

Felicitas Lea Kleinkopf

Text- und Data-Mining

Die Anforderungen digitaler Forschungsmethoden an ein innovations- und wissenschaftsfreundliches Urheberrecht



Nomos

Schriftenreihe des Archivs für
Medienrecht und Medienwissenschaft (UFITA)

herausgegeben von
Prof. Dr. Michael Grünberger, LL.M. (NYU)
Prof. Dr. Nadine Klass, LL.M. (Wellington)

Band 300

Felicitas Lea Kleinkopf

Text- und Data-Mining

Die Anforderungen digitaler Forschungsmethoden an ein
innovations- und wissenschaftsfreundliches Urheberrecht



Nomos

The book processing charge was funded by the Baden-Württemberg Ministry of Science, Research and Arts in the funding programme Open Access Publishing and the University of Freiburg.

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

Zugl.: Freiburg, Univ., Diss., 2022

1. Auflage 2022

© Felicitas Lea Kleinkopf

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): 978-3-8487-7561-3
ISBN (ePDF): 978-3-7489-3536-0

DOI: <https://doi.org/10.5771/9783748935360>



Onlineversion
Nomos eLibrary



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

Meiner Familie

Vorwort

Dass die Technologie des Text- und Data-Minings, die Digitalforschung im Allgemeinen sowie die Organisation von Forschungsdaten die Urheberrechtswissenschaft noch umfangreich beschäftigen wird, zeigt die Fülle an Literatur und Studien, die in der vorliegenden Untersuchung Einklang fand. Insbesondere die Evaluationsstudie des Bundesministeriums für Justiz aus April 2022 verdeutlicht den weiteren Diskussions- und teilweise auch den weiteren gesetzgeberischen Handlungsbedarf. Die vorliegende Arbeit strukturiert und analysiert die urheberrechtlichen Implikationen des Text- und Data-Minings nicht nur, sondern macht auch praktische Vorschläge, wie diese komplexe Urheberrechtsmaterie interessengerecht gestaltet werden kann. Sie wurde im Sommersemester 2022 von der Rechtswissenschaftlichen Fakultät der Albert-Ludwigs-Universität Freiburg als Dissertation angenommen und entstand während meiner Tätigkeit als akademische Mitarbeiterin am Zentrum für Angewandte Rechtswissenschaft (ZAR) am Karlsruher Institut für Technologie (KIT). Die Untersuchung wurde zwar Anfang April 2022 eingereicht, später erschienene Literatur konnte aber jedenfalls teilweise noch vor der Drucklegung berücksichtigt werden.

Mein herzlichster Dank gilt Herrn Prof. Dr. Thomas Dreier für die hervorragende Betreuung, für die zahlreichen hilfreichen Anregungen, die meine Arbeit stets voranbrachten, und für die wissenschaftlichen Freiräume, die er mir an seinem Institut ermöglichte. Zum Dank verpflichtet bin ich außerdem Herrn Prof. Dr. Maximilian Haedicke für die rasche Erstellung des Zweitgutachtens. Mein besonderer Dank gilt außerdem Herrn Dr. Thomas Pflüger, der mir die Möglichkeit einer überaus lehrreichen und spannenden Tätigkeit für das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg bot, durch die ich wertvolle Einblicke in das Gesetzgebungsverfahren zur Umsetzung der DSM-Richtlinie erhalten konnte. Diese Erfahrungen haben wesentlich zu meiner Doktorarbeit beigetragen. Den Herausgebern und dem Nomos-Verlag danke ich für die Aufnahme in die Schriftenreihe des Archivs für Urheber- und Medienrecht (UFITA) und die Betreuung der Publikation.

Meinen Kolleginnen und Kollegen am ZAR danke ich für die spannende und lehrreiche Zeit, insbesondere Olivia Hägle für den hilfreichen und produktiven Austausch über die Arbeit und über das Promovieren als

solches. Daneben gilt mein Dank meinen Freundinnen und Freunden für die großartige Unterstützung während meiner Promotion. Hanna Merkel hat mir mit ihrer technischen Expertise sehr geholfen, das interdisziplinäre Thema zu bewältigen. Julian Rosenberger, dir möchte ich ganz besonders für deine unermüdliche und wertvolle Diskussions- und Einsatzbereitschaft danken.

Zuletzt möchte ich meinen Eltern, Ulrike und Fred Kleinkopf, und meinen Schwestern, Rebecca Kleinkopf und Charlotte Zinngrebe, von Herzen meine Dankbarkeit ausdrücken. Ohne euren bedingungslosen und liebevollen Rückhalt sowie eure Unterstützung meines Werdegangs wäre für mich die Promotion nicht vorstellbar gewesen. Auch danke ich euch, dass ihr euch beim Korrekturlesen auf ein so alltagsfremdes juristisches Thema eingelassen habt. Euch ist die Arbeit deswegen gewidmet.

Karlsruhe, August 2022

Felicitas Lea Kleinkopf

Inhaltsverzeichnis

Abkürzungsverzeichnis	19
Einleitung	23
§ 1 Problemaufriss	23
§ 2 Herangehensweise	26
A. Grenzen der Untersuchung	26
B. Interdisziplinärer Ansatz	27
C. Gang der Untersuchung	27
Erster Teil: Text- und Data-Mining zwischen Technologie, Innovation und Wissenschaft	29
§ 3 Text- und Data-Mining zwischen Künstlicher Intelligenz und digitaler Forschung	30
A. Das Begriffsverständnis des Text- und Data-Minings	30
I. Das technische Begriffsverständnis	30
II. Das urheberrechtliche Begriffsverständnis	33
III. Begriffsfindung	35
B. Text- und Data-Mining zur Erkenntnisgewinnung	36
C. Text- und Data-Mining und Künstliche Intelligenz	38
I. Technische Grundlagen der Künstlichen Intelligenz	38
1. Begriffsentwicklung der Künstlichen Intelligenz	38
2. Die Technologie des maschinellen Lernens	41
3. Wissensbasierte Systeme und Expertensysteme	45
II. Überschneidungen von Text- und Data-Mining mit Technologien der Künstlichen Intelligenz	46
D. Ein Ablauf des Text- und Data-Minings	48
I. Datenauswahl und Selektierung	50
II. Die Vorverarbeitungsschritte	51
1. Digitalisierung und Textextrahierung	52
2. Segmentierung	53
3. Bereinigung	54
4. Metadaten und Annotationen	54
III. Das Korpus	56

IV. Die Text- und Data-Mining-Analyse im engeren Sinne	58
V. Das Analyseergebnis – Einordnung, Potenzial und Validität	60
1. Muster – Korrelation oder Kausalität?	62
2. Einordnung und Validität des Ergebnisses der TDM-Analyse	65
E. Weitere grundlegende Begriffsverständnisse	69
I. Begriffsverständnis von Algorithmus und Computerprogramm	69
II. Begriffsverständnis von Information, Wissen und Daten	69
III. Begriffsverständnis von Forschungsdaten	72
F. Zwischenfazit	73
§ 4 Ein Urheberrecht für die Wissenschaft?	73
A. Text- und Data-Mining zur wissenschaftlichen Forschung	74
I. Das Begriffsverständnis von Wissenschaft im Grundgesetz	74
II. Das Verständnis von Wissenschaft im Urheberrecht	76
III. Einordnung des Text- und Data-Minings	78
B. Ein Recht des Urhebers?	79
I. Der Urheber als Zentrum des Urheberrechts und Tendenzen zur Aufweichung	79
II. Die systematische Stellung von Schrankenbestimmungen	85
1. Das Verhältnis von geistigem Eigentum und Sozialbindung imdeutschen und europäischen Grundrechtsgefüge	85
2. Das Verhältnis von Werkschutz und Schranke im einfachen Recht	86
C. Die Besonderheiten des Wissenschaftsurheberrechts	89
I. Das urheberrechtliche Schutzniveau in Wissenschaftskontexten	89
1. Ein eingegrenzter Schutz wissenschaftlicher Werke	90
2. Die urheberrechtliche Ein- und Zuordnung von Forschungsdaten	98
II. Die besonderen Interessen im Wissenschaftsurheberrecht	101

III. Die Auswirkungen der Wissenschaftsfreiheit auf die Schränkenauslegung	106
D. Zwischenfazit	112
§ 5 Fortschritt durch Restriktion oder Restriktion des Fortschritts? – Die urheberrechtliche Steuerung von Innovation	113
A. Das urheberrechtliche Anreizparadigma	114
B. Keine Übertragbarkeit der Anreizthese auf Innovation	118
C. Mechanismen zur urheberrechtlichen Innovationsförderung	121
D. Stellungnahme	126
§ 6 Ergebnis des ersten Teils	127
Zweiter Teil: Forschung und Analyse mittels Text- und Data-Mining	129
§ 7 Die urheberrechtliche Relevanz des Text- und Data-Minings	130
A. Einordnung der Vorverarbeitungsschritte des TDMs	130
I. Beeinträchtigung der Verwertungsrechte	130
1. Digitalisierungsprozesse und Textextrahierungen	131
a. Begriffsverständnis von Vervielfältigungen, Bearbeitungen und anderen Umgestaltungen	131
b. Das Verhältnis von Vervielfältigungen, Bearbeitungen und anderen Umgestaltungen	133
c. Einordnung der Digitalisierung und Texterkennung	140
2. Segmentierungen	142
3. Bereinigungen	143
4. Annotationen	143
5. Die Nutzung von Werkteilen	145
6. Die Nutzung von Bearbeitungen	145
II. Beeinträchtigung von Urheberpersönlichkeitsrechten	146
1. Namensnennung, § 13 UrhG	146
2. Entstellungen, § 14 UrhG	146
III. Besonderheiten in Data-Mining-Prozessen	150
1. Datenbankwerke, §§ 4 Abs. 2, 2 Abs. 2 UrhG	151
a. Vervielfältigungen, § 16 Abs. 1 UrhG, und Bearbeitungen bzw. Umgestaltungen, § 23 Abs. 2 UrhG	152
b. Entstellungen, § 14 UrhG	154
2. Datenbanken, §§ 87a ff. UrhG	154
IV. Besonderheiten bei Web-Crawlern oder -Scrapern	155

B. Das Korpus	156
1. Schutzfähigkeit der Korpora	157
a. Schutzfähigkeit als Sprachwerk	157
b. Schutzfähigkeit als Sammelwerk	158
c. Schutzfähigkeit als Datenbankwerk	160
d. Schutzfähigkeit als Datenbank	161
2. Rechtsverletzungen durch das Zusammenfügen der Korpora	162
C. Die TDM-Analyse im engeren Sinne	163
D. Die Besonderheiten bei Nutzung von Technologien des maschinellen Lernens	166
E. Die Relevanz gemeinsamer Forschungsarbeiten	167
F. Das TDM als Ausnahme des Urheberrechts? – Übertragung analoger Grundsätze auf digitale Vorgänge	168
I. Das TDM als Mittel zum Gewinnen von urheberrechtlich nicht geschütztem Wissen und Information	171
II. Das „right to read“ und das „right to mine“ – Vergleichbarkeit des TDMs und des Werkgenusses	172
III. Grenzen der inhaltlichen Ausnahme des TDMs aus dem Urheberrecht	181
G. Heranziehung bestehender Erlaubnisnormen	182
I. Wissenschaftliche Forschung, § 60c UrhG	183
II. Zitate, § 51 UrhG	183
III. Vorübergehende Vervielfältigungshandlungen, § 44a UrhG	184
IV. Benutzung eines Datenbankwerks, § 55a UrhG	186
H. Zwischenfazit	187
§ 8 Die Bedeutung eines adäquaten Rechtsrahmens	188
A. Eine systematische Begründung eines gesetzgeberischen Auftrags	190
B. Negative Auswirkungen von Restriktion und Rechtsunsicherheit und Potenzial weitreichenderer Erlaubnisse	191
C. Die Vereinbarkeit mit dem Drei-Stufen-Test	196
D. Zwischenfazit	199

§ 9 Eine Analyse der Rechtsentwicklung in Deutschland und der Europäischen Union	200
A. Das Gesetz zum Urheberrecht in der Wissensgesellschaft	202
I. Privilegierte Technologien	203
II. Privilegierter Personenkreis	204
III. Erfasste Schutzgegenstände	205
IV. Rechtmäßiger Zugang	206
V. Erlaubte Verwertungshandlungen	213
1. Vervielfältigungen, § 16 UrhG	213
2. Bearbeitungen, § 23 UrhG	214
3. Weiterverwendung von Bearbeitungen	216
4. Öffentliche Zugänglichmachungen, § 19a UrhG	217
5. Nutzbarkeit von Datenbankwerken und Datenbanken	220
VI. Einschränkung des Urheberpersönlichkeitsrechts	221
1. Änderungsverbot, § 62 UrhG	221
2. Pflicht zur Quellenangabe, § 63 UrhG	223
VII. Die TDM-Analyse im engeren Sinne	224
VIII. Vertragliche und technische Abweichungsmöglichkeiten	224
IX. Vergütungspflicht, § 60h UrhG	225
X. Geltungsdauer	226
XI. Bewertung	226
B. Die Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt	228
I. Harmonisierungsgrad	229
II. Privilegierte Technologien	230
III. Privilegierte Personenkreise	230
1. Text- und Data-Mining zum Zwecke der wissenschaftlichen Forschung, Art. 3 DSM-RL	231
2. Text- und Data-Mining ohne Zweckbindung, Art. 4 DSM-RL	232
IV. Erfasste Schutzgegenstände	232
V. Erfasste Verwertungshandlungen	233
VI. Vergütung	234
VII. Vertragliche und technische Abweichungsmöglichkeiten	235
VIII. Bewertung	236
C. Das Gesetz zum Urheberrecht im digitalen Binnenmarkt	237
I. Privilegierte Technologien	238

II. Privilegierte Personenkreise	239
III. Erfasste Schutzgegenstände	240
IV. Rechtmäßiger Zugang	241
V. Erlaubte Verwertungshandlungen	242
1. Vervielfältigungen, § 16 UrhG	242
2. Bearbeitungen, § 23 UrhG	243
3. Weiterverwendung von Bearbeitungen	244
4. Öffentliche Zugänglichmachungen, § 19a UrhG	244
5. Nutzbarkeit von Datenbankwerken und Datenbanken	245
VI. Einschränkungen des Urheberpersönlichkeitsrechts	246
1. Änderungsverbot, § 62 UrhG	246
2. Pflicht zur Quellenangabe, § 63 UrhG	247
VII. Die TDM-Analyse im engeren Sinne	247
VIII. Vergütung, § 60h UrhG	248
IX. Vertragliche und technische Abweichungsmöglichkeiten	248
X. Geltungsdauer	252
XI. Bewertung	253
§ 10 Ergebnis des zweiten Teils	255
Dritter Teil: Der Zugang zu und die Nachnutzung von Text- und Data-Mining-Korpora	258
§ 11 Die urheberrechtliche Relevanz des Zugangs zu und der Nachnutzung von Text- und Data-Mining-Korpora	259
A. Begriffliche Abgrenzung von Zugang und Nachnutzung	260
B. Archivierung der Korpora	262
I. Vervielfältigung, § 16 Abs. 1 UrhG	262
II. Öffentliche Zugänglichmachung, § 19a UrhG	263
C. Die Bereitstellung der Korpora für die wissenschaftliche Überprüfung	263
I. Öffentliche Zugänglichmachung, § 19a UrhG, und öffentliche Wiedergabe als Innominatfall, § 15 Abs. 2 UrhG	263
II. Vervielfältigungen, § 16 Abs. 1 UrhG	264
D. Die Bereitstellung der Korpora für die wissenschaftliche Nachnutzung	265
I. Vervielfältigung, § 16 Abs. 1 UrhG	265

II. Öffentliche Zugänglichmachung, § 19a UrhG und öffentliche Wiedergabe als Innominatfall, § 15 Abs. 2 UrhG	266
E. Verletzung von Urheberpersönlichkeitsrechten	266
§ 12 Die Bedeutung des Zugangs und das Potenzial der Nachnutzbarkeit	267
A. Der Zugang als wissenschaftsimmanente Anforderung an das Urheberrecht	268
I. Wissenschaftstheoretische Betrachtung	269
II. Grundgesetzliche Betrachtung der Anforderungen der guten wissenschaftlichen Praxis	273
III. Negative Auswirkungen rechtlicher Defizite	275
IV. Zugang durch das Urheberrecht	277
B. Die Bedeutung von Forschungsdaten für das kulturelle Gedächtnis	278
I. Rechtliche und tatsächliche Besonderheiten digitaler Zugänglichmachung im Bereich von Gedächtnisinstitutionen	278
II. Kulturwissenschaftliche Betrachtung von Forschungsdaten	282
C. Vorzüge der Nachnutzbarkeit	287
I. Effektivere Nutzung öffentlicher Gelder	287
II. Auswirkungen auf die Datenqualität und das Erkenntnispotenzial	288
III. Anreizwirkung	290
D. Anforderungen an den Umgang mit TDM-Korpora und ihre Grenzen	290
I. Maßstabsbildung	290
II. Vereinbarkeit mit dem Drei-Stufen-Test	292
1. Archivierung und Überprüfbarkeit	292
2. Nachnutzbarkeit	294
E. Zwischenfazit	296
§ 13 Analyse der Rechtsentwicklung in Deutschland und in der Europäischen Union	297
A. Das Gesetz zum Urheberrecht in der Wissensgesellschaft	297
I. Aufbewahrung der Korpora	298
II. Zugriff zu Überprüfungszwecken	299
1. Öffentliche Zugänglichmachung, § 19a UrhG	299
2. Vervielfältigungen, § 16 Abs. 1 UrhG	300

3. Überprüfungen nach Forschungsabschluss	302
4. Umfang der erlaubten Überprüfungen	304
III. Verfügbarkeit für Anschlussforschungen	304
IV. Stellungnahme	306
B. Die Richtlinie über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt	307
I. Aufbewahrung der Korpora	307
II. Zugriff zu Überprüfungszwecken	308
III. Verfügbarkeit für Anschlussforschungen	309
IV. Stellungnahme	310
C. Das Gesetz zum Urheberrecht im digitalen Binnenmarkt	310
I. Aufbewahrungsmöglichkeit der Korpora	311
1. Löschpflicht i. R. d. § 44b UrhG	311
2. Aufbewahrung der Korpora i. R. d. § 60d UrhG	311
3. Rechtsfolgen	315
II. Zugriff zur Überprüfung der wissenschaftlichen Forschung	318
1. Öffentliche Zugänglichmachung, § 19a UrhG	318
2. Vervielfältigungen, § 16 Abs. 1 UrhG	319
3. Zugriff nach Forschungsabschluss	319
4. Umfang der erlaubten Überprüfungen	320
III. Verfügbarkeit für Anschlussforschungen	320
IV. Stellungnahme	321
D. Schlussfolgerungen	323
§ 14 Die Ermöglichung einer Nachnutzbarkeit	325
A. Technik über dem Recht – Technische Ansätze zur Überwindung urheberrechtlicher Hemmnisse im Bereich des Text- und Data-Minings	326
B. Ein Recht für die Technik – Nutzung des geltenden Rechts für die Nachnutzbarmachung von Text- und Data-Mining-Korpora	330
I. Kombinierbarkeit urheberrechtlicher Schranken	330
II. Heranziehung der Terminal-Schranke, § 60e Abs. 4 UrhG	332
1. Anwendbarkeit	332
2. Eignung	334
III. Heranziehung der Wissenschaftsschranke, § 60c UrhG	334
1. Anwendbarkeit	335
2. Eignung	337

C. Stellungnahme	338
§ 15 Ergebnis des dritten Teils	339
Vierter Teil: Entwicklung eines innovations- und wissenschaftsfreundlichen Rechtsrahmens	341
§ 16 Schlussfolgerungen	341
A. Zusammenfassende Thesen	341
B. Normierungsvorschlag	348
§ 17 Ausblick	351
Literaturverzeichnis	353

Abkürzungsverzeichnis

Computerprogramm-RL	Richtlinie 2009/24/EG des Europäischen Parlaments und des Rates vom 23. April 2009 über den Rechtsschutz von Computerprogrammen (ABl. L 111 S. 16)
Datenbank-RL	Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken, zuletzt geändert durch Art. 24 ÄndRL (EU) 2019/790 vom 17.4.2019 (Abl. L 130 S. 92)
DNBG	Gesetz über die Deutsche Nationalbibliothek
DNG	Datennutzungsgesetz
DSGVO	Datenschutz-Grundverordnung (VO (EU) 2016/679)
DSM-RL	Richtlinie (EU) 2019/790 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 (ABl. L 130 S. 92, ber. L 259 S. 86)
Erw.	Erwägungsgrund
IFG	Informationsfreiheitsgesetz
InfoSoc-RL	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 (ABl. L 167 S. 10, ber. 2002 L 6 S. 71), zuletzt geändert durch Art. 24 Abs. 2 ÄndRL (EU) 2019/790 vom 17.4.2019 (ABl. L 130 S. 92)
IWG	Informationsweiterverwendungsgesetz
KDD	Wissenserkenntung in Datenbanken (englisch: Knowledge Discovery in Databases)
KI	Künstliche Intelligenz

KI-Verordnungsentwurf	Vorschlag für eine Verordnung des Europäischen Parlamentes und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, Brüssel, den 21.4.2021, COM(2021) 206 final
KNN	Künstliche neuronale Netze
Online-SatCab-RL	Richtlinie (EU) 2019/789 des Europäischen Parlaments und des Rates vom 17. April 2019 mit Vorschriften für die Ausübung von Urheberrechten und verwandten Schutzrechten in Bezug auf bestimmte Online-Übertragungen von Sendunternehmen und die Weiterverbreitung von Fernseh- und Hörfunkprogrammen und zur Änderung der Richtlinie 93/83/EWG des Rates
PSI-RL	Richtlinie (EU) 2019/1024 des Europäischen Parlaments und des Rates vom 20. Juni 2019 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (Neufassung)
RBÜ	Revidierte Berner Übereinkunft zum Schutz von Werken der Literatur und Kunst
Schutzdauer-RL	Richtlinie 2006/116/EG des Europäischen Parlaments und des Rates vom 12. Dezember 2006 über die Schutzdauer des Urheberrechts und bestimmter verwandter Schutzrechte (ABl. L 372 S. 12), zuletzt geändert durch Art. 1 ÄndRL 2011/77/EU vom 27.9.2011 (ABl. L 265 S. 1, ber. 2013 L 117 S. 23)
UrhDaG	Gesetz über die urheberrechtliche Verantwortlichkeit von Diensteanbietern für das Teilen von Online-Inhalten (Urheberrechts-Diensteanbieter-Gesetz), verabschiedet durch das Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31.05.2021 BGBl. I S. 1204, 1215 Nr. 27, ausgegeben zu Bonn am 4. Juni 2021
UrhDBMG	Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31.05.2021, BGBl. 2021 I Nr. 27, ausgegeben zu Bonn am 4. Juni 2021

UrhWissG	Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft) vom 01.09.2017, BGBl. 2017 I Nr. 61, ausgegeben zu Bonn am 7. September 2017
WCT	WIPO-Urheberrechtsvertrag vom 20. Dezember 1996, BGBl. 2003 II S. 754, 755, ausgegeben zu Bonn am 18. August 2003
WPPT	WIPO-Vertrag über Darbietungen und Tonträger vom 20. Dezember 1996, BGBl. 2003 II S. 754, ausgegeben zu Bonn am 18.08.2003

Im Übrigen sei auf die im Literaturverzeichnis erläuterten Abkürzungen der zitierten Literatur verwiesen. Daneben orientiert sich die Untersuchung an den in der Rechtswissenschaft gebräuchlichen Abkürzungen nach *Kirchner*, Abkürzungen für Juristen, 2. Auflage, Berlin 1993.

Einleitung

§ 1 Problemaufriss

Mit computerlinguistischen Mitteln wurde *Joanne K. Rowling* im Jahr 2013 als wahre Autorin der Romane enttarnt, die sie unter ihrem Pseudonym *Robert Galbraith* veröffentlichte.¹ Mithilfe der Software „Deep-L“² ist es möglich, Texte aus 26 Sprachen in einer erstaunlichen Qualität übersetzen, die von der Vereinigung *OpenAI* entwickelte Software GPT-3 vermag es, ganze Texte generieren³ und das „Datenleak“ der sog. „Panama Papers“ konnte zu investigativjournalistischen Zwecken nach Inhalten und Zusammenhängen durchsuchbar gemacht werden.⁴ Das sind nur einige wenige Beispiele, die gemeinsam haben, dass sie darauf basieren, dass Computer lernen, die natürliche Sprache zu verstehen und unmittelbare inhaltliche Schlussfolgerungen als Text- oder Datengrundlagen ziehen. Letztlich basieren alle diese Beispiele auf der Technologie des sog. Text- und Data-Minings.

Das Text- und Data-Mining, im Folgenden als TDM abgekürzt, ist ein digitales Analyseverfahren, bei dem Datensätze, bestehend aus (teil-)strukturierten Rohdaten oder Texten, aber auch Bildern oder Musikstücke auf unterschiedliche Muster bzw. Zusammenhänge untersucht werden. Begrifflich bezieht sich der englische Begriff „to mine“ ursprünglich auf den Prozess des Abbaus von Bodenschätzen.⁵ Im übertragenen Sinne geschieht dies auch beim TDM: Aus digitalen Ressourcen werden Erkenntnisse gewonnen.⁶ Dafür ist in der Regel erforderlich, dass die Daten zu einem einheitlichen Datensatz vorverarbeitet wurden, der das sog. Korpus bildet. Nur dann sind die den Datensätzen immanenten Muster für diese Algorithmen sichtbar. Dazu werden umfangreiche Arbeitsschritte vorge-

1 B. Zimmer, Rowling und "Galbraith": an authorial analysis (16.07.2013).

2 <https://www.deepl.com/translator>.

3 Lobe, Frankfurter Allgemeine Sonntagszeitung 20.09.2020, S. 36.

4 Munzinger, Panama und Paradise Papers (21.06.2018); Mukhopadhyay/Ghosh, SSRN Journal 2020.

5 So die Oxford English Learner's Dictionaries sowie das PONS Wörterbuch, <https://www.oxfordlearnersdictionaries.com/definition/english/mining>; <https://de.pons.com/%C3%BCbersetzung/englisch-deutsch/mine>.

6 Schölkopf, Causality for Machine Learning (24.11.2019), S. 1 f.

nommen, in denen die Texte oder Daten, die analysiert werden sollen, vereinheitlicht und mit weiteren Informationen, die auf das konkrete Analyseziel angepasst sind, angereichert werden. Die Technologie des TDMs kann also als Forschungsmethode, d. h. als Werkzeug zur Beantwortung wissenschaftlicher Fragestellungen, eingesetzt werden. Weiter ist die Technologie in vielen Fällen Grundlage von Systemen, die zur sog. Künstlichen Intelligenz, abgekürzt als KI, zählen.⁷ Insgesamt hat die Technologie des TDMs eine große Bedeutung für Innovation und Wissenschaft. Nicht ohne Grund bezeichnet die DSM-RL aus dem Jahr 2019 das Verfahren als die „vorherrschende Technik der Digitalwirtschaft“.⁸

Wie viele andere digitale Prozesse auch ist das TDM wesentlich vom Urheberrecht reglementiert, obwohl die rechtlichen Probleme, die das TDM aufwirft, lange vorwiegend aus datenschutzrechtlicher Hinsicht diskutiert wurden⁹ und urheberrechtliche Aspekte erst nach 2010 Aufmerksamkeit erfuhren.¹⁰ Wenn also Ressourcen Gegenstand von TDM-Analysen werden sollen, die durch das UrhG geschützt sind, stellt sich die Frage, ob und in welchem Umfang ihre Nutzung zulässig ist. Das gilt sowohl für die algorithmusbasierte Analyse als auch für die Aufbereitung der genutzten Ressourcen. Mittels spezifischer Schrankenbestimmungen, von denen die erste auf deutscher Ebene schon vor dem Tätigwerden auf EU-Ebene geschaffen wurde (§ 60d UrhG a. F.) und die in Umsetzung der Vorgaben der späteren Art. 3 und 4 der DSM-RL reformiert wurde (§§ 44b, 60d UrhG), sollte das TDM auch an urheberrechtlichen Schutzgegenständen ermöglicht werden.¹¹ Insbesondere die DSM-RL zeigt dabei eine wertungsmäße

7 *Truyens/van Eecke*, CLSR 2014, S. 153, 154, die darüber hinaus auf die engen Verknüpfungen zum Datenbank-Management verweisen.

8 DSM-RL Erw. 8.

9 *Truyens/van Eecke* in: European Language Resources Association (Hrsg.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation* ({LREC}'14), S. 2182, 2182.

10 *Truyens/van Eecke* in: European Language Resources Association (Hrsg.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation* ({LREC}'14), S. 2182, 2182; Europäische Kommission, Mitteilung der Europäischen Kommission über Inhalte im digitalen Binnenmarkt, COM(2012) 789 final (18.12.2012), S. 4.

11 Die Gesetzesbegründung des UrhWissG enthält hierzu die Angabe: „Die Vorschrift ermöglicht es, auf gesetzlicher Grundlage Werke mit Inhalten aller Art automatisiert auszuwerten, z. B. Werke mit Texten, Daten, Bildern, Tönen oder audiovisuellen Inhalten, um damit nicht kommerzielle wissenschaftliche Forschung zu betreiben. Schlagwortartig wird dieser Vorgang häufig als sogenanntes Text und Data Mining bezeichnet. Die Regelung hat keinen Vorläufer im bisherigen deutschen Recht.“, BT-Drs. 18/12329, S. 40; die DSM-RL will über die zuvor bestehenden unionsrechtlichen

ge Privilegierung des TDMs, indem sie die Erlaubnisse des TDMs für die Mitgliedsstaaten verpflichtend ausgestaltet und auch das innovationspolitische Potenzial des TDMs würdigt.¹²

Ob es auf Grundlage der verschiedenen Normfassungen – insbesondere auf Grundlage der aktuellen Rechtslage – möglich ist, das Potenzial des TDMs auch bei der Analyse urheberrechtlicher Schutzgegenstände zu nutzen, bedarf einer genauen Analyse, denn die Regulierung digitaler Prozesse wirft in einem Urheberrecht, das weitestgehend einem analogen Zeitalter entstammt, verschiedene Probleme auf. In Bezug auf das TDM äußern sich diese Probleme etwa bei der Beantwortung der Frage danach, ob das TDM mit dem Werkgenuss vergleichbar ist, der traditionell urheberrechtlich freigestellt ist, für den aber in digitalen Umgebungen andere Mechanismen gelten. Gerade im Bereich der wissenschaftlichen Forschung sind außerdem besondere Anforderungen an einen Rechtsrahmen zu beachten, die die Einhaltbarkeit der wissenschaftlichen Eigengesetzlichkeiten, z. B. in Gestalt der Anforderungen der guten wissenschaftlichen Praxis,¹³ aber auch das Erfordernis von Rechtssicherheit, betreffen. Diese wissenschaftlichen Eigengesetzlichkeiten beziehen sich nicht nur auf die Beachtung wissenschaftlicher Grundsätze während des Forschungsprozess als solchen, sondern erfordern insbesondere eine Transparenz der wissenschaftlichen Forschung, d. h. auch der Forschungsergebnisse.

Letztlich verfolgt die Arbeit das Ziel, die Rechtsentwicklung auf ihre Eignung zur Regulierung des technischen Prozesses des TDMs zu überprüfen, Defizite aufzuzeigen und basierend auf einer teleologischen, systematischen sowie interdisziplinären Argumentation gezielte Rechtsanpassungen vorzuschlagen.

Möglichkeiten insbesondere rechtssichere Bedingungen schaffen: „Die Rechtsunsicherheit im Hinblick auf Text und Data Mining sollte beseitigt werden, indem für Hochschulen und andere Forschungsorganisationen sowie für Einrichtungen des Kulturerbes eine verbindliche Ausnahme für das ausschließliche Recht auf Vervielfältigung, aber auch auf das Recht, Entnahmen aus einer Datenbank zu untersagen, eingeführt wird.“, DSM-RL Erw. 11 S. 1.

12 So betont sie in Erw. 18 S. 1: „Verfahren des Text und Data Mining haben nicht nur im Zusammenhang mit der wissenschaftlichen Forschung hohe Bedeutung, sondern sie werden auch in großem Umfang sowohl von privaten als auch öffentlichen Einrichtungen eingesetzt, um große Datenmengen in verschiedenen Lebensbereichen und zu unterschiedlichen Zwecken zu analysieren, auch für staatliche Dienste, komplexe unternehmerische Entscheidungen und die Entwicklung neuer Anwendungen oder Technologien.“.

13 Z. B. kodifiziert in Deutsche Forschungsgemeinschaft, Leitlinien zur Sicherung guter wissenschaftlicher Praxis (2019).

§ 2 Herangehensweise

A. Grenzen der Untersuchung

Bei der urheberrechtlichen Betrachtung des TDMs kommt eine Vielzahl an Einzelfragen auf, die eine Eingrenzung des Untersuchungsgegenstands erfordert. Die vorliegende Untersuchung beschränkt sich auf die Analyse des Rechtsrahmens in Bezug auf das TDM bei der Nutzung von Schriftwerken, § 2 Abs. 1 Nr. 1, Abs. 2 UrhG, Datenbankwerken, §§ 4 Abs. 2, 2 Abs. 2 UrhG, sowie von Datenbanken, § 87a UrhG, die als Leistungsschutzrechte ins UrhG aufgenommen wurden, für das TDM. Nicht untersucht werden hingegen Besonderheiten, die sich bei der Analyse von Werken der bildenden Kunst oder Musik ergeben.¹⁴ In Bezug auf das TDM in Web-Kontexten, das durch Web-Crawler oder -Scraper durchgeführt bzw. mittels jener praktiziert wird, ergaben sich seit den 2000er-Jahren zahlreiche Gerichtsverfahren in kommerziellen Kontexten, in denen zumeist Gegenstand war, dass Webrobots Daten aus Webseiten abriefen und in Meta-Datenbanken einpflegten. Diese Fragestellungen werden nur referenziell zur urheberrechtlichen Betrachtung des TDMs thematisiert, nicht aber hinsichtlich eigener Rechtsfragen, die durch sie aufgeworfen werden.¹⁵ Nicht weiter analysiert werden Fragen nach einem Schutz der Ausgabe aus dem Bereich des maschinellen Lernens, z. B. sog. generativer Modelle, die vermeintliche Werke oder Schutzgegenstände, z. B. in einem bestimmten künstlerischen Stil erschaffen können.¹⁶ Ebenso wenig wird thematisiert, ob und inwieweit die Konzeption von Forschungsfragen urheberrechtlich schutzfähig ist.

14 Vgl. dazu allerdings *Rack/Frieler* in: Gräfe/Telemedicus e.V. (Hrsg.), Tagungsband zur Sommerkonferenz 2019: Über den Tellerrand, S. 27.

15 Thematisiert wurde in diesem Kontext in erster Linie die Frage nach dem Bestehen eines virtuellen Hausrechts und die Bezüge zur Informationsfreiheit, vgl. v. *Schönfeld*, Screen Scraping und Informationsfreiheit, A. *Schmidt*, Virtuelles Hausrecht und Webrobots; *Elteste*, CR 2015, S. 447; *Gausling*, CR 2021, S. 609; *Golla/v. Schönfeld*, K&R 2019, S. 15 ff.; *Schapiro/Żdanowiecki*, MMR 2015, S. 497 ff.

16 Das untersuchen *Käde*, Kreative Maschinen und Urheberrecht, S. 173 ff.; *Grätz*, Künstliche Intelligenz im Urheberrecht, S. 45 ff.; *Dornis*, GRUR 2019, S. 1252, 1256 ff.; *Ehinger/Stiemerling*, CR 2018, S. 761 ff.; *Hetmank/Lauber-Rönsberg*, GRUR 2018, S. 574 ff.; *Lauber-Rönsberg*, GRUR 2019, S. 244 ff. Im Bereich sprachbasierter Werke ist es z. B. möglich, automatische Abstracts zu erstellen oder Texte im Stil bestimmter Autoren verfassen zu lassen, zu Textgeneratoren *Gräfe/J. Kahl*, MMR 2021, S. 121 ff.; *Lobe*, Frankfurter Allgemeine Sonntagszeitung 20.09.2020, S. 36.

B. Interdisziplinärer Ansatz

Die Arbeit untersucht das TDM als Methode der wissenschaftlichen Forschung und der Datenanalyse anhand urheberrechtlicher Schutzgegenstände. Zentrale Untersuchungsgegenstände bilden deswegen das urheberrechtliche Regelungsregime im europäischen Mehrebenensystem, das Verhältnis von Schutzrecht und Schranke sowie die Besonderheiten des Wissenschaftsurheberrechts. Daneben werden auch grundrechtliche Wertungen miteinbezogen, die einerseits das Verhältnis des Schutzes des Eigentums zu Inhalts- und Schrankenbestimmungen, Art. 14 Abs. 1 S. 1, 2 GG und andererseits den Einfluss der Freiheit von Wissenschaft und Forschung auf urheberrechtliche Schrankenbestimmungen, Art. 5 Abs. 3 S. 1 Alt. 2 GG, betreffen.

Daneben verfolgt die Arbeit einen interdisziplinären Ansatz, bei dem Grundprinzipien und Erkenntnisse anderer Fachdisziplinen für die juristische Betrachtung nutzbar gemacht werden: Zur Erfassung des technischen Verfahrens wird in größerem Umfang informationstechnische Literatur gewürdigt. Weiter wird die Bedeutung des TDM-Verfahrens und der Forschungsdaten des TDMs – der Korpora – aus wissenschaftstheoretischer und kulturwissenschaftlicher Perspektive betrachtet. Zur Erfassung der ökonomischen Auswirkung des Urheberrechts auf Innovation bedarf es außerdem der Berücksichtigung rechtsökonomischer Erwägungen eingeschlossen der Auswertung einiger weniger empirischer Studien, die diese Problematik konkret in Bezug auf das TDM untersuchen.

C. Gang der Untersuchung

Der erste Teil der Arbeit bildet die Grundlage der Untersuchung und stellt dar, was das TDM ist und welche Bedeutung es als Forschungsmethode, für das maschinelle Lernen sowie weitere Teilbereiche der künstlichen Intelligenz hat. Anschließend nähert sich die Untersuchung einem Ablauf des TDMs an, der in gewissem Umfang auf verschiedene Anwendungen des TDMs übertragen werden kann. Herausgearbeitet wird daneben, wie die Muster einzuordnen sind, die das TDM hervorbringt und worin ihr besonderes Potenzial liegt. Im weiteren Fortgang widmet sich der erste Teil der Arbeit vorwiegend der Steuerungswirkung von wissenschaftlicher Forschung sowie Innovation durch das Urheberrecht, die später dazu nutzbar gemacht wird, den Rechtsrahmen für das TDM zu beurteilen.

Der zweite Teil der Arbeit geht dem Forschungs- und Analyseprozess des TDMs in seiner urheberrechtlichen Einordnung und der Frage, ob es aus übergeordneten Gründen freigestellt oder jedenfalls freizustellen ist, nach. Auf der Analyse des Urheberrechtsrahmens des TDM-Prozesses aufbauend wird die Rechtsentwicklung in Bezug auf ihre Eignung und Vollständigkeit analysiert.

Im dritten Teil stehen die Abläufe im Vordergrund, die im Anschluss an die eigentlichen Forschungsarbeiten stattfinden – die Möglichkeit wissenschaftlichen Überprüfung durch eine transparente Archivierung und Zugänglichkeit für die Überprüfungen und zur weiteren Nutzung der Korpora für Anschlussforschungen, über deren Definition, Unterteilung und Bezeichnung weitestgehend Uneinigkeit herrscht. Aus den Rückschlüssen, die diesbezüglich aus der Wissenschaftsfreiheit sowie aus der Wissenschaftstheorie und der Kulturwissenschaft gezogen werden, wird ein Maßstab eines geeigneten Rechtsrahmens gebildet. Schließlich wird das geltende Recht daraufhin überprüft, inwieweit es die gesetzten Maßstäbe erfüllt. Alternativ werden Ansätze vorgestellt, wie Korpora über den explizit normierten Umfang hinaus nachnutzbar gemacht werden können. Dabei stehen sich schließlich technische und rechtliche Ansätze gegenüber, die gleichermaßen relevant in technischen Sachverhalten und ihrer urheberrechtlichen Regulierung sind.

Im vierten Teil der Arbeit werden die Erkenntnisse aus der Arbeit in Thesenform zusammengefasst, die schließlich in konkreten Verbesserungsvorschlägen münden. Zuletzt wagt die Arbeit einen Ausblick.

Erster Teil: Text- und Data-Mining zwischen Technologie, Innovation und Wissenschaft

Das TDM dient einerseits als eigenständiges Analyseverfahren, das genutzt werden kann um Erkenntnisse über potenziell große Datensätze zu erlangen. Andererseits stellt es die Schlüsseltechnologie für das maschinelle Lernen und andere Teilbereiche des Forschungsfeldes der Künstlichen Intelligenz dar. Im nachfolgenden § 3 werden deswegen sowohl begriffliche als auch technische Grundlagen von TDM und der Künstlichen Intelligenz, eingeschlossen derjenigen Technologien, die für das TDM besonders relevant sind, erläutert. In Bezug auf das Begriffsverständnis des TDM wird zuerst aus technischer und dann aus juristischer Perspektive untersucht, welche Teile des umfangreichen technischen Prozesses unter das TDM zu fassen sind bzw. was „das“ TDM als solches ist. Daraus resultierend wird ein Begriffsverständnis für die Arbeit gebildet. § 3 stellt daneben auch einen exemplarischen Forschungsprozess vor. Anschließend werden die Muster, die das TDM hervorbringt, mittels einer wissenschaftstheoretischen Herangehensweise kritisch beleuchtet.

Sodann thematisiert § 4 das Verhältnis von Wissenschaft und Urheberrecht, das für die Forschungsmethode des TDMs relevant ist. In diesem Bereich wirken sich bestimmte urheberrechtliche Grundwertungen auf eine Art und Weise aus, die sich vom sonstigen Urheberrecht unterscheidet und die dazu genutzt werden können, wissenschaftliche Forschung zu steuern.

Gerade aufgrund der Relevanz für das Training von KI-Modellen hat das TDM auch eine ökonomische Bedeutung, die es nach Bestrebungen auf europäischer und nationaler Ebene zu nutzen gilt. Aus diesem Grund legt § 5 der Arbeit in § 5 neben dem Verständnis der Steuerungswirkung bzgl. wissenschaftlicher Forschung auch ein Verständnis der Steuerungswirkung des Urheberrechts bzgl. Innovation zugrunde.

§ 3 Text- und Data-Mining zwischen Künstlicher Intelligenz und digitaler Forschung

A. Das Begriffsverständnis des Text- und Data-Minings

I. Das technische Begriffsverständnis

Unter dem Data-Mining und dem Text-Mining wurde ursprünglich nur die automatisierte algorithmusbasierte Analyse von Daten- oder Textbeständen verstanden, d. h. der eigentliche automatisierte Auswertungsvorgang, nicht hingegen mehrstufige Abläufe mit umfangreichen Vorverarbeitungsschritten wie der Erstellung eines Datensatzes, der als Korpus bezeichnet wird, die vor der eigentlichen Datenanalyse stattfinden.¹⁷ Für die Analyse sind auch die englischen Bezeichnungen der „*knowledge extraction*“ (Wissensextrahierung), der „*information discovery*“ (Informationsentdeckung) bzw. des „*information harvestings*“ und des „*data pattern processings*“ (Daten-Muster-Verarbeitung) gebräuchlich.¹⁸ Mehrstufige Verfahren kamen seit den 1990er-Jahren als sog. „*Knowledge Discovery in Databases*“ bzw. als Wissensentdeckung in Datenbanken (abgekürzt als KDD) ins Gespräch. Hintergrund war, dass eine zielgerichtete Datenvorbereitung entscheidende Einflüsse auf Analyseergebnisse hat, sie sich aber als äußerst arbeitsintensiv darstellt.¹⁹ Mit dem KDD wurde erstmals ausdrücklich das Ziel verfolgt, Wissen in mehrstufigen Verfahren aus größeren Datengrundlagen zu extrahieren.²⁰ Fayyad, Piatetsky-Shapiro und Smyth, die das Begriffsverständnis des TDMs prägten, stellten die folgende Definition vor: „*Wissensentdeckung in Datenbanken ist der nicht triviale Prozess der Identifikation gültiger, neuer, potenziell nützlicher und schlussendlich verständlicher*

17 „In our view, KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data.“, Fayyad/Piatetsky-Shapiro/Smyth, AI Mag 1996, S. 37, 39, in Bezug auf Text-Mining „Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools“, Feldman/Sanger, The Text Mining Handbook, S. 1; ebenso Beierle/Kern-Isberner, Methoden wissensbasierter Systeme, S. 145; Jo, Text Mining, S. 3.

18 Fayyad/Piatetsky-Shapiro/Smyth, AI Mag 1996, S. 37, 39; Leimeister, Einführung in die Wirtschaftsinformatik, S. 178.

19 Beierle/Kern-Isberner, Methoden wissensbasierter Systeme, S. 145.

20 Fayyad/Piatetsky-Shapiro/Smyth, AI Mag 1996, S. 37, 39.

Muster in (großen) Datenbeständen“.²¹ Das KDD wird als ein mehrstufiges Verfahren beschrieben, das es beinhaltet, die Daten zielgerichtet auszuwählen, sie zu vereinheitlichen und zu bereinigen, eine Analysemethode zielgerichtet auszuwählen, sie der eigentlichen algorithmusbasierten Analyse zu unterziehen sowie die Analyseergebnisse im Anschluss auszuwerten.²² In Abgrenzung dazu solle das Data Mining die algorithmusbasierte Mustererkennung ohne die Vorverarbeitungsschritte bezeichnen.²³ Heute wird das TDM in der Durchführungsweise, die Gegenstand dieser Untersuchung ist, hingegen synonym zum KDD verstanden,²⁴ wobei die TDM-Analyse nach wie vor den Kern des Prozesses ausmachen soll.²⁵

Weiter wird vom sog. „*Information Retrieval*“ gesprochen, wenn es darum geht, dass Informationen aus Datenbeständen extrahiert werden sollen. Das Information Retrieval und das Data-Mining werden teilweise dadurch unterschieden, dass das Data-Mining der Auffindung unbekannter Strukturen dient, während das Information Retrieval gerade die Informationen auffinden soll, die bereits bestehen.²⁶ Ob diese Abgrenzung geeignet ist, ist in Zweifel zu ziehen, schließlich sind, wie später erläutert wird, alle Muster, die aus den Datenbeständen extrahiert werden, in diesen Datenbeständen bereits angelegt.

Weitere Unterfälle sind das sog. „*Web-Mining*“ bzw. das „*Web-Crawling*“ bzw. „-*Scraping*“ oder „-*Harvesting*“, bei denen das Web von sog. „*Web-Crawlern*“, d. h. speziellen Computerprogrammen, durchsucht wird, die wiederum Daten auslesen und sie „*kratzen*“, d. h. herunterladen.²⁷ Aus technischer Perspektive rufen die Web-Crawler je nach Auftrag eine Viel-

21 Fayyad/Piatetsky-Shapiro/Smyth, AI Mag 1996, S. 37, 40 f.; Fayyad/Piatetsky-Shapiro/Smyth in: Fayyad (Hrsg.), *Advances in knowledge discovery and data mining*, S. 1, 6; übersetzt aus dem Englischen von Wrobel/Joachims/Morik in: Görz/Schneeberger/Schmid (Hrsg.), *Handbuch der Künstlichen Intelligenz*, S. 405, 409.

22 Beierle/Kern-Isbener, *Methoden wissensbasierter Systeme*, S. 147 f.; Fayyad/Piatetsky-Shapiro/Smyth, AI Mag 1996, S. 37, 40 f.

23 Fayyad/Piatetsky-Shapiro/Smyth in: Fayyad (Hrsg.), *Advances in knowledge discovery and data mining*, S. 1, 4.

24 Leimeister, *Einführung in die Wirtschaftsinformatik*, S. 177 f.; Wrobel/Joachims/Morik in: Görz/Schneeberger/Schmid (Hrsg.), *Handbuch der Künstlichen Intelligenz*, S. 405, 405, 409.

25 Fayyad/Piatetsky-Shapiro/Smyth, AI Mag 1996, S. 37, 37.

26 Jannidis/Kohle/Rehbein, *Digital Humanities*, S. 268.

27 Jo, *Text Mining*, S. 13; v. Schönfeld, *Screen Scraping und Informationsfreiheit*, S. 24 f.; Elteste, CR 2015, S. 447, 447; Schrenk, *Webbots, spiders, and screen scrapers*, S. 227.

zahl an Webseiten auf, die eigentlich für menschliche Nutzer²⁸ konstruiert wurden,²⁹ und verarbeiten deren Inhalte.³⁰ Anschließend laden sie bestimmte Teile dieser Web-Präsenzen herunter, extrahieren in einem weiteren Schritt darin enthaltene Daten und speichern sie dann in einer Meta-Datenbank ab.³¹ Web-Crawler und Web-Scraper werden dergestalt unterschieden, dass Erstere auf statische Webseiten zugreifen, deren Inhalt sich nach Schließen der Webseite nicht ändert, wovon Letztere nicht abhängig sind, denn diese vermögen es, Formulare abzusenden und damit menschliches Verhalten zu imitieren.³² Das Web-Crawling oder -Scraping kann sowohl kommerziellen als auch wissenschaftlichen Zwecken dienen, z. B. der Auswertung von Daten aus sozialen Netzwerken.³³ Grundsätzlich basiert jede Suchmaschine auf Web-Crawlern.³⁴

Während *Fayyad*, *Piatetsky-Shapiro* und *Smyth* also davon ausgehen, dass das Data-Mining im engeren Sinne nur die algorithmusbasierte Datenanalyse ist und der Begriff des Wissensentdeckung in Datenbanken wiederum den Gesamtprozess meint, gehen andere davon aus, dass zum Data-Mining nur die Datenbeschaffung zählt.³⁵ Begrifflich treffen beide Verständnisse zu: Nach Daten zu schürfen hat eine andere Bedeutung als die Daten zu schürfen, also aus ihnen etwas zu generieren. Das gilt auch für das

28 Die vorliegende Arbeit bemüht sich um eine geschlechtsneutrale Formulierung. Wo dies nicht gelingt, sind mit der genutzten Form nicht Personen eines bestimmten Geschlechts angesprochen.

29 *Golla/v. Schönfeld*, K&R 2019, S. 15, 16; *H. Ernst/Jochen Schmidt/Beneken*, Grundkurs Informatik, S. 760 f.; *Gumm/Sommer et al.*, Einführung in die Informatik, S. 664 f. und deren Erläuterung des technischen Aufbaus von HTML-Webseiten.

30 *A. Schmidt*, Virtuelles Hausrecht und Webrobots, S. 15 ff.

31 *A. Schmidt*, Virtuelles Hausrecht und Webrobots, S. 17 f.; aus technischer Perspektive anschaulich *Schrenk*, Webbots, spiders, and screen scrapers, S. 174.

32 *A. Schmidt*, Virtuelles Hausrecht und Webrobots, S. 17 ff. Damit sind sie technisch anspruchsvoller aufgebaut. Die besondere technische Herausforderung liegt insofern erstens darin, dass Webseiten für menschliche Nutzer programmiert sind, *Golla/v. Schönfeld*, K&R 2019, S. 15, 15, und andererseits, dass insbesondere die Analyse dynamischer Webseiten besondere Herausforderungen birgt.

33 Ein bekanntes Beispiel, in dem Web Scraping zum Einsatz kam, ist die Cambridge Analytica Affäre, in der über die vormalig noch funktionsfähige Programmier-Schnittstelle von Facebook Nutzerdaten zu Wahlkampfzwecken ausgelesen wurden, *Golla/v. Schönfeld*, K&R 2019, S. 15, 15; *Kastrenakes*, Facebook will limit developers' access to account data (21.03.2018).

34 *A. Schmidt*, Virtuelles Hausrecht und Webrobots, S. 16 f.; *Schrenk*, Webbots, spiders, and screen scrapers, S. 173; *Abts/Mülder*, Grundkurs Wirtschaftsinformatik, S. 365. Der bekannteste Web-Crawler dürfte damit derjenige sein, den Google für seine Suchmaschine nutzt, vgl. *Golla/v. Schönfeld*, K&R 2019, S. 15, 16.

35 *Leimeister*, Einführung in die Wirtschaftsinformatik, S. 177.

Web-Mining, -Crawling oder -Scraping, die zwar der Datenbeschaffung dienen können, bei denen aber bereits die Auswahl dieser Daten auf einer TDM-Analyse basiert.³⁶

Das Text-Mining und das Data-Mining unterscheiden sich wiederum hauptsächlich in ihrem Untersuchungsobjekt. Diesbezüglich wird teilweise differenziert, weil Texte als unstrukturiert, während Daten nach mancher Auffassung als strukturiert gelten.³⁷ Doch auch Texte sind Daten im weiteren Sinne³⁸ und aufgrund ihrer logischen Binnenstruktur (der Aufbau von Texten, die Gliederung unter Überschriften und in Absätze sowie die Grammatik) nicht in diesem Sinne unstrukturiert.³⁹ Zudem ist gerade dann, wenn umfassende Vorverarbeitungsschritte durchgeführt werden, kaum eine sinnvolle Abgrenzung mehr möglich, denn während Daten in den Vorverarbeitungsschritten strukturiert werden, wird die Binnenstruktur von Texten teilweise aufgebrochen, wie an späterer Stelle ausführlich erläutert wird.⁴⁰

II. Das urheberrechtliche Begriffsverständnis

Im Urheberrecht hat sich der Begriff des Text- und Data-Minings als ein Sammelbegriff etabliert, der in den jeweiligen Gesetzesmaterialien umschrieben und teils auch legaldefiniert wird.

Die erste deutsche Norm, die das Text- und Data-Mining an urheberrechtlichen Schutzgegenständen ausdrücklich erlaubt, § 60d UrhG a. F., der zwischen dem 01.03.2021 und dem 06.06.2021 galt, verzichtet zwar auf eine Legaldefinition, bezieht sich aber ausweislich der Gesetzesbegründung auf *„einen mehrstufigen Prozess, bei dem große Datenmengen, legaldefiniert als Ursprungsmaterial, in digitaler Form gesammelt, aufbereitet und automatisiert nach bestimmten Merkmalen durchsucht und ausgewertet werden“*.⁴¹ Das Korpus, unter dem § 60d UrhG a. F. die Sammlung der zuvor ver-

36 Davon gehen auch Golla/v. Schönfeld, K&R 2019, S. 15, 17 f. aus.

37 Käde, Kreative Maschinen und Urheberrecht, S. 57; W. Zimmer, Ansturm der Algorithmen, S. 196; anderer Auffassung ist Jo, Text Mining, S. 3, demzufolge das Text-Mining ein Spezialfall des Data-Mining ist.

38 Dazu näher S. 38.

39 Heyer/Quasthoff/Wittig, Text Mining: Wissensrohstoff Text, S. 1; Jo, Text Mining, S. 3, W. Zimmer, Ansturm der Algorithmen, S. 196 f.; Gesellschaft für Informatik, Text Mining.

40 Vgl. dazu S. 48 ff.

41 BT-Drs. 18/12329, S. 40.

vielfältigten, normalisierten, strukturierten und kategorisierten Inhalte versteht, wird nach dem gesetzgeberischen Verständnis anschließend – u. U. mit eigens programmierten Computerprogrammen – ausgewertet.⁴² Diese extrahieren „statistische Häufigkeiten oder Korrelationen“, die schließlich für die wissenschaftliche Forschung zur Verfügung stehen.⁴³ Ziel des Gesamtprozesses sei es, „neue Erkenntnisse zu gewinnen“.⁴⁴ Dieses weit gefasste Verständnis lässt zunächst die Schlussfolgerung zu, der Gesetzgeber meine mit dem TDM den Gesamtprozess, d. h. die Datenaufbereitung und die -Analyse und vertrete insoweit ein an das KDD angelehntes Verständnis. Ein anderer Eindruck wird vermittelt, weil die Analyse-Software als „eigentliche Software für das sogenannte Text und Data Mining“ bezeichnet wird,⁴⁵ das wiederum die Schlussfolgerung zulässt, das TDM sei nur die algorithmusbasierte Korpusanalyse exklusive der Vorverarbeitungsschritte. An dieser Uneindeutigkeit zeigt sich der vorläufige Charakter der Norm, die in dem Bewusstsein, bald neues Unionsrecht umsetzen zu müssen, verabschiedet wurde.⁴⁶

Auch der DSM-RL vom 17.04.2019 liegt ein weites Verständnis des TDMs zugrunde, indem sie es in Art. 2 Nr. 2 DSM-RL als „eine Technik für die automatisierte Analyse von Texten und Daten in digitaler Form, mit deren Hilfe Informationen unter anderem — aber nicht ausschließlich — über Muster, Trends und Korrelationen gewonnen werden können“ legaldefiniert. Demnach dient das TDM dem Zweck, digitale Informationen computergestützt auszuwerten, um neue Erkenntnisse zu gewinnen und Trends zu erkennen.⁴⁷ Doch ist daraus nicht zwingend zu schlussfolgern, dass die DSM-RL mit dem Begriff des TDMs den Gesamtprozess meint, so erlauben Art. 3 und 4 DSM-RL Vervielfältigungen (und Entnahmen) für das TDM.⁴⁸ Das lässt den Schluss zu, dass das TDM die eigentliche Analyse ist,

42 BT-Drs. 18/12329, S. 40.

43 BT-Drs. 18/12329, S. 40.

44 BT-Drs. 18/12329, S. 22.

45 BT-Drs. 18/12329, S. 40.

46 BT-Drs. 18/12329, S. 40.

47 DSM-RL Erw. 8 S. 1, 2.

48 In der deutschen Übersetzung enthält Art. 3 Abs. 1 DSM-RL einen grammatikalischen Fehler: „Die Mitgliedstaaten sehen eine Ausnahme [...] für Vervielfältigungen und Entnahmen vor, die [...] zum Zwecke der wissenschaftlichen Forschung für die Text und Data Mining vorgenommen werden“. In Art. 4 Abs. 1 DSM-RL wird es demgegenüber klarer, dieser sieht Ausnahmen oder Beschränkungen „für zum Zwecke des Text und Data Mining“ vorgenommenen Vervielfältigungen bzw. Entnahmen vor. In der englischen Sprachfassung des Art. 3 Abs. 1 DSM-RL heißt es: „in order to carry out, for the purposes of scientific research, text and data mining“.

für die im Vorfeld bestimmte Handlungen – die Vorverarbeitungsschritte – erforderlich sind.

Die aktuellste Normfassung im deutschen Urheberrecht kombiniert eine reformierte Fassung des § 60d UrhG mit dem neu geschaffenen § 44b UrhG, der das TDM in Anlehnung an das Verständnis von Art. 2 Nr. 2 DSM-RL als „*die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen*“ definiert. „Für das Text und Data Mining“⁴⁹ dürfen wiederum bestimmte urheberrechtlich relevante Handlungen vorgenommen werden, §§ 44b Abs. 2, 60d Abs. 1 UrhG. Auch diesbezüglich liegt es also nahe, dass der Gesetzgeber unter dem TDM nur die algorithmusbasierte Analyse versteht und verschiedene Handlungen zu diesem Zwecke, d. h. die Vorverarbeitungsschritte, erlauben möchte.

Neben diesen Abgrenzungen deuten die gesetzgeberischen Formulierungen und die Ausführungen der DSM-RL darauf hin, dass sich die Normgeber auf das TDM als Forschungsmethode mit einem konkreten Analyseziel beziehen, wie sie heute z. B. in den digitalen Geisteswissenschaften (englisch: Digital Humanities) praktiziert wird.⁵⁰ Dennoch ist der Begriff so weit gefasst, dass auch eine große Bandbreite anderer Anwendungen, die auf der Technologie des TDM basieren, erfasst werden, wie an späterer Stelle diskutiert wird.⁵¹

III. Begriffsfindung

Aus technischer und rechtlicher Perspektive zeigt sich, dass mit dem Begriff des TDMs sowohl die eigentliche algorithmusbasierte Datenanalyse als auch der Gesamtprozess des TDMs inklusive der Datenvorverarbeitung adressiert sein kann. Gerade aus technischer Perspektive existiert eine Fülle wechselbezüglicher Begriffe mit weitreichenden inhaltlichen Überschneidungen. Die verschiedenen Definitionen haben im Wesentlichen gemeinsam, dass sie sich darauf beziehen, dass aus mehr oder minder strukturier-

49 Im Gesetz wird statt der in dieser Arbeit verwendeten Schreibweise – Text- und Data-Mining – eine Schreibweise ohne Bindestriche vorgenommen, das gilt sowohl im deutschen Recht als auch in der DSM-RL. Wenn aus dem Gesetz zitiert wird, werden deswegen keine Bindestriche genutzt.

50 Das wird ausführlich auf S. 36 ff. thematisiert.

51 Vgl. dazu den zweiten Teil der Arbeit, insbesondere § 9.

ten Daten ein Mehrwert gezogen wird, der in der Offenbarung neuer Erkenntnisse, unbekannter Muster oder Korrelationen liegen kann. Uneinigkeiten bestehen darin, welche Arbeitsschritte für das TDM begriffsprägend sind. Der Begriff des TDMs ist im Urheberrecht ebenso uneindeutig, aber doch tendenziell weit gefasst.

Trotz der Uneindeutigkeiten lässt sich eine gemeinsame Linie der juristischen und technischen Perspektive herausarbeiten, die mehrstufige Verfahren unter den Gesamtkomplex des TDMs fassen, bei dem große Datenbestände aufgebaut und aufbereitet werden, um sie anschließend auf Muster oder Strukturen hin zu analysieren.

Der folgenden Untersuchung wird ein Verständnis zugrunde liegen, das das TDM als algorithmusbasierte Methode zur Mustererkennung ansieht, das zu verschiedensten Zwecken genutzt werden kann und das eine umfassende Datenvorbereitung erfordern kann. Das TDM erfasst dabei in Anlehnung an den Begriff des KDDs nicht nur die eigentliche Datenanalyse, sondern den Gesamtprozess, wobei bei Bedarf zwischen den Vorverarbeitungsschritten und der algorithmusbasierten TDM-Analyse im eigentlichen Sinne differenziert wird. Wenn diese Analyse gemeint ist, wird dies durch die Formulierung „*TDM im engeren Sinne*“ oder „*TDM-Analyse im engeren Sinne*“ gesondert klargestellt. Das Text-Mining wird in dieser Untersuchung als Unterkategorie des Data-Minings angesehen, wobei die Begriffe nur in den Fällen getrennt verwendet werden sollten, wenn es auf ihre Differenzierung ankommt.

B. Text- und Data-Mining zur Erkenntnisgewinnung

Seine Anfänge fand das TDM als geisteswissenschaftliches Forschungswerkzeug bereits in den 1940er-Jahren, als der italienische Jesuit *Roberto Busa*, der noch im Rahmen seiner Dissertation zur Beantwortung seiner Forschungsfragen händisch Wörter zählte, in Kooperation mit dem heutigen Technologie- und Softwarekonzern, damals aber noch Lochkarten produzierenden, Unternehmen IBM, zur Erforschung unbewusster Formulierungen das Gesamtwerk von *Thomas von Aquin* systematisierte.⁵² Das Projekt sollte es ermöglichen, in umfangreichen Textsammlungen schnell bestimmte Phänomene oder Passagen aufzufinden.⁵³ Aus diesem Projekt

52 Jannidis/Kohle/Rehbein, Digital Humanities, S. 3.

53 Sasaki in: Klimpel/Euler (Hrsg.), Der Vergangenheit eine Zukunft, S. 84, 85. Ähnliche Ansätze gab es auch in den Vereinigten Staaten, diese konnten aber erst in

ist der *Index Thomisticus*,⁵⁴ ein noch heute erforschtes Korpus hervorgegangen, bestehend aus dem systematisierten, lemmatisierten und annotierten Gesamtwerk von *Thomas von Aquin* sowie aus Sekundärquellen. *Roberto Busa* musste also feststellen, dass technische Hilfe notwendig ist, um überhaupt bzw. in einem angemessenen Zeitrahmen und Budget die nicht offensichtlichen Zusammenhänge und Bedeutungen in größeren Textbeständen zu erkennen.

In Bezug auf die digitalen Geisteswissenschaften ist viel diskutiert, ob es sich um eine Querschnittsmaterie handelt, die computergestützt neue Erkenntnisse in geisteswissenschaftlichen Fragestellungen untersucht, oder um eine eigene Forschungsdisziplin.⁵⁵ Letzteres würde der Vielfalt der geisteswissenschaftlichen Disziplinen nicht gerecht.⁵⁶ Allerdings sind auch die Fragestellungen der digitalen Geisteswissenschaften im Vergleich zu denjenigen der traditionellen Geisteswissenschaften durch die digitalen

den 1960er Jahren realisiert werden, *Jannidis/Kohle/Rehbein*, Digital Humanities, S. 3.

54 Dieser ist heute auch online verfügbar, s. dazu <https://www.corpusthomicum.org/wintrode.html>.

55 *Sable* in: Baum/Stäcker (Hrsg.), Grenzen und Möglichkeiten der Digital Humanities: „Die DH erscheinen als voll ausgereiftes, eigenständiges Fach. Es kann kaum bezweifelt werden, dass wir es hier mit einer eigenständigen Disziplin zu tun haben. Allerdings scheint die Sache doch nicht ganz so einfach zu sein, weil diese Disziplin einen besonderen Charakter hat. Als »Brückenfach«, ihrem Wesen nach inter- oder transdisziplinär und zugleich viele spezialisiertere Ausprägungen umgreifend, bestimmt sie sich nämlich nicht nur in Abgrenzung von anderen Fächern, sondern gerade auch in ihrer Hinwendung zu ihnen.“. Anderer Auffassung sind *Jannidis/Kohle/Rehbein*, die die digitalen Geisteswissenschaften als die Summe aller Versuche definieren, „die Informationstechniken auf den Gegenstandsbereich der Geisteswissenschaften zu übertragen“, *Jannidis/Kohle/Rehbein*, Digital Humanities, S. 13, die aber auch die Eigenheiten der untersuchten Fragestellungen als kennzeichnend für die Anerkennung einer eigenen Forschungsdisziplin ansehen, *Jannidis/Kohle/Rehbein*, Digital Humanities, S. 13 f. Das soll gerade dann gelten, wenn Erkenntnisse getroffen werden sollen, die ohne die Mittel der digitalen Geisteswissenschaften nicht auffindbar wären. Gerade das Faktum, dass es nicht „die analogen Geisteswissenschaften“ gibt, spricht auf der anderen Seite dafür, auch die digitalen Geisteswissenschaften nicht als Forschungsdisziplin einzuordnen, *Jannidis/Kohle/Rehbein*, Digital Humanities, S. 15. Diese „Werkzeugtheorie“ spricht gleichwohl nicht gegen die Bedeutsamkeit oder das Potenzial der digitalen Geisteswissenschaften.

56 Das repräsentiert auch das Tagungsprogramm der Jahrestagungen des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd), die die wohl größten jährlichen Tagungen in diesem Themenfeld im deutschsprachigen Raum abhält sowie das Selbstverständnis dieses Verbands, Digital Humanities im deutschsprachigen Raum, Was sind digitale Geisteswissenschaften?

Mittel bestimmt, weswegen die digitalen Geisteswissenschaften auch nicht auf eine reine Methodik reduziert werden sollten.⁵⁷ Letztlich fällt es schwer, die digitalen Geisteswissenschaften begrifflich zu umreißen. Jedenfalls wird das TDM in den digitalen Geisteswissenschaften als Schlüsseltechnologie genutzt, um auch – aber nicht nur – sprachbasierte Inhalte auszuwerten.⁵⁸

C. Text- und Data-Mining und Künstliche Intelligenz

Wie bereits erläutert wurde, ist das TDM eng mit dem Forschungsfeld der KI verbunden. Im Folgenden werden zunächst deren Begriffsverständnis und anschließend verschiedene Ausprägungen bzw. Teilbereiche, die besonders relevant für das TDM sind, thematisiert, bevor das Verhältnis von TDM und KI erläutert werden kann.

I. Technische Grundlagen der Künstlichen Intelligenz

1. Begriffsentwicklung der Künstlichen Intelligenz

Der Begriff der Künstlichen Intelligenz stammt aus dem Jahr 1955 und wurde durch *McCarthy* und *Minsky* geprägt.⁵⁹ Die theoretischen Grundsteine für einzelne Technologien wurden aber bereits in den 1940er-Jahren gelegt.⁶⁰ Trotz dieser frühen Anfänge wurden besondere Fortschritte erst deutlich später aufgrund der fortschreitenden Rechenleistung von Großcomputern zu verzeichnet, was auch die Nutzung künstlicher neuronaler Netze bzw. des Deep Learnings ermöglichte.⁶¹

Der Begriff der KI sowie die Beantwortung der Frage, wann ein Programm als intelligent bezeichnet werden kann, ist bis heute äußerst

57 *Jannidis/Kohle/Rehbein*, Digital Humanities, S. 13 f.

58 „Zentraler Arbeitsbereich“, *Döhl*, RuZ 2020, S. 195, 198.

59 *McCarthy/Minsky/Rochester/Shannon*, AI Mag 2006, S. 12 ff.; *Pallay* in: Mainzer (Hrsg.), Philosophisches Handbuch Künstliche Intelligenz, S. 1, 13; *Kaplan*, Künstliche Intelligenz, S. 27 f.; *Lenzen*, Künstliche Intelligenz, S. 22.

60 *Pallay* in: Mainzer (Hrsg.), Philosophisches Handbuch Künstliche Intelligenz, S. 1, 5 f.; *Kaplan*, Künstliche Intelligenz, S. 27 ff.; *Negnevitsky*, Artificial intelligence, S. 2.

61 *Mikut*, at 2020, S. 295, 295 f.

strittig.⁶² Bzgl. der Bestimmung des Autonomiegrads von Systemen, die auf Technologien der KI basieren, wird an die Differenzierung zwischen der sog. „starken“ und „schwachen“ KI angeknüpft, die wiederum davon abhängt, ob die entsprechenden Modelle universell einsetzbar sind und ein Bewusstsein haben oder lediglich die Intelligenz des Menschen imitieren.⁶³ Es sei an dieser Stelle klargestellt, dass es eine „starke – also eine universell einsetzbare – KI nach dem aktuellen Stand der Technik nicht gibt.“⁶⁴ Die begrifflichen Schwierigkeiten zeigen sich auch in Definitionen oder Definitionsversuchen der deutschen und europäischen Normgeber oder Exekutivorgane. Die Europäische Kommission versteht unter der Künstlichen Intelligenz Systeme, die auf verschiedenen Analysen basierend

62 „Der Begriff der künstlichen Intelligenz ist dabei ebenso schillernd wie konturenlos“, Specht, GRUR 2019, S. 253, 253. Eine Sammlung an 70 Definitionen der Intelligenz bieten Legg/Hutter in: Goertzel/Wang (Hrsg.), *Advances in artificial general intelligence*, S. 17 ff., 2.

63 Hauck/Cevc, ZGE 2019, S. 135, 140 f. präzisieren den Begriff der schwachen KI hin zu einer spezifischen bzw. einer anwendungsbereichsbezogenen KI; zum Kriterium der Autonomie ausführlich Kirm/Müller-Hengstenberg, MMR 2014, S. 225, 225; Pieper, InTeR 2018, S. 9 ff., 11; Aust, *Das Zeitalter der Daten*, S. 26 f.; Grätz, *Künstliche Intelligenz im Urheberrecht*, S. 4 Fn. 31, 13 f.; Kaplan, *Künstliche Intelligenz*, S. 83; Käde, *Kreative Maschinen und Urheberrecht*, S. 41; Mainzer, *Künstliches Leben und virtuelle Agenten* (16.01.1997); in Bezug auf denkende Maschinen Strube/Ferstl/Konieczny/Ragni in: Görz/Schneeberger/Schmid (Hrsg.), *Handbuch der Künstlichen Intelligenz*, S. 21, 44. Auf letzteren Standpunkt stellt sich die deutsche Bundesregierung, Bundesregierung, *Strategie Künstliche Intelligenz der Bundesregierung* (2018), S. 4 f.

64 Hauck/Cevc, ZGE 2019, S. 135, 140 f.; Reichwald/Pfisterer, CR 2016, S. 208, 210 f.; Niederée/Neidl in Ebers/Heinze/Krügel, *Künstliche Intelligenz und Robotik* § 2 Rn. 2; Bauberger/B. Beck/Burchardt/Remmers in: Görz/Schmid/Braun (Hrsg.), *Handbuch der Künstlichen Intelligenz*, S. 907, 507; Lenzen, *Künstliche Intelligenz: Fakten, Chancen, Risiken*, S. 199 f.; Strube/Ferstl/Konieczny/Ragni in: Görz/Schneeberger/Schmid (Hrsg.), *Handbuch der Künstlichen Intelligenz*, S. 21, 44 betonen, dass bisher keine autonom denkenden Maschinen entwickelt werden konnten; ebenso Volland, *Die kreative Macht der Maschinen*, S. 14. Optimistisch äußert sich Pieper, InTeR 2018, S. 9 ff., 14: „Es ist anzunehmen, dass es in Zukunft autonome Systeme geben wird, die mit „Starker KI“ ausgestattet sind“, zweifelnder ist Grätz, *Künstliche Intelligenz im Urheberrecht*, S. 14: „Ungewiss ist, ob dies jemals gelingen wird“. Ein Forschungszweig der KI richtet sich nach der Suche dieser „künstlichen allgemeinen Intelligenz“, Lenzen, *Künstliche Intelligenz*, S. 32 f., 120: „Zu prognostizieren, wie lange es noch dauern wird, eine solche künstliche allgemeine Intelligenz zu realisieren, ist zu einem beliebten Gesellschaftsspiel geworden. Die Antworten reichen von »unmöglich« bis »unvermeidlich«, der Zeithorizont wird mal mit 20, mal mit 200 Jahren angegeben.“

(teil-)autonome, intelligente Verhaltensweisen zeigen.⁶⁵ Angeknüpft wird also an bestimmte Eigenschaften. Der KI-Verordnungsentwurf zur „*Festlegung harmonisierter Vorschriften für Künstliche Intelligenz*“ definiert ein System „*der Künstlichen Intelligenz*“ in ihrem Art. 3 Nr. 1 als „*Software, die mit einer oder mehreren der im Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren*“. Der Anhang dieses Verordnungsentwurfs benennt das maschinelle Lernen inklusive des Deep Learnings sowie logik- und wissensgestützte Systeme und statistische Ansätze. Auch die deutsche Bundesregierung wählt in ihrer KI-Strategie ein beispielbasiertes Verständnis und versteht jedenfalls unter schwacher KI – „*Deduktionssysteme [...], wissensbasierte Systeme [...], Musteranalysen und Mustererkennung [...], Robotik*“ und „*intelligente multimodale Mensch-Maschine-Interaktionen*“.⁶⁶ Mittels dieser technologiespezifischen Aufzählungen wird es vermieden, definitorisch zu bestimmen, was die KI eigentlich ausmacht.

Die Schwierigkeiten, die KI abstrakt zu definieren, sind mitunter darauf zurückzuführen, dass es kein allgemeingültiges Verständnis von „*Natürlicher Intelligenz*“ oder von „*Intelligenz*“ als solcher gibt.⁶⁷ Die Antwort, die der Duden auf die Frage nach dem Begriff der „*Intelligenz*“ liefert, ist für diese Kernproblematik bezeichnend: Er versteht darunter „*die Fähigkeit [des Menschen], abstrakt und vernünftig zu denken und daraus zweckvolles Handeln abzuleiten*“.⁶⁸ Das wirft die Frage auf, ob Intelligenz zwangsläufig mit dem Menschsein zusammenhängt. Das war auch das ursprüngliche Verständnis von Künstlicher Intelligenz, es galt also als intelligent, was sich zumindest als Mensch ausgeben konnte.⁶⁹ Jedenfalls die Lernfähigkeit

65 Europäische Kommission, Künstliche Intelligenz für Europa - COM(2018) 237 final, S. 1. Eine gute Übersicht zu den verschiedenen europäischen Bestrebungen für den Datenmarkt und die Bereiche der KI bietet *Jüngling*, MMR 2020, S. 440 ff.

66 Bundesregierung, Strategie Künstliche Intelligenz der Bundesregierung (November 2018), S. 4 f.

67 *Kaplan*, Künstliche Intelligenz, S. 15.

68 Duden, Definition der Intelligenz, abrufbar unter: <https://www.duden.de/rechtshilfe/Intelligenz>.

69 *McCarthy/Minsky/Rochester/Shannon*, AI Mag 2006, S. 12, 12; *Turing* in: *Carpenter/Doran*, A. M. Turing's ACE report of 1946 and other papers, S. 106 (123); *Pallay* in: *Mainzer* (Hrsg.), Philosophisches Handbuch Künstliche Intelligenz, S. 1, 9; *Turing* in: *Carpenter/Doran* (Hrsg.), A. M. Turing's ACE report of 1946 and other papers, S. 106 ff., 123; *Kaplan*, Künstliche Intelligenz, S. 15. In diesem Kontext sei auch der nach *Alan Turing* benannte *Turing Test* erwähnt, der in sei-