 (287).

senschaftlichen Sinne,⁷²⁰ sondern auch das „*right to train*“ beim Training von KI auf Basis von Werken.⁷²¹

III. Vervielfältigungen während der Trainingsanalyse

Während der Trainingsanalyse kommt es außerdem zu weiteren Vervielfältigungen der Trainingsdaten. Sie werden beispielsweise vorübergehend in den Arbeitsspeicher der für das Training genutzten Rechenarchitektur geladen. Vervielfältigt werden allerdings nicht die Werke selbst, sondern die Tensoren, also die numerischen Repräsentationen der Trainingsdaten, die im Rahmen der Datenvorverarbeitung angefertigt worden sind.⁷²² Diese sind keine Vervielfältigungsstücke der Werke. Ihre Vervielfältigung im Rahmen der Trainingsanalyse berührt demzufolge auch nicht das urheberrechtliche Ausschließlichkeitsrecht.⁷²³ Die Vervielfältigungen während der Trainingsanalyse sind vielmehr frei möglich. Dies steht im Einklang mit den identifizierten regulatorischen Freistellungsbedürfnissen.

IV. Gewichtungswerte und andere Parameter des trainierten Modells

Ziel des Trainingsvorgangs ist eine Konfiguration der im KNN enthaltenen, numerischen Parameter. Sie soll in einer Art und Weise erfolgen, die eine bestmögliche Qualität des vom KI-System generierten Outputs gewährleistet. Auch die Konfiguration der Parameter innerhalb des KNN sollte als Teil des Trainingsprozesses aus einer regulatorischen Perspektive durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte freigestellt sein.⁷²⁴

Ob die identifizierten Freistellungsbedürfnisse im vorliegenden Fall vom geltenden Urheberrecht erfüllt werden, hängt davon ab, ob die konfigurierten Parameter des KNN Vervielfältigungsstücke der zum Training verwen-

720 Zur Formulierung *Spindler*, ZGE 2018, 273 (283); im Kontext wissenschaftlicher Forschung *Raue*, ZUM 2019, 684 (686).

721 A. A. *Schack*, NJW 2024, 113 Rn. 9; im Kontext generativer KI-Systeme *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 103.

722 Siehe zum Feature Encoding unmittelbar oben unter 3. Teil B. I. 2.

723 A. A. wohl *Jäger*, Artificial Creativity?, S. 331 f., dessen Standpunkt auf der Annahme beruht, dass nicht nur numerische Repräsentationen, sondern stets die Werke selbst verarbeitet werden. Ebenfalls a. A. *Schack*, NJW 2024, 113 Rn. 4.

724 Vgl. oben unter 2. Teil C. I.

deten Werke im Sinne des § 16 UrhG sind. Der Vervielfältigungsbegriff wird dabei nicht rein technisch, sondern normativ anhand der vom Vervielfältigungsrecht geschützten Verwertungsinteressen der Urheber bestimmt.⁷²⁵ Eine Vervielfältigung liegt im konkreten Fall danach insbesondere dann nicht vor, wenn die eigenschöpferischen Züge des Werkes in den numerischen Parametern des trainierten KNN nicht mehr wiedererkennbar sind.⁷²⁶ Die objektiv zu bestimmende Wiedererkennbarkeit eines Werkes in einem Erzeugnis ist dabei von subjektiven Zugangsbeschränkungen zu unterscheiden.⁷²⁷ Für das Fehlen der Wiedererkennbarkeit ist nicht erforderlich, dass der neuen, veränderten Gestaltung, also den Parametern des trainierten KNN, selbst als Werk urheberrechtlicher Schutz zukommt.⁷²⁸ Die Schutzbereichsgrenze der Wiedererkennbarkeit besteht auch bei rein technischen Verarbeitungen wie der Konfiguration von Parametern innerhalb eines KNN.

Damit die schutzbegründenden Elemente der zum Training verwendeten Werke in den Parametern eines KNN noch wiedererkennbar sind, müssen sie dort zunächst noch tatsächlich und objektiv identifiziert werden können.⁷²⁹ Bei einer vollständigen Unkenntlichmachung der eigenschöpferischen Züge scheidet eine Wiedererkennbarkeit hingegen aus. Bei der Beurteilung kommt es auf die Wahrnehmung eines durchschnittlichen Betrachters an.⁷³⁰ Die Notwendigkeit einer technischen Rekonstruktion der Inhalte ist für die tatsächliche Wiedererkennbarkeit der schutzbegründenden Merkmale dabei unschädlich, solange die Wiederherstellung vom durchschnittlichen Rezipienten ohne besondere technische Fähigkeiten oder mit üblicherweise bekannten technischen Hilfsmitteln und Programmen vorgenommen werden kann.⁷³¹

Erst wenn ein durchschnittlicher Betrachter schutzbegründende Elemente in den Parametern eines KNN objektiv identifizieren kann, ist in einem zweiten Schritt zu untersuchen, ob es trotz einer objektiv-tatsächlichen Er-

725 Dazu m. w. N. auch zu abweichenden Ansichten siehe oben unter 3. Teil B. I. 1. a) bb).

726 Zur Herleitung genauer oben unter 3. Teil B. I. 1. a) bb).

727 Das verkennen *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 77 f.

728 Ausführlicher hierzu mit abweichenden Ansichten oben unter 3. Teil B. I. 2. a) aa) (4).

729 Dazu ausführlich m. w. N. oben unter 3. Teil B. I. 2. a) aa) (1).

730 Hierzu bereits siehe 3. Teil B. I. 2. a) aa) (2).

731 Str., ausführlich hierzu mit abweichenden Ansichten oben unter 3. Teil B. I. 2. a) aa) (3).

kennbarkeit der zum Training verwendeten Werke in den Parametern des KNN dennoch zu einer normativ-einschränkenden Auslegung der Wiedererkennbarkeitsgrenze mit der Folge kommen kann, dass die Parameter des KNN trotz der Identifizierbarkeit schutzbegründender Elemente dennoch keine Vervielfältigungsstücke im Sinne des § 16 UrhG beinhalten.

1. Objektive Identifizierbarkeit schutzbegründender Elemente in Parametern eines KNN

Ein durchschnittlicher Betrachter verfügt typischerweise nicht über die technischen Fähigkeiten, um die Architektur und Parameter eines KNN einzusehen. Prohibitiv wirkt bereits, dass die hierfür notwendigen Programme nicht üblicherweise bekannt sind. Hierin besteht auch ein Unterschied zur Speicherung von Werken auf Computerfestplatten, die über marktübliche, in der Regel auf einem Computer vorinstallierte und dem objektiven Durchschnittsbetrachter zur Verfügung stehende Programme typischerweise leicht wahrnehmbar gemacht werden können.⁷³² Selbst wenn der durchschnittliche Betrachter die Architektur und Parameter eines KNN einsehen könnte, sind im KNN selbst nur numerische Darstellungen, also einzelne Zahlenwerte sichtbar. Sie wurden auf Basis der zum Training genutzten Werke ermittelt. Ein durchschnittlicher Betrachter kann aus Angaben wie „ $w_1 = 0,5$; $w_2 = 1,8$; $w_3 = 21,9$ “ allerdings keine Schlüsse auf die schutzbegründenden Merkmale eines Werkes ableiten. Die Zahlenwerte vermitteln an den durchschnittlichen Betrachter also nicht die schöpferischen Merkmale eines zum Training genutzten Werkes, also beispielsweise einen Pinselstrich bestimmter Technik und Führung oder ein Sujet, in der die Persönlichkeit eines Urhebers zum Ausdruck kommen.⁷³³ Aufgrund dieser numerischen Repräsentation besteht auch keine Vergleichbarkeit zur Speicherung von Werken im Arbeitsspeicher eines Computers. Die Werke werden dort – vom nativen Maschinencode abgesehen – nicht als numerische Repräsentationen, sondern in Klarform vorübergehend abgelegt.⁷³⁴

732 Anders jedoch *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 77, die eine Gleichstellung beider Vorgänge vornehmen.

733 Darauf stellen auch *Käde*, ZUM 2024, 174 (177); *Konertz/Schönhof*, WRP 2024, 534 Rn. 8; *Kögel*, DSRITB 2023, 285 (288) ab.

734 A. A. *Jager*, Artificial Creativity?, S. 332.

2. Möglichkeiten der Invertierung und die Wiedererkennbarkeit

Zwar ist in der Computerwissenschaft noch nicht abschließend geklärt, ob und wenn ja zu welchem Grad eine Memorisierung von Werken in den Parametern eines trainierten KNN stattfindet und mit welcher Ergebnisqualität eine Invertierung, also eine Extraktion der zum Training genutzten Werke aus dem trainierten KNN, erfolgen kann.⁷³⁵ Selbst wenn dies möglich wäre, ist eine Invertierung allerdings technisch komplex.⁷³⁶ Die Invertierung könnte nicht mit für technische Laien bekannten Mitteln erfolgen, sondern setzte vielmehr besondere Fähigkeiten und Ressourcen voraus. Sie ist computerwissenschaftlichen Fachkreisen vorbehalten. Die Rekonstruktion von Werken aus den Parametern eines KNN ist für den durchschnittlichen Betrachter demzufolge nicht möglich. Infolgedessen ist unabhängig davon, ob eine Invertierung der KNN theoretisch möglich ist, nicht von einer technischen Rekonstruierbarkeit der zum Training genutzten Werke auszugehen, die zu einer tatsächlichen Identifizierbarkeit und damit zu einer Wiedererkennbarkeit im urheberrechtlichen Sinne führt.⁷³⁷

735 Dazu vgl. beispielhaft *Feldman*, arXiv:1906.05271 2019; *Nagarajan/Raffel/Goodfellow*, Paper zur NeurIPS Conference 2018; *Carlini u. a.*, arXiv: 2301.13188 2023, wobei es hier, wie der Auffand der Rekonstruktion zeigt, weniger um strukturelle, sondern um Einzelfallprobleme geht; aus der juristischen Literatur *Pesch/Böhme*, GRUR 2023, 997 (1001); *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 59 ff.

736 Insoweit zustimmend auch *Jager*, Artificial Creativity?, S. 331. Eindrucksvoll insbesondere das Beispiel von *Carlini u. a.*, arXiv: 2301.13188 2023: Die Autoren untersuchten 160 Millionen verwendete Trainingsdaten eines generativen KI-Systems und nutzten 350.000 Datenpaare, die im Trainingsdatensatz doppelt hinterlegt worden sind. Die hieraus extrahierten Bildbeschreibungen wurden jeweils 500 Mal als Prompt in das generative KI-System eingegeben. Die 175 Millionen hieraus generierten Inhalte wurden nunmehr mit den 350.000 zuvor ermittelten Datenpaaren verglichen. Hieraus ergaben sich 109 Übereinstimmungen, was einer Übereinstimmungsquote zum gesamten Trainingsdatensatz von 0,00000062 % entspricht. Für diese Berechnungen mussten alle Bildinhalte in Vektoren umgewandelt werden. Für das Experiment waren also fortgeschrittene technische Fähigkeiten und erhebliche Rechenkapazitäten notwendig.

737 Ein vergleichbares Konzept verfolgt auch *Leistner*, GRUR 2024, 1665 (1669). Andere Ansicht insbesondere *Jager*, der den Erfolg einiger Forschungsversuche zur Invertierung als ausreichend für eine generelle Wahrnehmbarmachung der zum Training verwendeten Werke erachtet, siehe *Jager*, Artificial Creativity?, S. 331. A. A. außerdem *Pesch/Böhme*, GRUR 2023, 997 (1005) sowie jüngst *Dornis*, CR 2024, 765 Rn. 22 ff. und insbesondere Rn. 26, der sich allerdings ebenfalls nicht mit der Funktion des urheberrechtlichen Vervielfältigungsrechts und dessen Folgen auseinandersetzt.

3. Identifizierbarkeit schutzbegründender Elemente im Output von KI und die Wiedererkennbarkeit

Sollten schutzbegründende Merkmale von zum Training genutzten Werken ausnahmsweise im Output eines KI-Systems enthalten sein, werden die schutzbegründenden Eigenschaften erst mit der Generierung dieses konkreten Outputs wieder tatsächlich, objektiv identifizierbar. Die Generierung des Outputs ist sowohl zeitlich als auch wirtschaftlich vom Training eines KI-Systems abgrenzbar.⁷³⁸ Sie geht außerdem auf einen eigenen Willensentschluss zurück. Die Entwicklung eines KI-Systems inklusive der Konfiguration der Parameter innerhalb des KNN und die Generierung konkreter, gegebenenfalls urheberrechtsverletzender KI-Erzeugnisse sind demzufolge voneinander unabhängig. Gegebenenfalls liegen auch zwei unabhängige Nutzungshandlungen vor.⁷³⁹ Hieraus folgt, dass nicht von der Generierung des Outputs auf eine Werknutzung beim Training der KI geschlossen werden kann.⁷⁴⁰ Eine Betrachtung hat vielmehr jeweils getrennt voneinander zu erfolgen. Die Identifizierbarkeit schutzbegründender Elemente von zum Training verwendeten Werken im Output eines KI-Systems spricht demzufolge nicht automatisch für eine urheberrechtsrelevante, erlaubnisbedürftige *Vervielfältigung* der zum Training verwendeten Werke in den Parametern des KNN.⁷⁴¹ Bei der Erzeugung von Output durch eine KI wird auch aus einer technischen Perspektive außerdem nicht die Gesamtheit der Parameter, sondern das Rechenergebnis des KNN, also der Aktivierungszustand der letzten Neuronenschicht, sichtbar gemacht. Dieser wurde *auf Basis* der Parameter des KNN errechnet.⁷⁴² Damit kann das Verhältnis von Parametern und Output eines KNN auch nicht mit klassischen Werkwiedergaben, beispielsweise durch einen Schallplattenspieler, verglichen werden. In diesem Fällen wird das untersuchte Vervielfältigungsstück (Schallplatte) und kein auf Basis des Vervielfältigungsstücks (Schalplatte) abgeleitetes Erzeugnis wahrnehmbar gemacht. Vielmehr ist die Generierung des Outputs eine eigene urheberrechtlich relevante Vervielfältigungs-

738 So auch *Leistner*, GRUR 2024, 1665 (1669).

739 So auch *Käde*, ZUM 2024, 174 (177); *Schwarz/Söbbing*, RD 2023, 415 Rn. 13.

740 A. A. *Dornis*, CR 2024, 765 Rn. 9; *Sesing-Wagenpfeil*, ZGE 2024, 212 (229 ff., 233 ff.).

741 Anders jedoch *Pesch/Böhme*, GRUR 2023, 997 (1005) mittels „geschickten Eingabeprompts“; *Konertz/Schönhof*, WRP 2024, 534 Rn. 8 am Ende; insbesondere auch *Sesing-Wagenpfeil*, ZGE 2024, 212 (229 f.), dessen Untersuchung maßgeblich auf der gegenteiligen Annahme gründet.

742 Insoweit technisch unzutreffende Grundannahme in *Sesing-Wagenpfeil*, ZGE 2024, 212 (229 ff.).

handlung im Sinne des § 16 UrhG, wenn in dem Output schutzbegründende Elemente wiedererkennbar sind. Diese Generierung ist für sich genommen sodann erlaubnispflichtig.⁷⁴³

Darüber hinaus erscheint ein zwingender Schluss von der Wahrnehmbarmachung eines Werkes auf das Vorliegen eines Vervielfältigungsstücks desselben Werkes jedenfalls zweifelhaft.⁷⁴⁴ Beispielsweise macht eine Textanleitung zur Reproduktion eines Werks der bildenden Kunst die schöpferischen Züge des Werkes nicht urheberrechtlich relevant wiedererkennbar. Somit qualifiziert sie sich nicht als Vervielfältigungsstück des Werkes. Dennoch ist die Anleitung geeignet, das urheberrechtlich geschützte Werk durch ihre Ausführung wahrnehmbar zu machen. Die Wahrnehmbarmachung eines Werkes kann, muss damit aber nicht zwingend auf Basis eines Vervielfältigungsstücks erfolgen.

Sollen die *Parameter* eines KNN als Vervielfältigungen von zum Training genutzten Werken im Sinne des § 16 UrhG eingeordnet werden, müssen ihre schutzbegründenden Merkmale vielmehr direkt bei einer Analyse des KNN und nicht des Outputs der KI wiedererkennbar sein. Dies ist, wie dargelegt wurde, bereits aufgrund der numerischen Darstellungsform und einer komplexen technischen Rekonstruktion nicht der Fall. Nach dem Stand der Technik existiert kein Lesegerät, also kein „Parameterspieler“, mit dessen Hilfe die in Parametern eines KNN repräsentierten Werke sichtbar gemacht werden können. Es kommt deswegen auch nicht darauf an, ob tatsächlich schutzbegründende Eigenschaften der Werke in den numerischen Parametern abgelegt werden.

4. Schlussfolgerungen

Alles in allem sind die schutzbegründenden Merkmale der zum Training genutzten Werke aus den vorgenannten Gründen damit nicht in den Parametern des trainierten KNN im urheberrechtlichen Sinne wiedererkennbar.⁷⁴⁵ Dementsprechend werden die Werke beim Trainingsprozess nicht in den Parametern des KNN im Sinne des § 16 UrhG vervielfältigt.⁷⁴⁶ Falls

743 So richtigerweise auch *Leistner*, GRUR 2024, 1665 (1669).

744 So jedoch *Dornis*, CR 2024, 765 Rn. 9, 11 ff.; *Sesing-Wagenpfeil*, ZGE 2024, 212 (229); *Kögel*, InTeR 2023, 179 (181).

745 Auf die mangelnde Erkennbarkeit stellt auch *Käde*, ZUM 2024, 174 (177) ab.

746 Tendenziell auch *Hofmann*, WRP 2024, 11 Rn. 12; zustimmend auch *Maamar*, ZUM 2023, 481 (483); *Lennartz/Kraetzig*, GRUR 2024, 1258 (1259); *Kraetzig*, GRUR 2022,

nach dem Training des KNN von einem technischen Fachexperten eine Invertierung durchgeführt, in deren Folge schutzbegründende Elemente zum Training verwendeter Werke objektiv identifizierbar gemacht werden, wird erst mit dieser Invertierung eine neues Vervielfältigungsstück im Sinne des § 16 UrhG hergestellt. Erst die Invertierung stellt infolgedessen eine erlaubnisbedürftige urheberrechtlich relevante Nutzungshandlung dar.

Im Ergebnis ist die Konfiguration der Parameter des KNN, letzter Schritt des Trainingsprozesses, damit durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte urheberrechtlich freigestellt. Die für diesen Teilaspekt des KI-Trainings im zweiten Teil der Untersuchung identifizierten Regelungsbedürfnisse werden durch das geltende Urheberrecht also erfüllt.

V. Fazit zur Durchführung des Trainings auf Basis von Werken

Weil die vier Teilaspekte des Trainingsvorgangs im Wesentlichen nicht in den Anwendungsbereich des urheberrechtlichen Vervielfältigungsrechts aus § 16 UrhG fallen, erfüllt das geltende Urheberrecht die für die Durchführung des KI-Trainings identifizierten Regelungsbedürfnisse weitestgehend. Etwas anderes gilt nur für die Normalisierung der urheberrechtlich geschützten Trainingsdaten. Sie ist zwar durch die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erlaubnisfrei gestellt. Dieser Regelungsmechanismus erweist sich aus einer regulatorischen Perspektive aber nicht als zweckmäßig für die Freistellung der Normalisierungshandlung. Das ist de lege ferenda durch eine Anpassung der Reichweite des urheberrechtlichen Vervielfältigungsrechts aus § 16 UrhG zu lösen.

1707 Rn. 7; ebenso *Käde*, ZUM 2024, 174 (177 f., 182); *Käde/v. Maltzan*, InTeR 2020, 201 (207); *Kögel*, DSRITB 2023, 285 (288); *Schwarz/Söbbing*, RD 2023, 415 Rn. 13; a. A. *Dornis*, CR 2024, 765 Rn. 6; *Sesing-Wagenpfeil*, ZGE 2024, 212 (242 f.); v. *Welser*, GRUR-Prax 2023, 516 Rn. 21; *Jäger*, Artificial Creativity?, S. 331; *Kleinkopf*, Text- und Data-Mining, S. 166 f.; dazwischen *Pesch/Böhme*, GRUR 2023, 997 (1004 f.), nach denen eine Vervielfältigung in den Parametern grundsätzlich nicht, im Einzelfall und im frühen Trainingsstadium allerdings stattfinden kann; *Konertz/Schönhof*, WRP 2024, 534 Rn. 8, die außer im Ausnahmefall keine Vervielfältigung annehmen.

C. Speicherung urheberrechtlich geschützter Inputdaten

Ein Freistellungsbedürfnis konnte im zweiten Teil der vorliegenden Untersuchung weiterhin für die Speicherung von urheberrechtlich geschützten Inputdaten ermittelt werden.⁷⁴⁷ Diese können beispielsweise mittels Sensortechnik wie Kameras oder Mikrofonen oder durch Webcrawler erfasst, aber auch manuell in ein KI-System eingegeben werden. Hierzu gehört auch die Verwendung von geschützten Werken als Prompt eines generativen KI-Systems. Von der Speicherung urheberrechtlich geschützter Trainingsdaten unterscheidet sie sich dadurch, dass in der Anwendungsphase des KI-Systems – in dieser werden die Inputdaten verarbeitet – nur vorübergehende Speicherungen der Werke notwendig sind.

Die Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten sollte aus einer regulatorischen Perspektive im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte erfolgen. Damit das Freistellungsbedürfnis erfüllt wird, darf die Speicherung von Werken als Inputdaten also nicht in den Anwendungsbereich des urheberrechtlichen Vervielfältigungsrechts (§ 16 UrhG) fallen.

I. Eingriff in das Vervielfältigungsrecht

Bei der Speicherung urheberrechtlich geschützter Inputdaten werden vorübergehende, digitale Kopien der betroffenen Werke hergestellt. Diese dienen im Anschluss als Berechnungsgrundlage des KI-Systems. Die Erfassung und Speicherung der Werke als Inputdaten fällt dabei nicht in den Anwendungsbereich des urheberrechtlichen Vervielfältigungsrechts aus § 16 UrhG, wenn trotz des digitalen Kopierens keine Vervielfältigungsstücke der Werke im Sinne des § 16 Abs. 1 UrhG hergestellt werden.

Wie bereits gezeigt werden konnte,⁷⁴⁸ ist der Vervielfältigungsbegriff als Rechtsbegriff nicht rein technischer, sondern normativer Natur.⁷⁴⁹ Seine

747 Dazu ausführlich siehe oben unter 2. Teil D.

748 Ausführlich zur Begründung und Ausgestaltung eines normativen, funktionsbezogenen Vervielfältigungsbegriffs siehe oben unter 3. Teil B. I. 1. a) bb).

749 Hugenholtz, EIPR 2000, 482 (485); sich hieran anschließend Schönberger, ZGE 2018, 35 (53); a. A. wohl Heerma, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; Chiou, JIPITEC 2019, 398 Rn. 6; außerdem Jäger, Artificial Creativity?, S. 332.

Bedeutung muss daher insbesondere auch unter Berücksichtigung des Zwecks des in § 16 UrhG umgesetzten Art. 2 InfoSoc-RL bestimmt werden. Hieraus folgt, dass nur solche Werkverwendungen auch erlaubnispflichtige Werknutzungen im Sinne von Vervielfältigungen sind, die die Verwertungsmöglichkeit der Urheber zumindest potenziell beeinträchtigen.⁷⁵⁰ Für eine Beeinträchtigung der Verwertungsinteressen genügt, dass ein unmittelbares oder mittelbares Vermarktungspotential aus der Werkverwendung entsteht und damit Substitutionsgefahr für das verwendete Werk hervorgerufen wird.⁷⁵¹

Im vorliegenden Fall spricht gegen eine Beeinträchtigung des Vermarktungspotentials der betroffenen Werke, dass die Werke nicht zum Genuss der schöpferischen Merkmale, sondern nur als Berechnungsgrundlage für ein KI-System und dementsprechend zur Erfüllung einer technischen Funktion abgespeichert werden. Je nachdem, welche Art von KI-System konkret eingesetzt wird, kann die Erfassung von Werken als Inputdaten auch beiläufig und zufällig sein. Außerdem müssen die Inputdaten eines KI-Systems mitunter aus einem konkreten Anwendungsszenario heraus erhoben und sofort verarbeitet werden können. Auch die Beziehung des jeweiligen Werks zu seiner Umwelt kann von Bedeutung sein. Lizenzverträge zur Nutzung von als Inputdaten erfassten Werken können demzufolge aus tatsächlichen Gründen in einigen Fällen nicht abgeschlossen werden.

Allerdings können generierte Inputdaten losgelöst von der Anwendung eines konkreten KI-Systems jedenfalls im Nachhinein dazu verwendet werden, um KI-Systeme zu trainieren. Für viele KI-Entwickler sind die von Systemnutzern generierten Inputdaten hierfür auch notwendig. Denn unter Umständen kann im Entwicklungsstadium des KI-Systems keine ausreichende Menge an Trainingsdaten hinreichender Diversität akquiriert werden. Die Verwendung von aus einem realen Anwendungsumfeld abgepeisten Inputdaten für die Optimierung oder das Training eines Systems

750 öst. OGH, Beschl. v. 23.02.2010 - 4 Ob 208/09 f, ZUM 2010, 629 (630); v. Ungern-Sternberg, in: Schrickler/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23; v. Ungern-Sternberg, GRUR 2012, 576 (580); Schönberger, ZGE 2018, 35 (53); Antoine, Verändernde Werknutzungen, S. 179 f.; in diese Richtung auch Grisse/Kaiser, ZUM 2021, 401 (404); grundsätzlich wie hier auch Antoine, ZGE 2023, 387 (398), wobei sie eine Einschränkung der funktionsbezogenen Auslegung auf die „Nutzungsfälle an den Rändern des urheberrechtlichen Ausschließlichkeitsrechts“ befürwortet, dazu sogleich. Zum Verhältnis zu den urheberpersönlichkeitsrechtlichen Bestimmungen siehe Antoine, ZGE 2023, 387 (401 ff.).

751 Siehe oben für weitere Details zur Schutzbereichsgrenze aus der funktionsbezogenen Auslegung des Vervielfältigungsbegriffs unter 3. Teil B. I. 1. a) bb).

ist auch verbreitet.⁷⁵² Daraus folgt, dass als Inputdaten generierte Werke zwar mitunter nicht als solche monetarisiert werden können. Dennoch besteht jedenfalls ein Vermarktungspotenzial der als Inputdaten generierten Werke zur Verwendung als Trainingsdaten. In dieser Form besteht auch Substitutionsgefahr für die Verwertungsmöglichkeiten des Urhebers. Alles in allem beeinträchtigt die Speicherung urheberrechtlich geschützter Inputdaten das Verwertungsinteresse betroffener Urheber damit zumindest potenziell. Hieraus folgt, dass nach einer funktionsbezogenen Auslegung des Vervielfältigungsbegriffs bei der Speicherung urheberrechtlich geschützter Inputdaten für KI Vervielfältigungsstücke erfasster Werke im Sinne des § 16 UrhG hergestellt werden. Damit findet, beispielsweise bei der Aufnahme von urheberrechtlich geschützten Werken der Baukunst durch die Kameras autonomer Fahrzeuge, ein Eingriff in das urheberrechtliche Ausschließlichkeitsrecht in Form des Vervielfältigungsrechts statt. Die Vervielfältigungen bei der Speicherung von Werken als Inputdaten sind demzufolge im Grundsatz erlaubnispflichtig. Diese Erlaubnis könnte zwar durch eine urheberrechtliche Schrankenbestimmung gewährt werden. Aus alledem folgt allerdings bereits, dass das UrhG das für die Speicherung von urheberrechtlich geschützten Inputdaten identifizierte Regelungsbedürfnis nach einer Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte in den §§ 15 ff. UrhG nicht erfüllt.

II. Anwendung urheberrechtlicher Schrankenbestimmungen

Nichtsdestotrotz ließe sich durch die Anwendung einer urheberrechtlichen Schrankenbestimmung zumindest das Ob des regulatorischen Freistellungsbedürfnisses erfüllen.

1. Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)

Solange die Werke nur als Berechnungsgrundlage des KI-Systems in dessen Anwendungsphase dienen sollen, ist eine vorübergehende Speicherung der Inhalte ausreichend. Diese könnte von der Schrankenbestimmung aus § 44a UrhG abgedeckt sein. Dafür muss die Vervielfältigung der Werke als

⁷⁵² Vgl. als Beispiele *Hoffmann/Vogt*, MMR-Aktuell 2020, 431466; *Shah*, AIwire v. 8.3.2023; *Stecklow/Cunningham/Jin*, Reuters v. 6.4.2023.

Inputdaten vorübergehend und flüchtig oder begleitend im Sinne des § 44a UrhG sowie integraler und wesentlicher Teil eines technischen Verfahrens sein. Ihr alleiniger Zweck muss sich außerdem darin erschöpfen, eine Übertragung in einem Netz zwischen Dritten durch einen Vermittler (§ 44a Nr. 1 UrhG) oder eine rechtmäßige Nutzung (§ 44b Nr. 2 UrhG) zu ermöglichen. Den Vervielfältigungen darf dabei selbst keine eigene wirtschaftliche Bedeutung zukommen.

a) Vorübergehende Vervielfältigung

Vorübergehend ist eine Vervielfältigung dann, wenn sie nach der Zweckerfüllung innerhalb des technischen Verfahrens zur Löschung vorgesehen ist.⁷⁵³ Zweck der Vervielfältigung der Werke ist, sie als Input eines KI-Systems zu verwenden. Die Vervielfältigungsstücke der erfassten Werke werden hierfür nach der Umwandlung in numerische Tensoren und spätestens nach der Eingabe dieser Tensoren in den Algorithmus nicht mehr benötigt. Denn der Algorithmus läuft im Weiteren eigenständig ab, eine Bezugnahme auf die Inputdaten ist nach der ersten Berechnungsstufe nicht mehr erforderlich. Insofern mit den gespeicherten Werken keine Sekundärzwecke wie beispielsweise das KI-Training verfolgt werden und sich der Datenlebenszyklus infolgedessen ausschließlich an der Verwendung der Werke als Inputdaten ausrichtet, ist daher davon auszugehen, dass die Vervielfältigungen nach der Erfüllung ihres Zwecks innerhalb des KI-Systems zur Löschung vorgesehen sind. Die Speicherung urheberrechtlich geschützter Inputdaten kann damit vorübergehend im Sinne von § 44b UrhG sein.⁷⁵⁴

b) Wesentlicher und integraler Bestandteil eines technischen Verfahrens

Ohne Inputdaten kann das KI-System keinen Output generieren, also seinen Anwendungszweck nicht erfüllen. Es besteht daher auch die notwendige Kausalbeziehung zwischen den Vervielfältigungshandlungen und dem

753 *Dustmann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 44a Rn. 9; vgl. EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

754 Soll eine Sekundärverwertung der erfassten Werke stattfinden, richtet sich die urheberrechtliche Zulässigkeit der Werknutzung nach den Besonderheiten der jeweiligen Sekundärverwertungsart. Dies ist aber nicht Gegenstand der vorliegenden Untersuchung.

technischem Verfahren.⁷⁵⁵ Die Vervielfältigungshandlungen werden zudem allein durch das KI-System beziehungsweise vorgelagerte und unmittelbar mit dem KI-System verbundene technische Verfahren ausgeführt.⁷⁵⁶ Demzufolge ist die Speicherung urheberrechtlich geschützter Inputdaten auch integraler und wesentlicher Bestandteil eines technischen Verfahrens, der Anwendung eines KI-Systems.

c) Flüchtig oder begleitender Charakter der Vervielfältigungen

Die Vervielfältigungen müssen darüber hinaus flüchtig oder begleitend sein. Flüchtig sind die Vervielfältigungen dann, wenn ihre Speicherdauer derart auf das für das technische Verfahren notwendige Maß beschränkt ist, dass die Vervielfältigungen nach Erfüllung ihres Zwecks innerhalb des technischen Verfahrens automatisch gelöscht werden und deshalb von einer besonders kurzweiligen Vervielfältigung gesprochen werden kann.⁷⁵⁷

Werden die Inputdaten nicht zu Sekundärzwecken verwendet, wovon im Rahmen der vorliegenden Untersuchung auszugehen ist, spricht nichts dagegen, dass die Vervielfältigungen nach ihrer Vorverarbeitung und Eingabe in das KI-System, also nach Erfüllung ihres Zwecks innerhalb des technischen Verfahrens, automatisch wieder gelöscht werden. Die Vervielfältigungen, die bei der Speicherung urheberrechtlich geschützter Inputdaten angefertigt werden, können demzufolge flüchtigen Charakters im Sinne des § 44a UrhG sein.

Solange mit den urheberrechtlich geschützten Inputdaten keine Sekundärzwecke verwirklicht werden sollen, dienen die Vervielfältigungen auch keinem anderen, eigenständigen Zweck als der Anwendung des KI-Systems. Sie sind unselbstständig in den Anwendungsvorgang des KI-Systems eingebunden. Die Speicherung urheberrechtlich geschützter Inputdaten ist daher auch begleitend zur Anwendung des KI-Systems im Sinne von § 44a UrhG.⁷⁵⁸ Beide Tatbestandsalternativen sind daher erfüllt.

755 Vgl. zur Voraussetzung EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

756 Vgl. zur Voraussetzung EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

757 EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 64 – Infopaq; EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA; KG, Urt. v. 30.04.2004 - 5 U 98/02, GRUR-RR 2004, 228 (231) – Ausschnittdienst.

758 Vgl. Zur Voraussetzung EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

d) Rechtmäßige Nutzung

Weil Zweck der Vervielfältigungen nicht die Übertragung von Inhalten zwischen Dritten ist, müssen die Vervielfältigungen für ein Eingreifen des Schrankentatbestands gemäß § 44a Nr. 2 UrhG ausschließlich dazu dienen, die rechtmäßige Nutzung der Werke zu ermöglichen. Rechtmäßig sind solche Nutzungen, die vom Rechtsinhaber zugelassen oder nicht vom urheberrechtlichen Ausschließlichkeitsrecht erfasst und Dritten damit gestattet sind.⁷⁵⁹ Rechtmäßige Nutzung ist insbesondere der rezeptive Werkgenuss, der von vornherein nicht Teil der urheberrechtlichen Verwertungsrechte und für Nutzer damit frei möglich ist.⁷⁶⁰ Der freie Werkgenuss kann auch in einer Informationsentnahme münden. Denn die im Werk enthaltenen semantischen Informationen genießen grundsätzlich ebenfalls keinen urheberrechtlichen Schutz.⁷⁶¹ Damit ist auch die Analyse eines Werkes, beispielsweise zu wissenschaftlichen Zwecken, als rechtmäßige Nutzung von § 44a Nr. 2 UrhG umfasst.⁷⁶²

Zweck der Vervielfältigungen, die bei der Speicherung urheberrechtlich geschützter Inputdaten stattfinden, ist die Ermöglichung der Anwendung von KI-Systemen. Urheberrechtlich geschützte Inputdaten werden dabei nach einer Umwandlung in numerische Darstellungsformen als Grundlage für Berechnungen innerhalb der KI genutzt. Das Berechnungsergebnis wird als Output ausgegeben. Werke als Berechnungsgrundlage für Algorithmen zu verwenden, ist eine besondere Form der Analyse der im Werk enthaltenen Informationen. Die Art der Verwendung ist demzufolge nicht dem urheberrechtlichen Ausschließlichkeitsrecht zugewiesen.⁷⁶³ Das gilt für eine händische Berechnungen, aber auch für automatisierte Prozesse.⁷⁶⁴ Schließlich kommt es für die Berechnung sogar mitunter nicht auf die schöpferischen Elemente des Werkes an.⁷⁶⁵ Die algorithmische Verwendung von Werken als Inputdaten einer KI als solche ist, unabhängig von dabei angefertigten Vervielfältigungshandlungen, daher nicht vom urheberrechtlichen Ausschließlichkeitsrecht umfasst. Demzufolge dienen die

759 ErwGr. 33 InfoSoc-RL.

760 *Dustmann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 44a Rn. 17; vgl. beispielhaft EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 171 – Football Association Premier League.

761 EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 44 – Infopaq II.

762 Vgl. ErwGr. 9 S. 2 DSM-RL

763 Wie hier zum KI-Training auch *Hofmann*, WRP 2024, 11 Rn. 14.

764 Vgl. ErwGr. 9 S. 2 DSM-RL, außerdem BT-Drs. 18/12329, S. 40.

765 *Jager*, Artificial Creativity?, S. 341.

Vervielfältigungen urheberrechtlich geschützter Werke zur Verwendung als Inputdaten eines KI-Systems ausschließlich einer rechtmäßigen Nutzung im Sinne von § 44a Nr. 2 UrhG.

e) Keine eigenständige wirtschaftliche Bedeutung

Darüber hinaus dürfen die Vervielfältigungen der Werke gemäß § 44a UrhG auch keine eigenständige wirtschaftliche Bedeutung haben. Das ist der Fall, wenn die Vervielfältigungshandlungen einen wirtschaftlichen Vorteil generieren, der über den aus der rechtmäßigen Nutzung resultierenden Vorteil hinausgeht.⁷⁶⁶ Hiergegen spricht, wenn die Vervielfältigungshandlungen untrennbarere Bestandteile der rechtmäßigen Nutzung sind. Das ist beispielsweise dann der Fall, wenn der Nutzer auf die Vervielfältigungshandlungen schon deswegen keinen Einfluss nehmen kann, weil sie ihm nicht bewusst sind.⁷⁶⁷

aa) Eigene wirtschaftliche Bedeutung unmittelbar aus der Nutzung der Werke als Inputdaten

Der durchschnittliche Nutzer eines autonomen Fahrzeugs ist sich beispielsweise nicht darüber bewusst, dass das Fahrzeug mit Kamerasensoren am Wegesrand befindliche Werke der Baukunst vervielfältigt. Urheberrechtlich geschützte Inputdaten können also unter Umständen ohne Kenntnis des KI-Systemnutzers generiert werden. Sofern der Nutzer im Beispielfall nicht den Einsatz des autonomen Fahrzeugs insgesamt unterlassen will, kann er auch keinen Einfluss darauf nehmen, ob urheberrechtlich geschützte Werke bei der Anwendung des KI-Systems vervielfältigt werden. In Fallkonstellationen wie dieser sprechen demzufolge Gründe dafür, dass die vorübergehenden Vervielfältigungshandlungen untrennbare Bestandteile der Anwendung des KI-Systems sind. Dann käme ihnen keine eigene wirtschaftliche Bedeutung zu.

Bei der Anwendung anderer Arten von KI-Systemen könnte der Umgang mit urheberrechtlich geschützten Inhalten für Nutzer allerdings sichtbarer

766 EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 175 – Football Association Premier League.

767 Vgl. EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 176 – Football Association Premier League.

sein. Denkbar ist das insbesondere beim Einsatz von Webcrawlern zur Speicherung von Inputdaten. Diese laden aus frei zugänglichen Internetquellen mitunter große Mengen an Text- oder Bildinhalten herunter. Hierbei wird zumindest billigend in Kauf genommen, dass urheberrechtlich geschütztes Material mitervielfältigt wird. Unbewusst sind die Vervielfältigungen in diesen Fallkonstellationen damit nicht. Nichtsdestotrotz ist die Speicherung urheberrechtlich geschützter Inputdaten unselbstständig in die Anwendung des KI-Systems integriert. Die Speicherung der Inputdaten ist für die Anwendung der KI notwendig. Die Daten können nach der Umwandlung in mathematische Tensoren auch wieder gelöscht werden. Die Vervielfältigungen sind demzufolge auch in dieser Fallkonstellation elementare Bestandteile der Anwendung der KI-Systeme, also der rechtmäßigen Nutzungen. Da es keinen etablierten Markt für Inputdaten gibt, spart sich der KI-Nutzer durch die Speicherung der Werke auch keine eigenen Aufwendungen. Grundsätzlich bringt die Speicherung der geschützten Inputdaten abgesehen von der Ermöglichung der KI-Anwendung weder für KI-Nutzer noch für Entwickler oder Anbieter unmittelbar einen über die rechtmäßige Nutzung hinausgehenden wirtschaftlichen Vorteil.

bb) Mittelbare eigene wirtschaftliche Bedeutung aus der Möglichkeit zur Sekundärverwertung von Inputdaten

Allerdings können die Inputdaten auch zu Sekundärzwecken verwendet werden. Hierzu lassen sich die Vervielfältigungsstücke ohne größeren technischen Aufwand dauerhaft abspeichern. Die Vervielfältigungsstücke können dann insbesondere zum Training von KI genutzt werden. Verbreitet ist beispielsweise der Einsatz der Inputdaten zur Optimierung des KI-Systems selbst.⁷⁶⁸ Auch denkbar ist, dass die generierten Vervielfältigungsstücke zu Protokollierungs- und Beweissicherungszwecken archiviert werden. Solche Sekundärverwertungen finden häufig bei autonomen Fahrzeugen statt. Sie können auch wirtschaftlich relevant sein. Das gilt insbesondere für die Verwendung der Inputdaten als Trainingsmaterial.⁷⁶⁹ Findet eine Sekundärnutzung der generierten Inputdaten statt, kommt den Vervielfältigungshandlungen infolgedessen eine aus der Sekundärnutzung der Werke

768 Vgl. als Beispiele *Hoffmann/Vogt*, MMR-Aktuell 2020, 431466; *Shah*, AIwire v. 8.3.2023; *Stecklow/Cunningham/Jin*, Reuters v. 6.4.2023.

769 Dazu siehe oben unter 2. Teil D. I. 1. sowie zum Training I. Teil B. II. 3.

resultierende eigene wirtschaftliche Bedeutung zu. Sie geht über den wirtschaftlichen Vorteil aus der primären Anwendung des KI-Systems hinaus.

Fraglich ist, ob bereits die *Möglichkeit* der Sekundärverwertung der generierten Inputdaten und damit der Begründung einer eigenständigen wirtschaftlichen Bedeutung der Vervielfältigungen neben der rechtmäßigen Nutzung, also der Anwendung des KI-Systems, ausreicht, damit die Schrankenbestimmung aus § 44a UrhG nicht anwendbar ist.⁷⁷⁰

Dafür spricht zunächst, dass es für Urheber nur einen geringen Unterschied macht, ob der wirtschaftliche Vorteile aus der Vervielfältigung unmittelbar realisiert wird oder nur mittelbar realisierbar ist. Dass die in § 44a UrhG umgesetzte InfoSoc-RL (Art. 5 Abs. 1 InfoSoc-RL) ein hohes Schutzniveau für Urheber gewährleisten will,⁷⁷¹ spricht dafür, den Urhebern bereits Schutz vor einer mittelbaren, also drohenden Beeinträchtigung ihrer Verwertungsinteressen zu bieten. Das wird auch bei der funktionsbezogenen Auslegung des Vervielfältigungsrechts deutlich.⁷⁷²

Allerdings ist Art. 5 Abs. 1 InfoSoc-RL und damit auch § 44a UrhG richtlinienkonform so auszulegen, dass die praktische Wirksamkeit der Schrankenbestimmung erhalten bleibt.⁷⁷³ Die Erfassung von potenziellen Weiterverwertungsmöglichkeiten unter dem Tatbestandsmerkmal der eigenen wirtschaftlichen Bedeutung im Sinne des § 44a UrhG würde dazu führen, dass eine Anwendung der Schrankenbestimmung in den allermeisten Fällen ausscheidet. Denn vorübergehende Vervielfältigungen können selbst aus dem Cache (flüchtiger Zwischenspeicher) eines Computers geborgen werden. Damit würde potenziell eine eigenständige Verwertung der Vervielfältigungen ermöglicht.⁷⁷⁴ Wenn selbst diese Vervielfältigungen nicht mehr unter § 44a UrhG gefasst werden könnten, verbliebe kaum noch Anwendungsspielraum für den Erlaubnistatbestand. Die praktische Wirksamkeit der Schrankenbestimmung wäre damit erheblich beeinträchtigt. Darüber hinaus hatte der Richtlinienggeber auch neue Medien als Anwendungsfall für die Schrankenbestimmung vor Augen.⁷⁷⁵ Diese wären der weiten Lesart zur Folge aber zumeist ebenfalls nicht von § 44a UrhG umfasst.

770 Zu dieser Frage im Kontext des Streaming bereits *Stieper*, MMR 2012, 12.

771 ErwGr. ErwGr. 4, 9 InfoSoc-RL.

772 Siehe oben unter 3. Teil B. I. 1. a).

773 EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 163 – Football Association Premier League.

774 Dazu genauer *Stieper*, MMR 2012, 12 (16).

775 ErwGr. 31 InfoSoc-RL.

Weiterhin sind potenzielle Verwertungsmöglichkeiten alternative Sachverhalte, deren Verwirklichung vom Inhaber des Vervielfältigungsstücks abhängt. Sie bei der Bewertung der eigenen wirtschaftlichen Bedeutung einer Vervielfältigung im Rahmen des § 44a UrhG einzubeziehen, bedeutete daher, die urheberrechtliche Zulässigkeit einer Werknutzung auf Basis eines hypothetischen Sachverhalts zu bewerten, dessen Verwirklichung nicht sicher ist. Dies kann keine Basis für eine rechtliche Beurteilung sein. Erst der Eingriff in den bestehenden technischen Ablauf, also die Absonderung der generierten, urheberrechtlich geschützten Inputdaten zum Zwecke einer tatsächlich stattfindenden Sekundärverwertung, hat eine neue, eigenständige Nutzungsmöglichkeit zur Folge.⁷⁷⁶ Ohne diesen Eingriff ist die Verwertungsmöglichkeit zu abstrakt, als dass sie Beurteilungsgrundlage sein könnte.

Darüber hinaus entgehen den Urhebern auch erst durch eine tatsächliche Sekundärverwertung, beispielsweise zum Training von KI, Lizenzentnahmen.⁷⁷⁷

cc) Schlussfolgerungen

Alles in allem sprechen damit durchgreifende Erwägungen insbesondere aus der Auslegung des in § 44a UrhG umgesetzten Art. 5 Abs. 1 Info-Soc-RL für eine enge Auslegung des Tatbestandsmerkmals der eigenen wirtschaftlichen Bedeutung. Bloße entfernt-hypothetische, im technischen Prozess nicht angelegte Verwertungsmöglichkeiten genügen demzufolge nicht, damit Vervielfältigungen neben der rechtmäßigen Nutzung eine eigene wirtschaftliche Bedeutung zukommt.⁷⁷⁸ Für die Speicherung von urheberrechtlich geschützten Inputdaten bedeutet das, dass die Vervielfältigungen erst dann eine eigene wirtschaftliche Bedeutung erlangen, wenn eine Sekundärverwertung der Inputdaten beispielsweise zum Training von KI durch eine Absonderung der Inhalte hierfür tatsächlich vorbereitet wird. Die eigentliche Speicherung urheberrechtlich geschützter Inputdaten

⁷⁷⁶ Vgl. *Stieper*, MMR 2012, 12 (16).

⁷⁷⁷ Vgl. zur Beeinträchtigung der Verwertungsinteressen durch die Verwendung als Trainingsdaten 2. Teil B. I. 1.

⁷⁷⁸ *Eichelberger*, K&R 2012, 393 (397); *Stieper*, MMR 2012, 12 (16); anders wohl *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44a Rn. 9; *Loewenheim*, in: *Schricker/Loewenheim*, Urheberrecht, UrhG § 44a Rn. 13; *Fangerow/Schulz*, GRUR 2010, 677 (680); tendenziell weiter wohl auch *Wiebe*, in: *Spindler/Schuster*, Recht der elektronischen Medien, UrhG § 44a Rn. 6.

hat neben der Anwendung des KI-Systems daher keine eigenständige wirtschaftliche Bedeutung im Sinne des § 44a UrhG.

f) Zwischenergebnis

Insgesamt sind die bei der Speicherung von Inputdaten angefertigten vorübergehenden Vervielfältigungen urheberrechtlich geschützter Werke grundsätzlich⁷⁷⁹ von der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG gedeckt.⁷⁸⁰

g) Erfassung rechtswidrig verfügbarer Werke

Bei der typischerweise automatisiert ablaufenden Sammlung von Inputdaten insbesondere mit Hilfe automatisierter Webcrawler kann nicht zweifelsfrei zwischen rechtmäßig und unrechtmäßig zugänglich gemachten Werken unterschieden werden. Unklar ist vor diesem Hintergrund, ob die Freistellungswirkung des § 44a Nr. 2 UrhG auch dann eingreift, wenn – insbesondere mit Hilfe von Webcrawlern – rechtswidrig zugänglich gemachte Werke zur Verwendung als Inputdaten vorübergehend vervielfältigt werden. Das im zweiten Teil der Untersuchung identifizierte Freistellungsbedürfnis differenziert nicht zwischen diesen Werkgruppen. Vielmehr ist eine uneingeschränkte Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten interessengerecht.

Auch wenn die vorübergehende Vervielfältigung von rechtswidrig öffentlich zugänglich gemachten Werken als Inputdaten für die Anwendung von KI-Systemen einer rechtmäßigen Nutzung im Sinne des § 44a Nr. 2 UrhG dient,⁷⁸¹ könnte die Freistellung dieser Vervielfältigungshandlungen nicht mit dem im Rahmen von § 44a UrhG zu berücksichtigenden Dreistufentest aus Art. 5 Abs. 5 InfoSoc-RL vereinbar sein. Das ist der Fall, wenn die vorübergehenden Vervielfältigungen eine Verringerung rechtmäßiger

779 Zum möglichen Sonderfall, der Anwendung generativer KI-Systeme, siehe unten unter 3. Teil C. II. 4.

780 So zum Web Scraping von Werken als Inputdaten für KI-Textgeneratoren im journalistischen Kontext auch *Gräfe/Kahl*, MMR 2021, 121 (124); die Möglichkeit sehen auch *Nägele/Apel*, in: Kaulartz/Braegelmann, *Rechtshandbuch Artificial Intelligence und Machine Learning*, Kap. 7.1 (KI und Urheberrecht) Rn. 32.

781 Dazu siehe unmittelbar oben.

Transaktionen zur Folge haben, die die betroffenen Urheber unzumutbar beeinträchtigen.⁷⁸² Da es keinen etablierten Markt gibt, auf dem im erheblichen Umfang Transaktionen über Inputdaten getätigt werden können,⁷⁸³ kommt eine Verringerung solcher rechtmäßiger Transaktionen aber bereits von vornherein kaum in Betracht. Darüber hinaus findet die Speicherung von Inputdaten in bestimmten Anwendungsszenarien automatisiert und in großem Umfang statt. In diesen Fällen ist es dem Datenverarbeiter nicht möglich, im Einzelnen die Rechtmäßigkeit der Zugänglichmachung der gespeicherten Werke zu überprüfen. Im Unterschied zu Fällen wie dem Linksetzen auf urheberrechtsverletzend öffentlich zugänglich gemachte Werke erfolgt auch keine weitere Verbreitung der rechtswidrigen Inhalte. Die Werknutzung beschränkt sich vielmehr auf eine technische Verarbeitung der Werke. Aus alledem folgt, dass der Dreistufentest gemäß Art. 5 Abs. 5 InfoSoc-RL im konkreten Fall nicht verletzt ist. Seine Anwendung führt demzufolge nicht zu einer einschränkenden Auslegung der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen, nach der nur rechtmäßig zugänglich gemachte Inhalte für die Speicherung von Werken als KI-Inputdaten verwendet werden dürfen. Auch urheberrechtswidrig öffentlich verfügbar gemachte Werke dürfen daher als Inputdaten für KI gemäß § 44a Nr. 2 UrhG vorübergehend vervielfältigt werden.⁷⁸⁴

2. Weitere Schrankenbestimmungen

Die Speicherung urheberrechtlich geschützter Inputdaten könnte daneben noch in den Anwendungsbereich weiterer Schrankenbestimmungen fallen. Hiervon sind aber keine umfangreicheren Freistellungswirkungen zu erwarten. Deswegen sollen die übrigen Schrankenbestimmungen nur Überblicksartig dargestellt werden.

a) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)

Die Speicherung urheberrechtlich geschützter Inputdaten könnte insbesondere auch von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfasst sein.

782 EuGH, Urt. v. 26.04.2017 - C-527/15, GRUR 2017, 610 Rn. 70 – Filmspeler.

783 Dazu bereits siehe oben unter 2. Teil D. I. 1.

784 A.A. Käde, Kreative Maschinen und Urheberrecht, S. 84 f.

aa) Anwendbarkeit der Schrankenbestimmung

Damit § 44b UrhG auch auf die Verarbeitung von urheberrechtlich geschützten Inputdaten Anwendung finden kann, muss auch die *Anwendung* einer KI unter den Text und Data Mining Begriff im Sinne des § 44b Abs. 1 UrhG subsumiert werden können. Text und Data Mining beschreibt gemäß § 44b Abs. 1 UrhG die automatisierte Analyse von einzelnen digitalen oder digitalisierten Werken, um „daraus“ Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen. Nach einer richtlinienkonformen Auslegung des Tatbestandsmerkmals genügt es hierbei, wenn die Informationen nicht unmittelbar *aus* dem analysierten Werk, sondern *mit dessen Hilfe* gewonnen werden. Notwendig, aber ausreichend ist demnach eine Kausalbeziehung zwischen der automatisierten Analyse eines Werkes und der Speicherung von Informationen.⁷⁸⁵

Bei der Verwendung von Werken als Inputdaten werden die Werke zunächst in numerische Tensoren umgewandelt. Dann können sie als Berechnungsgrundlage des Algorithmus verwendet werden. Dabei werden die einzelnen Bestandteile der Werke extrahiert und als Anknüpfungspunkte für die beim Training ermittelten Muster und Zusammenhänge verwendet. Auf ihrer Grundlage können statistische Vorhersagen getroffen werden. Sie sind im Output des KI-Systems repräsentiert. Die eingegebenen Werke werden also im Lichte der bereits ermittelten Muster und Zusammenhänge ausgewertet. In der Anwendungsphase eines KI-Systems findet damit eine automatisierte Analyse der Inputdaten im Sinne von § 44b UrhG statt.⁷⁸⁶ Ziel der automatisierten Analyse der Inputdaten ist die Generierung von Output. Dieser kann beispielsweise in Form von Klassifikationen, Vorhersagen oder auch Kreativinhalten⁷⁸⁷ ausgestaltet sein. Ziel des Analyseprozesses ist daher die Gewinnung einer Information durch den Output des KI-Systems.⁷⁸⁸ Bezugspunkt der Informationsgewinnung ist demzufolge auch hier

785 Dazu siehe oben unter 3. Teil A. III. 2. i) ee) (3) bb).

786 Zum Begriff der Analyse siehe *Kleinkopf*, Text- und Data-Mining, S. 30. Zum Tatbestandsmerkmal der Automatisierung vgl. *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. II.

787 Zur Anwendung der hier ermittelten Grundsätze auf generative KI-Systeme siehe unten 3. Teil C. II. 4. Zum Informationsgewinnungsbegriff und dessen Reichweite insbesondere im Kontext der Erzeugung fiktiver (Kreativ-)inhalte bereits oben unter 3. Teil A. III. 2. c).

788 So auch *Jager*, Artificial Creativity?, S. 351; a. A. *Schippan*, ZUM 2024, 670 (675), nach dem ein Output selbst dann keine gewonnene Information im Sinne der Bestimmung verkörpert, wenn er sich in einer Musteranalyse erschöpft.

die Erzeugung des Outputs durch das KI-System. Die Output-Informationen werden dabei durch, also mit Hilfe der automatisierten Analyse der Inputdaten generiert. Infolgedessen besteht auch die notwendige Kausalbeziehung zwischen der Analyse der urheberrechtlich geschützten Inputdaten und der Informationsgewinnung. Mithin ist auch die Verarbeitung von Werken als Inputdaten eines KI-Systems Text und Data Mining im Sinne des § 44b Abs. 1 UrhG.

Die Speicherung, also Vervielfältigung urheberrechtlich geschützter Inputdaten ist infolgedessen grundsätzlich⁷⁸⁹ auch von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 UrhG freigestellt.⁷⁹⁰ Da die Inhalte nach Eingabe in das KI-System gelöscht werden können, ist hierfür die Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG kein Hindernis. Auch bei der Vervielfältigung der Werke als Inputdaten ist jedoch ein gemäß § 44b Abs. 3 UrhG erklärter Nutzungsvorbehalt zu beachten.

bb) Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen an dritte Datenverarbeiter

Auch im Vorfeld der Speicherung urheberrechtlich geschützter Inputdaten können Zugangsinhaber Werkzugänge an Dritte weitergeben oder diesen Dritten Vorlagen zur weiteren Vervielfältigung von Werken zur Verfügung stellen, die sie auf Grundlage der Werkzugänge angefertigt haben. Die Dritten stellen im Anschluss hieran die eigenen Vervielfältigungsstücke der Werke her, die als Inputdaten verwendet werden sollen. Denkbar sind solche Fälle beispielsweise im Rahmen einer KI-gestützten Analyse von Werken im Auftrag eines Zugangsinhabers. Hierfür vermittelt dieser dem die Analyse ausführenden Datenverarbeiter Zugang zu den zu analysierenden Werken. Hiermit konkretisiert er den Analysegegenstand.

Auch in solchen Fällen stellt sich die Frage, inwieweit vom originären Zugangsinhaber begünstigte Dritte, die die Werke als Inputdaten vervielfältigen, rechtmäßigen Zugang zu den vom Zugangsinhaber vermittelten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben. Allgemein kann dabei

789 Zum möglichen Sonderfall, der Anwendung generativer KI-Systeme, siehe unten 3. Teil C. II. 4.

790 In diesem Sinne wohl auch *Jager*, *Artificial Creativity?*, S. 351; in einem engeren Kontext außerdem *Hofmann*, ZUM 2024, 166 (174), wobei in Einzelfällen eine Anwendung der Schrankenbestimmung abgelehnt wird; a. A. hingegen *Schippman*, ZUM 2024, 670 (675); wohl auch *Maamar*, ZUM 2023, 481 (487).

zwischen der Vermittlung des Zugangs selbst, beispielsweise durch die Weitergabe von Login-Daten zu einer geschützten Datenbank, und der Bereitstellung von seitens des Zugangsinhabers angefertigten Vervielfältigungsvorlagen unterschieden werden, auf deren Grundlage der Datenverarbeiter im Folgenden neue Werkstücke anfertigt. Die Vervielfältigung der urheberrechtlich geschützten Inputdaten kann dabei als Hilfstätigkeit für den Zugangsinhaber oder im Eigeninteressen des Datenverarbeiters ausgeführt werden.⁷⁹¹

Da die sich hieraus ergebenden vier Fallkonstellationen (Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen der Werke sowie jeweils im Rahmen einer Hilfstätigkeit oder im Eigeninteresse des Datenverarbeiters) den Weitergabe- beziehungsweise Vermittlungsszenarien bei der Speicherung von Werken als Trainingsdaten entsprechen, können die hierfür gewonnenen Erkenntnisse auf die vorliegenden Fallkonstellationen übertragen werden. Demzufolge ist die Tatbestandsvoraussetzung der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG so auszulegen, dass die Zugangsvermittlung für eine Speicherung von urheberrechtlich geschützten Inputdaten im Interesse des Zugangsinhabers, also als Hilfstätigkeit für diesen, vom Begriff der rechtmäßigen Zugänglichkeit mitumfasst ist. Datenverarbeiter, denen zur Herstellung von Vervielfältigungsstücken im Interesse des Zugangsinhabers Zugang zu Werken vermittelt wurde, haben demzufolge rechtmäßigen Zugang zu diesen geschützten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG.⁷⁹² Sie müssen die aufgestellten Zugangsvoraussetzungen nicht in ihrer eigenen Person erfüllen. Insbesondere müssen die also keine eigenen Zugangsdaten erwerben und nicht selbst ein Zugangsabonnement abschließen.⁷⁹³ Werden dem Datenverarbeiter statt Zugangsinformationen Vervielfältigungsvorlagen zur Verfügung gestellt, die durch den Zugangsinhaber auf Basis seiner Werkzugänge angefertigt wurden, hat der Datenverarbeiter bei der Herstellung seiner eigenen Vervielfältigungsstücke ebenfalls rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG.⁷⁹⁴ Auf das Verhältnis zwischen Zugangsinhaber und Rechteinhaber sowie den Ursprung der verwendeten Vervielfältigungsvorlagen kommt es nicht an.⁷⁹⁵ Das Tatbestandsmerkmal der rechtmäßigen Zugänglichkeit ist für den Verarbeiter der Inputdaten bereits deswegen er-

791 Vgl. zu den Fallgruppen ausführlicher oben unter 3. Teil A. III. 2. f) bb).

792 Vgl. oben unter 3. Teil A. III. 2. f) bb) (1).

793 Vgl. oben unter 3. Teil A. III. 2. f) bb) (2).

794 Vgl. oben unter 3. Teil A. III. 2. f) bb) (2).

795 Vgl. oben unter 3. Teil A. III. 2. f) bb) (2).

füllt, weil er selbst beim Zugriff auf die genutzten Vervielfältigungsvorlagen des Zugangsinhabers keine Zugangsschranken überwinden muss.

Anders sind jedoch die Fallkonstellationen zu beurteilen, in denen die Werke nicht im Interesse des Zugangsinhabers, sondern im Eigeninteresse des Datenverarbeiters vervielfältigt werden. Aufgrund einer Anwendung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL können die Datenverarbeiter für ihre Vervielfältigungshandlungen bei der Speicherung der urheberrechtlich geschützten Inputdaten dann nicht von der Schrankenbestimmung aus § 44b UrhG profitieren, wenn für sie objektiv erkennbar ist, dass die Werke einer zugangsgeschützten Quelle entstammen.⁷⁹⁶ Das ist stets der Fall, wenn einem Inputdatensammler der Zugang zu einer geschützten Quelle weitervermittelt wird, beispielsweise durch die Weitergabe von Login-Daten zu einer Datenbank.⁷⁹⁷ Eine Freistellung der Vervielfältigungshandlungen des Datenverarbeiters über § 44b UrhG kann dann nicht stattfinden. Eine Ausnahme hiervon muss nur dann gemacht werden, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, den Werkzugang an dritte Datenverarbeiter zum Zwecke des Text und Data Minings weiterzugeben.⁷⁹⁸ Werden dem im Eigeninteresse tätigen Datenverarbeiter durch den Zugangsinhaber Vervielfältigungsvorlagen für die Herstellung von Werkstücken zur Verfügung gestellt, kommt es für eine Freistellung seiner Vervielfältigungshandlungen nach den vorgenannten Grundsätzen auf die Umstände des konkreten Einzelfalls an.⁷⁹⁹ Eine Subsumtion seiner Vervielfältigungshandlungen zur Speicherung von urheberrechtlich geschützten Inputdaten unter § 44b UrhG ist aber grundsätzlich möglich.

Alles in allem können von Zugangsinhabern begünstigte Inputdatenverarbeiter wie beispielsweise Betreiber KI-gestützter Analysesysteme rechtmäßigen Zugang zu den vermittelten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben. Dabei kommt es darauf an, unter welche der vier herausgearbeiteten Fallgruppen die konkrete Sachverhaltskonstellation gefasst werden kann.

796 Vgl. oben unter 3. Teil A. III. 2. f) bb) (3).

797 Vgl. oben unter 3. Teil A. III. 2. f) bb) (iv).

798 Vgl. dazu ausführlich oben unter 3. Teil A. III. 2. f) bb) (3).

799 Vgl. oben unter 3. Teil A. III. 2. f) bb) (3).

cc) Fernwirkung des Nutzungsvorbehalts bei Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen an dritte Datenverarbeiter

Bei der durch Zugangsinhaber initiierten Vermittlung von Werkzügen an dritte Datenverarbeiter, beispielsweise durch eine Weitergabe von Login-Daten zu einer Datenbank, oder der Bereitstellung von auf Grundlage der Zugänge angefertigten Werkstücken als Vorlagen für weitere Vervielfältigungshandlungen durch Dritte stellt sich im Übrigen die Frage, ob und inwieweit ein gegenüber dem Zugangsinhaber erklärter Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG auch die Freistellung der Vervielfältigungshandlungen des dritten Datenverarbeiters beendet. Technisch ist dabei zwischen der echten Zugangsvermittlung, also insbesondere durch die Weitergabe von Login-Daten zu einer Datenbank, und der Anfertigung von Vervielfältigungen durch den Zugangsinhaber zu unterscheiden, die dem Datenverarbeiter als Vorlage für eigene Vervielfältigungshandlungen zur Verfügung gestellt werden. In beiden Fällen kann, muss der Datenverarbeiter keine Kenntnis vom erklärten Nutzungsvorbehalt haben. Dies ist vom jeweiligen Einzelfall und der konkreten Umsetzung der Vorbehaltserklärung abhängig.

Maßgeblich für die Beurteilung der Fernwirkung des Nutzungsvorbehalts ist die wiedergabebezogene Wirkung der Vorbehaltserklärung.⁸⁰⁰ Bei der Zurverfügungstellung von Vervielfältigungsvorlagen für den Datenverarbeiter führt sie dazu, dass gegenüber dem Zugangsinhaber erklärte Nutzungsvorbehalte, beispielsweise in den Nutzungsbedingungen einer Webseite, keine Wirkung zulasten des Datenverarbeiters entfalten, wenn die Vorbehaltserklärung nicht mit den vom Zugangsinhaber angefertigten Vervielfältigungsvorlagen verbunden ist. Denn nur diese Wiedergaben der betreffenden Werke werden vom Datenverarbeiter verwendet. Deswegen kann der Datenverarbeiter trotz eines ursprünglich erklärten Nutzungsvorbehalts für seine Vervielfältigungshandlungen von der Freistellungswirkung des § 44b Abs. 2 UrhG profitieren.⁸⁰¹ Hiergegen spricht auch nicht die Anwendung des Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL.⁸⁰² Einzelfallabhängig kann der Datenverarbeiter aber als Teilnehmer für die Urheberrechtsverletzung des Zugangsinhabers mitverantwortlich sein.

800 Dazu siehe oben unter 3. Teil A. III. 2. i) ff).

801 Dazu ausführlich siehe oben unter 3. Teil A. III. 2. i) ee) (3) hh) (10) (i).

802 Zur Anwendung des Dreistufentests oben ausführlich unter 3. Teil A. III. 2. i) ee) (3) hh) (10) (i).

Wird dem Datenverarbeiter lediglich der Zugang zu den Werken vermittelt, die vervielfältigt werden, so greift er für seine Vervielfältigungshandlungen auf die originären Werkstücke beziehungsweise öffentlichen Wiedergaben der Werke zurück. An diesen ist der Nutzungsvorbehalt erklärt worden. Dies führt dazu, dass auch der Datenverarbeiter bei der Vervielfältigung der vermittelten Inhalte die erklärten Nutzungsvorbehalte beachten muss. Ist eine wirksame Vorbehaltserklärung abgegeben, erlischt demzufolge auch für ihn die Schrankenwirkung des § 44b Abs. 2 UrhG.⁸⁰³ Im Ergebnis muss sich der Datenverarbeiter bei der Vermittlung eines Zugangs zu Werken infolgedessen selbst vergewissern, dass beispielsweise in den Nutzungsbedingungen keine wirksamen Nutzungsvorbehalte erklärt worden sind.⁸⁰⁴

b) Anwendungsspezifische Schrankenbestimmungen

Abhängig von der konkreten Fallkonstellation könnte die Speicherung urheberrechtlich geschützter Inputdaten für die Anwendung von KI-Systemen zudem insbesondere noch von den Schrankenbestimmungen aus §§ 48 (öffentliche Reden), 53 (Privatkopie), 59 (Werke an öffentlich verfügbaren Plätzen), 61 (verwaiste Werke) oder 61d UrhG (nicht verfügbare Werke) erfasst sein. Da diese Schrankenbestimmungen aber keine umfassende, fallgruppenunabhängige Freistellung der Speicherung von Werken als Inputdaten zur Folge haben, werden die Regelungen hier nicht näher beleuchtet.

3. Verhältnis der Schrankenbestimmungen aus § 44a Nr. 2 UrhG und § 44b UrhG

Die Speicherung, also Vervielfältigung von Werken als Inputdaten für die Anwendung von KI-Systemen wird im Grundsatz sowohl von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG als auch von der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG urheberrechtlich freigestellt. Die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining kann jedoch mit Hilfe

803 Dazu ausführlich siehe oben unter 3. Teil A. III. 2. i) ee) (3) hh) (10) (ii).

804 Vgl. zur Abgabe von Vorbehaltserklärungen in Nutzungsbedingungen siehe 3. Teil A. III. 2. i) aa).

eines Nutzungsvorbehalts im Sinne des § 44b Abs. 3 UrhG beendet werden. Fraglich ist vor diesem Hintergrund, ob die Erklärung des Nutzungsvorbehalts auch Auswirkungen auf die Freistellung der Vervielfältigungshandlungen zur Verwendung von Werken als KI-Inputdaten über § 44a Nr. 2 UrhG hat.

Denkbar wäre, dass die Erklärung des Nutzungsvorbehalts auf einer dinglichen Ebene auch die Schrankenwirkung des § 44a Nr. 2 UrhG beendet. Dafür spricht, dass der Nutzungsvorbehalt gegen die Vervielfältigung von Werken als Inputdaten praktisch leerlaufen würde, wenn eine Freistellung der im Rahmen des § 44b UrhG vorbehaltenen Nutzungshandlung auch über § 44a Nr. 2 UrhG erfolgen könnte. Allerdings beinhaltet ErwGr. 18 UAbs. 2 S. 3 DSM-RL eine Klarstellung, nach der sich die quasingliche Beendigungswirkung des Nutzungsvorbehalts nur auf die Schrankenbestimmung für Text und Data Mining erstreckt.⁸⁰⁵ Ohnehin kann der Urheber die Nutzung seines Werkes dinglich nicht begrenzen, wenn die Nutzungshandlung zu keinem Zeitpunkt seiner Rechtsposition zugeordnet war.⁸⁰⁶ Schrankenbestimmungen führen jedoch dazu, dass dem Urheber von vornherein nur ein beschränktes Recht gewährt wird.⁸⁰⁷ Sie gestalten den Umfang des Urheberrechts von Anfang an aus.⁸⁰⁸ Es findet keine nachträgliche Einschränkung einer zunächst umfassend gewährten Rechtsposition statt, wie es bei der Gewährung eines ausschließlichen Nutzungsrechts durch den Urheber der Fall ist. Weil die Schrankenbestimmungen deswegen nicht mit vertraglichen Nutzungsrechtseinräumungen des Urhebers vergleichbar sind, ist auch § 31 Abs. 1 S. 2 UrhG auf gesetzliche Erlaubnisse nicht anwendbar.⁸⁰⁹ Die Regelung sieht eine inhaltliche Beschränkbarkeit vertraglicher Nutzungsrechte vor. Auch eine analoge Anwendung scheidet

805 Dazu bereits ausführlicher mit einer abweichenden Ansicht zum Verständnis des Erwägungsgrunds unter 3. Teil A. III. 2. i) dd).

806 *Gräbig*, Abdingbarkeit und vertragliche Beschränkungen urheberrechtlicher Schranken, S. 59; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 204.

807 *Zech*, in: *Leible/Ohly/Zech*, Wissen - Märkte - Geistiges Eigentum, S. 187 (192); *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 129 ff.

808 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 129 ff.; a. A. *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236.

809 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 200 ff.; a. A. *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236 f.

aus, weil die Rechtsfragen mit Hilfe einer teleologischen Auslegung der urheberrechtlichen Schrankentatbestände beantwortet werden können. Damit fehlt es an einer Regelungslücke.⁸¹⁰ Zudem geht das UrhG gemäß § 32 Abs. 1 S. 2 UrhG davon aus, dass eine rechtsgeschäftliche Lizenz im Regelfall mit der Gewährung einer angemessenen Vergütung einhergeht. Im Gegensatz dazu sind die meisten gesetzlichen Schrankenbestimmungen vergütungsfrei ausgestaltet. Auch daran lässt sich erkennen, dass das Urheberrecht beiden Regelungskomplexen eine jeweils eigene Regelungskonzeption zugrunde legt. Das zeigt, dass zwischen den urheberrechtlichen Schrankentatbeständen und den Regelungen zu rechtsgeschäftlichen Nutzungsrechten grundsätzlich auch keine vergleichbare Interessenlage besteht. Im Einzelfall kann das mitunter anders beurteilt werden. Insbesondere gewährt das Urheberrecht aber eine absolute, quasi-dingliche Rechtsposition.⁸¹¹ Damit unterliegt es auch einem Typenzwang.⁸¹² Eine dingliche Beendigung der Schrankenwirkung kraft Disposition des Urhebers ist mit diesem Typenzwang grundsätzlich nicht vereinbar. Alles in allem folgt hieraus, dass eine quasi-dinglich wirkende Ausgestaltung des urheberrechtlichen Ausschließlichkeitsrechts durch Vertrag oder einseitige Erklärung des Rechtsinhabers insbesondere im Bereich der Schrankenbestimmungen grundsätzlich nicht möglich ist.⁸¹³ Dieser Rechtsgedanke spiegelt sich auch in § 137 S. 1 BGB wider.⁸¹⁴

Eine Ausnahme hiervon kann nur dann gemacht werden, wenn die quasi-dingliche Rechtsposition im UrhG ausdrücklich zugunsten einer fakultativen Dispositionsbefugnis geöffnet wird. Dies ist beispielsweise in

810 A. A. aber *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236 f.

811 *Loewenheim*, in: Schricker/Loewenheim, Urheberrecht, Einleitung zum UrhG Rn. 25; *Wolf/Wellenhofer*, Sachenrecht § 2 Rn. 14.

812 *Stieper*, in: Schricker/Loewenheim, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 56; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 204; *Hofmann*, UFITA 2014, 381 (388).

813 *Stieper*, in: Schricker/Loewenheim, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 56; *Lüft*, in: Wandtke/Bullinger, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 3; *Zech*, in: Leible/Ohly/Zech, Wissen - Märkte - Geistiges Eigentum, S. 187 (192); *Gräbig*, Abdingbarkeit und vertragliche Beschränkungen urheberrechtlicher Schranken, S. 59; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 204 f.; sich anschließend hieran *Hofmann*, UFITA 2014, 381 (388); vgl. zum allgemeinen Sachenrecht und dem Typenzwang *Wolf/Wellenhofer*, Sachenrecht § 3 Rn. 3; a. A. *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236 f.

814 *Hofmann*, UFITA 2014, 381 (388 f.).

§ 44b Abs. 3 UrhG und § 49 Abs. 1 S. 1 UrhG der Fall. Eine entsprechende Öffnungsklausel sieht § 44a UrhG jedoch nicht vor. Demzufolge kann die Freistellungswirkung der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen nicht fakultativ durch den Rechtsinhaber eingeschränkt werden. Ein gemäß § 44b Abs. 3 UrhG erklärter Nutzungsvorbehalt beendet demzufolge nicht quasi-dinglich die Wirksamkeit der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG. Vielmehr sind die Vervielfältigungshandlungen der Inputdatensammler auch dann gemäß § 44a Nr. 2 UrhG urheberrechtlich freigestellt, wenn ein wirksamer Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist.

4. Anwendung der Schrankenbestimmungen auf generative KI-Systeme

Fraglich ist jedoch, ob die herausgearbeiteten Ergebnisse zur Anwendung insbesondere der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a Nr. 2 UrhG) sowie der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) auch auf die Speicherung von Werken als Inputdaten für generative KI-Systeme übertragen werden können. Hierbei handelt es sich um KI-Systeme, die Kreativinhalte wie Texte oder Bilder erzeugen.

Einer parallelen Behandlung der Speicherung von Werken als Inputdaten für generative KI-Systeme würde entgegenstehen, wenn technische Unterschiede in der Verarbeitung der Inputdaten, die spezifische Art des von generativen KI-Systemen erzeugten Outputs oder eine Anwendung des urheberrechtlichen Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL eine abweichende Bewertung der vorgenommenen Vervielfältigungshandlungen notwendig machen.

a) Technische Perspektive auf die Verarbeitung als Inputdaten

Aus technischer Perspektive werden auch Inputdaten generativer KI-Systeme als Berechnungsgrundlage eines Algorithmus verwendet. Er besteht auch bei generativen KI-Systemen aus mathematischen Funktionen, die aus den Inputdaten abgeleitete numerische Aktivierungswerte verarbeiten. Hierfür werden die Inputdaten ebenfalls vorverarbeitet und insbesondere in numerische Tensoren umgewandelt.

Aus technischer Perspektive besteht demzufolge kein Unterschied zwischen der Verarbeitung von Werken als Inputdaten sonstiger sowie generativer KI-Systeme. Bei der Verwendung von Werken als Inputdaten generativer KI-Systeme werden die Werke daher insbesondere auch rechtmäßig im Sinne des § 44a Nr. 2 UrhG als Berechnungsgrundlage genutzt. Die technische Perspektive auf den Verarbeitungsprozess auf Eingabeseite rechtfertigt daher keine abweichende Beurteilung.

b) Erzeugung von Kreativinhalten im Output

Die Erzeugung gegebenenfalls fiktiver Kreativinhalte im Output generativer KI-Systeme könnte zu Auswirkungen darauf haben, ob eine Informationsgewinnung im Sinne von § 44b Abs. 1 UrhG stattfindet. Dies ist für die Subsumtion der KI-Anwendung unter den Text und Data Mining-Begriff der Schrankenbestimmung notwendig. Bereits an anderer Stelle konnte allerdings gezeigt werden, dass § 44b Abs. 1 UrhG nicht nach der Art der erzeugten Informationen unterscheidet. Insbesondere auch die Generierung von Kreativinhalten wie Bildern oder Texten ist als Informationsgewinnung erfasst. Auch eine Fiktion der gewonnenen Informationen ist unschädlich.⁸¹⁵ Die spezielle Art des Outputs generativer KI-Systeme hat damit keinen Einfluss darauf, dass die tatbestandlich notwendige Informationsgewinnung im Sinne von § 44b Abs. 1 UrhG stattfindet.

Zum anderen ist aber auch fraglich, ob die Erzeugung möglicherweise urheberrechtsverletzenden Outputs durch das generative KI-System schon die Freistellung der Vervielfältigung von Werken als Inputdaten über die Schrankenbestimmungen aus §§ 44a, 44b UrhG beenden muss.

Hiergegen spricht allerdings, dass das Urheberrecht als techniksensibles Recht traditionell zwischen verschiedenen technischen Verarbeitungsschritten differenziert und diese jeweils einer konkreten, individuellen Betrachtung unterzieht. Aus diesem Grunde beurteilt sich beispielsweise die Vervielfältigung im Arbeitsspeicher eines Computers eigenständig und bedarf der Anwendung einer Schrankenbestimmung (§ 44a Nr. 2 UrhG), auch wenn mit Hilfe dieser Vervielfältigung nur der private, urheberrechtlich freie Werkgenuss bezweckt wird. In gleicher Weise bedarf die Übertragung in einem Netz zwischen Dritten durch einen Vermittler einer gesonderten Schrankenbestimmung (§ 44a Nr. 1 UrhG), auch wenn der soziale Sinnge-

815 Dazu bereits oben unter 3. Teil A. III. 2. c).

halt der Fallkonstellation durch die Nutzung des Übermittlungsdienstes seitens eines Endnutzers und dessen urheberrechtskonformen oder rechtsverletzenden Verhaltens bestimmt wird. Die Generierung von möglicherweise urheberrechtsverletzenden Outputs ist ebenfalls eine eigene Vervielfältigungshandlung.⁸¹⁶ Sie ist daher im Lichte der dargestellten Regelungskonzeption ebenfalls eigenständig zu beurteilen.

§ 44a Nr. 1 UrhG, der die Werknutzung des Vermittlers unabhängig davon privilegiert, ob die Nutzer seines Übertragungsdienstes mitunter eine Urheberrechtsverletzung begangen haben, bringt darüber hinaus zum Ausdruck, dass die Rechtmäßigkeit einer Vorbereitungshandlung nicht zwingend das Schicksal der nachgelagerten Nutzung teilt. Ein entsprechender Mechanismus besteht im Urheberrecht nicht. Das ist auch bei der Anwendung generativer KI-Systeme und der urheberrechtlichen Bewertung einer Speicherung von Werken als Inputdaten – einer Vorbereitungshandlung – zu berücksichtigen. Die Rechtmäßigkeit der Speicherung von Werken als Inputdaten ist damit von der Rechtmäßigkeit der späteren Output-Generierung zu trennen. Umgekehrt bedeutet dies aber auch, dass ein urheberrechtsverletzender und sanktionsfähiger Output auch dann erzeugt werden kann, wenn die Verarbeitung des Werkes auf Inputebene rechtmäßig gewesen ist. Die Erzeugung des Outputs ist als eigenständige Nutzungshandlung eigenständig zu beurteilen.

Schließlich ist die Erzeugung konkreter Urheberrechtsverletzungen durch generative KI-Systeme einzelfallabhängig, sodass die urheberrechtliche Bewertung ebenfalls auf die jeweiligen Einzelfälle bezogen sein sollte.⁸¹⁷ Beispielsweise ist die Erzeugung von Abstracts mittels generativer KI-Systeme in gewissen Grenzen urheberrechtlich frei möglich, wenn im Abstract keine schutzbegründenden Elemente des Werks identifizierbar sind. Auch die Tatsache, dass im Output generativer KI-Systeme Kreativinhalte und mitunter vereinzelt sogar Urheberrechtsverletzungen erzeugt werden, verlangt deswegen keine abweichende Beurteilung.

c) Dreistufentest

Klärungsbedürftig bleibt damit, ob die Anwendung des Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL – gegebenenfalls in Verbindung mit Art. 7 Abs. 2 DSM-RL – einer Anwendung der einschlägigen Schrankenbestimmungen

816 Dazu bereits ausführlicher oben unter 3. Teil B. IV. 3.

817 Vgl. dazu bereits oben unter 2. Teil B. I. 2. d) aa).

auf die Speicherung von Werken als Inputdaten *generativer* KI-Systeme entgegensteht. Danach darf die Schrankenbestimmung nur in bestimmten Sonderfällen angewandt werden, in denen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt und die berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden. Letztendlich kommt es hierbei auf eine umfassende Abwägung von Urheber- und Nutzerinteressen an.⁸¹⁸

Im Rahmen dieser ist im Kontext der Speicherung von Werken als Inputdaten generativer KI-Systeme insbesondere zu berücksichtigen, dass im Output der KI-Systeme möglicherweise Urheberrechtsverletzungen erzeugt werden. Zudem besteht auch unabhängig von konkreten Rechtsverletzungen die Gefahr, dass Konkurrenzprodukte zu den als Inputdaten verwendeten Werken erstellt werden.

Beide Aspekte sind jedoch schon umfassend bei einer Abwägung von Urheber- und Nutzerinteressen im Kontext der Freistellung des Web Scrapings von Werken als Trainingsdaten berücksichtigt worden.⁸¹⁹ Sie konnten entkräftet werden und haben im Ergebnis nicht zu einem Überwiegen der Urheberinteressen geführt. Aus diesem Grund steht der Dreistufentest auch keiner Freistellung des Web Scrapings von Werken als Trainingsdaten entgegen.⁸²⁰ Die Interessenabwägung in der vorliegenden Fallkonstellation ist vergleichbar, sodass auch hier aus den oben genannten Gründen kein Überwiegen der Urheberinteressen festgestellt werden kann. Es kann deswegen angenommen werden, dass der Dreistufentest auch bei der Speicherung von Werken als Inputdaten für generative KI-Systeme nicht verletzt ist. Auch das rechtfertigt daher keine abweichende Beurteilung.

d) Schlussfolgerungen

Alles in allem sind keine Gründe dafür ersichtlich, die herausgearbeiteten Ergebnisse nicht auf die Speicherung von Werken als Inputdaten für den Einsatz generativer KI-Systeme anzuwenden. Auch diese Verwendungsform ist deswegen – sofern sie vorübergehend erfolgt und die weiteren Tatbestandsvoraussetzungen erfüllt sind – ebenfalls von den Schrankenbestimmungen aus § 44a Nr. 2 UrhG sowie § 44b UrhG freigestellt.

818 Vgl. zur methodischen Herleitung oben unter 3. Teil A. III. 2. d).

819 Siehe 2. Teil B. I. 2. d) aa)

820 Vgl. oben unter 3. Teil A. III. 2. d).

IV. Zwischenfazit zur Speicherung urheberrechtlich geschützter Inputdaten

Alles in allem werden bei der Speicherung urheberrechtlich geschützter Inputdaten Vervielfältigungsstücke von Werken hergestellt. Damit findet bei der Speicherung von Werken als Inputdaten für KI ein Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG statt. Das im zweiten Teil der Untersuchung identifizierte Regelungsbedürfnis nach einer Freistellung durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte wird *de lege lata* demzufolge nicht erfüllt. *De lege ferenda* ist dies durch eine Einschränkung des urheberrechtlichen Vervielfältigungsrechts in § 16 UrhG zu lösen.

Nichtsdestotrotz ist zumindest das Ob des Freistellungsinteresses im geltenden Urheberrecht durch die Anwendbarkeit von Schrankenbestimmungen abgebildet. Denn die Speicherung urheberrechtlich geschützter Inputdaten beispielsweise durch Kamerasensoren oder Webcrawler ist als vorübergehende Vervielfältigungshandlung von § 44a UrhG erfasst, solange die Inhalte nicht zur Verwirklichung von Sekundärzwecken aus der technischen Umgebung des KI-Systems extrahiert werden. Da auch die Verwendung von Werken als Inputdaten in der Anwendungsphase eines KI-Systems unter den Text und Data Mining-Begriff aus § 44b Abs. 1 UrhG gefasst werden kann, fallen die Vervielfältigungshandlungen darüber hinaus in den Anwendungsbereich der Schrankenbestimmung aus § 44b UrhG. Das gilt allerdings nur, solange kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist. Die Freistellungswirkungen der Schrankenbestimmungen aus §§ 44a, 44b UrhG greifen dabei unabhängig davon ein, welche Art von KI-System durch die Verwendung von Werken als Inputdaten konkret eingesetzt werden soll. Auch die Anwendung generativer KI kann auf Inputseite grundsätzlich unter die Schrankentatbestände gefasst werden. Der konkret generierte Output ist hiervon allerdings zu trennen.

D. Vorverarbeitung von urheberrechtlich geschützten Inputdaten durch Normalisierung und Feature Encoding

Auch Inputdaten müssen in der Regel vorverarbeitet, also insbesondere normalisiert und in ein numerisches Format umgewandelt werden, bevor

sie als Berechnungsgrundlage eines KI-Systems verwendet werden können.⁸²¹ Wie auch die Vorverarbeitung für das Training von KI sollten auch diese Vorverarbeitungsschritte aus regulatorischen Gesichtspunkten urheberrechtlich freigestellt sein. Dabei ist zweckmäßig, die Freistellung im geltenden Urheberrecht durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte abzubilden.⁸²²

I. Normalisierung der Inputdaten

Damit eine Freistellung der Vorverarbeitung entsprechend den identifizierten Regelungsbedürfnissen erfolgen kann, darf die Normalisierung der Inhalte nicht das urheberrechtliche Ausschließlichkeitsrecht beeinträchtigen.

1. Eingriff in das urheberrechtliche Verwertungsrecht

Da die Normalisierungsverfahren von Inputdaten einerseits und Trainingsdaten andererseits dem gleichen Zweck dienen sowie identische Techniken verwendet werden, besteht zwischen ihnen kein wesensmäßiger Unterschied. Mit der Normalisierung urheberrechtlich geschützter Trainingsdaten kann ein Datenverarbeiter in das Ausschließlichkeitsrecht der betroffenen Urheber in Form des Vervielfältigungsrechts aus § 16 UrhG eingreifen.⁸²³ Im Detail hängt das von den angewendeten Normalisierungsschritten ab. Solange schutzbegründende Elemente der verwendeten Werke in den normalisierten Erzeugnissen wiedererkennbar bleiben, haben hierauf auch Veränderungen des Werke im Normalisierungsverfahren keinen Einfluss. Vervielfältigungen im Sinne des § 16 UrhG, die im Zuge der Normalisierung gegebenenfalls hergestellt werden, sind daher erlaubnispflichtig. Diese Ergebnisse können auch auf die Normalisierung von Inputdaten übertragen werden. Das hat zur Folge, dass das urheberrechtliche Regelungsregime auch für die Normalisierung von Inputdaten den identifizierten Freistellungsinteressen nicht gerecht wird.

821 Zur Vorverarbeitung technisch ausführlicher siehe oben unter 1. Teil B. II. 2.

822 Zur Herleitung siehe oben unter 2. Teil E.

823 Dazu ausführlich siehe oben unter 3. Teil B. I. 1. a).

2. Anwendung urheberrechtlicher Schrankenbestimmungen

Sofern bei der Normalisierung der als Inputdaten verwendeten Werke ein Eingriff in das urheberrechtliche Vervielfältigungsrecht stattfindet, könnte dieser jedoch von einer urheberrechtlichen Schrankenbestimmung freigestellt werden.

a) Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)

Denkbar wäre zum einen, dass die Vervielfältigung von Werken im Rahmen der Normalisierung von Inputdaten durch die Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen gemäß § 44a UrhG freigestellt wird. Im Rahmen der vorliegenden Untersuchung wird dabei davon ausgegangen, dass keine Sekundärverwertung der normalisierten Werke stattfindet, die mitunter auch eine dauerhafte Speicherung der Inhalte notwendig macht.

aa) Grundvoraussetzungen des Schrankentatbestands

Wie auch bei der Speicherung der urheberrechtlich geschützten Inputdaten ist auch die Funktion der normalisierten Vervielfältigungsstücke nach Beendigung des Berechnungsvorgangs innerhalb des KI-Systems erfüllt. Nach Ablauf der konkreten KI-Anwendung können sie deswegen – von einer möglichen Sekundärverwertung der Inhalte abgesehen – gelöscht werden. Auch die bei der Normalisierung der Inputdaten möglicherweise angefertigten Vervielfältigungsstücke können daher vorübergehender Natur im Sinne des § 44a UrhG sein.⁸²⁴

Zwar kann der Algorithmus aus einer technischen Perspektive darüber hinaus auch auf der Grundlage nicht-normalisierter Inputdaten in Gang gesetzt werden. Dies kann aber zu einer Verringerung der Ausgabegenauigkeit

824 Vgl. zur Voraussetzung *Dustmann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 44a Rn. 9; außerdem EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

des KI-Systems führen.⁸²⁵ Eine Normalisierung der Daten ist daher für eine effiziente und einwandfreie Funktion des KI-Systems notwendig. Damit sind auch die bei der Normalisierung von Inputdaten angefertigten Vervielfältigungsstücke integraler und wesentlicher Bestandteil der Anwendung eines KI-Systems.⁸²⁶ Da normalisierte Vervielfältigungsstücke grundsätzlich auch automatisiert nach Ablauf des Algorithmus gelöscht werden können und neben der Anwendung des KI-Systems keinen eigenständigen Zweck erfüllen, sind sie darüber hinaus auch sowohl flüchtiger als auch begleitender Natur.⁸²⁷ Schließlich ermöglicht das Normalisierungsverfahren mit der informationsbezogenen Analyse der in der Anwendungsphase der KI als Input vorgesehenen Werke auch eine rechtmäßige Nutzung im Sinne des § 44a Nr. 2 UrhG.⁸²⁸

bb) Eigene wirtschaftliche Bedeutung der Vervielfältigungen

Grundsätzlich kommt den Vervielfältigungshandlungen bei der Normalisierung der Inputdaten auch keine eigenständige wirtschaftliche Bedeutung zu, die einer Anwendung des § 44a UrhG entgegensteht.⁸²⁹ Dass die Möglichkeit besteht, die normalisierten Inputdaten potenziell auch zu Sekundärzwecken wie insbesondere dem Training von KI zu verwenden, führt nicht zu einer eigenständigen wirtschaftlichen Bedeutung der Vervielfältigungshandlungen, solange keine Absonderung aus dem technischen Verfahren zur Verwirklichung von Sekundärverwertung stattfindet.⁸³⁰ Die bei der Normalisierung der Werke vorgenommenen Änderungen sind vielmehr außerdem rein technisch bedingt. Außerhalb der Anwendung des KI-Systems erlangen sie keine eigenständige Bedeutung. Auch die im Rah-

825 Werden die eingegebenen Zahlenwerte zu groß oder klein, können sie nicht mehr in unterscheidungsfähige Aktivierungswerte der einzelnen Neuronen übersetzt werden, siehe *Paaß/Hecker*, Künstliche Intelligenz, S. 102.

826 Vgl. zur Voraussetzung EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 36 – PRCA/NLA.

827 Vgl. zur Flüchtigkeit EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 64 – Infopaq; EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA; KG, Urt. v. 30.04.2004 - 5 U 98/02, GRUR-RR 2004, 228 (231) – Ausschnittendienst. Zum begleitenden Charakter einer Vervielfältigung vgl. EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 43 – PRCA/NLA.

828 Dazu ausführlich oben unter 3. Teil C. II. 1. d).

829 Dazu ausführlich oben unter 3. Teil C. II. 1. e).

830 Dazu bereits ebenfalls siehe oben unter 3. Teil C. II. 1. e)

men des Normalisierungsverfahrens stattfindenden Änderungen an den digitalen Vervielfältigungsstücken führen demzufolge nicht dazu, dass die geänderten Vervielfältigungsstücke als solche *neben* der rechtmäßigen Nutzung, also der Anwendung des KI-Systems, verwertet werden sollen.⁸³¹ Eine eigene wirtschaftliche Bedeutung der Vervielfältigungshandlungen folgt deswegen auch nicht aus der Veränderung der betreffenden Werke im Rahmen der Normalisierung. Alles in allem ist deswegen nicht davon auszugehen, dass den Vervielfältigungshandlungen bei der Normalisierung urheberrechtlich geschützter Inputdaten neben Anwendung der KI eine eigene wirtschaftliche Bedeutung zukommt.⁸³²

cc) Schlussfolgerungen zur grundsätzlichen Anwendbarkeit von § 44a Nr. 2 UrhG

Alles in allem erfüllen die Vervielfältigungshandlungen, die bei der Normalisierung von urheberrechtlich geschützten Inputdaten für die Anwendung von KI stattfinden können, alle Voraussetzungen des § 44a Nr. 2 UrhG. Sie sind daher grundsätzlich als vorübergehende Vervielfältigungshandlung von der Schrankenbestimmung aus § 44a Nr. 2 UrhG freigestellt. Eine Sekundärverwertung der als Inputdaten normalisierten Werke, die mitunter auch eine dauerhafte Speicherung notwendig macht, ist hier aber nicht berücksichtigt.

dd) Verändernde Vervielfältigungen und das Änderungsverbot (§ 62 UrhG)

Bei der Normalisierung der Inputdaten können weiterhin veränderte Vervielfältigungsstücke hergestellt werden. Wie auch bei der Vorverarbeitung der zum Training vervielfältigten Werke ist auch hier fraglich, ob diese Änderungen vor dem Hintergrund des in § 62 UrhG verankerten Änderungsverbots durchgeführt werden können. Entscheidend ist hierfür aufgrund

831 Davon, dass eine Veränderung an einem Werk automatisch zu einer eigenen wirtschaftlichen Bedeutung der Vervielfältigung führt, geht aber in unzutreffender Weise der EuGH aus, siehe EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 53 – Infopaq II. Im Anschluss hieran auch *Dregelies*, GRUR 2024, 1484 (1485).

832 Zu den Anforderungen vgl. EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn.175 – Football Association Premier League. Kritisch allerdings *Maamar*, ZUM 2023, 481 (487).

des in § 62 Abs. 1 S. 2 UrhG angelegten Verweises auf § 39 UrhG, ob die Änderungen zur Erfüllung des jeweiligen Schranken Zwecks erforderlich und zumutbar sind.⁸³³ Darüber hinaus muss in richtlinienkonformer Auslegung aber auch berücksichtigt werden, ob die der Schrankenregelung zugrundeliegende Richtlinienbestimmung verändernde Vervielfältigungshandlungen urheberrechtlich freistellen wollte. Im Anwendungsbereich des § 44a UrhG kommt es hierbei auf Art. 5 Abs. 1 InfoSoc-RL an.⁸³⁴

Dafür, dass Art. 5 Abs. 1 lit. b InfoSoc-RL für die Ermöglichung rechtmäßiger Nutzungen auch verändernde Vervielfältigungshandlungen freistellt, spricht, dass gemäß ErwGr. 33 S. 3 Hs. 2 InfoSoc-RL eine Veränderung von Informationen im Kontext des Art. 5 Abs. 1 lit. a InfoSoc-RL in bestimmten Fallkonstellationen unzulässig sein soll. In Bezug auf die Ermöglichung rechtmäßiger Nutzungen über Art. 5 Abs. 1 lit. b InfoSoc-RL kann der Richtlinie ein solcher Aussagegehalt hingegen nicht entnommen werden. Aus der ausdrücklichen Regelung zur Veränderung von Inhalten im Kontext des Art. 5 Abs. 1 lit. a InfoSoc-RL kann demzufolge geschlussfolgert werden, dass eine Veränderung der genutzten Werke im Rahmen des Art. 5 Abs. 1 lit. b InfoSoc-RL nicht zu einem Ausbleiben der Freistungswirkung führen soll.⁸³⁵ Dies muss im Regelungsbereich des § 44a Nr. 2 UrhG im Wege richtlinienkonformer Auslegung auch bei der Anwendung des § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG berücksichtigt werden.

Darüber hinaus ist der Schrankenzweck des § 44a Nr. 2 UrhG respektive Art. 5 Abs. 1 lit. b InfoSoc-RL, rechtmäßige Nutzungen zu ermöglichen. Eine solche ist auch die Verwendung von Werken als Berechnungsgrundlage einer KI.⁸³⁶ Hierfür müssen die Werke im Rahmen des vorgeschalteten Normalisierungsverfahrens verändert werden.⁸³⁷ Die Herstellung der veränderten Vervielfältigungsstücke im Normalisierungsverfahren ist dem-

833 OLG Hamburg, Urt. v. 05.06.1969 - 3 U 21/69, GRUR 1970, 38 (39) – Heintje; Engels, in: Götting/Lauber-Rönsberg/Ahlberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 62 Rn. 10 f.; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 62 Rn. 6, 12; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 62 Rn. 10. Zur Herleitung dieser Voraussetzung siehe oben unter 3. Teil B. I. 1. b) bb) (2).

834 Zur Herleitung dieser Voraussetzung siehe oben unter 3. Teil B. I. 1. b) bb) (2).

835 Zu einer anderen Sachverhaltskonstellation und methodisch eingebettet in die eigene wirtschaftliche Bedeutung der verändernden Vervielfältigungshandlungen abweichende Ansicht EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 53 – Infopaq II, wobei das Gericht hier einem Fehlschluss unterliegt. Siehe dazu oben unter 3. Teil D. I. 2. a) bb).

836 Dazu siehe 3. Teil C. II. 1. d).

837 Dazu oben unter 3. Teil D. I. 2. a) aa).

zufolge auch zur Erfüllung des mit § 44a Nr. 2 UrhG respektive Art. 5 Abs. 1 lit. b InfoSoc-RL verfolgten Schranken Zwecks erforderlich. Da die Werknutzung nicht über die rein technische Verarbeitung der veränderten Werkstücke hinausgeht und insbesondere keine menschliche Rezeption der veränderten Inhalte stattfindet, sind die verändernden Werknutzungen für betroffene Rechteinhaber darüber hinaus auch zumutbar.

Beide Gründe sprechen im Ergebnis dafür, dass die Herstellung der veränderten Vervielfältigungsstücke im Rahmen der Normalisierung von Inputdaten zur Anwendung von KI, die § 44a Nr. 2 UrhG dem Grunde nach freistellt, gemäß § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG vom urheberrechtlichen Änderungsverbot ausgenommen ist. Auch unter Berücksichtigung der stattfindenden Veränderungen an den Werken findet demzufolge eine Freistellung der Nutzungshandlung über § 44a Nr. 2 UrhG statt.

b) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)

Vervielfältigungen von Werken bei der Normalisierung von Inputdaten sind, wie auch die wesensmäßig verwandten Vervielfältigungen bei der Normalisierung von urheberrechtlich geschützten Trainingsdaten, darüber hinaus von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 UrhG gedeckt. Hierfür ist ebenfalls unschädlich, dass bei der Normalisierung der Inputdaten unter Umständen veränderte Vervielfältigungsstücke der Werke hergestellt werden.⁸³⁸

3. Zwischenergebnis zur Normalisierung der Inputdaten

Weil das Vervielfältigungsrecht bei der Normalisierung von urheberrechtlich geschützten Inputdaten beeinträchtigt werden kann, erfüllt das geltende Urheberrecht auch bei der Normalisierung von als Inputdaten verwendeten Werken nicht die im zweiten Teil der vorliegenden Untersuchung identifizierten Regelungsbedürfnisse. Dennoch greifen die Schrankenbestimmungen sowohl für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG als auch für das Text und Data Mining aus § 44b UrhG ein. Dies hat zur Folge, dass es für die Normalisierung von als Inputdaten

838 Dazu ausführlich vgl. oben unter 3. Teil B. I. 1. b) bb).

verwendeten Werken keiner rechtsgeschäftlichen Erlaubnis des Urhebers bedarf. Demzufolge wird der geltende Rechtsrahmen in Bezug auf die Normalisierung urheberrechtlich geschützter Inputdaten zumindest dem Ob des identifizierten Freistellungsbedürfnisses gerecht.

II. Feature Encoding

Nach der Normalisierung werden die erstellten Eingabedaten in Tensoren umgewandelt. Dieser Vorgang wird als Feature Encoding bezeichnet. Tensoren sind numerische Darstellungen der Eingabedaten. Sie sind erforderlich, weil nur diese zahlenbasierten Repräsentationen im KNN als statistischen Modell verarbeitet werden können.

Wie die Normalisierung ist auch das Feature Encoding urheberrechtlich geschützter Inputdaten mit dem Feature Encoding zum Training verwendeter Werke wesensverwandt. Deswegen können die für den Umgang mit Trainingsdaten entwickelten Grundsätze auch im vorliegenden Fall angewendet werden.⁸³⁹ Aus diesem Grund ist davon auszugehen, dass die beim Feature Encoding urheberrechtlich geschützter Inputdaten verwendeten Werke aufgrund ihrer numerischen Repräsentation in den Tensoren mangels tatsächlicher Identifizierbarkeit nicht mehr im urheberrechtlichen Sinne wiedererkennbar sind. Damit werden sie in den Tensoren nicht im Sinne des § 16 UrhG vervielfältigt. Dies führt dazu, dass das Feature Encoding der geschützten Inputdaten nicht das urheberrechtliche Ausschließlichkeitsrecht berührt. Das gilt auch für Vervielfältigungen der Tensoren im Rahmen der Anwendung des KI-Systems. Hinsichtlich des Feature Encodings wird der geltende Rechtsrahmen den identifizierten Freistellungsbedürfnissen daher gerecht. Eine urheberrechtlich relevante Vervielfältigung der Werke findet erst bei der Rückumwandlung der Tensoren in menschlich wahrnehmbare Werkrepräsentationen statt.

III. Zwischenfazit zur Vorverarbeitung urheberrechtlich geschützter Inputdaten

Es erwies sich als zweckmäßig, dass die Vorverarbeitung urheberrechtlich geschützter Inputdaten insgesamt im Wege einer unmittelbar begrenzenden

⁸³⁹ Dazu insgesamt siehe oben unter 3. Teil B. I. 2.

Ausgestaltung der urheberrechtlichen Verwertungsrechte erlaubnisfrei gestellt sein sollte. Während der Normalisierung als Inputdaten verwendeter Werke können – je nach angewendetem Verfahren – jedoch Vervielfältigungen im Sinne des § 16 UrhG hergestellt werden. Infolgedessen besteht ein regulatorisches Defizit hinsichtlich der Normalisierung von Inputdaten.

Allerdings ermöglicht eine Anwendung sowohl der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG als auch der Schrankenregelung für Text und Data Mining nach § 44b UrhG eine Freistellung der urheberrechtlich relevanten Normalisierungsschritte. Damit wird das geltende Urheberrecht jedenfalls dem Ob des regulatorischen Interesses gerecht. Letztere ist jedoch nur anwendbar, wenn kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt wurde. De lege ferenda könnte eine Anpassung der Reichweite des Vervielfältigungsrechts in § 16 UrhG Abhilfe schaffen.

Im Gegensatz hierzu werden die für das anschließende Feature Encoding der urheberrechtlich geschützten Inhalte ermittelten Regelungsbedürfnisse vom geltenden Rechtsrahmen vollständig erfüllt. Die Erstellung der numerischen Werkrepräsentationen greift nicht in das Vervielfältigungsrecht aus § 16 UrhG ein.

E. Zusammenfassung des dritten Teils

Für das Web Scraping urheberrechtlich geschützter Trainingsdaten hält das geltende Urheberrecht mit § 44b UrhG eine Schrankenbestimmung bereit, deren Freistellungswirkung die Nutzungshandlung dem Grunde nach umfasst. Der Einsatz von Webcrawlern zur Vervielfältigung von Werken für das Training von KI bedarf daher grundsätzlich keiner rechtsgeschäftlichen Erlaubnis. Dies deckt sich insoweit auch mit den regulatorischen Freistellungsbedürfnissen, die im zweiten Teil der Untersuchung identifiziert worden sind. Das gilt auch für den persönlichen Anwendungsbereich der Schrankenbestimmung. Er umfasst auch Web Scraping-Aktivitäten kommerzieller Trainingsdatensammler. Sowohl bereits angefertigte Werkstücke als auch Zugänge zu Werkdatenbanken können bei Einhaltung bestimmter Voraussetzungen auch an Dritte Datenverarbeiter weitergegeben werden, ohne dass die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining-Aktivitäten für diese Dritten erlischt. Hierfür kommt es

auf die Auslegung des Tatbestandsmerkmals der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG an.

Dennoch führen sowohl die in § 44b Abs. 2 S. 2 UrhG geregelte Löschungspflicht, die Vergütungsfreiheit der Nutzung sowie die in § 44b Abs. 3 UrhG angelegte Möglichkeit, einen die Schrankenwirkung beendenden Nutzungsvorbehalt zu erklären, dazu, dass das geltende Urheberrecht hinter den regulatorischen Freistellungsbedürfnissen zurückbleibt. Der ex nunc wirkende Nutzungsvorbehalt kann dabei auch in natürlicher Sprache abgegeben werden. Wirksamkeitsvoraussetzung ist aber, dass die Erklärung in einer Art und Weise digital hinterlegt ist, die eine unmittelbare Verarbeitung durch Computer ermöglicht. Sie muss darüber hinaus sowohl angemessen als auch ausdrücklich erfolgt sein. Für eine interessengerechte Freistellung des Web Scrapings von Werken zum Training von KI sind alles in allem Anpassungen im urheberrechtlichen Rechtsrahmen vorzunehmen.

Das eigentliche Training von KI-Systemen auf Basis von Werken kann in vier Teilschritte unterteilt werden: Zunächst müssen die als Trainingsdaten gesammelten Werke vorverarbeitet, also für die maschinelle Verarbeitung aufbereitet werden. Hierzu gehören die Normalisierung, also die Transponierung auf eine vergleichbare Skala, sowie das Feature Encoding, also die Umwandlung der Werke in numerische Tensoren. Nur diese numerischen Repräsentationen der Werke können durch das KNN verarbeitet werden. An die Vorverarbeitung der Trainingsdaten schließt sich die eigentliche Trainingsanalyse an. Bei ihr werden ggf. vorübergehende Vervielfältigungen der in numerische Tensoren umgewandelten Trainingsdaten angefertigt. Der Trainingsprozess mündet in der Anpassung der Gewichtungswerte und anderen Parameter innerhalb eines KNN.

Im Rahmen der beschriebenen Teilschritte des Trainingsprozesses können KI-Entwickler allein bei der Normalisierung der als Trainingsdaten verwendeten Werke möglicherweise in das urheberrechtliche Ausschließlichkeitsrecht eingreifen. Ob Vervielfältigungsstücke im Sinne des § 16 UrhG hergestellt werden, hängt von den im Einzelnen angewendeten Normalisierungsverfahren ab. Alle übrigen Teilschritte des Trainingsprozesses fallen nicht in den Anwendungsbereich eines Verwertungsrechts. Damit erfüllt das geltende Urheberrecht hier im Wesentlichen die regulatorischen Bedürfnisse. Nach ihnen sollte das Training im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Ausschließlichkeitsrechte freigestellt sein. Die während der Normalisierung urheberrechtlich geschützter Trainingsdaten möglicherweise stattfindenden Vervielfältigungshandlungen sind im Übrigen zwar durch die Schrankenbestimmung für

Text und Data Mining aus § 44b UrhG erlaubnisfrei gestellt. Hierdurch wird das Urheberrecht jedoch nur dem *Ob*, nicht aber dem *Wie* der im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnisse gerecht. Hinsichtlich der Normalisierung von Trainingsdaten macht dies de lege ferenda Anpassungen der Rechtslage notwendig.

Auch die Speicherung urheberrechtlich geschützter Inputdaten, beispielsweise durch Sensortechnik wie Kameras oder Webcrawler, bedarf aus einer regulatorischen Perspektive der Freistellung im Wege einer unmittelbar beschränkenden Ausgestaltung der urheberrechtlichen Verwertungsrechte. Zwar ist die Speicherung als Inputdaten verwendeter Werke de lege lata sowohl über die Schrankenbestimmung für Text und Data Mining, § 44b UrhG, als auch über die Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG erlaubnisfrei gestellt. Dennoch wird dies, wie auch bei der Normalisierung urheberrechtlich geschützter Trainingsdaten, nicht dem *Wie* des identifizierten Freistellungsbedürfnisses gerecht. Daher bedarf es auch hier de lege ferenda einer Anpassung des Urheberrechts. Durch sie muss die Speicherung urheberrechtlich geschützter Inputdaten für eine Verarbeitung durch KI im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte aus dem Schutzbereich des urheberrechtlichen Ausschließlichkeitsrechts ausgenommen werden.

Für die gleichermaßen notwendige Vorverarbeitung der Inputdaten gilt im Wesentlichen das für die Vorverarbeitung von Trainingsdaten Gesagte. Zwar ist die Normalisierung urheberrechtlich geschützter Inputdaten auch von der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen, § 44a Nr. 2 UrhG, erlaubnisfrei gestellt. Im Ergebnis gilt aber auch hier, dass die Vorverarbeitung inklusive der Normalisierung der als Inputdaten verwendeten Werke aus einer regulatorischen Perspektive durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte freigestellt sein sollte. Die Normalisierung urheberrechtlich geschützter Inputdaten, Teil des Vorverarbeitungsprozesses, kann aber auch hier in das Vervielfältigungsrecht betroffener Urheber aus § 16 UrhG eingreifen. Auch in Bezug auf die Vorverarbeitung urheberrechtlich geschützter Inputdaten wird das Freistellungsbedürfnis daher de lege lata nicht vollständig erfüllt. Damit müssen de lege ferenda auch in Bezug auf die Vorverarbeitung, konkret die Normalisierung urheberrechtlich geschützter Inputdaten Anpassungen im UrhG erfolgen.

Vierter Teil: Perspektiven für das Urheberrecht

Beim Vergleich des geltenden urheberrechtlichen Regelungsregimes für die Verwendung von Werken als Trainings- und Inputdaten mit den im zweiten Teil der Untersuchung identifizierten Regelungsbedürfnissen ergeben sich mehrere Regelungsdefizite. De lege ferenda sind deswegen Anpassungen des urheberrechtlichen Rechtsrahmens notwendig. Im Folgenden werden Änderungsvorschläge für das Urheberrechts- sowie das Verwertungsgesellschaftengesetz erarbeitet, mit deren Hilfe die Regelungsdefizite behoben werden können.⁸⁴⁰

A. Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern

Die Speicherung von urheberrechtlich geschützten Trainingsdaten mit Hilfe von Webcrawlern sollte aus einer regulatorischen Perspektive durch eine Schrankenbestimmung erlaubnisfrei gestellt sein. Interessengerecht ist dabei, dass die Regelung sowohl nicht-kommerziell als auch kommerziell tätige Trainingsdatensammler privilegiert. Ein Opt-out-Mechanismus ist nicht erforderlich. Urheber sollen für die Vervielfältigungen ihrer Werke zum Training von KI allerdings eine angemessene Vergütung erhalten.⁸⁴¹

Das geltende Urheberrecht sieht mit der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG eine Regelung vor, deren Freistellungswirkung sich nach richtlinienkonformer Auslegung des Tatbestands auch auf das automatisierte Sammeln urheberrechtlich geschützter Trainingsdaten mit Hilfe von Webcrawlern erstreckt. Allerdings erfüllt die Schrankenbestimmung wegen der in § 44b Abs. 2 S. 2 UrhG vorgesehenen Löschungspflicht trotzdem nicht die herausgearbeiteten regulatorischen Anforderungen. Gleiches gilt für den in § 44b Abs. 3 UrhG vorgesehenen Opt-out-Me-

840 Änderungsvorschläge jeweils durch Fettdruck hervorgehoben.

841 Zu den identifizierten Regelungsbedürfnissen insgesamt siehe 2. Teil B.

chanismus und die de lege lata ermöglichte vergütungsfreie Nutzung der Werke.⁸⁴²

I. Reichweite der Schrankenbestimmung

Da § 44b UrhG das Web Scraping von Werken als Trainingsdaten für KI dem Grunde nach freistellt, ist es ausreichend, die notwendigen Anpassungen durch punktuelle Änderungen des § 44b UrhG umzusetzen. Zusätzlich zum herausgearbeiteten Änderungsbedarf ist dabei zu berücksichtigen, dass sich die Freistellungswirkung des § 44b UrhG für das Sammeln von Werken als Trainingsdaten erst nach richtlinienkonformer Auslegung des Schrankentatbestands bestimmen lässt. Zur Erhöhung der Rechtssicherheit bei der Speicherung von Werken als Trainingsdaten für KI sollte daher außerdem eine Klarstellung in den Regelungstext aufgenommen werden. Unter Berücksichtigung des gesamten Änderungsbedarfs könnte die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG de lege ferenda wie folgt aussehen:

§ 44b UrhG-E: Text und Data Mining⁸⁴³

- (1) Text und Data Mining ist die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, ~~um daraus~~ **mit deren Hilfe** Informationen insbesondere über Muster, Trends und Korrelationen ~~zu gewinnen~~ **gewonnen werden können. Hierzu gehört auch das Training von auf künstlichen neuronalen Netzen basierenden Systemen maschinellen Lernens (Systeme künstlicher Intelligenz).**
- (2) Zulässig sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining. Die Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.
- (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.
- (4) **Absatz 2 Satz 2 und Absatz 3 gelten nicht, wenn die Vervielfältigungen für das Training von Systemen künstlicher Intelligenz angefertigt worden sind.**

842 Zur geltenden Urheberrechtslage für die Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern siehe insgesamt 3. Teil A.

843 Hervorhebung steht für neu einzufügenden Normtext.

II. Angemessene Vergütung für Urheber

Darüber hinaus ist es interessengerecht, den vom Web Scraping für das KI-Training betroffenen Urhebern für die stattfindenden Vervielfältigungshandlungen eine angemessene Vergütung zu gewähren.

1. Vergütungspflicht und Vergütungssystem

Problematisch ist dabei allerdings, dass weder betroffene Urheber noch Trainingsdatensammler aufgrund der Automatisierung der Datensammlung, oftmals fehlender Metadaten zu den erfassten Inhalten und eines Kontrolldefizits regelmäßig Kenntnis darüber haben, ob es überhaupt zu urheberrechtlich relevanten Vervielfältigungshandlungen gekommen ist und welche Werke genutzt worden sind. Es fehlt daher regelmäßig an einer Konkretisierung des Vergütungsgläubigers. Der einzuführende gesetzliche Vergütungsanspruch muss aus diesem Grund durch Verwertungsgesellschaften geltend gemacht werden.⁸⁴⁴ Darüber hinaus ist die Abrechnung von Einzelnutzungen vor dem Hintergrund der skizzierten praktischen Schwierigkeiten und der Größe der erstellten Datensätze ebenfalls nur erschwert möglich. Der Vergütungsanspruch muss deswegen pauschaliert werden. Wichtig ist hierbei aber, keine Überkompensation der Urheber zulasten der Trainingsdatensammler zu verursachen.

Auch die Hersteller- und Betreibervergütung (§§ 54 ff. UrhG), mit Hilfe derer Urhebern eine gesetzliche Vergütung unter anderem für gemäß § 53 UrhG hergestellte Privatkopien ihrer Werke zuteil wird, bedient sich solcher Pauschalierungsverfahren. Denn auch bei der Herstellung von Privatkopien haben Urheber im Regelfall keine Kenntnis darüber, ob und wie viele Privatkopien ihrer Werke angefertigt wurden. Die Hersteller und Betreiber von Speichermedien oder Kopiergeräten, Vergütungsschuldner, können außerdem nicht wissen, inwieweit ihre Hilfsmittel für urheberrechtlich relevante Nutzungshandlungen verwendet worden sind. Aus diesen Gründen wird die Hersteller- und Betreibervergütung ebenfalls über Verwertungsgesellschaften abgewickelt. Im Ergebnis bestehen wesentliche Parallelen zwischen der Hersteller- und Betreibervergütung aus den §§ 54 ff. UrhG und der gesetzlichen Vergütungspflicht, die für die Vervielfältigung von Werken zum Training von KI eingeführt werden sollte. Letz-

844 Ebenso Geiger, IIC 2024, 1129 (1153).

tere ist daher nach dem Vorbild der Hersteller- und Betreibervergütung auszugestalten. Eine Regelung könnte danach wie folgt aussehen:

§ 44c UrhG-E: Vergütung für die automatisierte Vervielfältigung von Werken aus Internetquellen zum Training von Systemen künstlicher Intelligenz

- (1) *Lässt die Art eines im Internet öffentlich zugänglich gemachten Werkes eine nach § 44b Absatz 2 UrhG erlaubte automatisierte Vervielfältigung für das Training von Systemen künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E) erwarten, so hat der Urheber des Werkes gegen den Trainingsdatensammler Anspruch auf Zahlung einer angemessenen Vergütung.*
- (2) *Trainingsdatensammler ist, wer Inhalte im Geltungsbereich des Urheberrechtsgesetzes für das Training von Systemen künstlicher Intelligenz vervielfältigt, bei denen erwartet werden kann, dass sie ganz oder teilweise nach diesem Gesetz schutzfähig sind. Das umfasst insbesondere diejenigen, die online zugängliche Inhalte erstmals auf Datenträger speichert, als auch diejenigen, die einen vorverarbeiteten Trainingsdatenkörper von einem Dritten erlangt.*
- (3) *Maßgebend für die Vergütungshöhe ist insbesondere, zum Training wie vieler Systeme künstlicher Intelligenz das Werk genutzt wird und welche Aufgabe mit dem jeweiligen System künstlicher Intelligenz erfüllt werden soll. Dabei ist zu berücksichtigen, inwieweit technische Schutzmaßnahmen nach § 95a auf die betreffenden Werke angewendet werden. Die Vergütung darf die Trainingsdatensammler nicht unzumutbar beeinträchtigen; sie muss in einem wirtschaftlich angemessenen Verhältnis zum Wert des angefertigten Trainingsdatenkörpers als solchem und dem System künstlicher Intelligenz stehen, das auf Basis des Trainingsdatenkörpers entwickelt worden ist.*
- (4) *Die Vergütung ist so zu gestalten, dass sie auch mit Blick auf die Vergütungspflicht anderer Trainingsdatensammler insgesamt angemessen ist. Mehrere an einem Trainingsvorgang beteiligte Trainingsdatensammler haften als Gesamtschuldner.*

§ 44d UrhG-E: Meldepflicht

- (1) *Wer Inhalte im Geltungsbereich des Urheberrechtsgesetzes als Trainingsdaten speichert, bei denen erwartet werden kann, dass sie ganz oder teilweise nach diesem Gesetz schutzfähig sind, ist dem Urheber gegenüber verpflichtet, monatlich bis zum zehnten Tag nach Ablauf jedes Kalendermonats schriftlich Mitteilung zu leisten über:
 1. *Die Menge und Art der vervielfältigten Daten,*
 2. *Die Anzahl der Systeme künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E), die mit den Daten trainiert werden sollen, sowie*
 3. *soweit ersichtlich, die Aufgabe des jeweiligen Systems künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E).**
- (2) *Kommt der Meldepflichtige seiner Meldepflicht nicht, nur unvollständig oder sonst unrichtig nach, kann der doppelte Vergütungssatz verlangt werden.*
- (3) *Die Meldung erfolgt an eine gemeinsame Empfangsstelle der Verwertungsgesellschaften. § 54h Absätze 3 bis 5 gelten entsprechend.*

§ 44e UrhG-E: Verwertungsgesellschaften; Handhabung der Mitteilungen

- (1) *Der Anspruch nach § 44c kann nur durch eine Verwertungsgesellschaft geltend gemacht werden.*
- (2) *Jedem Berechtigten steht ein angemessener Anteil an den nach § 44b gezahlten Vergütungen zu. Soweit Werke mit technischen Maßnahmen gemäß § 95a geschützt sind, werden sie bei der Verteilung der Einnahmen nicht berücksichtigt.*
- (3) *Das Deutsche Patent- und Markenamt kann Muster für die Mitteilungen nach § 44d im Bundesanzeiger bekannt machen. Werden Muster bekannt gemacht, sind diese zu verwenden.*
- (4) *Die Verwertungsgesellschaften und die Empfangsstelle dürfen die gemäß § 44b erhaltenen Angaben nur zur Geltendmachung der Ansprüche nach § 44c verwenden.*

2. Vergütungshöhe und ihre Konkretisierung durch Verwertungsgesellschaften

Für eine Pauschalisierung der angemessenen Vergütung müssen Annahmen beispielsweise darüber getroffen werden, bei wie vielen der für das Training gespeicherten Daten (Beispiel: Textinhalte) aus einer statistischen Perspektive davon ausgegangen werden kann, dass sie urheberrechtlich schutzfähig sind (Beispiel: Eine Auswertung mehrerer repräsentativer Trainingsdatensätzen könnte ergeben, dass durchschnittlich 83 % der gesammelten Textdaten urheberrechtlich schutzfähige Sprachwerke sind). Um Tarife für die Vervielfältigung von Werken zum Training von KI aufzustellen, sind demzufolge empirische Voruntersuchungen notwendig. Die hierfür erforderliche Regelung muss im Verwertungsgesellschaftengesetz umgesetzt werden. Sie könnten wie folgt aussehen:

§ 40a VGG-E: Gestaltung der Tarife für die Entwicklung von Systemen künstlicher Intelligenz (§ 44b Absatz 1 Satz 2 UrhG)

- (1) *Die Höhe der Vergütung für das Vervielfältigen von Werken zum Training von Systemen künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E) bestimmt sich nach § 44c UrhG-E. Die Verwertungsgesellschaften stellen hierfür Tarife auf Grundlage einer empirischen Untersuchung aus einem Verfahren gemäß § 93 auf. § 38 Satz 2 bleibt unberührt.*
- (2) *Die Pflicht zur Tarifaufstellung entfällt, wenn zu erwarten ist, dass der dafür erforderliche wirtschaftliche Aufwand außer Verhältnis zu den zu erwartenden Einnahmen stehen würde.*

Daneben sind noch weitere redaktionelle Folgeanpassungen im Verwertungsgesellschaftengesetz notwendig, die hier nicht näher dargestellt werden.

Im Übrigen bestimmt sich die Höhe der angemessenen Vergütung nach dem wirtschaftlichen Wert der Nutzung.⁸⁴⁵ Bei der Vervielfältigung eines Werks als Trainingsdatum für die Entwicklung von KI wird die Vergütung demzufolge danach bestimmt, welcher wirtschaftliche Vorteil von der Nutzung des Werkes als Trainingsdatum ausgeht. Für die Konkretisierung dieses Vorteils liefert § 44c Abs. 3 UrhG-E Anhaltspunkte. Danach kommt es insbesondere darauf an, zum Training wie vieler KI-Systeme der betreffende Trainingsdatenkörper verwendet werden soll. Denn der wirtschaftliche Vorteil, der aus den Trainingsdaten folgt, steigert sich proportional hierzu. Darüber hinaus ist von Bedeutung, welche Aufgaben das entwickelte KI-System erfüllen wird. Berücksichtigt werden kann beispielsweise aber auch, ob mit der KI direkte Konkurrenzprodukte zu den genutzten Trainingsdaten generiert werden sollen. Maßgebend ist auch, welche Qualität die genutzten Inhalte haben und welche Popularität ihnen zukommt.⁸⁴⁶ Denn je hochqualitativer und populärer die zum Training genutzten Inhalte, desto wahrscheinlicher ist auch der wirtschaftliche Erfolg der durch das KI-System erschaffenen Konkurrenzprodukte.⁸⁴⁷ Die Informationen, die zur Bestimmung der von den Trainingsdatensammlern zu leistenden Vergütung benötigt werden, sind von der Meldepflicht im Sinne des § 44d UrhG-E erfasst.

Der Vergütungsanspruch darf allerdings nicht zu einer Überkompensation der betroffenen Urheber führen. Daher ist die Vergütung gemäß § 44c Abs. 4 UrhG-E so auszugestalten, dass sie auch mit Blick auf die Vergütungspflicht anderer Trainingsdatensammler insgesamt angemessen ist. Denkbar ist beispielsweise, dass ein intermediärer Trainingsdatensammler eine im Vergleich zum eigenhändig sammelnden KI-Entwickler geringere Vergütung zahlen muss, die sich vor allem nach der Anzahl der KI-Systeme bemisst, die mit den Daten trainiert werden sollen (vgl. § 44c Abs. 3 UrhG-E). Das bedeutet, dass er für jede Datenübertragung eine Vergütung zu leisten hätte. Die Vergütung der KI-Entwickler, welche die Trainingsdatensätze erworben haben, könnte danach bemessen werden, welche Aufgabe mit dem jeweiligen KI-System erfüllt werden soll (vgl. § 44c Abs. 3 UrhG-E). Auch sie müsste im Vergleich zur Ein-Personen-Konstellation

845 Zur Vergütungshöhe bei vergütungspflichtigen Schrankenbestimmungen: *Pflüger*, Gerechter Ausgleich und angemessene Vergütung, S. 145. Für die Tarife der Verwertungsgesellschaften ergibt sich das aus § 39 Abs. 1 S. 1 VGG.

846 *Geiger*, IIC 2024, 1129 (1154).

847 *Geiger*, IIC 2024, 1129 (1154); ganz nach dem Motto: „garbage in, garbage out“, vgl. *de la Durantaye*, ZUM 2023, 645 (647).

reduziert werden. Im Ergebnis würde Urhebern dadurch eine angemessene Vergütung zuteilwerden, die interessengerecht zwischen intermediärem Trainingsdatensammler und KI-Entwickler aufgeteilt ist. Da gemäß § 44c Abs. 4 UrhG-E eine gesamtschuldnerische Haftung vorgesehen ist, könnte die Vergütung insgesamt vom intermediären Trainingsdatensammler geleistet werden, der bei der Veräußerung der Trainingsdatensätze einen Teil der wirtschaftlichen Belastung an die KI-Entwickler weiterreichen kann.

3. Erwartung der Vervielfältigung als Voraussetzung für die Vergütung (§ 44c Abs. 1 UrhG-E)

Voraussetzung des Vergütungsanspruchs ist gemäß § 44c Abs. 1 UrhG-E, dass die Art des im Internet veröffentlichten Werkes eine nach § 44b Absatz 2 UrhG erlaubte Vervielfältigung für das Training von KI erwarten lässt. Diese § 54 Abs. 1 UrhG nachempfundene Regelung trägt der Pauschalisierungsnotwendigkeit Rechnung, die der Vergütungsanspruch für die Speicherung von Werken als Trainingsdaten mit Hilfe von Webcrawlern mit sich bringt.⁸⁴⁸

Eine Vervielfältigung kann gemäß §§ 44c Abs. 1 UrhG-E erwartet werden, wenn die jeweilige Werkart üblicherweise zum Training von KI verwendet wird.⁸⁴⁹ Ob sich die Vervielfältigung eines Werkes erwarten lässt, hängt damit entscheidend vom Trainingsdaten- und KI-Markt sowie dem Stand der Technik bei der KI-Entwicklung ab. Eine Beurteilung kann im Rahmen der empirischen Untersuchung gemäß § 40a VGG-E erfolgen. Sie sollte aufgrund des Fortschreitens der Entwicklung und Anwendung von KI regelmäßig überprüft werden.

Inwieweit die Vervielfältigung eines konkreten Werkes zum Training von KI wahrscheinlich ist, ist für die Begründung der Vergütungspflicht hingegen nicht von Bedeutung.⁸⁵⁰ Sie kann allerdings bei Bestimmung der Vergütungshöhe berücksichtigt werden.⁸⁵¹ Die Vervielfältigungswahr-

848 Zur Pauschalisierungsnotwendigkeit siehe oben unter 4. Teil A. II. 1. Zu § 54 UrhG vgl. außerdem *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 17.

849 Vgl. zur entsprechenden Voraussetzung in § 54 UrhG *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 17.

850 Vgl. *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 18.

851 Vgl. *Loewenheim/Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 54 Rn. 18.

scheinlichkeit wird insbesondere dadurch beeinflusst, dass das Werk durch Zugangsbeschränkungen geschützt oder das Web Scraping beispielsweise mit Hilfe des Robots Exclusion Standards vorbehalten worden ist. Die Urheber sollten bei der Meldung der Werke demzufolge entsprechende Angaben tätigen müssen.

4. Trainingsdatensammler als Vergütungsschuldner (§ 44c Abs. 2 UrhG-E)

Vergütungsschuldner sind gemäß § 44c Abs. 1, 2 UrhG-E Trainingsdatensammler. Dies ist, wer Inhalte für das Training von KI speichert, bei denen erwartet werden kann, dass sie ganz oder teilweise urheberrechtlich schutzfähig sind.

Hervorzuheben ist dabei, im Sinne der Legaldefinition nicht vorausgesetzt wird, dass *tatsächlich* Werke für das Training von KI vervielfältigen werden. Konkrete Nutzungshandlungen bei der Erstellung und Verarbeitung der Trainingsdatensätze sollen nach dem der Vergütungspflicht zugrundeliegenden Pauschalisierungsgedanken gerade nicht dargelegt und bewiesen werden müssen. Der Nachweis könnte von einem Urheber auch kaum erbracht werden. Der Pauschalisierung der Vergütung entsprechend muss der Kreis der Vergütungsschuldner zunächst weit gefasst werden. Ob und inwieweit erwartet werden kann, dass bei der Speicherung von Inhalten für das Training von KI insgesamt oder teilweise Werke erfasst werden, bestimmt sich nach Art und Umfang der abgespeicherten Daten sowie danach, wie viele der erfassten Inhalte aus einer statistischen Perspektive typischerweise urheberrechtlich geschützt sind.⁸⁵² Die hierfür notwendigen empirischen Untersuchungen sind von Verwertungsgesellschaften vorzunehmen.

Der Kreis der Vergütungsschuldner umfasst insbesondere diejenigen, die online zugängliche Inhalte zur eigenen Verwendung erstmals auf Datenträger abspeichern. Hierunter fällt das Web Scraping von Trainingsdaten für die Entwicklung von KI. Darüber hinaus sind nach dem Wortlaut des Regelungsvorschlags aber auch Beteiligte erfasst, die vorverarbeitete Trainingsdatenkorpora von Dritten erlangen. Denn auch sie vervielfältigen die Inhalte im Sinne des § 44c Abs. 2 UrhG-E. Auch nachgelagerte KI-Entwickler, denen die Trainingsdatensätze von intermediären Trainingsdatensammlern zur Verfügung gestellt werden, sind damit Trainingsdatensammler im Sinne des § 44c

852 Dazu bereits siehe oben unter 4. Teil A. II. 2.

Abs. 2 UrhG-E. Hierdurch wird auch dann eine Vergütung für Urheber sichergestellt, wenn die Trainingsdatensätze zwischen KI-Entwicklern weitergegeben werden. Von der Regelung außerdem erfasst werden solche KI-Entwickler, die ihre Trainingsdatenkorpora aus dem Ausland beziehen, wo gegebenenfalls keine Vergütung betroffener Urheber stattfindet.⁸⁵³

Hinterfragt werden kann allerdings, ob die Vergütungspflicht nicht direkt an die Tätigkeit der KI-Entwickler beziehungsweise -Anbieter oder gar die Nutzung von KI-Systemen angeknüpft werden könnte.⁸⁵⁴ Dafür spricht, dass insbesondere Anbieter von KI-Systemen, die erkennbar auf dem Markt auftreten, von Urhebern und Verwertungsgesellschaften mit geringerem Aufwand identifiziert und adressiert werden können. Dies erleichtert die Durchsetzung des Vergütungsanspruchs. Allerdings werden Vervielfältigungen urheberrechtlich geschützter Trainingsdaten gemäß § 16 UrhG nur im Vorfeld des Trainings von KI vorgenommen. Vor allem sind, unabhängig von einer technischen Rekonstruierbarkeit, keine Vervielfältigungsstücke der zum Training genutzten Werke dauerhaft im KNN abgelegt.⁸⁵⁵ Bei einer Anknüpfung der Vergütungspflicht an die KI-Nutzer, -Entwickler oder -Anbieter kann vor diesem Hintergrund nicht sichergestellt werden, dass urheberrechtlich relevante Nutzungen der zum Training verwendeten Werke in Deutschland stattgefunden haben. Hierauf hätte auch die Anerkennung eines *domaine public payant*, einer auf eine Vergütungspflicht beschränkten Ausformung eines kollektiven Urheberrechts sämtlicher Urheber,⁸⁵⁶ keinen Einfluss.⁸⁵⁷ Denn auch diese kollektiviert Rechtsposition wäre nicht berührt, wenn keine Nutzung der Werkesamtheit im Inland stattfände. Ohne eine Beeinträchtigung des nationalen Urheberrechts als verfassungsrechtlich gewährleistetes Eigentumsrecht besteht jedoch kein legitimer Zweck

853 Im Rahmen der Privatkopievergütung erfüllt die Händler- und Importeurshaftung (§ 54b UrhG) diese Funktion, siehe *Wiebe*, in: Spindler/Schuster, Recht der elektronischen Medien, UrhG § 54b Rn. 2.

854 Ein solches Vergütungsmodell, welches an den Betrieb von KI-Systemen anknüpft, die geeignet sind, Werkschaffen zu substituieren, schlagen *Senfleben*, IIC 2023, 1535 (1549 ff.); *Jacques/Flynn*, GRUR Int 2024, 1137 (1141, 1148) vor. Die generierten Einnahmen sollten sodann beispielsweise in eine Fondsstruktur zur Unterstützung Kreativer fließen, vgl. instruktiv *Jacques/Flynn*, GRUR Int 2024, 1137 (1148).

855 Dazu im Detail siehe oben 3. Teil B. IV.

856 So insbesondere noch vorgesehen im Entwurf des Urheberrechtsgesetzes von 1965 BT-Drs. IV/270, S. 15 ff. wobei der Vorschlag auf Anregung des Rechtsausschusses allerdings gestrichen worden ist. Dazu siehe BT-Drs. IV/3401, S. 13; Vgl. zur Historie und Funktionsweise allgemein *Dillenz*, GRUR Int 1983, 920; *Marzetti*, GRUR Int. 2019, 343.

857 Mit einem solchen Ansatz *Senfleben*, IIC 2023, 1535 (1552 f.).

dafür, die Freiheitsrechte der KI-Nutzer, -Entwickler und -Anbieter durch eine Vergütungspflicht einzuschränken. Um eine Vergütungspflicht verfassungsrechtlich rechtfertigen zu können, ist damit ein Eingriff in das nationale urheberrechtliche Ausschließlichkeitsrecht notwendig. Er muss zumindest erwartbar sein. Eine Verlagerung der Vergütungspflicht auf andere Beteiligte als den Trainingsdatensammler wie insbesondere Anbieter oder Nutzer von KI-Systemen ist deswegen aus verfassungsrechtlichen Gründen nicht möglich.

Ohnehin wäre es dogmatisch inkonsistent, die Vergütungsbedürftigkeit einer Nutzungshandlung (Vervielfältigung zum Training) durch die Vergütung einer anderen, urheberrechtlich eigenständigen Nutzungshandlung (Generierung urheberrechtlich relevanten Outputs) zu realisieren. Es ist vielmehr stets klar zwischen den verschiedenen Nutzungshandlungen bei der Entwicklung und dem Einsatz von KI zu trennen. Auch die Passivlegitimation für Urheberrechtsverletzungen beim Training einerseits und bei der Anwendung einer KI andererseits könnte unterschiedliche Verantwortlichkeiten begründen.

Weiterhin ist davon auszugehen, dass intermediäre Trainingsdatensammler die anfallende Vergütung an nachgelagerte KI-Entwickler weiterreichen. Bringen diese das KI-System nicht selbst in Verkehr, haben sie ebenfalls die Möglichkeit, die durch die Vergütungspflicht entstehenden Kosten auf den KI-Anbieter umzulegen. Damit kommt es zu einer automatischen Zuordnung der wirtschaftlichen Belastung zu dem Beteiligten, der wirtschaftlich erheblich von der Nutzung der Werke zum Training von KI profitiert. Das ist ökonomisch interessengerecht. Aus diesen Gründen ist eine Anknüpfung der Vergütungspflicht an die Tätigkeit der Trainingsdatensammler notwendig.⁸⁵⁸

III. Unionsrechtliche Perspektive auf den Regelungsvorschlag

Fraglich ist jedoch, ob der entwickelte Regelungsvorschlag mit den unionsrechtlichen Grundlagen des Urheberrechts vereinbar ist. Hierfür muss er insbesondere mit Art. 4 DSM-RL vereinbar sein. Er wird in § 44b UrhG

858 Andere Ansicht aber *Schack*, NJW 2024, 113 Rn. 26, nach dem die Vergütungspflicht an die Generierung des KI-Outputs angeknüpft werden sollte.

umgesetzt.⁸⁵⁹ Die in § 44b Abs. 1 S. 2 UrhG-E eingefügte Klarstellung ist darüber hinaus an Art. 2 Nr. 2 DSM-RL zu messen.

1. Klarstellung zum KI-Training

Wie auch § 44b UrhG adressieren Art. 2 Nr. 2 und Art. 4 DSM-RL das Sammeln urheberrechtlich geschützter Trainingsdaten nicht ausdrücklich. Dennoch ist es von der in Art. 2 Nr. 2 DSM-RL normierten Legaldefinition des Text und Data Mining umfasst.⁸⁶⁰ Die Rechtsänderung hat nur klarstellenden Charakter. Der nationale Gesetzgeber ist hinsichtlich der konkreten Ausgestaltung seiner Rechtsnormen jedoch weitestgehend frei, solange das in der Richtlinie vorgegebene Ziel erreicht wird (vgl. Art. 288 Abs. 3 AEUV). Durch die klarstellende Formulierung wird der Anwendungsbeereich der Regelung jedoch nicht verändert. Ihre Einfügung ist demzufolge mit dem Unionsrecht vereinbar.

2. Verzicht auf Löschungspflicht und Opt-Out-Mechanismus

Der zur Umsetzung in § 44b Abs. 4 UrhG-E vorgeschlagene Verzicht auf das Löschungserfordernis und den Opt-out-Mechanismus bei der Vervielfältigung urheberrechtlich geschützter Trainingsdaten könnte jedoch mit dem Unionsrecht nicht vereinbar sein.

Die Rechtsänderung muss sich zum einen an Art. 4 DSM-RL messen lassen. Löschungspflicht und Opt-out-Mechanismus sind hier in Art. 4 Abs. 2, 3 DSM-RL geregelt. In den Bestimmungen der DSM-RL ist hingegen keine Ausnahme für das Sammeln von Trainingsdaten vorgesehen.

Der DSM-RL kommt in Bezug auf ihre Schrankenbestimmungen allerdings nur mindestharmonisierende Wirkung zu.⁸⁶¹ Gemäß Art. 25 DSM-RL können für Arten und Bereiche der Nutzung, für die die Ausnahmen und Beschränkungen der Richtlinie gelten, umfassendere Bestimmungen erlassen werden, insofern diese mit den in der Datenbank- und InfoSoc-RL vorgesehenen Ausnahmen und Beschränkungen vereinbar sind. Der nationale Gesetzgeber könnte zugunsten der Trainingsdatensammler daher auch ein Mehr an Freistellungswirkung realisieren, insofern die eingeführ-

859 BT-Drs. 19/27426, S. 50.

860 Dazu 3. Teil A. III. 2. a) bis e).

861 *Stieper*, GRUR 2020, 1 (3 f.).

ten Schrankenbestimmungen mit Art. 6 Abs. 2 Datenbank-RL sowie Art. 5 InfoSoc-RL in Einklang stehen.

Da § 44b Abs. 4 UrhG-E des Regelungsentwurfs für Vervielfältigungen von Werken sämtlicher Werkarten und nicht nur für Datenbankwerke gilt, kommt nur Art. 5 InfoSoc-RL als unionsrechtliche Grundlage in Betracht. Hier könnte auf den ersten Blick zwar auf Art. 5 Abs. 3 lit. a InfoSoc-RL zurückgegriffen werden. Die dort vorgesehene Schrankenwirkung kann jedoch nur zum Zwecke der wissenschaftlichen Forschung eingreifen. § 44b Abs. 4 UrhG-E, der auch kommerziell tätige Trainingsdatensammler freistellt, übersteigt den durch Art. 5 Abs. 3 lit. a InfoSoc-RL gewährten Freistellungsspielraum demzufolge. Im Ergebnis wird die Erweiterung der Schrankenbestimmung durch § 44b Abs. 4 UrhG-E daher auch nicht durch Bestimmungen der InfoSoc-RL abgedeckt.

Mangels Grundlage in den urheberrechtlichen Richtlinien des Unionsgesetzgebers wäre der vorgeschlagene Verzicht auf die Löschungspflicht und den Opt-out-Mechanismus daher nicht mit dem europäischen Urheberrecht vereinbar. Der Umsetzung des entwickelten Regelungsvorschlags müsste demzufolge eine europäische Normgebungsinitiative vorangehen.

3. Einführung einer Vergütungspflicht für das Sammeln von Trainingsdaten

Klärungsbedürftig ist darüber hinaus, ob jedenfalls die in § 44c UrhG-E vorgeschlagene Vergütungspflicht für das Web Scraping urheberrechtlich geschützter Trainingsdaten vor dem Hintergrund des Art. 4 DSM-RL umgesetzt werden könnte.

a) Vergütungspflicht als solche

Art. 4 DSM-RL sieht keinen gerechten Ausgleich der Urheber für die Vervielfältigungshandlungen vor. Entscheidend ist damit, ob Art. 4 DSM-RL es den Mitgliedstaaten offenlässt, eine Vergütungspflicht für das Mining von Werken als KI-Trainingsdaten einzuführen.

Hiergegen spricht, dass Art. 5 Abs. 4 DSM-RL (Schrankenbestimmung für die Nutzung von Werken und sonstigen Schutzgegenständen für digitale und grenzüberschreitende Unterrichts- und Lehrtätigkeiten) ausdrücklich eine Vergütungspflicht anordnet. Aus dem Fehlen einer entsprechenden Regelung in Art. 4 DSM-RL könnte abgeleitet werden, dass Urhebern

im Umkehrschluss kein gerechter Ausgleich für das Text und Data Mining gewährt werden darf.⁸⁶² Darüber hinaus ging der Richtliniengeber jedenfalls davon aus, dass das Text und Data Mining für wissenschaftliche Zwecke nur einen minimalen Schaden für Urheber verursache und daher vergütungsfrei vorzunehmen sei.⁸⁶³ Dieser Gedanke ließe sich verallgemeinern und auf das allgemeine Text und Data Mining beziehungsweise das Sammeln urheberrechtlich geschützter Trainingsdaten übertragen.⁸⁶⁴ Nicht zuletzt werden beim Text und Data Mining und auch beim eigentlichen KI-Training auf Basis von Werken nur die mit den Schöpfungen übermittelten semantischen Informationen genutzt. Sie sind nicht Teil des urheberrechtlichen Schutzbereichs.⁸⁶⁵ Dies macht es unwahrscheinlich, dass der Richtliniengeber eine Vergütung der Nutzung hat ermöglichen wollen.⁸⁶⁶

Allerdings geht ErwGr. 17 DSM-RL, der die Vergütungsfreiheit des Text und Data Mining zum Zwecke wissenschaftlicher Forschung begründet, auf den im Jahr 2016 veröffentlichten Richtlinienentwurf der EU-Kommission zurück.⁸⁶⁷ Dieser sah eine Schrankenbestimmung für das Text und Data Mining allein zum Zwecke nicht-kommerziell tätiger Forschungsorganisationen vor.⁸⁶⁸ Zwar wurde die Schrankenbestimmung auf Initiative des Europäischen Parlaments im Gesetzgebungsprozess auf allgemeines Text und Data Mining ausgeweitet. Dies brachte auch die Einfügung zusätzlicher Erwägungsgründe mit sich.⁸⁶⁹ Die Formulierung des ErwGr. 17 DSM-RL (ex. ErwGr. 13 DSM-RL-E) wurde allerdings ausdrücklich auf das Text und Data Mining zu wissenschaftlichen Zwecken beschränkt. Diese Historie macht deutlich, dass der Richtliniengeber vor dem Hintergrund einer nur geringfügigen Beeinträchtigung der Urheberinteressen nur von einer Vergütungsfreiheit des Text und Data Mining zu wissenschaftlichen Zwecken (Art. 3 DSM-RL) ausgegangen ist. Ein allgemeiner Rechtsgedanke, der auch das allgemeine, in Art. 4 DSM-RL geregelte Text und Data Mining erfasst, lässt sich aus dem Erwägungsgrund daher nicht ableiten. Der Richtlinie kann demzufolge insgesamt kein Rechtsgedanke entnommen

862 BT-Drs. 19/27426, S. 88; *Raue*, ZUM 2020, 172 (173); *Raue*, ZUM 2021, 793 (795).

863 ErwGr. 17 DSM-RL.

864 *Kleinkopf*, Text- und Data-Mining, S. 234 f.; so auch *Stieper*, ZUM 2021, 776 (778), wobei sich die Aussage hier nur auf eine rechtspolitische Dimension erstreckt.

865 *Heesen/Jüngels*, RuZ 2021, 45 (53).

866 *Heesen/Jüngels*, RuZ 2021, 45 (53).

867 Vgl. ErwGr. 13 DSM-RL-E, siehe COM(2016) 593 final.

868 Art. 3 DSM-RL-E, siehe COM(2016) 593 final.

869 Siehe zum Parlamentsbeschluss in erster Lesung: P8_TA-PROV(2019)0231.

werden, nach dem das allgemeine Text und Data Mining vergütungsfrei ausgestaltet sein sollte.⁸⁷⁰ Ein Umkehrschluss aus ErwGr. 17 DSM-RL und die Entstehungsgeschichte des Erwägungsgrundes legen vielmehr einen gegenteiligen Schluss nahe.

Weiterhin bringt der Richtliniengeber in ErwGr. 36 InfoSoc-RL zum Ausdruck, dass die Mitgliedstaaten bei der spezifischen Umsetzung der Schrankentatbestände dort einen gerechten Ausgleich für Urheber vorsehen sollen, wo es im konkreten Fall insbesondere aufgrund des Dreistufentests notwendig ist; auch, wenn in Art. 5 InfoSoc-RL ein solcher Ausgleich nicht vorgesehen ist.⁸⁷¹ Hieraus kann ein allgemeiner Rechtsgrundsatz des Richtliniengebers beim Umgang mit urheberrechtlichen Schrankenbestimmungen abgeleitet werden, nach dem die Einführung einer Vergütungspflicht innerhalb der allgemeinen unionsrechtlichen Grenzen grundsätzlich den Mitgliedstaaten freisteht. Dort, wo er diesen durchbrechen wollte, hat er es ausdrücklich kenntlich gemacht. Das ist beispielsweise in ErwGr. 17 DSM-RL, nicht aber in Bezug auf Art. 4 DSM-RL der Fall.

Nicht zuletzt zielt das Text und Data Mining zwar darauf ab, die den vervielfältigten Werken zugrundeliegenden Informationen zu verarbeiten. Das ändert jedoch nichts daran, dass nicht nur die Informationen, sondern die gesamten Werke inklusive ihres eigenschöpferischen Gehalts vervielfältigt werden.

Insgesamt sprechen damit die besseren Gründe dafür, dass die DSM-RL den Mitgliedstaaten die Einführung einer Vergütungspflicht für das in Art. 4 DSM-RL geregelte allgemeine Text und Data Mining freistellt.⁸⁷²

b) Trainingsdatensammler als Vergütungsschuldner

Auch die Zuweisung der Kostenlast zu den Trainingsdatensammlern müsste mit den unionsrechtlichen Bestimmungen vereinbar sein.

870 So auch *Wandtke*, NJW 2019, 1841 (1842).

871 Zur Notwendigkeit einer angemessenen Vergütung der Urheber siehe oben unter 2. Teil B. II. 1.

872 *Wandtke*, MMR 2017, 367 (368); *Wandtke*, NJW 2019, 1841 (1842); *Württemberg/Freischem*, GRUR 2019, 1140 (1142); halten es zwar für unionsrechtlich möglich, aber für rechtlich nicht zielführend: *Stieper*, GRUR 2020, 1 (4); *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (29). Andere Ansicht *Hauk/Pflüger*, ZUM 2020, 383 (385); *Heesen/Jüngels*, RuZ 2021, 45 (53); *Raue*, ZUM 2021, 793 (795); *Kleinkopf*, Text- und Data-Mining, S. 234 f.; BT-Drs. 19/27426, S. 88.

Weder die InfoSoc-RL noch die DSM-RL treffen selbst Aussagen darüber, welchem Beteiligten eine Vergütungspflicht für Vervielfältigungshandlungen beim Sammeln von urheberrechtlich geschützten Trainingsdaten auferlegt werden kann. Ziel der Vergütung ist aber der Ausgleich wirtschaftlicher Nachteile, die Urheber durch freigestellte Werknutzungen erleiden.⁸⁷³ Im Umkehrschluss sollte die Belastung bei dem Beteiligten verortet werden, der die Nutzungshandlung im urheberrechtlichen Sinne vornimmt.⁸⁷⁴ Denn er zieht die Vorteile aus der Werknutzung, dessen Kehrseite den Nachteil für betroffene Urheber begründet. Bestehen praktische Schwierigkeiten bei der Inanspruchnahme des Werknutzers, kann die Vergütungspflicht allerdings auch auf einen anderen Beteiligten verlagert werden. In diesen Fällen muss jedoch die Möglichkeit bestehen, die Vergütungsleistung an den eigentlichen Werknutzer weiterzureichen sowie einen Erstattungsmechanismus zu bedienen, wenn die Vergütung tatsächlich nicht geschuldet gewesen ist.⁸⁷⁵

Die vorgeschlagene Vergütungspflicht knüpft an den Trainingsdatensammler an. Dieser stellt die Vervielfältigungen für das Training von KI her. Er ist demzufolge Werknutzer. Die Anknüpfung an die Trainingsdatensammler ist infolgedessen mit den unionsrechtlichen Anforderungen vereinbar.

c) Zwischenergebnis

Alles in allem ist die Einführung der in §§ 44c ff. UrhG-E umzusetzenden Vergütungspflicht für das Web Scraping von urheberrechtlich geschützten Trainingsdaten auf Basis von § 44b UrhG unionsrechtlich zulässig.

IV. Zusammenfassung und kritische Würdigung des Regelungsvorschlags

Insgesamt können die identifizierten Regelungsbedürfnisse für das Web Scraping von Werken als Trainingsdaten insbesondere durch die Anfügung eines vierten Absatzes in der Schrankenbestimmung für Text und Data Mi-

873 ErwGr. 35 InfoSoc-RL.

874 Für die Privatkopievergütung EuGH, Urt. v. 24.03.2022 - C-433/20, GRUR 2022, 558 Rn. 43 – Austro-Mechana/Strato.

875 Für die Privatkopievergütung EuGH, Urt. v. 24.03.2022 - C-433/20, GRUR 2022, 558 Rn. 45, 47 – Austro-Mechana/Strato.

ning aus § 44b UrhG verwirklicht werden. § 44b Abs. 4 UrhG-E bestimmt, dass weder Löschungspflicht noch Opt-out-Mechanismus bei Vervielfältigungshandlungen für das Sammeln von Trainingsdaten eingreifen sollen. Außerdem ist eine Klarstellung zur Anwendbarkeit der Schrankenbestimmung auf das Sammeln urheberrechtlich geschützter Trainingsdaten für die Entwicklung von KI zweckmäßig. Sie sollte als Satz 2 in § 44b Abs. 1 UrhG eingefügt werden.

Die Bereichsausnahme von Löschungserfordernis und Opt-out-Mechanismus kann aufgrund der Harmonisierungswirkung der DSM-RL aber nicht ohne weiteres im nationalen Urheberrecht umgesetzt werden. Hierfür fehlt es an einer entsprechenden Regelung oder Öffnungsklausel auf unionsrechtlicher Ebene. Der Umsetzung des § 44b Abs. 4 UrhG-E muss demzufolge eine europäische Rechtssetzungsinitiative vorangehen.

Die gesetzliche Vergütung, die Urheber für die Verwendung ihrer Werke als Trainingsdaten erhalten sollen, kann durch Einfügung der §§ 44c bis 44e UrhG-E umgesetzt werden. Die Einfügung des Vergütungsmechanismus ist auch unionsrechtlich zulässig. Er ist nach dem Vorbild der Hersteller- und Betreibervergütung (§§ 54 ff. UrhG) ausgestaltet. Dies bietet sich an, weil auch die Hersteller- und Betreibervergütung Pauschalisierungen bei Vergütungshöhe und Anspruchsinhaberschaft vorsieht. Beides ist auch im vorliegenden Fall notwendig. Vorteilhaft an einer Umsetzung nach dem Vorbild der Hersteller- und Betreibervergütung ist außerdem, dass den Verwertungsgesellschaften, die die Vergütungsansprüche geltend machen sollen (§ 44e Abs. 1 UrhG-E), die grundlegenden Mechanismen bekannt sind.

B. Speicherung urheberrechtlich geschützter Inputdaten

Die Speicherung urheberrechtlich geschützter Inputdaten wird von der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a UrhG erfasst.⁸⁷⁶ Dies gilt jedenfalls so lange, wie die erfassten Inhalte nicht zur Verwirklichung von Sekundärzwecken abesondert und die Werke tatsächlich nur vorübergehend gespeichert werden. Außerdem sind die Vervielfälti-

876 Dazu 3. Teil D. II. 1.

gungshandlungen in der Anwendungsphase des KI-Systems auch von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG gedeckt.⁸⁷⁷ Damit sind sie im Grundsatz urheberrechtlich erlaubnisfrei möglich.

Regulatorisch zweckmäßig ist allerdings, die Speicherung von Werken als Inputdaten für KI durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte, also bereits in den §§ 15 ff. UrhG definitorisch aus dem Ausschließlichkeitsrecht betroffener Urheber auszuklammern.⁸⁷⁸

I. Unmittelbar begrenzende Ausgestaltung des urheberrechtlichen Vervielfältigungsrechts

Damit die Speicherung urheberrechtlich geschützter Inputdaten durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte freigestellt wird, muss eine Anpassung des Vervielfältigungsrechts aus § 16 UrhG erfolgen. Eine entsprechende Regelung könnte wie folgt aussehen:

§ 16 UrhG-E: Vervielfältigungsrecht⁸⁷⁹

- (1) *Das Vervielfältigungsrecht ist das Recht, Vervielfältigungsstücke des Werkes herzustellen, gleichviel ob vorübergehend oder dauerhaft, in welchem Verfahren und in welcher Zahl.*
- (2) *Eine Vervielfältigung ist auch die Übertragung des Werkes auf Vorrichtungen zur wiederholbaren Wiedergabe von Bild- oder Tonfolgen (Bild- oder Tonträger), gleichviel, ob es sich um die Aufnahme einer Wiedergabe des Werkes auf einen Bild- oder Tonträger oder um die Übertragung des Werkes von einem Bild- oder Tonträger auf einen anderen handelt.*
- (3) **Keine Vervielfältigung ist die vorübergehende Zwischenspeicherung von Werken zu dem ausschließlichen Zweck, sie als Eingabedaten (Art. 3 Nr. 33 VO [EU] 2024/1689) für die Anwendung eines Systems künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG-E) zu verwenden.**

Ziel der Regelung ist dabei die Einschränkung des über § 95a UrhG gewährten Schutzes für technische Schutzmaßnahmen, die gegen die Speicherung von Inputdaten gerichtet sind.⁸⁸⁰ Demzufolge kommt es nicht entscheidend auf die mit § 16 Abs. 3 UrhG-E hervorgerufene Darlegungs- und Beweislastverteilung an. Aufgrund der negativen Formulierung des

877 Dazu 3. Teil D. II. 2. a).

878 Dazu im Detail siehe oben unter 2. Teil D. II.

879 Hervorhebung steht für neu einzufügenden Normtext.

880 Siehe im Detail unter 2. Teil D. II. 2.

Regelungsvorschlags wäre der Nutzer für die Freistellungsvoraussetzungen im Rahmen des § 16 Abs. 3 UrhG-E darlegungs- und ggf. beweisbelastet.

II. Unionsrechtliche Perspektive auf den Regelungsvorschlag und Schlussfolgerungen

Auch das Vervielfältigungsrecht ist durch Art. 2 InfoSoc-RL unionsrechtlich harmonisiert. Nationale Bestimmungen dürfen jedenfalls nicht hinter dem durch das Unionsrecht gewährten Schutzniveau zurückbleiben. Dies wäre bei einer Einschränkung des Vervielfältigungsrechts entsprechend § 16 Abs. 3 UrhG-E allerdings der Fall. Eine solche Rechtsänderung wäre daher unionsrechtswidrig. Auch die regulatorisch notwendigen Anpassungen am urheberrechtlichen Vervielfältigungsrecht könnten deswegen nur umgesetzt werden, wenn ihr eine europäische Rechtssetzungsinitiative vorgeht.

C. Normalisierung urheberrechtlich geschützter Trainings- und Inputdaten als Teil der Datenvorverarbeitung

Bei der Einordnung der Vorverarbeitung urheberrechtlich geschützter Trainings- sowie Inputdaten in den Rechtsrahmen des geltenden Urheberrechts muss zwischen der Normalisierung und dem Feature Encoding der Werke unterschieden werden. Bei der Normalisierung von Werken können – je nach eingesetztem Normalisierungsverfahren – Vervielfältigungshandlungen gemäß § 16 Abs. 1 UrhG stattfinden. Diese sind sowohl von der Schrankenbestimmung für vorübergehende Vervielfältigungen (§ 44a Nr. 2 UrhG) als auch von der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) gedeckt und daher im Grundsatz urheberrechtlich erlaubnisfrei möglich. Beim nicht-expressiven Feature Encoding der Werke fehlt es hingegen bereits an der Herstellung eines Vervielfältigungsstücks im Sinne des § 16 Abs. 1 UrhG. Auch diese unmittelbar aus der Reichweite der urheberrechtlichen Verwertungsrechte folgende Rechtslage macht das Feature Encoding im Ergebnis erlaubnisfrei möglich.

Sowohl das Feature Encoding als auch die Normalisierung von als Input- sowie Trainingsdaten generierten Werken sollte aus einer regulatorischen Perspektive jedoch durch eine unmittelbar begrenzende, definitonische Ausgestaltung des Vervielfältigungsrechts aus dem urheberrechtlichen Schutzbereich ausgeklammert sein.⁸⁸¹ Für die Normalisierung der Werke bedarf es damit trotz der Anwendbarkeit mehrerer Schrankenbestimmungen de lege ferenda einer Anpassung des urheberrechtlichen Regelungsregimes.

I. Unmittelbar begrenzende Ausgestaltung des urheberrechtlichen Vervielfältigungsrechts

Damit die Normalisierung urheberrechtlich geschützter Werke durch eine unmittelbar begrenzende Ausgestaltung des Vervielfältigungsrechts freigestellt wird, muss eine Anpassung in § 16 UrhG erfolgen. Ziel der Regelung ist die mittelbare Einschränkung des über § 95a UrhG erfolgenden Schutzes von gegen die Normalisierungsschritte eingesetzten technischen Schutzmaßnahmen.⁸⁸² Auch hier kommt es demzufolge nicht auf die aus der Regelung folgende Darlegungs- und Beweislastverteilung an. Da sich aus dem Freistellungsbedürfnis für die Speicherung von Werken als Inputdaten bereits Änderungsbedarf in Bezug auf § 16 UrhG ergeben hat,⁸⁸³ können beide Anpassungsvorschläge wie folgt zusammengeführt werden:

§ 16 UrhG-E: Vervielfältigungsrecht⁸⁸⁴

- (1) *Das Vervielfältigungsrecht ist das Recht, Vervielfältigungsstücke des Werkes herzustellen, gleichviel ob vorübergehend oder dauerhaft, in welchem Verfahren und in welcher Zahl.*
- (2) *Eine Vervielfältigung ist auch die Übertragung des Werkes auf Vorrichtungen zur wiederholbaren Wiedergabe von Bild- oder Tonfolgen (Bild- oder Tonträger), gleichviel, ob es sich um die Aufnahme einer Wiedergabe des Werkes auf einen Bild- oder Tonträger oder um die Übertragung des Werkes von einem Bild- oder Tonträger auf einen anderen handelt.*
- (3) **Keine Vervielfältigung sind folgende Werkverwendungen:**
 1. **die vorübergehende Zwischenspeicherung von Werken zu dem ausschließlichen Zweck, sie als Eingabedaten (Art. 3 Nr. 33 VO [EU] 2024/1689) für die**

881 Dazu im Detail siehe oben unter 2. Teil C. sowie E.

882 Siehe im Detail unter 2. Teil C. II. 2. sowie 2. Teil D. II. 2.

883 Dazu siehe oben unter 4. Teil B. I.

884 Hervorhebung steht für neu einzufügenden Normtext.

- Anwendung eines Systems künstlicher Intelligenz (§ 44b Abs. 1 S. 2 UrhG) zu verwenden, und**
- 2. die Transponierung von als Trainings- oder Eingabedaten (Art. 3 Nr. 33 VO [EU] 2024/1689) für ein System künstlicher Intelligenz erfassten Werken auf eine vergleichbare Skala (Normalisierung).**

II. Unionsrechtliche Perspektive auf den Regelungsvorschlag und Schlussfolgerungen

Auch bei einer Freistellung der Normalisierung von urheberrechtlich geschützten Trainings- sowie Inputdaten muss allerdings berücksichtigt werden, dass das urheberrechtliche Vervielfältigungsrecht durch Art. 2 InfoSoc-RL harmonisiert ist. Eine dem § 16 Abs. 3 UrhG-E entsprechende Rechtsänderung wäre daher, wie bereits für die Speicherung von Werken als Inputdaten gezeigt werden konnte, unionsrechtswidrig. Auch der Anpassung des urheberrechtlichen Regelungsregimes für die Normalisierung von Werken zur Verwendung als Trainings- und Inputdaten müsste deswegen eine europäische Rechtssetzungsinitiative vorausgehen.

Fünfter Teil: Zusammenfassung und Schlussfolgerungen aus der Untersuchung

Urheberrechtlich geschützte Werke werden in erheblichem Umfang für das Training von KI, also von auf künstlichen neuronalen Netzen basierenden Systemen maschinellen Lernens, benötigt. Außerdem können sie als Inputdaten und damit als Berechnungsgrundlage für die Anwendung von KI-Systemen im konkreten Einzelfall dienen. Entwicklung sowie Einsatz von KI und das urheberrechtliche Regelungsregime stehen daher in einer engen Beziehung zueinander. Vor diesem Hintergrund muss untersucht werden, welche Rahmenbedingungen das Urheberrecht für die Verwendung von Werken als Trainings- und Inputdaten zur Verfügung stellt. Fraglich ist außerdem, ob Perspektiven für den urheberrechtlichen Rechtsrahmen aufgezeigt werden können. Konkret sind die aufgeworfenen Fragestellungen dabei in Bezug auf die praktisch relevante Speicherung urheberrechtlich geschützter Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern, die eigentliche Durchführung des Trainings auf Basis von Werken, die Speicherung urheberrechtlich geschützter Inputdaten für die Anwendung von KI-Systemen in konkreten Anwendungsszenarien sowie die notwendige Vorverarbeitung dieser gewonnenen Rohdaten in den Blick zu nehmen. Um Perspektiven für den urheberrechtlichen Rechtsrahmen aufzuzeigen, müssen für die aufgezeigten Fallgruppen außerdem regulatorische Freistellungsbedürfnisse identifiziert werden. Aus einem Vergleich dieser Freistellungsbedürfnisse, also des Soll-Zustands, mit der urheberrechtlichen *Lex lata*, dem Ist-Zustand des Regelungsregimes, lassen sich Änderungsbedarfe im Urheberrecht ermitteln. Auf Basis dieser Änderungsbedarfe können anschließend konkrete Anpassungsvorschläge für den urheberrechtlichen Regelungskomplex erarbeitet werden.

Die Speicherung urheberrechtlich geschützter Trainingsdaten mit Hilfe von Webcrawlern sollte urheberrechtlich freigestellt werden. Grund hierfür ist, dass die Innovationsinteressen der Trainingsdatensammler sowie bestehende praktische Lizenzierungsschwierigkeiten das nur geringfügig beeinträchtigte Verwertungsinteresse betroffener Urheber überwiegen. Das Freistellungsinteresse wird dem Grunde nach durch die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfüllt. Dies ergibt sich aus einer richtlinienkonformen Auslegung des Schrankentatbestands. Die Re-

gelung privilegiert auch Vervielfältigungshandlungen kommerziell tätiger Trainingsdatensammler. Dennoch können nicht sämtliche der identifizierten Regelungsbedürfnisse mit der Schrankenbestimmung aus § 44b UrhG erfüllt werden. Denn zum einen ist eine Kompensation der freigestellten Speicherung urheberrechtlich geschützter Trainingsdaten zugunsten betroffener Urheber mittels eines gesetzlichen Vergütungsanspruchs notwendig. Einen solchen Vergütungsanspruch sieht § 44b UrhG aber nicht vor. Darüber hinaus sind der in § 44b Abs. 3 UrhG angelegte Opt-out-Mechanismus, der durch Vorbehaltserklärungen in natürlicher Sprache ausgelöst werden kann, sowie die in § 44b Abs. 2 S. 2 UrhG vorgesehene Löschungspflicht für Trainingsdatensammler nicht interessengerecht. Im Einzelfall kann eine Freistellungswirkung zugunsten von Trainingsdatensammlern zwar auch über die Schrankenbestimmung für die Nutzung von Werken als unwesentliches Beiwerk aus § 57 UrhG erreicht werden. Das strukturelle Freistellungsinteresse kann hierdurch *de lege lata* allerdings nicht erfüllt werden. Die Schrankenbestimmung für Text und Data Mining ist *de lege ferenda* deswegen durch einen § 44b Abs. 4 UrhG-E zu ergänzen. Er soll für das Sammeln urheberrechtlich geschützter Trainingsdaten eine Bereichsausnahme von der Löschungspflicht und dem Opt-out-Mechanismus vorsehen. Da § 44b UrhG den Art. 4 DSM-RL in nationales Recht umsetzt, kann das allerdings nur im Zusammenspiel mit einer vorangehenden europäischen Rechtssetzungsinitiative erfolgen. Der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG ist mit §§ 44c bis 44e UrhG-E darüber hinaus eine gesetzliche Vergütungspflicht für die Vervielfältigung von Werken zum Training von KI anzufügen. Sie ist nach dem Vorbild der Hersteller- und Betreibervergütung (§§ 54 ff. UrhG) auszugestalten. Die Einführung dieses Vergütungsmechanismus ist unionsrechtlich bereits jetzt zulässig.

Auch für die eigentliche Durchführung des KI-Trainings, also die Anpassung der numerischen Parameter des KNN auf Basis der zum Training gesammelten Werke, konnte ein regulatorisches Freistellungsinteresse ermittelt werden. Sie ist tatsächlich und rechtlich von der Aggregation der Trainingsdaten abzugrenzen. Zweckmäßig ist dabei eine vergütungsfreie Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte. Damit das Freistellungsinteresse im geltenden Urheberrecht abgebildet wird, müssen alle vier Teilaspekte des Trainings diese regulatorischen Bedürfnisse erfüllen. Das ist bei der Umwandlung der Trainingsdaten in numerische Tensoren, dem sogenannten Feature Encoding, der eigentlichen Analyse der Trainingswerke im Rechen-

prozess, hierbei angefertigter, vorübergehender Kopien der in Tensoren umgewandelten Trainingsdaten sowie insbesondere auch der Anpassung der Gewichtungswerte und anderen Parameter innerhalb des KNN der Fall. Denn im Rahmen dieser Teilschritte werden keine Vervielfältigungsstücke der Werke im Sinne des § 16 UrhG angefertigt. Das gilt selbst dann, wenn einzelne Werke in den Strukturen des KNN subsymbolisch gespeichert und im Einzelfall wieder rekonstruiert werden können. Diese Aspekte des Trainingsvorgangs berühren demzufolge nicht den Schutzbereich des urheberrechtlichen Ausschließlichkeitsrechts. Etwas anderes gilt nur für die Normalisierung von urheberrechtlich geschützten Trainingsdaten, also das Transponieren der Werke auf eine vergleichbare Skala. Je nach Normalisierungsverfahren und Werkart können hierbei Vervielfältigungshandlungen im Sinne des § 16 UrhG vorgenommen werden. Auch wenn die Normalisierung der Trainingsdaten von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfasst und damit freigestellt ist, wird der geltende urheberrechtliche Rechtsrahmen damit nicht dem *Wie* des identifizierten Regelungsbedürfnisses gerecht. Als Folge dessen ist § 16 UrhG durch einen § 16 Abs. 3 UrhG-E zu ergänzen. Er nimmt unter anderem die Normalisierung von als Trainingsdaten erfassten Werken vom Vervielfältigungsbegriff aus. Da das Vervielfältigungsrecht unionsrechtlich harmonisiert ist, muss dem ebenfalls eine europäische Rechtssetzungsinitiative vorangehen. Eine isoliert nationale Anpassung des UrhG ist im konkreten Fall nicht möglich.

Ebenfalls von § 16 Abs.3 UrhG-E umfasst werden sollte die Normalisierung von als Inputdaten für die Anwendung von KI-Systemen gespeicherten Werken. Sie muss de lege lata ebenfalls unter den Vervielfältigungsbegriff aus § 16 UrhG gefasst werden. Auch wenn die Normalisierungsvervielfältigungen der als Inputdaten gespeicherten Werke ebenfalls so lange von § 44b UrhG privilegiert und damit freigestellt werden, wie kein wirksamer Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist, löst dies vor dem Hintergrund des regulatorischen Freistellungsinteresses ein Regelungsdefizit im geltenden urheberrechtlichen Rechtsrahmen aus. Solange die Inputdaten nicht zur Verwirklichung von Sekundärzwecken abgesondert und nur vorübergehend vervielfältigt werden, ergibt sich eine Erlaubnis der Normalisierung urheberrechtlich geschützter Inputdaten weiterhin aus der Schrankenbestimmung für vorübergehende Vervielfältigungen, § 44a UrhG. Damit kann das regulatorische Interesse nach einer Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte allerdings gleichsam nicht erfüllt werden.

Im Übrigen können die für die Vorverarbeitung von Trainingsdaten herausgearbeiteten Grundsätze auf die Vorverarbeitung von Werken übertragen werden, die als Input für KI-Systeme verwendet werden sollen. Das führt dazu, dass bei der Umwandeln als Input verwendeter Werke in numerische Repräsentationen, also Tensoren, keine Vervielfältigungsstücke im Sinne des § 16 UrhG hergestellt werden. Das Feature Encoding der als Inputdaten verwendeten Werke greift demzufolge nicht in das urheberrechtliche Ausschließlichkeitsrecht ein. Diesbezüglich besteht infolgedessen kein Konflikt zwischen den identifizierten regulatorischen Freistellungsinteressen und dem geltenden Urheberrecht. Für das Feature Encoding der urheberrechtlich geschützten Inputdaten ist daher keine Anpassung des urheberrechtlichen Rechtsrahmens erforderlich.

Auch die vorgelagerte Speicherung urheberrechtlich geschützter Inputdaten beispielsweise durch Sensoren wie Kameras oder mit Hilfe von Webcrawlern sollte urheberrechtlich freigestellt sein. Infolge einer Zusammenschau verschiedener Begründungsansätze erweist sich die Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte als interessengerecht. Diesem Regelungsbedürfnis wird der geltende Rechtsrahmen jedoch nicht gerecht. Bei der Speicherung urheberrechtlich geschützter Inputdaten werden vorübergehende Vervielfältigungen hergestellt, die in das ausschließliche Vervielfältigungsrecht der Urheber nach § 16 UrhG eingreifen. Die Anwendung einer KI kann allerdings sowohl unter den Begriff des Text und Data Mining (§ 44b Abs. 1 UrhG) subsumiert als auch als rechtmäßige Nutzung (§ 44a Nr. 2 UrhG) eingeordnet werden. Solange mit den Werkstücken keine Sekundärverwertung bezweckt wird, die eine dauerhafte Speicherung der Inhalte notwendig macht, sind die Vervielfältigungshandlungen daher von den Schrankenbestimmungen für vorübergehende Vervielfältigungen aus § 44a UrhG sowie für Text und Data Mining aus § 44b UrhG erlaubnisfrei gestellt. Letzteres aber ist aber nur dann der Fall, wenn kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist. Dennoch wird mit der Anwendung der Schrankenbestimmungen nur dem *Ob*, nicht aber dem *Wie* der identifizierten Freistellungsbedürfnisse Rechnung getragen. Infolgedessen ist in dem neu einzufügenden § 16 Abs. 3 UrhG-E festzulegen, dass bei der vorübergehenden Zwischenspeicherung von Werken zu dem ausschließlichen Zweck, sie als Inputdaten für die Anwendung eines KI-Systems zu verwenden, keine Vervielfältigungsstücke im Sinne des § 16 Abs. 1 UrhG hergestellt werden. Aufgrund der unionsrechtlichen Vollharmonisierung des Vervielfältigungsrechts

fältigungsrechts kann auch das allerdings nur im Zusammenspiel mit einer unionsrechtlichen Rechtssetzungsinitiative erfolgen.

Alles in allem stellt das Urheberrecht einen Rechtsrahmen zur Verfügung, der sowohl das Training als auch den Einsatz von KI auf Basis urheberrechtlich geschützter Trainings- beziehungsweise Inputdaten grundsätzlich ermöglicht. Von besonderer Bedeutung hierfür ist die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG. Maßgebend für die urheberrechtliche Erlaubnisfreiheit des Trainings ist auch, dass keine Vervielfältigungsstücke der zum Training verwendeten Werke dauerhaft im KNN im Sinne des § 16 Abs.1 UrhG hinterlegt werden. Ein Eingriff in das urheberrechtliche Vervielfältigungsrecht findet hierdurch demzufolge nicht statt. Hierfür kommt es nicht entscheidend auf die technische Funktionsweise des KI-Trainings an. Bedeutung erlangt vielmehr ein nicht rein-technisches, sondern normatives Verständnis des urheberrechtlichen Vervielfältigungsbegriffs.

Insbesondere der auch in natürlicher Sprache erklärbarer Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG sowie das Fehlen einer Vergütungspflicht für die Speicherung von Werken als Trainingsdaten zugunsten betroffener Urheber, aber auch andere Bestandteile des urheberrechtlichen Regelungsregimes führen allerdings dazu, dass sich die aus den Sachverhaltskonstellationen extrahierbaren Interessenslagen nicht vollständig im geltenden Urheberrecht abbilden. Nationaler und europäischer Gesetzgeber sind deswegen angehalten, den urheberrechtlichen Rechtsrahmen für das Training und die Anwendung von KI einer Revision zu unterziehen. Aufgrund des unionsrechtlichen Mehrebenensystems kommt es hierbei insbesondere auf den europäischen Richtliniengeber an. Der nationale Gesetzgeber kann jedoch bereits jetzt ein Vergütungsmodell für die Nutzung von Werken als Trainingsdaten für KI-Entwicklung einführen, ohne dass dies unionsrechtlich vorgezeichnet sein müsste. Bei dieser Gelegenheit sollte er in § 44b Abs.1 UrhG auch eine Klarstellung dahingehend einfügen, dass das KI-Training unter den Text und Data Mining-Begriff der Regelung fällt. Dabei kann er auch die sich aus Art. 4 Abs.3 DSM-RL ergebenden Voraussetzungen der Angemessenheit und Ausdrücklichkeit von Nutzungsvorbehaltserklärungen im nationalen Regelungstext umsetzen. Sie können aktuell nur im Wege einer richtlinienkonformen Auslegung des Tatbestands berücksichtigt werden.

Literaturverzeichnis

- Ahlberg, Hartwig/Götting, Horst-Peter/Lauber-Rönsberg, Anne (Hrsg.):* Beck'scher Online-Kommentar Urheberrecht, 43. Edition Stand 15. Februar 2024, München 2024
- Akinci, Nick/Heidrich, Jörg:* Der digitale Schöpfer: Urheberrechtsfragen im Zeitalter der künstlichen Intelligenz, in: IPRB 2023, S. 270–275
- Alpaydin, Ethem:* Machine learning: the new AI, Cambridge 2016
- Alwosheel, Ahmad/Van Cranenburgh, Sander/Chorus, Caspar G.:* Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis, in: Journal of Choice Modelling 2018, S. 167–182
- Anderhuber, Friedrich/Pera, Franz/Streicher, Johannes (Hrsg.):* Waldeyer - Anatomie des Menschen, Berlin/Boston 2012
- Antoine, Lucie:* Verändernde Werknutzungen: Computerprogramme und der urheberrechtliche Interessenausgleich, Tübingen 2023
- Antoine, Lucie:* Der Monismus des deutschen Urheberrechts und die unionsrechtlichen Verwertungsrechte – ein gordischer Knoten?, in: ZGE 2023, S. 387–413
- Apel, Simon:* Anmerkung zu EuGH, Urt. v. 29.7.2019 - C-476/17, Pelham u.a./Hütter u.a. [„Metall auf Metall“], in: MMR 2019, S. S.601–603
- Bachmann, Hans-Gert:* Vom Erz zum Metall (Kupfer, Silber, Eisen) – Die chemischen Prozesse im Schaubild, in: Archäologie in Deutschland 1993, S. 35–40
- Baktash, Jawid Ahmad/Davodi, Mursal:* Gpt-4: A Review on Advancements and Opportunities in Natural Language Processing, in: arXiv preprint 2023, abrufbar unter: <http://arxiv.org/abs/2305.03195> (zuletzt abgerufen am 18.11.2024)
- Ballard, Dana Harry/Brown, Christopher M.:* Computer vision, Englewood Cliffs 1982
- Barudi, Malek (Hrsg.):* Das neue Urheberrecht, Baden-Baden 2021
- Baumann, Malte:* Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld, in: NJW 2023, S. 3673–3678
- Bäcker, Kerstin:* Der Text und Data Mining Vorbehalt: Technische Umsetzung der Maschinenlesbarkeit, in: Lausen Rechtsanwälte Blog v. 22.7.2024, abrufbar unter: <https://www.lausen.com/blog/der-text-und-data-mining-vorbehalt-technische-umsetzung-der-maschinenlesbarkeit/> (zuletzt abgerufen am 6.12.2024)
- Beck, Susanne:* Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, in: JR 2009, S. 225–230
- Beck, Susanne/Kusche, Carsten/Valerius, Brian (Hrsg.):* Digitalisierung, Automatisierung, KI und Recht, 1. Auflage, Baden-Baden 2020
- Beck, Susanne/Kusche, Carsten/Valerius, Brian (Hrsg.):* Digitalisierung, Automatisierung, KI und Recht, Baden-Baden 2020
- Becker, Maximilian:* Ein modernes Urheberrecht. Von der Nutzungshandlung zum digitalen Lebensbereich, in: ZGE 2016, S. 239–288
- Becker, Maximilian:* Das Urheberrecht als Trostpreis für den Menschen?, in: GRUR 2024, S. 505–514

- Bell, Jason*: Machine learning: hands-on for developers and technical professionals, Indianapolis (IN, USA) 2015
- Bergström, Svante*: Die neue schwedische Gesetzgebung über das Urheberrecht, in: GRUR Ausl 1962, S. 364–381
- Bischoffshausen, Albrecht*: Die ökonomische Rechtfertigung der urheberrechtlichen Schutzfrist: Analyse der Schutzfrist de lege lata und de lege ferenda aus historischer, dogmatischer und rechtsökonomischer Sicht, Baden-Baden 2013
- Bisges, Marcel*: Ökonomische Analyse des Urheberrechts, in: ZUM 2014, S. 930–938
- Bitkom e.V.*: BITKOM Positionspapier Künstliche Intelligenz: Mehr Sachlichkeit bitte, abrufbar unter: <https://www.bitkom.org/sites/main/files/file/import/BITKOM-Positionspapier-Kuenstliche-Inteligenz.pdf> (zuletzt abgerufen am 16.10.2024)
- Bomhard, David*: KI-Training mit fremden Daten – IP-rechtliche Herausforderungen rund um § 44b UrhG, in: DSRITB 2023, S. 255–269
- Bomhard, David*: Text und Data Mining auf Grundlage von Webcrawling und Web-scraping, in: InTeR 2023, S. 174–179
- Bomhard, David/Siglmüller, Jonas*: AI Act – das Trilogergebnis, in: RD i 2024, S. 45–55
- Borges, Georg*: Haftung für selbstfahrende Autos - Warum eine Kausalhaftung für selbstfahrende Autos gesetzlich geregelt werden sollte, in: CR 2016, S. 272–280
- Borges, Georg*: Rechtliche Rahmenbedingungen für autonome Systeme, in: NJW 2018, S. 977–982
- Borges, Georg*: Haftung für KI-Systeme - Konzepte und Adressaten der Haftung, in: CR 2022, S. 553–561
- Borges, Georg/Hilber, Marc (Hrsg.)*: Beck'scher Online-Kommentar IT-Recht, 14. Edition Stand 1.4.2024, München 2024
- Bräutigam, Peter/Rücker, Daniel (Hrsg.)*: E-Commerce – Rechtshandbuch, München 2017
- Brockman, John (Hrsg.)*: Was sollen wir von Künstlicher Intelligenz halten? die führenden Wissenschaftler unserer Zeit über intelligente Maschinen, 2. Auflage, Frankfurt am Main 2017
- Brockmann, John (Hrsg.)*: Was sollen wir von Künstlicher Intelligenz halten?, Frankfurt a.M. 2017
- Brockmeyer, Henning*: Text und Data Mining: eine rechtsökonomische Analyse der neuen Schranken im Urheberrecht, München 2022
- Brown, Tom B./Mann, Benjamin/Ryder, Nick/Subbiah, Melanie/Kaplan, Jared/Dhariwal, Prafulla/Neelakantan, Arvind/Shyam, Pranav/Sastry, Girish/Askill, Amanda/Agarwal, Sandhini/Herbert-Voss, Ariel/Krueger, Gretchen/Henighan, Tom/Child, Rewon/Ramesh, Aditya/Ziegler, Daniel M./Wu, Jeffrey/Winter, Clemens/Hesse, Christopher/Chen, Mark/Sigler, Eric/Litwin, Mateusz/Gray, Scott/Chess, Benjamin/Clark, Jack/Berner, Christopher/McCandlish, Sam/Radford, Alec/Sutskever, Ilya/Amorei, Dario*: Language Models are Few-Shot Learners, arXiv preprint 2020, abrufbar unter: <http://arxiv.org/abs/2005.14165> (zuletzt abgerufen am 21.10. 2024)
- Brox, Hans/Walker, Wolf-Dietrich*: Allgemeiner Teil des BGB, 48. Auflage, München 2024

- Brunotte, Nico*: Virtuelle Assistenten - Digitale Helfer in der Kundenkommunikation, in: CR 2017, S. 583–589
- Buchalik, Barbaba/Gehrmann, Christine*: Von Nullen und Einsen zu Paragraphen: Der AI Act, ein Rechtscode für Künstliche Intelligenz, in: CR 2024, S. 145–153
- Bundesministerium für Wirtschaft und Energie (Hrsg.)*: Einsatz von KI in der Deutschen Wirtschaft – Stand der KI-Nutzung im Jahr 2019, Berlin 2020, abrufbar unter: <https://www.bmwk.de/Redaktion/DE/Publikationen/Wirtschaft/einsatz-von-ki-deutsche-wirtschaft.pdf>? (zuletzt abgerufen am: 16.10.2024)
- Burchardi, Sophie*: Risikotragung für KI-Systeme, Zur Zweckmäßigkeit einer europäischen Betreiberhaftung, in: EuZW 2022, S. 685–691
- Bünthe, Oliver*: Artstation: Künstler protestieren gegen KI-Kunstwerke und KI-Trainings, in: heise online v. 16.12.2022, abrufbar unter: <https://www.heise.de/news/Artstation-Kuenstler-protestieren-gegen-KI-Kunstwerke-7397005.html> (zuletzt abgerufen am 16.10.2024)
- Büscher, Wolfgang/Erdmann, Willi/Fuchs, Andreas/Jänich, Volker Michael/Loschelder, Michael/McGuire, Mary-Rose (Hrsg.)*: Rechtsdurchsetzung – Rechtsverwirklichung durch materielles Recht und Verfahrensrecht: Festschrift für Hans-Jürgen Ahrens zum 70. Geburtstag, Köln 2016 (zitiert als: Bearbeiter, in: FS Ahrens)
- v. Caemmerer, Ernst/Fischer, Robert/Nüßgens, Karl/Schmidt, Reimer (Hrsg.)*: Festschrift für Fritz Hauß zum 70. Geburtstag, Karlsruhe 1978 (zitiert als: FS Hauß)
- Calliess, Christian/Ruffert, Matthias (Hrsg.)*: EUV/AEUV – Kommentar, 6. Auflage, München 2022
- Carey-Wilson, Thomas/Simperl, Elena/Das, Arunav/Reklos, Ioannis*: Understanding data governance in AI: A lifecycle perspective – Report of the Open Data Institute v. 27.6.2024, abrufbar unter: <https://theodi.org/insights/reports/understanding-data-governance-in-ai-a-lifecycle-perspective> (zuletzt abgerufen am 22.11.2024)
- Carlini, Nicholas/Hayes, Jamie/Nasr, Milad/Jagielski, Matthew/Sehwag, Vikash/Tramèr, Florian/Balle, Borja/Ippolito, Daphne/Wallace, Eric*: Extracting Training Data from Diffusion Models, in: arXiv preprint 2023, abrufbar unter: <http://arxiv.org/abs/2301.13188> (zuletzt abgerufen am 21.10.2024)
- Chiou, Theodoros*: Copyright lessons on Machine Learning: what impact on algorithmic art?, in: JIPITEC 2019, S. 398–411
- Chopra, Samir/White, Laurence F.*: A legal theory for autonomous artificial agents, Ann Arbor 2011
- Cianciolo, Anna T./Sternberg, Robert J.*: Intelligence: a brief history, Malden (MA, USA) 2004
- Coase, Ronald*: The Problem of Social Cost, in: The Journal of Law and Economics 1960, S. 1–44
- Cocoru, Diana/Boehm, Mirko*: An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, 2016, abrufbar unter: <https://www.openforumeuropa.org/wp-content/uploads/2016/05/TDM-Paper-Diana-Cocoru-and-Mirko-Boehm.pdf> (zuletzt abgerufen am 17.10.2024)

- Conrad, Albrecht: Erweiterte kollektive Lizenzen – Legal Transplant oder Common Challenge? - Impulsvortrag im Rahmen des Online-Symposiums »Erweiterte Kollektive Lizenzen im Urheberrecht« des Instituts für Urheber- und Medienrecht in Kooperation mit dem Institut für Europäisches Medienrecht am 11.2.2022 in München, in: ZUM 2022, S. 436–444
- Conrad, Albrecht/Schubert, Tobias: How to Do Things with Code, in: GRUR 2018, S. 350–358
- Czerlinski, Jean/Goldstein, Daniel/Gigerenzer, Gerd/Goodie, Adam: Simple Heuristics That Make Us Smart, in: Behavioral and Brain Sciences 2000, S. 727–780
- Czychowski, Christian/Nordemann, Axel/Nordemann, Jan Bernd (Hrsg.): Urheberrecht, 13. Auflage, Stuttgart 2024 (zitiert als: Bearbeiter, in: Fromm/Nordemann, Urheberrecht)
- Datenethikkommission der Bundesregierung: Gutachten 2019, abrufbar unter: <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf> (zuletzt abgerufen am 17.10.2024)
- Denga, Michael: Deliktische Haftung für künstliche Intelligenz, in: CR 2018, S. 69–78
- Deutsch, Erwin: Das neue System der Gefährdungshaftungen: Gefährdungshaftung, erweiterte Gefährdungshaftung und Kausal-Vermutungshaftung, in: NJW 1992, S. 73–77
- Dietterich, Tom: Overfitting and undercomputing in machine learning, in: ACM Comput. Surv. 1995, S. 326–327
- Dillenz, Walter: Überlegungen zum Domaine Public Payant, in: GRUR Int 1983, S. 920–926
- Dornis, Tim W.: Die „Schöpfung ohne Schöpfer“ - Klarstellung zur „KI-Autonomie“ im Urheber- und Patentrecht, in: GRUR 2021, S. 784–792
- Dornis, Tim W.: Generative KI, urheberrechtliche Vervielfältigung und öffentliche Zugänglichmachung – Teil 1: Das Modellinnere, in: CR 2024, S. 765–772
- Dornis, Tim W./Stober, Sebastian: Urheberrecht und Training generativer KI-Modelle: Technologische und juristische Grundlagen, Baden-Baden 2024
- Dregelies, Max: KI-Training unter dem AI Act, in: GRUR 2024, S. 1484–1493
- Dreier, Thomas: Überlegungen zur Revision des Schrankenatalogs der Richtlinie 2001/29/EG, in: GRUR Int. 2015, S. 648–657
- Dreier, Thomas: Die Schlacht ist geschlagen – ein Überblick, in: GRUR 2019, S. 771–779
- Dreier, Thomas/Hugenholtz, P. B. (Hrsg.): Concise European Copyright Law, 2. Auflage, Alphen aan den Rijn, The Netherlands 2016
- Dreier, Thomas/Schulze, Gernot (Hrsg.): Urheberrechtsgesetz, 8. Auflage, München 2025
- Ducato, Rossana/Strowel, Alain: Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out, CRIDES Working Paper Series No. 1/2021, abrufbar unter: <https://ssrn.com/abstract=3829858> (zuletzt abgerufen am 17.10.2024)
- Dudenredaktion (Hrsg.): Duden: das Bedeutungswörterbuch, 5. Auflage, Berlin 2018
- Dudenredaktion (Hrsg.): Duden: das Fremdwörterbuch, 12. Auflage, Berlin 2020

- Dudenredaktion (Hrsg.): Duden: das Herkunftswörterbuch*, 6. Auflage, Berlin 2020
- de la Durantaye, Katharina: Allgemeine Bildungs- und Wissenschaftsschranke*, Münster 2014
- de la Durantaye, Katharina: Neues Urheberrecht für Bildung und Wissenschaft – eine kritische Würdigung des Gesetzentwurfs*, in: GRUR 2017, S. 558–567
- de la Durantaye, Katharina: Weit und kollektiv – Vergriffene Werke und kollektive Lizenzvergabe mit erweiterter Wirkung nach der DSM-RL*, in: GRUR 2020, S. 7–14
- de la Durantaye, Katharina: „Garbage in, garbage out“ – Die Regulierung generativer KI durch Urheberrecht*, in: ZUM 2023, S. 645–660
- de la Durantaye, Katharina/Hofmann, Franz: Regulierungsansätze, -defizite und -trends im Urheberrecht 2021*, in: ZUM 2021, S. 873–881
- de la Durantaye, Katharina/Kuschel, Linda: Regelungen zu nicht verfügbaren Werken – Kommentar zum Diskussionsentwurf des BMJV v. 24.6.2020 zu einem Zweiten Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes*, in: ZUM 2020, S. 717–728
- Döbel, Inga/Leis, Miriam/Vogelsang, Manuel Molina/Neustroev, Dmitry/Petzka, Henning/Riemer, Annamaria/Rüping, Stefan/Voss, Angelika/Wegele, Martin/Weltz, Juliana: Maschinelles Lernen: Eine Analyse zu Kompetenzen, Forschung und Anwendung*, München 2018
- Döhmgen gen. Spieker, Indra: Zur Zukunft systemischer Digitalisierung – Erste Gedanken zur Haftungs- und Verantwortungszuschreibung bei informationstechnischen Systemen*, in: CR 2016, S. 698–704
- Ebers, Martin/Heinze, Christian/Krüger, Tina/Steinrötter, Björn (Hrsg.): Künstliche Intelligenz und Robotik*, München 2020
- Ehinger, Patrick/Stiemerling, Oliver: Die urheberrechtliche Schutzfähigkeit von Künstlicher Intelligenz am Beispiel von Neuronalen Netzen*, in: CR 2018, S. 761–769
- Ehmann, Eugen/Selmayr, Martin (Hrsg.): Datenschutz-Grundverordnung*, 3. Auflage, München 2024
- Eichelberger, Jan: Vorübergehende Vervielfältigungen und deren Freistellung zur Ermöglichung einer rechtmäßigen Werknutzung im Urheberrecht*, in: K&R 2012, S. 393–398
- Eidenmüller, Horst: Effizienz als Rechtsprinzip: Möglichkeiten und Grenzen der ökonomischen Analyse des Rechts*, 4. Auflage, Tübingen 2015
- Ensthaler, Jürgen: Zum neuen Verhältnis zwischen Rechtswissenschaft und Technik*, in: ZRP 2022, S. 55–58
- Epping, Volker: Grundrechte*, 9. Auflage, Berlin/Heidelberg 2021
- Epping, Volker/Hillgruber, Christian (Hrsg.): Beck'scher Online-Kommentar zum Grundgesetz*, 58. Edition Stand 15.6.2024, München 2024
- Ertel, Wolfgang: Grundkurs Künstliche Intelligenz: eine praxisorientierte Einführung*, 4. Auflage, Wiesbaden 2016

- European Copyright Society (Hrsg.):* General Opinion on the EU Copyright Reform Package, abrufbar unter: <https://europeancopyrightsocietydotorg.files.wordpress.com/2015/12/ecs-opinion-on-eu-copyright-reform-def.pdf> (zuletzt abgerufen am 17.10.2024)
- Expert Group on Liability and New Technologies:* Liability for artificial intelligence and other emerging digital technologies, Luxemburg 2019
- Fangerow, Kathleen/Schulz, Daniela:* Die Nutzung von Angeboten auf www.kino.to – Eine urheberrechtliche Analyse des Film-Streamings im Internet, in: GRUR 2010, S. 677–682
- Faust, Florian/Schäfer, Hans-Bernd (Hrsg.):* Zivilrechtliche und rechtsökonomische Probleme des Internet und der künstlichen Intelligenz, Tübingen 2019
- Feldman, Vitaly:* Does Learning Require Memorization? A Short Tale about a Long Tail, in: arXiv preprint 2019, abrufbar unter: <http://arxiv.org/abs/1906.05271> (zuletzt abgerufen am 21.10.2024)
- Fenton, Norman E./Biemann, James:* Software Metrics: A Rigorous and Practical Approach, 3. Auflage, Boca Raton (FL, USA) 2014
- Fezer, Karl-Heinz:* Nochmals: Kritik an der ökonomischen Analyse des Rechts, in: JZ 1988, S. 223–228
- Finke, Moritz:* Werk und Veränderung: Verwertungsrechte an veränderten Gestaltungen im Urheberrecht, Tübingen 2022
- Fisher, William W.:* Property and Contract on the Internet, in: Chicago-Kent Law Review 1998, S. 1203–1256
- Flechsig, Norbert P.:* Rechtmäßige private Vervielfältigung und gesetzliche Nutzungsgrenzen: Zur Frage, in welchem Umfang privat hergestellte Vervielfältigungsstücke einer außerprivaten Nutzung zugeführt werden dürfen und zur Beweislast im Urheberverletzungsprozeß, in: GRUR 1993, S. 532–538
- Folkers, Andreas:* Steuerung eines autonomen Fahrzeugs durch Deep Reinforcement Learning, Wiesbaden 2019
- Fortune Business Insights (Hrsg.):* AI Training Dataset Market Size, Share & Industry Analysis, By Type (Text, Audio, Image, Video, and Others), By Deployment Mode (On-Premises and Cloud), By End-Users (IT and Telecommunications, Retail and Consumer Goods, Healthcare, Automotive, BFSI, and Others), and Regional Forecast, 2025-2032, 26.5.2025, abrufbar unter: <https://www.fortunebusinessinsights.com/ai-training-dataset-market-109241> (zuletzt abgerufen am 10.6.2025)
- Gao, Leo/Biderman, Stella/Black, Sid/Golding, Laurence/Hoppe, Travis/Foster, Charles/Phang, Jason/He, Horace/Thite, Anish/Nabeshima, Noa/Presser, Shawn/Leahy, Connor:* The Pile: An 800GB Dataset of Diverse Text for Language Modeling, arXiv preprint 2020, abrufbar unter: <http://arxiv.org/abs/2101.00027> (zuletzt abgerufen am 21.10.2024)
- Gausling, Tina:* Künstliche Intelligenz im digitalen Marketing: Datenschutzrechtliche Bewertung KI-gestützter Kommunikations-Tools und Profiling-Maßnahmen, in: ZD 2019, S. 335–341
- Gausling, Tina:* Kommerzialisierung öffentlich-zugänglicher Informationen im Wege des Data Scraping, in: CR 2021, S. 609–614

- Geiger, Christophe*: Der urheberrechtliche Interessenausgleich in der Informationsgesellschaft – Zur Rechtsnatur der Beschränkungen des Urheberrechts, in: GRUR Int 2004, S. 815–821
- Geiger, Christophe*: Elaborating a Human Rights-Friendly Copyright Framework for Generative AI, in: IIC 2024, S. 1129–1165
- Geiger, Christophe/Frosio, Giancarlo/Bulayenko, Oleksandr*: Text and Data Mining in the Proposed Copyright Reform: Marking the EU Ready for an Age of Big Data? - Legal Analysis and Policy Recommendations, in: IIC 2018, S. 814–845
- Geiger, Christophe/Griffiths, Jonathan/Hilty, Reto/Suthersanen, Uma*: Erklärung für eine ausgewogene Auslegung des Drei-Stufen-Tests im Urheberrecht, in: GRUR Int 2008, S. 822–825
- Gerecke, Martin*: Social Media und Recht: Einige urheberrechtliche Gedanken zu generativen KI-Modellen, in: GRUR-Prax 2023, S. 381–384
- Gernhardt, Franz*: Urheberrechtsverletzungen durch künstliche Intelligenz am Beispiel der bildenden Künste; Werk ohne Autor einmal anders, in: GRUR-Prax 2023, S. 69–71
- v. *Gierke, Otto*: Deutsches Privatrecht Band 1: Allgemeiner Teil u. Personenrecht, Leipzig 1895
- Ginsburg, Jane C.*: „Fair Use for Free, or Permitted-but-Paid?“
- Gless, Sabine/Janal, Ruth*: Hochautomatisiertes und autonomes Autofahren – Risiko und rechtliche Verantwortung, in: JR 2016, S. 561–575
- Gless, Sabine/Seelmann, Kurt (Hrsg.)*: Intelligente Agenten und das Recht, Baden-Baden 2016
- Godt, Christine*: Regulative Einbettung der Immaterialgüterrechte, in: ZGE 2014, S. 279–301
- Goldmann, Michael*: Die Berechnung des Schadensersatzanspruchs vor und nach Umsetzung der Durchsetzungsrichtlinie, in: WRP 2011, S. 950–971
- Grabitz, Eberhard*: Der Grundsatz der Verhältnismäßigkeit in der Rechtsprechung des Bundesverfassungsgerichts, in: AöR 1973, S. 568–616
- Griffiths, Jonathan/Synodinou, Tatiana/Xalabarder, Raquel*: Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market, in: GRUR Int. 2023, S. 22–36
- Grise, Karina/Kaiser, Carola*: Freiheit ab Unkenntlichkeit? – Die Bedeutung der (fehlenden) Wiedererkennbarkeit für das urheberrechtliche Vervielfältigungsrecht, in: ZUM 2021, S. 401–415
- Grum, Marcus/Sultanow, Eldar/Friedmann, Daniel/Ullrich, André/Gronau, Norbert*: Tools des Maschinellen Lernens, Berlin 2021
- Gräbig, Johannes*: Abdingbarkeit und vertragliche Beschränkungen urheberrechtlicher Schranken, Baden-Baden 2011
- Gräbig, Johannes*: Abdingbarkeit urheberrechtlicher Schranken, in: GRUR 2012, S. 331–337

- Gräfe, Hans-Christian/Kahl, Jonas: KI-Systeme zur automatischen Texterstellung: Urheber- und medienrechtliche Einordnung von Textgeneratoren in Journalismus und E-Commerce, in: MMR 2021, S. 121–126
- Grätz, Axel: Künstliche Intelligenz im Urheberrecht: eine Analyse der Zurechnungskriterien und der Prinzipien der Verwandten Schutzrechte vor dem Hintergrund artifiziieller Erzeugnisse, Wiesbaden 2021
- Grünberger, Michael: Bedarf es einer Harmonisierung der Verwertungsrechte und Schranken: Ein Beitrag zur Entwicklung dogmatischer Bausteine eines umweltsensiblen Urheberrechts, in: ZUM 2015, S. 273–290
- Grünberger, Michael: Die Entwicklung des Urheberrechts im Jahr 2022, in: ZUM 2023, S. 309–360
- Guidotti, Riccardo/Monreale, Anna/Ruggieri, Salvatore/Turini, Franco/Giannotti, Fosca/Pedreschi, Dino: A Survey of Methods for Explaining Black Box Models, in: ACM Comput. Surv. 2019, S. 1–42
- Gupta, Sonali/Bhatia, Komal Kumar: A Comparative Study of Hidden Web Crawlers, in: IJCTT 2014, S. 111–118
- Görz, Günther (Hrsg.): Einführung in die künstliche Intelligenz, 2. Auflage, Bonn 1995
- Görz, Günther/Schmid, Ute/Braun, Tanya (Hrsg.): Handbuch der Künstlichen Intelligenz, 6. Auflage, Berlin/Boston 2021
- Görz, Günther/Schmid, Ute/Braun, Tanya (Hrsg.): Handbuch der Künstlichen Intelligenz, Berlin/Boston 2021
- Guibault, Lucie/Handke, Christian/Hargreaves, Ian/Martens, Bertin/Valcke, Peggy: Standardisation in the area of innovation and technological development, notably in the field of text and data mining, 2014, abrufbar unter: <https://op.europa.eu/s/w/ZDX> (zuletzt abgerufen am 21.10.2024)
- Güllich, Hans-Peter: Fuzzy-Expertensysteme zur Beurteilung von Kreditrisiken, Wiesbaden 1997
- Günther, Jan-Philipp: Roboter und rechtliche Verantwortung: eine Untersuchung der Benutzer- und Herstellerhaftung, München 2016
- Haberstumpf, Helmut: Die freie Benutzung darf nicht sterben, in: ZUM 2020, S. 810–819
- Haberstumpf, Helmut: Die freie Benutzung lebt! - Besprechung von BGH, Urteil vom 7.4.2022 – I ZR 222/20 – Porsche 911 (ZUM 2022, 547) und OLG Hamburg, Urteil vom 28.4.2022 – 5 U 48/05 (ZUM 2022, 563), in: ZUM 2022, S. 795–806
- Hacker, Philipp: Ein Rechtsrahmen für KI-Trainingsdaten, in: ZGE 2020, S. 239–271
- Hamann, Hanjo: Der Nutzungsvorbehalt für KI-Training in der Rechtsgeschäftslehre der Maschinenkommunikation - Dogmatische und praktische Schwächen von Art. 4 Abs. 3 DSM-RL und § 44b Abs. 3 UrhG, in: ZGE 2024, S. 113–168
- Handke, Christian/Girard, Yann/Mattes, Anselm: Fördert das Urheberrecht Innovation? Eine empirische Untersuchung, 2015, abrufbar unter: <https://bit.ly/3MBWF6t> (zuletzt abgerufen am 18.10.2024)
- Handke, Christian/Guibault, Lucie/Vallbé, Joan-Josep: Copyright's impact on data mining in academic research, in: Manage Decis Econ 2021, S. 1999–2016

- Harte-Bavendamm, Henning/Henning-Bodewig, Frauke (Hrsg.):* Gesetz gegen den unlauteren Wettbewerb (UWG): mit Preisangabenverordnung und Geschäftsgeheimnisgesetz: Kommentar, 5. Auflage, München 2021
- Hau, Wolfgang/Poseck, Roman (Hrsg.):* Beck'scher Online-Kommentar zum BGB, 71. Edition Stand 1.8.2024, München 2024
- Hauk, Stefanie/Pflüger, Thomas:* Die Umsetzung der DSM-RL in das Urheberrecht – Neujustierung der urheberrechtlichen Rahmenbedingungen für Hochschulen, Forschungs- und Kultureinrichtungen in der digitalen Welt, in: ZUM 2020, S. 383–392
- Hasselblatt, Gordian N. (Hrsg.):* Münchener Anwalts-Handbuch Gewerblicher Rechtsschutz, 6. Auflage, München 2022
- He, Kaiming/Zhang, Xiangyu/Ren, Shaoqing/Sun, Jian:* Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, arXiv preprint 2015, abrufbar unter: <http://arxiv.org/abs/1502.01852> (zuletzt abgerufen am 21.10.2024)
- Heaven, Will Douglas:* These pop songs were written by OpenAI's deep-learning algorithm, in: MIT Technology Review 2021, abrufbar unter: <https://www.technologyreview.com/2020/05/01/1000942/pop-songs-katy-perry-elvis-openai-neural-network-deep-learning-algorithm/> (zuletzt abgerufen am 18.10.2024)
- Heesen, Hendrik/Jüngels, Leonie:* Der Regierungsentwurf der Text und Data Mining-Schranken (§§ 44b, 60d UrhG-E), in: RuZ 2021, S. 45–55
- Heine, Robert:* Generative KI: Nutzungsrechte und Nutzungsvorbehalt, in: GRUR-Prax 2024, S. 87–90
- Henke, Hannes:* E-Books im Urheberrecht: Kollision von Buchkultur und digitaler Wissensgesellschaft, Göttingen 2018
- Hennemann, Moritz:* Interaktion und Partizipation: Dimensionen systemischer Bindung im Vertragsrecht, Tübingen 2020
- Hesse, Konrad:* Grundzüge des Verfassungsrechts der Bundesrepublik Deutschland, Neudr. der 20. Aufl. Auflage, Heidelberg 1999
- Hetmank, Sven/Lauber-Rönsberg, Anne:* Künstliche Intelligenz – Herausforderungen für das Immaterialgüterrecht, in: GRUR 2018, S. 574–582
- Hilgendorf, Eric (Hrsg.):* Robotik im Kontext von Recht und Moral, Baden-Baden 2013
- Hilty, Reto M/Moscon, Valentina:* Modernisation of the EU Copyright Rules - Position Statement of the Max Planck Institute for Innovation and Competition: Max Planck Institute for Innovation & Competition Research Paper No. 17-12, 2017, abrufbar unter: <https://doi.org/10.17617/2.2470998> (zuletzt abgerufen am 21.10.2024)
- Hilty, Reto M./Peukert, Alexander (Hrsg.):* Interessenausgleich im Urheberrecht, Baden-Baden 2004
- Hippner, Hajo (Hrsg.):* Handbuch Data Mining im Marketing: Knowledge Discovery in Marketing Databases, 1. Auflage, Braunschweig/Wiesbaden 2001
- Hirschberg, Julia/Manning, Christopher D:* Advances in natural language processing, in: Science 2015, S. 261–266
- Hoeren, Thomas/Sieber, Ulrich/Holznapel, Bernd (Hrsg.):* Handbuch Multimedia-Recht, 62. Edition Stand 6.2024, München 2024

- Hoffmann, Hanna/Vogt, Verena*: Unbedacht überwacht – Entwicklungen aus der Welt der Sprachassistenten, in: MMR-Aktuell 2020, S. 431-466
- Hofmann, Franz*: Wechselwirkungen zwischen subjektiven Rechten: Eine Analyse des Zusammenspiels von Sacheigentum, Vertragsrecht, Immaterialgüterrechten und Persönlichkeitsrechten am Beispiel des Urheberrechts, in: UFITA 2014, S. 381-406
- Hofmann, Franz*: Grundsatz der Technikneutralität im Urheberrecht? Zugleich Gedanken zu einem more technological approach, in: ZGE 2016, S. 482-512
- Hofmann, Franz*: Der Unterlassungsanspruch als Rechtsbehelf, Tübingen 2017
- Hofmann, Franz*: Das Allgemeininteresse an der Verfügbarkeit von Internet im Spannungsverhältnis zum Schutz von Urheberrechten - Zugleich Anmerkung zu EuGH, Urt. v. 15.9.2016 – C-484/14 (Tobias McFadden ./ Sony Music Entertainment Germany GmbH) sowie zum Dritten Gesetz zur Änderung des Telemediengesetzes (neues WLAN-Gesetz – 3. TMGÄndG), in: GPR 2017, S. 176-182
- Hofmann, Franz*: Kontrolle oder nachlaufender Rechtsschutz – wohin bewegt sich das Urheberrecht?, in: GRUR 2018, S. 21-29
- Hofmann, Franz*: Zehn Thesen zu Künstlicher Intelligenz (KI) und Urheberrecht, in: WRP 2024, S. 11-18
- Hofmann, Franz*: Retten Schranken Geschäftsmodelle generativer KI-Systeme? Vortrag im Rahmen des Symposiums »Generative K.I. und das Urheberrecht – Eine komplizierte Beziehung« des Instituts für Urheber- und Medienrecht am 10.11.2023 in München, in: ZUM 2024, S. 166-174
- Hohagen, Gisbert*: Die Freiheit der Vervielfältigung zum eigenen Gebrauch, München 2004
- Hohlweck, Martin*: Eckpfeiler der mittelbaren Verantwortlichkeit von Plattformbetreibern in der Rechtsprechung Vortrag auf dem Symposium »Internetplattformen – Aktuelle Herausforderungen der digitalen Ökonomie an das Urheber- und Medienrecht« am 4.11.2016 in München, in: ZUM 2017, S. 109-114
- Homar, Philipp*: Enge Handlungsspielräume für das Sampling: Anmerkung zu EuGH, Urteil vom 29.7.2019 - C-476/17 - Pelham u.a./Hüter u.a. (ZUM 2019, 738), in: ZUM 2019, S. 731-737
- Hope, Karen (Hrsg.)*: Annual Report on European SMEs 2020/2021: digitalisation of SMEs, 2021, abrufbar unter: https://op.europa.eu/publication/manifestation_identifier/PUB_EAAK21001ENN (zuletzt abgerufen am 21.10.2024)
- Hugenholtz, P. Bernt*: Caching and Copyright: The Right of Temporary Copying, in: EIPR 2000, S. 482-493
- Iglesias, Maria/Shamuilia, Sharon/Anderberg, Amanda*: Intellectual Property and Artificial Intelligence - A literature review, Luxemburg 2021
- Intellectual Property Office (GB)*: Modernising Copyright: A modern, robust and flexible framework, 2012, abrufbar unter: http://data.parliament.uk/DepositedPapers/Files/DEP_2012-1936/ModernisingCopyrightGovernmentresponse.pdf (zuletzt abgerufen am 21.10.2024)
- Irrgang, Bernhard/Klawitter, Jörg (Hrsg.)*: Künstliche Intelligenz, Stuttgart 1990

- Ismer, Roland/Reimer, Ekkehart/Rust, Alexander/Waldhoff, Christian (Hrsg.):* Territorialität und Personalität - Festschrift für Moris Lehner zum 70. Geburtstag, Köln 2019
- Jacobsen, Jonas/Hartmann, Mattias:* „Maschinenlesbarkeit“ des Rechteevorbehalts im neuen § 44b UrhG, in: MMR-Aktuell 2021, S. 441332
- Jacques, Sabine/Flynn, Mathew:* Protecting Human Creativity in AI-Generated Music with the Introduction of an AI-Royalty Fund, in: GRUR Int. 2024, S. 1137–1149
- Jäger, Lucas Anton:* Artificial Creativity? - Zur urheberrechtlichen Verortung von Systemen „künstlicher Intelligenz“
- Jarass, Hans-D.:* Charta der Grundrechte der Europäischen Union: unter Einbeziehung der sonstigen Grundrechtsregelungen des Primärrechts und der EMRK, 4. Auflage, München 2021
- Jarass, Hans-D./Pieroth, Bodo (Hrsg.):* Grundgesetz für die Bundesrepublik Deutschland: Kommentar, 16. Auflage, München 2020
- Jauernig, Othmar (Hrsg.):* Bürgerliches Gesetzbuch, 19. Auflage, München 2023
- Jordan, M. I./Mitchell, T. M.:* Machine learning: Trends, perspectives, and prospects, in: Science 2015, S. 255–260
- Joshi, Aravind J.:* Natural Language Processing, in: Science 1991, S. 1242–1249
- Kaindl, Hermann:* Problemlösen durch heuristische Suche in der Artificial Intelligence, Wien/New York 1989
- Kaulartz, Markus/Braegelman, Tom (Hrsg.):* Rechtshandbuch Artificial Intelligence und Machine Learning, München 2020
- Kersten, Jens:* Menschen und Maschinen Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen, in: JZ 2015, S. 1–8
- Kirchgässner, Gebhard:* Homo oeconomicus: the economic model of individual behavior and its applications in economics and other social sciences, New York 2008
- Kleinemenke, Manuel:* Fair Use im deutschen und europäischen Urheberrecht?: Eine rechtsvergleichende Untersuchung zur Flexibilisierung des urheberrechtlichen Schrankenkatalogs nach dem Vorbild der US-amerikanischen Fair Use-Doktrin, Baden-Baden 2013
- Kleiner, Cornelius:* Die elektronische Person: Entwurf eines Zurechnungs- und Haftungssubjekts für den Einsatz autonomer Systeme im Rechtsverkehr, Baden-Baden 2021
- Kleinkopf, Felicitas:* Text- und Data-Mining: die Anforderungen digitaler Forschungsmethoden an ein innovations- und wissenschaftsfreundliches Urheberrecht, Baden-Baden 2022
- Klimpel, Paul:* Neue Chancen für das kulturelle Erbe, in: Bibliotheksdienst 2020, S. 559–564
- Kluge, Vanessa/Müller, Anne-Kathrin:* Autonome Systeme, in: InTeR 2017, S. 24–31
- Konertz, Roman/Schönhof, Raoul:* Das technische Phänomen „Künstliche Intelligenz“ im allgemeinen Zivilrecht: Eine kritische Betrachtung im Lichte von Autonomie, Determinismus und Vorhersehbarkeit, Baden-Baden 2020

- Konertz, Roman/Schönhof, Raoul*: Vervielfältigungen und die Text- und Data-Mining-Schranke beim Training von (generativer) Künstlicher Intelligenz: Über die Codierung von Daten und Informationen in Künstlichen Neuronalen Netzen, in: WRP 2024, S. 289–296
- Konertz, Roman/Schönhof, Raoul*: Rechtsfolgen der Urheberrechtsverletzung bei generativer Künstlicher Intelligenz: Über die Möglichkeit des „Vergessens“ in Neuronalen Netzen, in: WRP 2024, S. 534–541
- Kraetzig, Viktoria*: Zensururheberrecht, in: GRUR 2022, S. 1707–1711
- Kreutz, Oliver*: Der Webseitenutzungsvertrag – Fiktion oder unbekanntes Rechtsgeschäft?, in: ZUM 2018, S. 162–168
- Krüger, Wolfgang/Rauscher, Thomas (Hrsg.)*: Münchener Kommentar zur ZPO, 6. Auflage, München 2020
- Kuhlen, Rainer/Lewandowski, Dirk/Semar, Wolfgang/Womser-Hacker, Christa (Hrsg.)*: Grundlagen der Informationswissenschaft, Berlin/Boston 2023
- Kuschel, Linda/Asmussen, Sven/Golla, Sebastian J. (Hrsg.)*: Intelligente Systeme - intelligentes Recht: GRUR Junge Wissenschaft Hamburg 2020/2021, Baden-Baden 2021
- Käde, Lisa*: Kreative Maschinen und Urheberrecht: die Machine Learning-Werkschöpfungskette vom Training über Modellschutz bis zu Computational Creativity, Baden-Baden 2021
- Käde, Lisa*: Do You Remember? – Enthalten KI-Modelle Vervielfältigungen von Trainingsdaten, lassen sich diese gezielt rekonstruieren und welche Implikationen hat das für das Urheberrecht? Vortrag im Rahmen des Symposions »Generative K.I. und das Urheberrecht – Eine komplizierte Beziehung« des Instituts für Urheber- und Medienrecht am 10.11.2023 in München, in: ZUM 2024, S. 174–183
- Käde, Lisa*: Wann ist ein Nutzungsvorbehalt gegenüber Text- und Data Mining maschinenlesbar und dem Rechtsinhaber zurechenbar?, in: CR 2024, S. 599–603
- Käde, Lisa/v. Maltzan, Stephanie*: Algorithmen, die nicht vergessen - Sicherheitslücken in Machine-Learning-Modellen und deren Bedeutung für den Schutz der Daten und der Urheberrechte, in: InTeR 2020, S. 201–208
- Kowalla, Michalina*: Protection of Press Publishers in the Age of Generative AI – In Search of Legal Remedies to Adapt to the Pace of Technology, in: ICC 2024, S. 1604–1623
- Kögel, Daniel*: Urheberrechtliche Implikationen bei der Verwendung kreativer und generativer Künstlicher Intelligenz, in: DSRITB 2023, S. 285–303
- Kögel, Daniel*: Urheberrechtliche Implikationen bei der Verwendung kreativer und generativer künstlicher Intelligenz, in: InTeR 2023, S. 179–185
- Köhler, Helmut*: BGB Allgemeiner Teil: ein Studienbuch, 47. Auflage, München 2023
- Köhler, Helmut/Bornkamm, Joachim/Feddersen, Jörn (Hrsg.)*: Gesetz gegen den unlauteren Wettbewerb, 40. Auflage, München 2022
- Larenz, Karl/Canaris, Claus-Wilhelm*: Methodenlehre der Rechtswissenschaft, 3. Auflage, Berlin/Heidelberg 1995
- Lauber-Rönsberg, Anne*: Urheberrecht und Privatgebrauch: Eine rechtsvergleichende Untersuchung des deutschen und des britischen Rechts, Baden-Baden 2011

- Lauber-Rönsberg, Anne*: Autonome „Schöpfung“ – Urheberschaft und Schutzfähigkeit, in: GRUR 2019, S. 244–253
- Lauber-Rönsberg, Anne*: Reform des Bearbeitungsrechts und neue Schrankenregelung für Parodien, Karikaturen und Pastiche, in: ZUM 2020, S. 733–740
- Leible, Stefan/Ohly, Ansgar/Zech, Herbert (Hrsg.)*: Wissen - Märkte - Geistiges Eigentum, Tübingen 2010
- Leistner, Matthias*: Leistner: „Ende gut, alles gut“ ... oder „Vorhang zu und alle Fragen offen“?, in: GRUR 2019, S. 1008–1015
- Leistner, Matthias*: TDM und KI-Training in der Europäischen Union, in: GRUR 2024, S. 1665–1675
- Leistner, Matthias/Metzger, Axel*: The EU Copyright Package: A Way Out of the Dilemma in Two Stages, in: IIC 2017, S. 381–384
- Lennartz, Jannis/Kraetzig, Viktoria*: AI Learning as Pastiche, in: GRUR 2024, S. 1258–1264
- Leupold, Andreas/Wiebe, Andreas/Glossner, Silke (Hrsg.)*: Münchener Anwaltshandbuch IT-Recht, 4. Auflage, München 2021
- Lin, Tsung-Yi/Maire, Michael/Belongie, Serge/Bourdev, Lubomir/Girshick, Ross/Hays, James/Perona, Pietro/Ramanan, Deva/Zitnick, C. Lawrence/Dollár, Piotr*: Microsoft COCO: Common Objects in Context, arXiv preprint 2015, abrufbar unter: <http://arxiv.org/abs/1405.0312> (zuletzt abgerufen am 21. 10.2024)
- Loewenheim, Ulrich (Hrsg.)*: Handbuch des Urheberrechts, 3. Auflage, München 2021
- Lohmann, Melinda*: Ein europäisches Roboterrecht – überfällig oder überflüssig?, in: ZRP 2017, S. 168–171
- Ludyga, Hannes*: Urheberrechtlicher Schutz von Tweets, in: AfP 2017, S. 284–286
- Longpre, Shayne/Yauney, Georgy/Reif, Emily/Lee, Katherine/Roberts, Adam/Zoph, Barret/Zhou, Denny/Wei, Jason/Robinson, Kevin/Mimno, David/Ippolito, Daphne*: A Pretrainer’s Guide to Training Data: Measuring the Effects of Data Age, Domain Coverage, Quality, & Toxicity, in: NAACL 2024, S. 3245–3276
- Löbner, Sebastian*: Semantik: eine Einführung, 2. Auflage, Berlin/Boston 2015
- Maamar, Niklas*: Computer als Schöpfer, Tübingen 2021
- Maamar, Niklas*: Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen, in: ZUM 2023, S. 481–491
- Mann, William*: Text Generation, in: Computational Linguistics 1982, S. 62–69
- Margoni, Thomas*: 21 for 2021: Computational Uses – CREATE, 2020, CREATE blog v. 8.10.2021, abrufbar unter: <https://www.create.ac.uk/blog/2021/10/08/21-for-2021-computational-uses> (zuletzt abgerufen am 21.10.2024)
- Margoni, Thomas*: Saving research: Lawful access to unlawful sources under Art. 3 CDSM Directive?, Kluwer Copyright Blog v. 22.12.2023, abrufbar unter: <https://copyrightblog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/> (zuletzt abgerufen am 21.10.2024)
- Margoni, Thomas/Kretschmer, Martin*: A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology, in: GRUR Int. 2022, S. 685–701

- Martini, Mario/Kolain, Michael/Neumann, Katja/Rehorst, Tobias/Wagner, David:* Datenhoheit - Annäherung an einen offenen Leitbegriff, in: MMR-Beil. 2021, S. 3–23
- Marzetti, Maximiliano:* Paying for works in the public domain? The “domaine public payant” in the 21st century, in: GRUR Int. 2019, S. 343–351
- Maslej, Nestor/Fattorini, Loredana/Brynjolfsson, Erik/Etchemendy, John/Ligett, Katrina/Lyons, Terah/Manyika, James/Ngo, Helen/Niebles, Juan Carlos/Parli, Vanessa/Shoham, Yoav/Wald, Russell/Clark, Jack/Perrault, Raymond:* The AI Index 2023 Annual Report, Stanford (CA, USA) 2023, abrufbar unter: https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf (zuletzt abgerufen am 21.10.2024)
- Maslej, Nestor/Fattorini, Loredana/Perrault, Raymond/Parli, Vanessa/Reuel, Ankam/Brynjolfsson, Erik/Etchemendy, John/Ligett, Katrina/Lyons, Terah/James, Manyika/Niebles, Juan Carlos/Shoham, Yoav/Wald, Russell/Clark, Jack:* The AI Index 2024 Annual Report, Stanford (CA, USA) 2024, abrufbar unter: https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf (zuletzt abgerufen am 21.10.2024)
- Matthias, Andreas:* Automaten als Träger von Rechten, 2. Auflage, Berlin 2010
- Mayinger, Samantha Maria:* Die künstliche Person: Untersuchung rechtlicher Veränderungen durch die Installation von Softwareagenten im Rahmen von Industrie 4.0, unter besonderer Berücksichtigung des Datenschutzrechts, Frankfurt am Main 2017
- McCarthy, John/Minsky, Marvin L./Rochester, Nathaniel/Shannon, Claude E.:* A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, in: AI Magazine 2006, S. 12–14
- McKinsey (Hrsg.):* The economic potential of generative AI: The next productivity frontier, 14.6.2023, abrufbar unter: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> (zuletzt abgerufen am 10.6.2025)
- Menzli, Amal:* Tokenization in NLP: Types, Challenges, Examples, Tools, 11.8.2023, abrufbar unter: <https://neptune.ai/blog/tokenization-in-nlp> (zuletzt abgerufen am 21.10.2024)
- Meyer, Jürgen/Hölscheidt, Sven (Hrsg.):* Charta der Grundrechte der Europäischen Union, 5. Auflage, Baden-Baden 2019
- Meyer, Stephan:* Künstliche Intelligenz und die Rolle des Rechts für Innovation, in: ZRP 2018, S. 233–238
- Meys, Romain:* Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?, in: GRUR Int. 2020, S. 457–473
- Michaelis, Robert:* Persönlichkeitsrechtliche Befugnisse im deutschen Urheberrecht und droit moral des französischen Rechts, Berlin 1926
- Michalski, Ryszard Stanislaw/Carbonell, Jaime Guillermo/Mitchell, Tom M. (Hrsg.):* Machine learning: an artificial intelligence approach, Berlin Heidelberg 1984

- Mishra, Saurabh*: Data Lifecycle of Machine Learning Project, in: Medium/towards data science v. 12.2.2022, abrufbar unter: <https://medium.com/analytics-vidhya/data-lifecycle-of-machine-learning-project-83b2e656911d> (zuletzt abgerufen am 22.11.2024)
- Mitchell, Tom M.*: Machine Learning, New York 1997
- Morgan, Charles*: Responsible AI: a global policy framework, 2019, abrufbar unter: <https://www.itechlaw.org/publication/responsible-ai-a-global-policy-framework-2019-first-edition> (zuletzt abgerufen am 21.10.2024)
- Mueller, Sebastian*: Big Data Analysen, Berlin/Boston 2018
- Musielak, Hans-Joachim/Voit, Wolfgang (Hrsg.)*: Zivilprozessordnung: ZPO, 19. Auflage, München 2022
- Müller-Hengstenberg, Claus/Kirn, Stefan*: Intelligente (Software-)Agenten: Eine neue Herausforderung unseres Rechtssystems – Rechtliche Konsequenzen der „Verselbstständigung“ technischer Systeme, in: MMR 2014, S. 307–313
- Müller-ter Jung, Marco/Rexin, Lewin*: Rechtliche Anforderungen an intelligentes und automatisiertes Technologiescouting, in: CR 2023, S. 169–179
- Nagarajan, Vaishnavh/Raffel, Colin/Goodfellow, Ian J.*: Theoretical Insights into Memorization in GANs, in: Paper zur NeurIPS Conference 2018, abrufbar unter: <https://colinraffel.com/publications/idlt2018theoretical.pdf> (zuletzt abgerufen am 21.10.2024)
- Nolte, Georg*: Informationsmehrwertdienste und Urheberrecht, Baden-Baden 2009
- Novelli, Claudio/Casolari, Federico/Hacker, Philipp/Spedicaato, Giorgio/Floridi, Luciano*: Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity, in: Computer Law & Security Review (Bd. 55) 2024, Artikelnr. 106066
- Nuernbergk, Christian/Neuberger, Christoph (Hrsg.)*: Journalismus im Internet, 2. Auflage, Wiesbaden 2018
- NVIDIA Corporation (Hrsg.)*: Normalize Operator — NVIDIA DALI 1.29.0 documentation, abrufbar unter: <https://docs.nvidia.com/deeplearning/dali/user-guide/docs/examples/general/normalize.html> (zuletzt abgerufen am 21.10.2024)
- Ohly, Ansgar*: Urheberrecht im digitalen Binnenmarkt – Die Urheberrechtsnovelle 2021 im Überblick, in: ZUM 2021, S. 745–755
- Ory, Stephan/Sorge, Christoph*: Schöpfung durch Künstliche Intelligenz?, in: NJW 2019, S. 710–713
- Oster, Jan*: Haftung für Persönlichkeitsrechtsverletzungen durch Künstliche Intelligenz, in: UFITA 2018, S. 14–52
- Otero, Gonzalez*: Machine Learning Models Under the Copyright Microscope: Is EU Copyright Fit for Purpose?, in: GRUR Int. 2021, S. 1043–1055
- Ott, Stephan*: Bildersuchmaschinen und Urheberrecht – Sind Thumbnails unerlässlich, sozial nützlich, aber rechtswidrig?, in: ZUM 2009, S. 345–354
- Paal, Boris/Pauly, Daniel (Hrsg.)*: Kommentar zu Datenschutz-Grundverordnung und Bundesdatenschutzgesetz, 3. Auflage, München 2021
- Paafß, Gerhard/Hecker, Dirk*: Künstliche Intelligenz: was steckt hinter der Technologie der Zukunft?, Wiesbaden/Heidelberg 2020

- Papastefanou, Stefan*: Anmerkung zu EuGH, Urt. v. 29.7.2019 - C-476/17, Pelham u.a./Hütter u.a. [„Metall auf Metall“], in: CR 2019, S. 600–602
- Papastefanou, Stefan*: DALL-E 2 – Ist das Urheberrecht noch zeitgemäß? Rechtliche Herausforderungen für das Urheberrecht anhand der aktuellen Entwicklungen des text-to-image Modells „DALL-E 2“, in: CR 2023, S. 1–8
- Paton, Mark/Morton, Jeremy*: Copyright Protection for Software Written By Software, in: CRi 2011, S. 8–13
- Pesch, Paulina Jo/Böhme, Rainer*: Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern, in: GRUR 2023, S. 997–1007
- Peukert, Alexander*: Urheberrecht und verwandte Schutzrechte: ein Studienbuch, 19. Auflage, München 2023
- Peukert, Christian/Windisch, Margaritha*: The Economics of Copyright in the Digital Age, in: Center for Law & Economics Working Paper Series ETH Zürich 2023, abrufbar unter: <http://hdl.handle.net/20.500.11850/639999> (zuletzt abgerufen am 21.10.2024)
- Pflüger, Claudius*: Gerechter Ausgleich und angemessene Vergütung: Dispositionsmöglichkeiten bei Vergütungsansprüchen aus gesetzlichen Lizenzen, Baden-Baden 2017
- Podszun, Rupprecht/Grünberger, Michael*: Ein more technological approach für das Immaterialgüterrecht?, in: ZGE 2014, S. 269–270
- Pukas, Jonathan*: Die Repräsentativität von Verwertungsgesellschaften nach § 51b VGG, in: ZUM 2021, S. 912–921
- Pukas, Jonathan*: KI-Trainingsdaten und erweiterte kollektive Lizenzen: Generierung von Werken als KI-Trainingsdaten auf Basis erweiterter kollektiver Lizenzen, in: GRUR 2023, S. 614–621
- Pukas, Jonathan*: Haftungszuweisung für Urheberrechtsverletzungen durch generative KI-Systeme, in: ZGE 2024, S. 38–69
- Puppe, Frank*: Einführung in Expertensysteme, 2. Auflage, Berlin/New York 1991
- Rasmussen, Christoffer Bøgelund/Kirk, Kristian/Moeslund, Thomas B.*: The Challenge of Data Annotation in Deep Learning—A Case Study on Whole Plant Corn Silage, in: Sensors 2022, S. 1596
- Raue, Benjamin*: Die dreifache Schadensberechnung: eine Untersuchung zum deutschen und europäischen Immaterialgüter-, Lauterkeits- und Bürgerlichen Recht, Baden-Baden 2017
- Raue, Benjamin*: Das Urheberrecht der digitalen Wissen(schaft)sgesellschaft, in: GRUR 2017, S. 11–19
- Raue, Benjamin*: Text und Data Mining - Die neue Urheberrechtsschranke des § 60d UrhG, in: CR 2017, S. 656–662
- Raue, Benjamin*: Das subjektive Vervielfältigungsrecht – eine Lösung für den digitalen Werkgenuss?, in: ZGE 2017, S. 514–538
- Raue, Benjamin*: Free Flow of Data? The Friction Between the Commission’s European Data Economy Initiative and the Proposed Directive on Copyright in the Digital Single Market, in: IIC 2018, S. 379–383

- Raue, Benjamin*: Rechtssicherheit für datengestützte Forschung Die Text-und-Data-Mining-Schranken in Art. 3 und 4 DSM-Richtlinie, in: ZUM 2019, S. 684–693
- Raue, Benjamin*: Die geplanten Text und Data Mining-Schranken (§§ 44b und 60d UrhG-E) Kommentar zum Diskussionsentwurf des BMJV v. 15.1.2020 zur Umsetzung der DSM-RL, in: ZUM 2020, S. 172–175
- Raue, Benjamin*: Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ URHG § 44b, URHG § 60d UrhG), in: ZUM 2021, S. 793–802
- Rauer, Nils/Bibi, Alexander*: Extended Collective Licencing in the DSM Directive – An Opportunity to Make Art. 17 DSM Directive Work?, in: GRUR Int. 2022, S. 113–122
- Redeker, Helmut*: IT-Recht, 7. Auflage, München 2020
- Reimer, Franz*: Juristische Methodenlehre, 2. Auflage, Baden-Baden 2020
- Reinemann, Susanne/Remmert, Frank*: Urheberrechte an User-generated Content, in: ZUM 2012, S. 216–227
- Rey, Günter Daniel/Wender, Karl F.*: Neuronale Netze: eine Einführung in die Grundlagen, Anwendungen und Datenauswertung, 3. Auflage, Bern 2018
- Rich, Elaine*: Artificial Intelligence, New York 1983
- Rieber-Mohn, Thomas*: Harmonising Anti-Circumvention Protection with Copyright Law: The Evolution from WCT to the Norwegian Anti-Circumvention Provisions, in: IIC 2006, S. 182–195
- Riehm, Thomas*: Nein zur ePerson! Gegen die Anerkennung einer digitalen Rechtspersönlichkeit, in: RD 2020, S. 42–48
- Riesenhuber, Karl (Hrsg.)*: Europäische Methodenlehre, 4. Auflage, Berlin/Boston 2021
- Ringlage, Philipp*: Haftungskonzepte für autonomes Fahren - „ePerson“ und „RmbH“?, Baden-Baden 2021
- Romero, Alberto*: GPT-3 – A Complete Overview, in: Medium/towards data science v. 24.5.2021, abrufbar unter: <https://towardsdatascience.com/gpt-3-a-complete-overview-190232eb25fd> (zuletzt abgerufen am 21.10.2024)
- Rosengrün, Sebastian*: Künstliche Intelligenz zur Einführung, Hamburg 2021
- Runco, Mark A./Jaeger, Garrett J.*: The Standard Definition of Creativity, in: Creativity Research Journal 2012, S. 92–96
- Runge, Philipp*: Die Vereinbarkeit einer Content-Flatrate für Musik mit dem Drei-Stufen-Test, in: GRUR Int. 2007, S. 130–137
- Russell, Stuart J./Norvig, Peter/Davis, Ernest/Edwards, Douglas*: Artificial intelligence: a modern approach, Boston u.a. 2016
- Sag, Matthew*: Copyright and Copy-Reliant Technology, in: NULR 2009, S. 1607–1682
- v. Savigny, Friedrich Carl*: Das Obligationenrecht als Theil des heutigen römischen Rechts / Zweiter Band, Berlin 1853
- v. Savigny, Friedrich Carl*: System des heutigen römischen Rechts, Berlin 1840
- Säcker, Franz Jürgen/Rixecker, Roland/Oetker, Hartmut/Limberg, Bettina (Hrsg.)*: Münchener Kommentar zum BGB: Band 1, 9. Auflage, München 2021
- Säcker, Franz Jürgen/Rixecker, Roland/Oetker, Hartmut/Limberg, Bettina (Hrsg.)*: Münchener Kommentar zum BGB: Band 2, 9. Auflage, München 2022

- Säcker, Franz Jürgen/Rixecker, Roland/Oetker, Hartmut/Limperg, Bettina (Hrsg.): Münchener Kommentar zum BGB: Band 3, 9. Auflage, München 2022
- Säcker, Franz Jürgen/Rixecker, Roland/Oetker, Hartmut/Limperg, Bettina (Hrsg.): Münchener Kommentar zum BGB: Band 8, 8. Auflage, München 2020
- Schack, Haimo: Das Persönlichkeitsrecht der Urheber und ausübenden Künstler nach dem Tode, in: GRUR 1985, S. 352–361
- Schack, Haimo: Schutz digitaler Werke vor privater Vervielfältigung – zu den Auswirkungen der Digitalisierung auf § 53 UrhG, in: ZUM 2002, S. 497–511
- Schack, Haimo: Urheberrechtliche Schranken für Bildung und Wissenschaft, in: ZUM 2016, S. 266–284
- Schack, Haimo: Das neue UrhWissG – Schranken für Unterricht, Wissenschaft und Institutionen, in: ZUM 2017, S. 802–808
- Schack, Haimo: Schack: Schutzgegenstand, „Ausnahmen oder Beschränkungen“ des Urheberrechts, in: GRUR 2021, S. 904–909
- Schack, Haimo: Vervielfältigung und/oder Bearbeitung oder freie Benutzung?, in: ZGE 2023, S. 263–272
- Schack, Haimo: Auslesen von Webseiten zu KI-Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda, in: NJW 2024, S. 113–117
- Schafer, Burkhard/Komuves, David/Zatarain, Jesus Manuel Niebla/Diver, Laurence: A fourth law of robotics? Copyright and the law and ethics of machine co-production, in: ArtifIntell Law 2015, S. 217–240
- Schaub, Renate: Interaktion von Mensch und Maschine, in: JZ 2017, S. 342–349
- Scheurle, Klaus-Dieter/Mayen, Thomas (Hrsg.): Telekommunikationsgesetz, 3. Auflage, München 2018
- Shippan, Martin: Der Einsatz von Künstlicher Intelligenz bei der Verarbeitung von journalistischen Inhalten, in: ZUM 2024, S. 670–679
- Schmidt, Robert F/Lang, Florian/Heckmann, Manfred (Hrsg.): Physiologie des Menschen, Berlin/Heidelberg 2010
- Schreiber, Sven: Natürliche Intelligenz Teil 1: Neuronen und Synapsen - alles nur ein organischer Computer?, in: c't 1987, S. 98–101
- Schricker, Gerhard/Loewenheim, Ulrich/Ohly, Ansgar (Hrsg.): Urheberrecht: Kommentar, 6. Auflage, München 2020
- Schricker, Gerhard/Ohly, Ansgar (Hrsg.): Perspektiven des geistigen Eigentums und Wettbewerbsrechts: Festschrift für Gerhard Schricker zum 70. Geburtstag, München 2005
- Schulze, Gernot: Die freie Benutzung im Lichte des EuGH-Urteils „Pelham“
- Schulze, Reiner (Hrsg.): Bürgerliches Gesetzbuch, 12. Auflage, Baden-Baden 2024
- Schulze, Reiner/Staudenmayer, Dirk (Hrsg.): EU Digital Law, Baden-Baden 2020
- Schwarz, Alexander/Söbbing, Thomas: Urheberrechtliche Grenzen für lernende künstliche Intelligenz, in: RD i 2023, S. 415–419
- Schäfer, Hans-Bernd/Ott, Claus: Lehrbuch der ökonomischen Analyse des Zivilrechts, 6. Auflage, Berlin 2021

- Schönberger, Daniel*: Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML), in: ZGE 2018, S. 35–58
- Schöneich, Anja*: Der Begriff der Dinglichkeit im Immaterialgüterrecht, Baden-Baden 2017
- Searle, John R.*: Minds, brains, and programs, in: Behavioral and Brain Sciences 1980, S. 417–424
- Senftleben, Martin*: Generative AI and Author Remuneration, in: IIC 2023, S. 1535–1560
- Senftleben, Martin/Margoni, Thomas/Antal, Daniel/Balázs, Bodó/van Gompel, Stef/Handke, Christian/Kretschmer, Martin/Poort, Joost/Quintais, João/Schwemer, Sebastian*: Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive, in: JIPITEC 2022, S. 67–86
- Sesing-Wagenpfeil, Andreas*: KI und Kunst – Zwei Thesen zur urheberrechtlichen Schutzfähigkeit computergenerierter Werke: Zurechnung als bewegliches System, Bewältigung von Schutzrechtsungewissheit als Regulierungsaufgabe, in: CR 2022, S. 749–756
- Sesing-Wagenpfeil, Andreas*: Trainierte KI-Modelle als Vervielfältigungsstücke im Sinne des Urheberrechts, in: ZGE 2024, S. 212–268
- Shah, Agam*: How Tesla Uses and Improves Its AI for Autonomous Driving, in: AIwire v. 8.3.2023, abrufbar unter: <https://www.aiwire.net/2023/03/08/how-tesla-uses-and-improves-its-ai-for-autonomous-driving/> (zuletzt abgerufen am 13.12.2024)
- Shavell, Steven*: Foundations of Economic Analysis of Law, Cambridge, MA (USA) 2004
- Siglmüller, Jonas/Gassner, Daniel*: Softwareentwicklung durch Open-Source-trainierte KI - Schutz und Haftung, in: RDl 2023, S. 124–132
- Sobel, Benjamin*: Artificial Intelligence's Fair Use Crisis, in: CJLA 2017, S. 45–97
- Solon, Olivia*: Facial recognition's 'dirty little secret': Social media photos used without consent, in: NBC News v.12.3.2019, abrufbar unter: <https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921> (zuletzt abgerufen am 21.10.2024)
- Sommer, Martin*: Haftung für autonome Systeme: Verteilung der Risiken selbstlernender und vernetzter Algorithmen im Vertrags- und Deliktsrecht, 1. Auflage, Baden-Baden 2020
- Sosnitzer, Olaf*: Das Internet der Dinge - Herausforderung oder gewohntes Terrain für das Zivilrecht?, in: CR 2016, S. 764–772
- Specht-Riemenschneider, Louisa*: Diktat der Technik: Regulierungskonzepte technischer Vertragsinhaltsgestaltung am Beispiel von Bürgerlichem Recht und Urheberrecht, Baden-Baden 2019
- Specht-Riemenschneider, Louisa*: Zum Verhältnis von (Urheber-)Recht und Technik, in: GRUR 2019, S. 253–259

- SPIEGEL (Hrsg.): Künstliche Intelligenz: Christie's erzielt mit KI-Gemälde 432.500 Dollar*, in: DER SPIEGEL v.26.10.2018, abrufbar unter: <https://www.spiegel.de/netzwelt/web/kuenstliche-intelligenz-christie-s-erzielt-mit-ki-gemaelde-432-500-dollar-a-1235226.html> (zuletzt abgerufen am 21.10. 2024)
- Spindler, Gerald: Europäisches Urheberrecht in der Informationsgesellschaft*, in: GRUR 2002, S. 105–120
- Spindler, Gerald: Roboter, Automation, künstliche Intelligenz, selbststeuernde Kfz – Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen*, in: CR 2015, S. 766–776
- Spindler, Gerald: Text und Data Mining – urheber- und datenschutzrechtliche Fragen*, in: GRUR 2016, S. 1112–1120
- Spindler, Gerald: Text- und Datamining im neuen Urheberrecht und in der europäischen Diskussion*, in: ZGE 2018, S. 273–300
- Spindler, Gerald: Die neue Urheberrechts-Richtlinie der EU, insbesondere „Upload-Filter“ - Bittersweet?*, in: CR 2019, S. 277–291
- Spindler, Gerald/Schuster, Fabian (Hrsg.): Recht der elektronischen Medien, 4. Auflage*, München 2019
- Staats, Robert: Kollektive Lizenzvergabe mit erweiterter Wirkung nach Art.12 der DSM-Richtlinie – eine sinnvolle Lösung für Deutschland?*, in: ZUM 2019, S. 703–712
- Staats, Robert: Kollektive Lizenzen mit erweiterter Wirkung nach § 51 VGG-E: Kommentar zum Diskussionsentwurf des BMJV v. 24.6.2020 zu einem Zweiten Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes*, in: ZUM 2020, S. 728–733
- Stecklow, Steve/Cunningham, Waylon/Jin, Hyunjoo: Tesla workers shared sensitive images recorded by customer cars*, in: Reuters v. 6.4.2023, abrufbar unter: <https://www.reuters.com/technology/tesla-workers-shared-sensitive-images-recorded-by-customer-cars-2023-04-06/> (zuletzt abgerufen am 13.12.2024)
- Steinbrecher, Judith: Steinbrecher: Die EU-Urheberrechtsrichtlinie aus Sicht der Digitalwirtschaft*, in: MMR 2019, S. 639–643
- Stieper, Malte: Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts*, Tübingen 2009
- Stieper, Malte: Rezeptiver Werkgenuss als rechtmäßige Nutzung: Urheberrechtliche Bewertung des Streaming vor dem Hintergrund des EuGH-Urteils in Sachen FAPL/Murphy*, in: MMR 2012, S. 12–17
- Stieper, Malte: Fan Fiction als moderne Form der Pastiche - Zur Abgrenzung von Vervielfältigung, Bearbeitung und freier Benutzung im Lichte des Unionsrechts*, in: AfP 2015, S. 301–305
- Stieper, Malte: Der Trans Europa Express ist aus Luxemburg zurück - auf dem Weg zu einer Vollharmonisierung der urheberrechtlichen Schranken*, in: ZUM 2019, S. 713–720
- Stieper, Malte: Die Umsetzung von Art. 17 VII DSM-RL in deutsches Recht (Teil 1)*, in: GRUR 2020, S. 699–708
- Stieper, Malte: Das Verhältnis der verpflichtenden Schranken der DSM-RL zu den optionalen Schranken der InfoSoc-RL*, in: GRUR 2020, S. 1–7

- Stieper, Malte*: Die Schranken des Urheberrechts im Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, in: ZUM 2021, S. 776–785
- Stieper, Malte/Denga, Michael*: Die Reichweite des EU-Urheberrechts nach der KI-VO, in: GRUR 2024, S. 1473–1484
- Streinz, Rudolf (Hrsg.)*: EUV/AEUV: Vertrag über die Europäische Union, Vertrag über die Arbeitsweise der Europäischen Union, Charta der Grundrechte der Europäischen Union, Kommentar, 3. Auflage, München 2018
- Strobl, Henrike*: Virtuelle Welten, reale Rechte: Die Durchsetzung des Urheberrechts im Metaverse, in: ZUM 2023, S. 492–500
- Sucker, Reinhard*: Der digitale Werkzeuggenuss im Urheberrecht, Tübingen 2014
- Söbbing, Thomas*: Fundamentale Rechtsfragen künstlicher Intelligenz (AI Law), Frankfurt a. M. 2019
- Söbbing, Thomas*: Künstliche neuronale Netze, in: MMR 2021, S. 111–116
- Thouvenin, Florent/Peukert, Alexander/Jaeger, Thomas/Geiger, Christophe (Hrsg.)*: Kreation Innovation Märkte - Creation Innovation Markets: Festschrift Reto M. Hilty, Berlin/Heidelberg 2024 (zitiert als: Bearbeiter, in: FS Hilty)
- Timmermann, Daniel*: Legal Tech-Anwendungen: Rechtswissenschaftliche Analyse und Entwicklung des Begriffs der algorithmischen Rechtsdienstleistung, Baden-Baden 2020
- Trapova, Alina/Gervassis, Nicholas*: UKIPO's public consultation on AI and IP – text and data mining (Part 2), in: Kluwer Copyright Blog v. 16.3.2022, abrufbar unter: <http://copyrightblog.kluweriplaw.com/2022/03/16/ukipos-public-consultation-on-ai-and-ip-text-and-data-mining-part-2/> (zuletzt abgerufen am 21.10. 2024)
- Touvron, Hugo/Lavril, Thibaut/Izcard, Gautier/Martinet, Xavier/Lachaux, Marie-Anne/Lacroix, Timothée/Rozière, Baptiste/Goyal, Naman/Hambro, Eric/Azhar, Faisal/Rodriguez, Aurelien/Joulin, Armand/Grave, Edouard/Lample, Guillaume*: LLaMA: Open and Efficient Foundation Language Models, arXiv preprint 2023, abrufbar unter: <http://arxiv.org/abs/2302.13971> (zuletzt abgerufen am 21.10.2024)
- Trauboth, Heinz*: Software-Qualitätssicherung: konstruktive und analytische Maßnahmen, München/Wien 1993
- Trayer, Martin*: Technische Schutzmassnahmen und elektronische Rechtswahrnehmungssysteme: die Umsetzung von Art. 6 und 7 der EU-Urheberrechts-richtlinie in deutsches Recht und der Schutz des Nutzers urheber-rechtlicher Werke, Baden-Baden 2003
- Triaille, Jean-Paul/Meeûs d'Argenteuil, Jérôme de/Francquen, Amélie de*: Study on the legal framework of text and data mining (TDM), 2014, abrufbar unter: <https://op.europa.eu/s/wV6d> (zuletzt abgerufen am 21.10.2024)
- Trumpke, Felix*: Exklusivität und Kollektivierung: Das skandinavische Modell der Erweiterten Kollektiven Lizenz (Extended Collective Licensing), München 2016
- Turing, A. M.*: COMPUTING MACHINERY AND INTELLIGENCE, in: Mind 1950, S. 433–460
- Ulmann, Bernd (Hrsg.)*: Big Data Analysen, Berlin 2018

- Ulmer, Eugen: Urheber- und Verlagsrecht, 3. Auflage, Berlin/Heidelberg 1980
- Ulmer-Eilfort, Constanze/Obergfell, Eva Inés (Hrsg.): Verlagsrecht, 2. Auflage, München 2021
- v. Ungern-Sternberg, Joachim: Urheberrechtliche Verwertungsrechte im Lichte des Unionsrechts, in: GRUR 2012, S. 1198–1205
- v. Ungern-Sternberg, Joachim: Urheberrechtlicher Werknutzer, Täter und Störer im Lichte des Unionsrechts, zugleich Besprechung zu EuGH, Urt. v. 15. 3. 2012 – C-162/10 – Phonographic Performance (Ireland), und Urt. v. 15. 3. 2012 – C-135/10 – SCF, in: GRUR 2012, S. 576–582
- v. Ungern-Sternberg, Joachim: Verwendungen des Werkes in veränderter Gestalt im Lichte des Unionsrechts, in: GRUR 2015, S. 533–539
- v. Ungern-Sternberg, Joachim: Die Rechtsprechung des EuGH und des BGH zum Urheberrecht und zu den verwandten Schutzrechten im Jahre 2015, in: GRUR 2016, S. 321–338
- v. Ungern-Sternberg, Joachim: Zum Recht auf Anerkennung der Urheberschaft, in: GRUR 2017, S. 760–765
- v. Ungern-Sternberg, Joachim: Die Rechtsprechung des EuGH und des BGH zum Urheberrecht und zu den verwandten Schutzrechten im Jahre 2020, in: GRUR 2021, S. 1–18
- v. Ungern-Sternberg, Joachim: Die Rechtsprechung des EuGH und des BGH zum Urheberrecht und zu den verwandten Schutzrechten im Jahre 2022, in: GRUR 2022, S. 1777–1790
- Vesala, Juha: Developing Artificial Intelligence-Based Content Creation: Are EU Copyright and Antitrust Law Fit for Purpose?, in: IIC 2023, S. 351–380
- Wagner, Kristina: Schaffensfreiräume und Reichweite der InfoSoc-RL, in: MMR 2020, S. 728–733
- Wahlster, Wolfgang/Winterhalter, Christoph: Deutsche Normungsroadmap Künstliche Intelligenz des DIN e.V. und der DKE, 2020, abrufbar unter: <https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf> (zuletzt abgerufen am 21.10.2024)
- Waite, Maurice (Hrsg.): Oxford English Dictionary, 7. Auflage, Oxford 2012
- Walz, Christian: Das Ziel der Auslegung und die Rangfolge der Auslegungskriterien, in: ZJS 2010, S. 482
- Wandtke, Artur-Axel: Urheberrecht in der Reform oder wohin steuert das Urheberrecht?, in: MMR 2017, S. 367–373
- Wandtke, Artur-Axel: Grundsätze der Richtlinie über das Urheberrecht im digitalen Binnenmarkt, in: NJW 2019, S. 1841–1847
- Wandtke, Artur-Axel/Bullinger, Winfried (Hrsg.): Urheberrecht, 6. Auflage, München 2022
- Wandtke, Artur-Axel/Hauck, Ronny: „Scherz, Satire, Ironie und tiefere Bedeutung“: Die §§ 23 und 51a im Referentenentwurf zur Neufassung des UrhG, in: GRUR Prax 2020, S. 542–544

- Wandtke, Artur-Axel/Holzapfel, Henrik*: Ist § 31 IV UrhG noch zeitgemäß?, in: GRUR 2004, S. 284–293
- Welser, Marcus v.*: Generative KI und Urheberrechtsschranken, in: GRUR-Prax 2023, S. 516–520
- Weltorganisation für geistige Eigentum (WIPO)*: WIPO Technology Trends 2019: Artificial Intelligence, 2019, abrufbar unter: <https://www.wipo.int/publications/en/details.jsp?id=4386> (zuletzt abgerufen am 21.10.2024)
- Wettig, Steffen*: Vertragsschluss mittels elektronischer Agenten: eine interdisziplinäre Untersuchung mit juristischem Schwerpunkt unter Einbeziehung internationaler, technischer und philosophischer Aspekte, Berlin 2010
- Wiebe, Andreas*: Die elektronische Willenserklärung: kommunikationstheoretische und rechtsdogmatische Grundlagen des elektronischen Geschäftsverkehrs, Tübingen 2002
- Wimmers, Jörg/Barudi, Malek*: Der Mythos vom Value Gap, in: GRUR 2017, S. 327–338
- Wolf, Manfred/Wellenhofer, Marina*: Sachenrecht, 39. Auflage, München 2024
- Wolff, Heinrich Amadeus/Brink, Stefan/v. Ungern-Sternberg, Antje (Hrsg.)*: Beck'scher Online-Kommentar Datenschutzrecht, 48. Edition Stand 1.5.2024, München 2024
- Wulf, Julia/Löbeth, Tim-Jonas*: Text und Data Mining: Wenn gewolltes und geschaffenes Recht auseinanderfallen, in: GRUR 2024, S. 737–741
- Württemberg, Gert/Freischem, Stephan*: Stellungnahme des GRUR-Fachausschusses für Urheber- und Verlagsrecht zur Umsetzung der EU-RL im Urheberrecht (DSM-RL [EU] 2019/790 und Online-SatCab-RL [EU] 2019/789), in: GRUR 2019, S. 1140–1166
- Yse, Diego Lopez*: Text Normalization for Natural Language Processing (NLP), in: Medium/towards data science v.17.2.2021, abrufbar unter: <https://towardsdatascience.com/text-normalization-for-natural-language-processing-nlp-70a314bfa646> (zuletzt abgerufen am 21.10.2024)
- Zech, Herbert*: Information als Schutzgegenstand, Tübingen 2012
- Zech, Herbert*: Künstliche Intelligenz und Haftungsfragen, in: ZfPW 2019, S. 198–219
- Zech, Herbert*: Empfehlen sich Regelungen zu Verantwortung und Haftung beim Einsatz Künstlicher Intelligenz?, in: NJW-Beil 2022, S. 33–38
- Zech, Herbert*: Haftung für Trainingsdaten Künstlicher Intelligenz, in: NJW 2022, S. 502–507
- Zurth, Patrick*: Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht: eine dogmatische Analyse der Rechtsnatur und der vertraglichen Gestaltungsmöglichkeiten, Tübingen 2016

