



Ausarbeitung

Künstliche Intelligenz und Machine Learning
Eine urheberrechtliche Betrachtung

Künstliche Intelligenz und Machine Learning Eine urheberrechtliche Betrachtung

Aktenzeichen: WD 10 - 3000 - 67/18
Abschluss der Arbeit: 23.10.2018
Fachbereich: WD 10: Kultur, Medien und Sport

Die Wissenschaftlichen Dienste des Deutschen Bundestages unterstützen die Mitglieder des Deutschen Bundestages bei ihrer mandatsbezogenen Tätigkeit. Ihre Arbeiten geben nicht die Auffassung des Deutschen Bundestages, eines seiner Organe oder der Bundestagsverwaltung wieder. Vielmehr liegen sie in der fachlichen Verantwortung der Verfasserinnen und Verfasser sowie der Fachbereichsleitung. Arbeiten der Wissenschaftlichen Dienste geben nur den zum Zeitpunkt der Erstellung des Textes aktuellen Stand wieder und stellen eine individuelle Auftragsarbeit für einen Abgeordneten des Bundestages dar. Die Arbeiten können der Geheimschutzordnung des Bundestages unterliegende, geschützte oder andere nicht zur Veröffentlichung geeignete Informationen enthalten. Eine beabsichtigte Weitergabe oder Veröffentlichung ist vorab dem jeweiligen Fachbereich anzuzeigen und nur mit Angabe der Quelle zulässig. Der Fachbereich berät über die dabei zu berücksichtigenden Fragen.

Inhaltsverzeichnis

1.	Einleitung	4
1.1.	Begriffsbestimmungen	5
1.1.1.	„Künstliche Intelligenz“	5
1.1.2.	„Machine Learning“	5
1.2.	Reichweite der Untersuchung	6
2.	Urheberrechtlich relevante Nutzung beim Training mit urheberrechtlich geschützten Werken	6
2.1.	Zulässigkeit der Nutzungen	8
2.2.	Gemeinfreiheit	8
2.3.	Lizenzen	9
2.3.1.	Kommerzielle Lizenzen	9
2.3.1.1.	Einzellizenzen	9
2.3.1.2.	Einkauf fertiger Korpora	9
2.3.2.	Freie Lizenzen, Creative Commons (CC)	10
2.3.3.	Kostenlose Einräumung von Lizenzen durch AGB, User Generated Content	11
2.4.	Zwischenergebnis	12
2.5.	Schrankenbestimmungen	12
2.5.1.	Text-and-Data-Mining-Schranke	13
2.5.2.	Kritik	14
2.5.2.1.	Beschränkung auf nichtkommerzielle Forschung	15
2.5.2.2.	Vergütungspflicht	16
2.5.2.3.	Löschpflicht	16
3.	Urheberrechtlicher Schutz von Algorithmen und Computerprogrammen	17
4.	Urheberrechtlicher Schutz von durch KI erzeugten Ergebnissen	18
4.1.	Urheberrechte	18
4.2.	Verwandte Schutzrechte	20
4.2.1.	Lichtbildschutz	20
4.2.2.	Datenbankschutz	20
4.2.3.	Einführung eines Immaterialgüterrechts für KI-produzierte Schöpfungen?	22
5.	Fazit und Ausblick	22

1. Einleitung

Obwohl der Begriff der künstlichen Intelligenz (im Folgenden „KI“) bzw. Artificial Intelligence (AI) schon im Jahr 1956 das erste Mal in der Wissenschaft eingeführt wurde¹ und in den vergangenen Jahrzehnten schon öfter durch relativ „primitive“ Anwendungen abseits der großen Öffentlichkeit vor allem bei Technik-Nerds von sich reden machte, hat sich nach zwei sogenannten „KI-Wintern“ in den siebziger und neunziger Jahren erst in den letzten Jahren eine enorme Entwicklung abgezeichnet, so dass Forschung, Wirtschaft, Gesellschaft und auch die Politik mit Fragen der KI verstärkt konfrontiert werden. Die grundlegenden informatischen Prinzipien von z.B. neuronalen Netzen, die die Grundlage von KI-Anwendungen bilden, sind dabei nicht neu. Die gesteigerte Rechenleistung von Computern in Kombination mit „Big Data“, also der schier unendlichen Verfügbarkeit von digitalen Daten, und das Interesse der großen Internet- und Technologieunternehmen wie Google, Microsoft, IBM, SAP, Yandex, Alibaba oder Baidu, die das wirtschaftliche Potenzial von KI erkannt haben und Milliarden in diese Technologie investieren, haben die Entwicklung zuletzt aber regelrecht befeuert und machen erstmals auch praktische Anwendungen möglich. Der Google Vorstandsvorsitzende Sundar Pichai erklärte im Januar 2018 in einem Interview gegenüber dem amerikanischen Nachrichtensender MSNBC, dass KI wohl die wichtigste Entwicklung der Gegenwart sei und für die Menschheit tiefgreifender sein könne als die Entdeckung der Elektrizität oder des Feuers, wobei er gleichzeitig auch vor Risiken warnte.²

Dabei ist der potenzielle Anwendungsbereich von KI nicht auf „Spielereien“ wie Computer, die beim Schach Großmeister übertrumpfen und Smartphone-Apps, die Gesichter auf Fotos erkennen und modifizieren, beschränkt, sondern erstreckt sich auf nahezu alle Bereiche des Lebens. Schon heute findet künstliche Intelligenz bei der Entwicklung von autonomen Fahrzeugen, in der Robotik, in der medizinischen Forschung und Diagnostik Anwendung, wird im Online-Marketing eingesetzt, kontrolliert Aktienfonds oder wickelt in Form von Chatbots automatisiert Kundendienstanfragen in Unternehmen ab. Grundlage dafür ist die Fähigkeit von KI, große Datenmengen automatisiert zu analysieren, darin Strukturen zu erkennen, diese zu verstehen und zur Problemlösung anzuwenden.³

Neben erheblichen gesellschaftlichen, wirtschaftlichen, ethischen und technischen Fragen stellt die KI aber auch das Recht vor neue Herausforderungen, von denen in dieser Ausarbeitung aber nur die urheberrechtlichen eine Rolle spielen sollen.

1 Überschrift eines Projektantrags des Informatikers *John McCarthy* auf einer Konferenz der Universität Dartmouth im Jahr 1956.

2 *Clifford* „Google CEO: A.I. is more important than fire or electricity“ veröffentlicht am 1. Februar 2018 auf [cnbc.com](https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html), abrufbar unter <https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html> (letzter Zugriff 23.10.2018).

3 Gründerszene Lexikon, abrufbar unter <https://www.gruenderszene.de/lexikon/begriffe/kuenstliche-intelligenz> (letzter Zugriff 23.10.2018).

1.1. Begriffsbestimmungen

Künstliche Intelligenz, maschinelles Lernen, Algorithmen, Deep Learning, neuronale Netze, Big Data – in diesem Bereich der Informatik gibt es viele technische Fachbegriffe und Buzzwords, die sich nur schwer definieren lassen.⁴ Dass es verschiedenste Technologien, Konzepte und Anwendungsformen gibt und fortwährend neue hinzukommen, erschwert das Verständnis noch - gerade wenn man über kein detailliertes technisch-informatisches Fachwissen verfügt. Deshalb erscheint es sachgerecht, zunächst die grundlegendsten Prinzipien solcher intelligenten Systeme zu betrachten, ohne den Versuch einer allgemeingültigen Definition zu wagen.

1.1.1. „Künstliche Intelligenz“

KI kann zunächst als Überbegriff für Anwendungen verstanden werden, bei denen Maschinen menschenähnliche Fähigkeiten wie Lernen, Urteilen oder Problemlösen erlangen. Beim maschinellen Lernen können Computer durch eigene Erfahrungen und/oder aus großen Datenmengen lernen, Aufgaben immer besser auszuführen, sich also selbst effizienter zu machen.⁵ Der Programmcode wird vom Entwickler mit eigenen Auswahl- und Entscheidungsermessen ausgestattet und kann die Entscheidung über In- und Output selbst treffen. Das geschieht innerhalb einer menschlich bestimmten Rahmenprogrammierung.⁶

Man muss sich im Klaren sein, dass es in der aktuellen Diskussion vor allem um sogenannte „schwache“ KIs oder Expertensysteme geht, die sehr spezifische eng definierte Aufgaben erledigen und in diesem Bereich enorme Leistungen vollbringen können, die über die menschlichen hinausgehen können. Von einer „starken“ KI, also einer universellen künstlichen Intelligenz, die auch in ihrer Generalität mit dem menschlichen Geist vergleichbar wäre, sei man (sollte dies überhaupt jemals möglich sein) laut Experten aber noch Jahrzehnte entfernt.⁷

1.1.2. „Machine Learning“

Maschinelles Lernen geschieht in der Regel anhand von Datensätzen, die entweder schon vorgegebene Outputs enthalten oder in denen ein Algorithmus selbst Muster erkennen muss. Oft werden dabei mehrere Systeme kombiniert, wobei das eine Ergebnisse liefert, die dann vom zweiten System bewertet werden und diese Bewertung wiederum an das erste System zurückgespielt wird, so dass es seinen Prozess verbessern und „lernen“ kann. Mit mehreren solchen Schleifen entstehen dann immer bessere Ergebnisse. Die benötigten Daten können strukturiert in Tabellen

4 Eine lesenswerte Begriffskritik und -einordnung von *Herberger*, der schon auf die sprachliche Unschärfe hinweist, die sich aus den unterschiedlichen Bedeutungen der Wörter „artificial“ und „intelligence“ im Englischen ergeben: *Herberger* „Künstliche Intelligenz und Recht“ in NJW 2018, 2845.

5 *Schick* „Was ist künstliche Intelligenz?“, erschienen am 18. September 2018 bei SAP News, abrufbar unter <https://news.sap.com/germany/2018/03/was-ist-kuenstliche-intelligenz/> (letzter Zugriff am 23.10.2018).

6 *Denga* „Deliktische Haftung für künstliche Intelligenz“ in CR 2018, 69.

7 So *Dr. Aljoscha Burchard* in seinem Vortrag vom 15.10.2018 in der Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“ des Deutschen Bundestages.

vorliegen oder unstrukturiert, wobei alle Formen von Daten wie Bilder, Texte, Sprache aus unterschiedlichsten Quellen verwendet werden können, solange sie irgendwie maschinenlesbar sind oder gemacht werden können (Datenbanken, Forschungsergebnisse, Internetinhalte, digitalisierte Bücher).⁸ Von der Qualität der Daten hängt in hohem Maße die Qualität der Ergebnisse ab. Dies ist besonders wichtig, da aufgrund der Komplexität der Prozesse Fehler nur schwer nachvollzogen oder erkannt werden können und etwa um einen „algorithmic bias“, also antrainierte „Vorurteile“ zu vermeiden.

Es ist auch möglich, eine KI ohne vorhandene Datensätze zu trainieren. So wurden z.B. im Projekt „Alpha Go Zero“ von Google zwei KIs nur mit den grundlegenden Regeln des Brettspiels „Go“, das als besonders komplex gilt, programmiert und spielten dann ohne weiteren menschlichen Einfluss millionenfach gegeneinander. Sie trainierten sich auf diese Weise selbst. Nach nur 21 Tagen hatten die Systeme das Niveau der besten menschlichen Spieler erreicht und konnten die Großmeister Lee Sedol und Ke Jie besiegen und hatten dabei auch völlig neuartige revolutionäre Spielzüge angewandt, die die Profis in Erstaunen versetzten.⁹

1.2. Reichweite der Untersuchung

Für das Urheberrecht sind vor allem die Datensätze, anhand derer die KIs trainiert werden, die sogenannten Korpora (siehe unten), die Ergebnisse der KI aber auch die konkrete programmatische Ausgestaltung der KI selbst relevant, so dass diese im Rahmen dieser Ausarbeitung näher betrachtet werden. Ein besonderes Augenmerk gilt dabei der neuen Text and Data Mining Schranke. Systeme, die wie das letztgenannte ohne Ursprungsdaten auskommen, sind urheberrechtlich bis auf den Schutz des Programmcodes selbst nicht relevant.

2. Urheberrechtlich relevante Nutzung beim Training mit urheberrechtlich geschützten Werken

Eine KI-Software muss in der Regel mit Daten trainiert werden. Je mehr (korrekte) Daten für das Training zur Verfügung stehen, desto besser werden auch die Ergebnisse der KI. Diese Sammlung von Daten, woraus die Software Informationen ableitet, wird „Korpus“ oder englisch „Dataset“ genannt. Für die KI-Entwicklung sind diese Korpora sehr kostbar und ihre Erstellung teils sehr ressourcenintensiv. Da, wie oben bereits erörtert, nahezu alle Arten von Daten für das Training einer KI technisch in Frage kommen, können auch nahezu alle von Urheber- oder verwandten Schutzrechten geschützte Formen von Daten berührt sein, also insbesondere aber nicht abschließend Sprachwerke, Musik, Lichtbildwerke, Filmwerke, Datenbankwerke wissenschaftliche oder technische Darstellungen i.S.v. § 2 Urheberrechtsgesetz (UrhG)¹⁰ oder verwandte Schutzrechte von Datenbanken sui generis, Lichtbildern, Laufbildern usw. i.S.v. § 70 ff UrhG. Während der Urheberrechtsschutz von Werken gem. § 2 UrhG aus einer persönlichen geistigen Schöpfung des

8 *Schick*, a.a.O.

9 *Hassabis/Silver* „AlphaGo Zero: Learning from scratch“, abrufbar unter <https://deepmind.com/blog/alphago-zero-learning-scratch/> (letzter Zugriff 23.10.2018).

10 Urheberrechtsgesetz vom 9. September 1965 (BGBl. I S. 1273), das zuletzt durch Artikel 1 des Gesetzes vom 1. September 2017 (BGBl. I S. 3346) geändert worden ist.

Urhebers resultiert, werden durch die verwandten Leistungsschutzrechte, die keine derartige geistige Leistung voraussetzen, vor allem technische Leistungen und wirtschaftliche Investitionen geschützt.¹¹ Zu beachten ist, dass der abstrakte Informationsgehalt an sich grundsätzlich keinem urheberrechtlichen Schutz unterliegt. Geschützt ist nur dessen konkrete Gestaltung etwa in einem Text oder einem Lichtbild.¹²

Wenn urheberrechtlich geschützte Werke (und nicht etwa gemeinfreie oder nicht schutzfähige Daten wie wissenschaftliche Rohdaten) genutzt werden, erfolgen die urheberrechtlich relevanten Nutzungen in der Regel nicht beim Training selbst, sondern vor allem schon im Vorfeld bei der Erstellung der Korpora. So dürfte die Datenbeschaffung, die meist ein (automatisches massenhaftes) Speichern der Ursprungsdaten erfordert, eine Vervielfältigung gem. § 16 UrhG darstellen.¹³ Ebenso stellen die anschließende Normalisierung der Daten in eine für den Algorithmus lesbare Form und die darauf folgende Zusammenfassung in einem Datenbanksystem (meist XML-Dateien) weitere Vervielfältigungshandlungen und nicht etwa Bearbeitungen oder Umwandlungen i.S.d. § 23 UrhG dar, da das Werk als solches bei einer bloßen Formatänderung nicht geändert wird.¹⁴

Die anschließende automatisierte Analyse oder Auswertung der Daten selbst sowie die Kategorisierung mit Metadaten sind hingegen keine urheberrechtsrelevanten Handlungen.¹⁵ Der Aufruf und das reine (Aus-)Lesen der Informationen sind als bestimmungsgemäßer Werkzeuggebrauch grundsätzlich nicht vom Urheberrecht erfasst.¹⁶

Damit ist, was die Vervielfältigung betrifft, der Anwendungsbereich für die Regelungen des Urheberrechtsgesetzes eröffnet.¹⁷

11 *Fechner* „Medienrecht“, 19. Aufl., 5. Kapitel, Rn. 110.; *Dreier* in *Dreier/Schulze* „Urheberrechtsgesetz“, 6. Aufl., Vorb. zu §§ 70 ff., Rn. 1-2.

12 *Hubmann/Rehbinder/Peukert* „Urheberrecht und verwandte Schutzrechte“, 18. Aufl., Rn. 192.

13 *Ebenda*, Rn. 542.

14 *Kroitzsch/Götting* in „BeckOK Urheberrecht“, 21. Edition, § 16, Rn. 13; *Spindler* „Text und Data Mining – urheber- und datenschutzrechtliche Fragen“, GRUR 2016, 1112 f.

15 Siehe Erwägungen zum „Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (Urheberrechts-Wissensgesellschafts-Gesetz - UrhWissG)“, Drucksache 312/17, Seite 40, abrufbar unter <http://dipbt.bundestag.de/dip21/brd/2017/0312-17.pdf> (letzter Zugriff 23.10.2018).

16 *Ungern-Sternberg* in: *Schricker/Loewenheim*, Urheberrecht, 5. Aufl. 2017, § 15 Rn. 188.

17 Es sind ausweislich der Erwägungen zum Gesetzentwurf auch technische Lösungen des „Text and Data Mining“ denkbar oder gar existent, die ohne eine Herstellung eines Korpus auskommen. Ob bei diesen Prozessen dennoch urheberrechtsrelevante Handlungen erfolgen, kann hier ohne genaue Kenntnis der technischen Abläufe solcher Systeme nicht beantwortet werden.

2.1. Zulässigkeit der Nutzungen

Grundsätzlich steht das Recht zur Nutzung und Verwertung von geschützten Werken nur dem Urheber oder Inhaber des jeweiligen Leistungsschutzrechts zu (§§ 11 ff UrhG). Verletzungen dieser Rechte können straf- und zivilrechtlich geahndet werden (§§ 97 ff UrhG). Um dies zu vermeiden, bedarf die Nutzung von geschützten Materialien einer Erlaubnis. Dazu können die Urheber oder Inhaber von Leistungsschutzrechten Nutzungs- und Verwertungsrechte durch Rechtsgeschäft Dritten übertragen oder einräumen.¹⁸ Wer bei einem konkreten Werk Verfügungsberechtigt ist, muss im Einzelfall ermittelt werden. So kann dies der Urheber/Leistungsschutzinhaber selbst sein, eine Verwertungsgesellschaft (z.B. GEMA, VG Wort, VG Bild-Kunst) oder ein anderer Dritter, dem derivativ ein Nutzungs- und Verwertungsrecht eingeräumt wurde. Einzelrechte können je nach vertraglicher Ausgestaltung jeweils durch (Unter-)Lizenzen weitergeben werden,¹⁹ wohingegen der persönlichkeitsrechtliche Kern des Urheberrechts vom Urheber (oder seinen Rechtsnachfolgern) prinzipiell nicht übertragen oder aufgegeben werden kann (§ 29 UrhG).²⁰

2.2. Gemeinfreiheit

Das Urheberrecht an Werken erlischt gem. § 64 UrhG in der Regel 70 Jahre nach dem Tod des Urhebers, die verwandten Schutzrechte haben unterschiedliche Geltungsdauern und Fristen (z.B. Datenbanken spätestens 15 Jahre nach Erscheinen gem. § 87d UrhG, Lichtbilder 50 Jahre nach Erscheinen gem. § 72 UrhG). Danach gelten diese Werke als gemeinfrei und können frei verwendet werden, also auch zu einem Korpus zusammengeführt und KI damit trainiert werden.

Es gibt auch Portale, die solche (teils historischen) gemeinfreien Inhalte sammeln und sie in normalisierter Form als XML-Dateien vor allem für die algorithmenbasierte Sprachforschung anbieten.²¹ Hierbei ist jedoch zu beachten, dass die Anbieter solcher Datenbanken wiederum mindestens ein Leistungsschutzrecht sui generis gem. §§ 87a ff. UrhG an der Datensammlung erwerben.

Von Beginn an freie Werke, wie es sie z.B. im angloamerikanischen Rechtsraum durch den völligen Rechtsverzicht des Urhebers gibt, sind im deutschen Recht aus o.g. Gründen (§ 29 UrhG) nicht möglich.²²

Erwähnt werden müssen noch die sogenannten verwaisten Werke gem. § 61 UrhG. Das sind Werke aus Sammlungen von öffentlichen Bibliotheken, Archiven usw., bei denen der Urheber oder Nutzungsrechtsinhaber auch durch eine sorgfältige Suche nicht (mehr) festgestellt werden

18 *Fechner*, a.a.O., 5. Kapitel, Rn. 35.

19 *Schulze in Dreier/Schulze*, a.a.O. § 31, Rn. 11.

20 *Spautz/Götting* in BeckOK Urheberrecht, 21. Edition, § 29 UrhG, Rn. 5.

21 Beispiele: Historisch: <http://www.deutschestextarchiv.de/>; Open Data: <https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/>; <https://skymind.ai/wiki/open-datasets>.

22 *Kroitzsch/Götting* in BeckOK Urheberrecht, UrhG § 11 Rn. 2.

kann. Diese sind mitnichten gemeinfrei und Nutzungen nur unter den engen Voraussetzungen des § 61 UrhG zulässig.²³

2.3. Lizenzen

Bei der Verwendung urheber- oder leistungsschutzrechtlich geschützter Werke oder Titel ist grundsätzlich eine Lizenz des Urhebers, Herstellers oder des Verwertungsrechtsinhabers nötig.²⁴ Es gibt jedoch verschiedene Formen von Lizenzen, die für die Erstellung von Korpora eine praxisrelevante Rolle spielen und die im Folgenden dargestellt werden sollen.

2.3.1. Kommerzielle Lizenzen

Durch Vertrag kann der Urheber oder Inhaber eines verwandten Schutzrechts anderen die Verwertungsrechte oder Nutzungsrechte einräumen (§ 31 UrhG). Inhaber der Verwertungsrechte können wiederum Unterlizenzen vergeben usw.²⁵ In aller Regel werden die Rechteinhaber dies nicht unentgeltlich tun sondern eine Gegenleistung verlangen. Dies ist im Gesetz als Regelfall vorgesehen: Gem. § 32 UrhG hat der Urheber für die Einräumung von Nutzungsrechten einen Anspruch auf eine angemessene Vergütung.

2.3.1.1. Einzellizenzen

Möchte ein Entwickler nun also einen Korpus mit geschützten Quellen erstellen, könnte er von den jeweiligen Rechteinhabern Lizenzen für die Nutzung jedes einzelnen Werks erwerben. Dies dürfte sich aber angesichts der für das KI-Training benötigten enormen Menge an Daten schwierig gestalten: Es müssten zunächst die einzelnen Rechteinhaber identifiziert und kontaktiert werden. Dann müsste mit jedem verhandelt und ein Lizenzvertrag geschlossen werden. Manchmal ist dies direkt der Urheber oder Hersteller, manchmal ist dies eine Verwertungsgesellschaft oder ein Verlag, vielleicht kann ein Urheber oder Rechteinhaber überhaupt nicht identifiziert werden. Es wird schnell deutlich, dass diese Methode organisatorisch wie wirtschaftlich kaum leistbar sein dürfte, vor allem wenn massenhaft auf im Web verfügbare Inhalte zugegriffen wird.

2.3.1.2. Einkauf fertiger Korpora

Es gibt eine Reihe von kommerziellen Anbietern, die fertige Korpora zur Lizenzierung anbieten.²⁶ Diese können urheberrechtlich geschützte Werke enthalten, deren Lizenzierung schon der Korpusanbieter erledigt hat. Aber auch Korpora, die keine geschützten Werke enthalten, genießen als Datenbankwerke oder Datenbanken sui generis den Schutz des UrhG. Der Aufwand, sich selbst

23 *Fechner*, a.a.O., 5. Kapitel, Rn. 108.

24 *Ebenda* Rn. 35 f.

25 *Schulze in Dreier/Schulze*, a.a.O. § 35 Rn. 8.

26 Beispiel: <https://www.clickworker.de/maschinelles-lernen-ki-kuenstliche-intelligenz/>.

mit jedem Rechteinhaber auseinandersetzen zu müssen, entfällt so für die Entwickler, sie haben für den fertigen Korpus nur einen Vertragspartner und müssen nur an diesen eine Gebühr zahlen.

2.3.2. Freie Lizenzen, Creative Commons (CC)

Zwar können Urheber oder Inhaber verwandter Schutzrechte auf diese nicht verzichten, sie können aber die freie Nutzung erlauben. Rechtlich stellt dies ein Angebot ad incertas personas zum Abschluss eines unentgeltlichen Lizenzvertrags dar, der durch die Nutzung des Werks konkludent angenommen wird.²⁷ Oft werden solche unentgeltlichen Nutzungseinräumungen in der Reichweite eingeschränkt oder an Bedingungen geknüpft. So kann sich die Lizenz nur auf bestimmte Nutzungsformen beschränken wie die Vervielfältigung, die Bearbeitung oder die öffentliche Zurschaustellung. Teils ist die Verwendung nur unter namentlicher Nennung des Urhebers oder Verwendung nur für nichtkommerzielle Zwecke erlaubt.

Um sowohl Urhebern, die ihre Werke frei anbieten möchten, aber auch möglichen Verwendern die einfache und rechtssichere Nutzung zu ermöglichen, wurde das Creative Commons System entwickelt.²⁸ Dieses enthält sechs fest definierte Lizenzvereinbarungen, die als allgemeine Geschäftsbedingungen in den Lizenzvertrag einfließen²⁹ und die ein Urheber frei benutzen kann, um sein Werk unkompliziert der Allgemeinheit zur Verfügung zu stellen. Die vier einzelnen Bausteine, die den sechs Lizenztypen zugrunde liegen, sind BY (Namensnennung des Urhebers), SA (Share Alike = Weiterverbreitung nur unter denselben Lizenzbedingungen), ND (Keine Bearbeitungen erlaubt) und NC (nur nichtkommerzielle Nutzung).

Die CC-Lizenzen sind dabei weltweit, nicht-exklusiv, dauerhaft und unwiderruflich und berechtigen den Lizenznehmer, geschützte Werke aller Medien und Formate zu vervielfältigen, anzuzei-gen, aufzuführen, öffentlich wiederzugeben, zu bearbeiten und zu verbreiten. Der Nutzer verpflichtet sich, bei jedem Werkstück, das er verbreitet oder wiedergibt, eine Kopie der oder einen Link zu den Lizenzbedingungen beizufügen. Außerdem muss immer der Autor genannt werden.³⁰ Bei einem Verstoß gegen die in der Lizenz enthaltenen Nutzungsbedingungen kommt es zum Erlöschen der Lizenz und zum Wiederaufleben der gesetzlichen urheberrechtlichen Regelungen.³¹

27 Wagner „Aktuelle Möglichkeiten und rechtliche Probleme der Creative-Commons Lizenzmodelle“ in MMR 2017, 216 (219).

28 Einzelheiten zu Creative Commons unter <https://de.creativecommons.org/index.php/was-ist-cc/> (letzter Zugriff 23.10.2018), sieht auch entsprechender „Aktueller Begriff“ der WD Nr. 98/09, abrufbar unter https://www.bundestag.de/blob/190910/6c1568b0478adf56e0b3b8c9e54e728c/creative_commons-data.pdf (letzter Zugriff 23.10.2018).

29 OLG Köln, Urteil vom 31.10.2014, Az. 6 U 60/14, veröffentlicht in MMR 2015, 331.

30 Wiebe in *Spindler/Schuster*, Recht der elektronischen Medien, 3. Auflage 2015, § 31 UrhG, Rn. 21.

31 OLG Köln, siehe Fn. 29.

Die Lizenzbedingungen sind auf Englisch verfasst und an verschiedenste Rechtsordnungen angepasst, so dass auch eine internationale Verwendung möglich ist. Erläuterungen sind in verschiedenen Sprachen verfügbar. Das OLG Köln hat dazu entschieden, dass die Verwendung der englischen Fassung des AGB-Textes einer Wirksamkeit nicht entgegensteht.³²

Ein weiterer Vorteil der CC-Lizenzen ist, dass der Hinweis auf die erlaubten Nutzungsformen als maschinenlesbare Metadaten hinter dem CC-Icon liegt, so dass z.B. mittels einer Suchmaschine schnell frei nutzbare und passende CC-lizenzierte Inhalte gefunden werden können.

In der neusten Version der CC-Lizenzen 4.0 wurde erstmals das Text und Data Mining, also die automatisierte Erstellung eines Korpus³³, explizit in die Regelungen aufgenommen.³⁴ So ist die Verwendung von CC-lizenziertem Material für Text-and-Data Mining grundsätzlich erlaubt. Einzige Einschränkung ist die Verwendung für kommerzielle Zwecke, wenn diese explizit aus der Lizenz ausgenommen sind (Attribut NC).

Durch die weite Verbreitung von CC-Inhalten im Internet (z.B. die komplette Wikipedia) und Möglichkeit der gezielten Auffindbarkeit durch Suchmaschinen ist die Verwendung zur Erstellung von Korpora für KI-Entwickler sehr attraktiv. Auch kommerzielle Verwender können durch Ausschluss des Attributs „NC“ einfach Daten auffinden, die sie legal verwenden können.

2.3.3. Kostenlose Einräumung von Lizenzen durch AGB, User Generated Content

Die führenden Entwickler von KI-Systemen sind die großen Internetkonzerne wie Facebook, Google oder Alibaba und das nicht ohne Grund: Zusätzlich zu den enormen wirtschaftlichen Ressourcen, die es ihnen erlauben die besten Entwickler und Forscher einzustellen und die leistungsfähigsten Rechenzentren zu bauen, können sie auf nahezu unendliche Datenmengen zum Training ihrer KI-Entwicklungen zurückgreifen. Natürlich hätten diese Konzerne die Mittel, um im großen Stil Daten und geschütztes Material zu erwerben bzw. zu lizenzieren, das müssen sie aber nicht: Ihre Milliarden Nutzer überlassen Ihnen das Material im Gegenzug zur Nutzung der Dienste kostenlos.

Zum einen erheben die Konzerne urheberrechtlich nicht geschützte Daten wie Bewegungs- oder Verbindungsdaten, Nutzungsstatistiken und Benutzerprofile, was vor allem datenschutzrechtliche Relevanz hat.

Die Nutzer überlassen diesen Diensten aber auch anderes, oft urheberrechtlich eigentlich geschütztes eigenes Material mit Zustimmung zu deren Nutzungsbedingungen: Fotos, die bei Facebook oder Instagram geteilt oder in der Cloud gespeichert werden. Texte und Kommentare, die gepostet oder in einen Übersetzungsservice geladen werden, Videos die bei Youtube hochgeladen werden. All diese Daten stehen den Konzernen zur Verfügung und sie können diese zur Erstellung von Korpora und dem Training von KI nutzen.

32 Ebenda.

33 Siehe zur Definition unten Punkt 2.5.1.

34 Einzelheiten abrufbar unter: https://wiki.creativecommons.org/wiki/Content_mining#Table_1: Summary_of_CC_License_Permissions_for_Text_and_Data_Mining (letzter Zugriff 23.10.2018).

So enthalten zum Beispiel die Nutzungsbedingungen von Google folgende Regelungen, die sich so oder so ähnlich auch bei den meisten anderen Online-Diensten finden:

*„Wenn Sie Inhalte in oder über unsere Dienste hochladen [...] räumen Sie Google [...] das Recht ein, diese Inhalte weltweit zu **verwenden**, zu **hosten**, zu **speichern**, zu **vervielfältigen**, zu **verändern**, **abgeleitete Werke daraus zu erstellen** [...] Diese von Ihnen im Rahmen dieser **Lizenz** gewährten Rechte dienen [u.A.] zur **Entwicklung neuer Dienste**. [...] Achten Sie darauf, dass Sie über die notwendigen Rechte verfügen, um uns eine entsprechende Lizenz für alle Inhalte zu erteilen, die Sie in unsere Dienste hochladen.“³⁵*

2.4. Zwischenergebnis

Es zeigt sich, dass der Lizenzerwerb mit Ausnahme der Einräumung von Nutzungsrechten per AGB für die Erstellung von Korpora mit geschütztem Material hohe Transaktionskosten und Aufwände verursacht, was vor allem kleinere Unternehmen wie Start-Ups, den Mittelstand aber auch die oft nur mit beschränkten finanziellen Mitteln ausgestattete wissenschaftliche Forschung vor Probleme stellen kann. Nicht für jeden Anwendungsbereich ist ein Training mit frei verfügbarem Material geeignet.

2.5. Schrankenbestimmungen

Das Urheberrecht wird nicht schrankenlos gewährleistet. Es gibt eine Vielzahl von Vorschriften, die eine freie Benutzung von geschützten Werken ohne Zustimmung der Rechteinhaber ermöglichen, auch wenn sie grundsätzlich dem Urheberrecht unterfallen.³⁶ Für das Training von KI kommen jedoch die meisten der Urheberrechtsschranken nicht in Betracht.

Von keiner Relevanz dürfte die Schranke aus § 44a UrhG sein, die eine technisch notwendige flüchtige Vervielfältigung erlaubt, wenn dieser keine eigenständige wirtschaftliche Bedeutung zukommt. Von dieser Regelung sind reine Übertragungs- und Caching-Vorgänge erfasst, die z.B. computerintern im Arbeitsspeicher anfallen oder bei Telekommunikationsunternehmen, die Daten an den Endkunden weiterleiten. Die Nutzung von geschützten Inhalten für das KI-Training ist davon auch bei einer „flüchtigen“ technischen Lösung wie screen scraping direkt aus einem Stream nicht erfasst, da der Analyse und Auswertung der Inhalte in der Regel eine eigenständige wirtschaftliche Bedeutung zukommen dürfte.³⁷

Relevant könnte aber z.B. die Ausnahme von amtlichen Werken wie Drucksachen des Deutschen Bundestages oder Gerichtsurteilen aus dem Schutz gem. § 5 UrhG sein, die frei genutzt werden können, so z.B. auch für den Einsatz zum Training von KI-Anwendungen.

35 Auszug aus Google Nutzungsbedingungen, Punkt „Ihre Inhalte in unseren Diensten“, abrufbar unter <https://policies.google.com/terms?hl=de&gl=ZZ#toc-protection> (letzter Zugriff 23.10.2018).

36 *Fechner*, a.a.O., 5. Kapitel, Rn. 84.

37 *Koch* „Digital Humanities & Compliance (Teil 2): Urheberrechtliche Grundlagen der automatisierten Datenveredelung durch skaleninvariante Merkmalstransformation („Cale invariant feature transform“ – SIFT)“, AnwZert ITR 5 / 2018 Anm. 2.

2.5.1. Text-and-Data-Mining-Schranke

Am relevantesten und sozusagen maßgeschneidert für datenbasierte Forschung und KI-Anwendungen ist die im März 2018 neu eingeführte Schranke für das sogenannte „Text and Data Mining“ (im Folgenden „TDM“) in § 60d UrhG. Diese Gesetzänderung wurde vom bundesdeutschen Gesetzgeber bereits mit Hinblick auf die nächste EU-Urheberrechtsreform eingeführt, für die eine ähnliche Schrankenregelung diskutiert wird.

Regelungsinhalt des § 60d UrhG ist, dass zur automatischen Auswertung für die nichtkommerzielle wissenschaftliche Forschung die automatisierte und systematische Vervielfältigung einer Vielzahl von Werken (Ursprungsmaterial) zulässig ist, um daraus insbesondere durch Normalisierung, Strukturierung und Kategorisierung einen auszuwertenden Korpus zu erstellen. Weiter darf dieser Korpus einem bestimmt abgegrenzten Kreis von Dritten für die Dauer einer wissenschaftlichen Überprüfung öffentlich zugänglich gemacht werden (sog. Peer Review). Danach muss der Korpus gelöscht werden. Gem. Absatz 3 dürfen Korpus und Ursprungsmaterial allerdings wissenschaftlichen Bibliotheken oder Archiven zur dauerhaften Aufbewahrung und ggf. erneuten wissenschaftlichen Überprüfung übermittelt werden. In Absatz 2 wird schließlich klargestellt, dass eine dem Absatz 1 entsprechende Auswertung von Datenbankwerken oder unwesentlichen Teilen Datenbanken als übliche Benutzung gilt und damit in der Regel zulässig ist.

Letztlich normiert § 60h UrhG, dass der Nutzer bei der Verwendung von urheberrechtlich geschützten Inhalten zu TDM-Zwecken eine angemessene pauschale Vergütung an die Rechteinhaber zu zahlen hat, wenn die Nutzung für TDM nicht explizit erlaubt wurde. Die Vergütung kann nur von einer Verwertungsgesellschaft geltend gemacht werden.

Im Einzelnen bedeutet dies:

Die Daten von allen Quellen, auf die (legaler) Zugriff besteht, dürfen zu TDM-Zwecken vervielfältigt werden. Dies können u.a. allgemein zugängliche Internetinhalte, kostenpflichtige (Wissenschafts-) Datenbanken, für die ein Abo besteht, Pressearchive oder Bibliotheken sein. Für dieses Prinzip steht die in diesem Zusammenhang oft zitierte Aussage „The right to read implicates the right to mine“.³⁸ Die Schranke gewährt jedoch keinen Anspruch auf Zugang zu Daten, die für den Nutzer nicht bereits zugänglich sind, sie setzt also einen bereits gesetzlich oder durch Lizenz eingeräumten Zugang voraus.³⁹

Das durch die Schranke gewährte Recht zur Vervielfältigung darf von den Inhabern nicht ausgeschlossen werden und geht insoweit anderslautenden Nutzungsbedingungen vor.⁴⁰ Die einzige Einschränkung ist, dass die Quelle wie z.B. eine Webseite oder Onlinedatenbank nicht in ihrer strukturellen Integrität belastet werden darf, also dass z.B. durch massenhaften Zugriff dermaßen die Bandbreite eines Servers ausgelastet wird, dass ein normaler Zugriff für andere Benutzer

38 Raue „Text and Data Mining – Die neue Urheberrechtsschranke des § 60d UrhG“ in CR 2017, 656.

39 Hubmann/Rehbinder/Peukert, a.a.O. Rn. 542; Koch, a.a.O.

40 Dreier in Schulze/Dreier „Urheberrechtsgesetz“, 6. Aufl. 2018, § 60d Rn. 15.

nicht mehr möglich ist.⁴¹ So könnte ein Anbieter einer Datenbank eine spezielle Schnittstelle (API) zum Zwecke des TDM anbieten, die dann benutzt werden müsste.

Eine erhebliche Einschränkung erfährt die TDM-Schranke durch die Begrenzung ihres Anwendungsbereichs auf die nichtkommerzielle wissenschaftliche Forschung. Der Begriff der wissenschaftlichen Forschung ist europarechtlich bedingt weit auszulegen und erfasst jede methodisch systematische Tätigkeit mit dem Ziel in nachprüfbarer Weise neue Erkenntnisse zu gewinnen. Dabei kommt es auf den Status der forschenden Person nicht an, so dass sowohl Privatpersonen aber auch institutionalisierte Forschungsinstitute berechtigt sind.⁴² Nichtkommerziell heißt in diesem Zusammenhang, dass der mit der Tätigkeit verfolgte Zweck keine kommerziellen, sondern nur gemeinnützige Interessen verfolgen darf. Ob die Forschung mit privaten (Dritt-)Mitteln finanziert ist, ist insoweit unerheblich. Sie darf nur nicht dem Zweck dienen, Waren oder Dienstleistungen zu entwickeln und diese zu vermarkten.⁴³

Die Begrenzung auf die nichtkommerzielle Forschung ist dabei europarechtlich durch Art. 5 Abs. 3 lit. a InfoSoc-RL⁴⁴ vorgegeben. Allerdings bestehen im Rahmen der anstehenden EU-Urheberrechtsreformen Bestrebungen, auf diese Anforderung künftig zu verzichten.

2.5.2. Kritik

Obwohl die Bestrebungen, für das TDM einen rechtssicheren Rahmen zu schaffen, grundsätzlich begrüßt werden, äußern Vertreter aus der Wirtschaft⁴⁵ aber auch aus der Wissenschaft⁴⁶, die ja durch die Regelungen bevorzugt werden soll, teils massive Kritik an der konkreten Ausgestaltung

41 Ebenda Rn. 16.

42 *Raue*, a.a.O. S. 656 f.

43 *Hubmann/Rehbinder/Peukert*, a.a.O. Rn. 540.

44 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.

45 Siehe z.B. Bitkom „Stellungnahme zu Text-and-Data-Mining und der urheberrechtlichen Schrankenregelung in Art. 3 des DSM- Urheberrecht-Richtlinienentwurfs“ vom 15.02.2018, abrufbar unter <https://www.bitkom.org/Bitkom/Publikationen/Bitkom-Stellungnahme-Text-and-Data-Mining.html> (letzter Zugriff 23.10.2018) und *Koshwitz* (Allied for Startups) „The EU just told data mining startups to take their business elsewhere“, abrufbar unter <https://www.euractiv.com/section/digital/opinion/the-eu-just-told-data-mining-startups-to-take-their-business-elsewhere/> (letzter Zugriff 23.10.2018).

46 Siehe Stellungnahme des Max Planck Institute for Innovation and Competition, abrufbar unter https://www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI-Position-Paper_TDM_2017-01-14_corr_def.pdf (letzter Zugriff 23.10.2018) und Stellungnahme der European Research and Innovation community “Securing Europe’s leadership in the data economy by revising the Text and Data Mining (TDM) exception“, abrufbar unter http://eare.eu/assets/uploads/2017/09/Open_Letter_on_TDM_to_the_Council_26_September_2017-1.pdf (letzter Zugriff 23.10.2018).

der Schranke. Teils ist umstritten, ob die Schaffung dieses Erlaubnistatbestands überhaupt notwendig war.⁴⁷

2.5.2.1. Beschränkung auf nichtkommerzielle Forschung

Zentraler Kritikpunkt in der aktuellen Diskussion um den europäischen Richtlinienentwurf ist die Begrenzung des Anwendungsbereichs der Schranke auf die nichtkommerzielle wissenschaftliche Forschung. Es wird in diesem Zusammenhang argumentiert, dass die entscheidenden Innovationstreiber der digitalen Wirtschaft im Bereich TDM insgesamt und speziell KI nicht nur die wissenschaftlichen Institute, sondern auch und vor allem die Privatwirtschaft sei. Innovative Start-Ups und mittelständische Unternehmen, die die deutsche und europäische digitale Entwicklung bedeutend voranbringen könnten, würden im internationalen Wettbewerb unangemessen benachteiligt.⁴⁸

Gerade für kleine innovative Unternehmen, die über keine Rechtsabteilungen oder Mittel für ausgiebige juristische Beratung haben, sei die Rechtsunsicherheit⁴⁹ und die damit verbundene Gefahr von Rechteinhabern mit Rechtsstreitigkeiten überzogen zu werden, ein großes Problem.

Fraglich ist auch, wie mit Auftragsforschung und Kooperationen von Hochschulen und privaten Unternehmen umzugehen sein wird. Drittmittelgeber, die für die Finanzierung von wissenschaftlicher Forschung an den deutschen Hochschulen eine wichtige Rolle spielen⁵⁰, dürften weit weniger investieren, wenn die Forschungsergebnisse sich nicht in neuen Produkten oder Geschäftsmodellen verwerten lassen könnten.⁵¹

Angesichts der enormen Bedeutung, die KI für die Zukunft eingeräumt wird, bestehen deshalb Bedenken, dass Deutschland und die EU wissenschaftlich wie wirtschaftlich in nicht mehr korri-

47 *de la Durantaye* „Neues Urheberrecht für Bildung und Wissenschaft – eine kritische Würdigung des Gesetzentwurfs“ GRUR 2017, 558 (561).

48 *Raue*, a.a.O. S. 656 f.

49 Aufgrund des enormen Transferaufwands, den die Lizenzbeschaffung darstellt, sind in der Praxis viele gerade von Start-Ups verwendete Korpora mutmaßlich eher als halblegal anzusehen, da die Daten oft aus dem Internet geschürft werden. Zwar ist die praktische Gefahr beim Speichern von Inhalten aus dem Web „erwischt“ zu werden zwar eher gering. Aber natürlich stellt diese Praxis der Datenbeschaffung ein erhebliches rechtliches und damit auch wirtschaftliches Risiko für die Verwender dar, was den Unternehmern z.B. spätestens in einer due-diligence-Prüfung durch einen Investor Schwierigkeiten bereiten kann.

50 Im Jahr 2016 trug die Privatwirtschaft etwa 20% zu den Drittmiteleinahmen der deutschen Hochschulen bei. Quelle: Statistisches Bundesamt, Finanzen der Hochschulen 2016, S. 28, abrufbar unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/BildungKulturFinanzen/FinanzenHochschulen2110450167004.pdf?_blob=publicationFile (letzter Zugriff 23.10.2018).

51 *Spindler* „Text and Data Mining – urheber- und datenschutzrechtliche Fragen“ GRUR 2016, 1112 (1118).

gierbarer Weise auf Jahrzehnte hinweg zurückfallen könnten, wenn Unternehmen in andere Länder abwandern, in denen es keine vergleichbaren rechtlichen Hindernisse beim Text and Data Mining gibt.⁵²

Doch nicht nur für die KI-Entwicklung sei der Ausschluss kommerzieller Nutzung aus der TDM-Regelung problematisch: Auch der Journalismus sei heutzutage in Form des „Datenjournalismus“ immer mehr auf automatisierte Datenauswertung angewiesen. Die Offenlegung von Wirtschaftsskandalen wie im Zuge der sogenannten Panama-Papers⁵³, die auf der Auswertung riesiger geleakter Datenmengen beruhten, deren manuelle Analyse unmöglich sei, sei durch die aktuelle Regelung erheblich erschwert.⁵⁴

2.5.2.2. Vergütungspflicht

Ebenfalls kritisiert wird die vorgesehene Vergütungspflicht. Wie eingangs erwähnt schützt das Urheberrecht nicht die in einem Werk enthaltene Information an sich sondern lediglich ihre konkrete schöpferische Ausgestaltung. Relevant für das TDM sei aber nur die Information, die aber eben nur zwangsläufig in einer urheberrechtlich geschützten Verpackung zur Verfügung stünde. Dass es nur zu einer technisch notwendigen, aber sonst eigentlich keiner nennenswerten Nutzung der Werke käme, könne vor dem Hintergrund, dass für die Anwendung der Schranke bereits ein Zugangsrecht bestehen müsse, eine Vergütungspflicht nicht rechtfertigen.⁵⁵ Inhaltenanbieter hätten zudem die Möglichkeit, den freien Zugang zu ihren Titeln einzuschränken, Anbieter von kostenpflichtigen Archiven oder Datenbanken könnten die „zusätzliche“ Nutzung in ihr bestehendes Gebührenmodell einpreisen.

Gerade vor dem Hintergrund, dass die Korpora nach dem Abschluss der Forschungsarbeiten und ihrer wissenschaftlichen Überprüfung gelöscht werden müssen, sei die Vergütungspflicht unverständlich. Fraglich ist zudem, welche Verwertungsgesellschaft für die Geltendmachung der Ansprüche zuständig ist und wie die Höhe der Vergütung ermittelt werden soll.

2.5.2.3. Löschpflicht

Es wird kritisiert, dass die Korpora, die mit erheblichem Aufwand und unter Einsatz von (möglicherweise staatlichen) Forschungsgeldern erstellt werden, nach dem Abschluss des konkreten

52 *Raue*, a.a.O. S. 656 f.

53 Beispiel: Im Jahr 2016 wurden Journalisten der Süddeutschen Zeitung Daten über ein System von Briefkastenfirmen zugespielt, über die teilweise illegale Finanzaktivitäten im großen Stil abgewickelt wurden. Das Datenpaket mit einer Größe von 2,6 Terabyte beinhaltete über 11,5 Millionen Dokumente, die erst maschinenlesbar gemacht und kategorisiert werden mussten und dann algorithmisch analysiert wurden, wobei die SZ mit einem internationalen Redaktionsnetzwerk zusammenarbeitete. Eine manuelle Bearbeitung und Auswertung einer solchen Datenmenge wäre kaum möglich gewesen.

54 *Raue*, a.a.O. S. 656 f.

55 Pflüger/Hinte „Das Urheberrechts-Wissensgesellschafts-Gesetz aus Sicht von Hochschulen und Bibliotheken“ ZUM 2018, 153 (160).

Projekts gelöscht werden müssen. Korpora seien oft für viele verschiedene Forschungszwecke geeignet. Deshalb sei es eine Verschwendung von Ressourcen, dass diese Korpora nicht für andere Forschungsziele verfügbar seien.⁵⁶

Von anderer Seite wird dagegen eingewandt, dass die Gefahr bestehe, dass die Korpora mit den rechtlich geschützten Werken als Inhalt, sollten sie weiter verfügbar sein, als Quelle für eine illegale Weiterverbreitung genutzt werden könnten.⁵⁷

3. Urheberrechtlicher Schutz von Algorithmen und Computerprogrammen

Unabhängig von der Frage ihrer Patentierbarkeit⁵⁸ können Algorithmen oder Software allgemein als Computerprogramme gem. § 69a UrhG urheberrechtlich geschützt sein.⁵⁹ Ideen und Grundsätze, die einem Computerprogramm oder einzelnen seiner Elemente zugrunde liegen, sind zwar urheberrechtlich nicht schutzfähig (§ 69a Abs. 2 UrhG), wenn sie nicht über allgemeingültige mathematische bzw. informatische Grundregeln hinausgehen. Geschützt ist aber die „kreative“ Leistung eines Programmierers⁶⁰, also die konkrete Quellcode-Gestaltung durch den Entwickler als konkreter Ausdruck eines Werkes, nicht hingegen der bloße Informationsinhalt.

Auch generelle Funktionalitäten eines Computerprogramms können nicht durch das Urheberrecht geschützt werden. Diese müssen der Allgemeinheit vielmehr zu weiterem Werkschaffen als Allgemeingut offen bleiben. Das Allgemeingut verdient als solches auch keinen Schutz, weil es nicht auf eine individuelle Schöpfung zurückgeht.⁶¹

56 *Raue*, a.a.O. S. 656 f.

57 Siehe Erwägungsgründe zum Regierungsentwurf des UrhWissG, BT-Drucks. 18/12329, S.41.

58 Siehe zur patentrechtlichen Frage: *Baldus* in *Auer-Reinsdorff/Conrad*, IT- und Datenschutzrecht, Teil B. Immaterialgüterrecht § 5 Rechtsschutz von Computerprogrammen und digitalen Inhalten Rn. 107-114.; *Schaub* „Interaktion von Mensch und Maschine“, JZ 2017, 342 (346).

59 Übersichtlich zur allgemeinen Schutzfähigkeit von Computerprogrammen: *Hoeren/Vossen* „Softwareverletzung – Missverständnisse bei der Feststellung der Schutzfähigkeit von Computerprogrammen“ K&R 2018, 79.

60 Zu beachten ist, dass durch eine Sonderregelung in § 69b UrhG nicht der Programmierer als Urheber selbst das Nutzungs- und Verwertungsrecht erwirbt, sondern, für den Fall, dass die Entwicklung auf einem Arbeitsverhältnis beruht, der Arbeitgeber die Verwertungsrechte per Gesetz erwirbt. In der Praxis bedeutet das keinen großen Unterschied, da Arbeitnehmer in der Regel durch eine Klausel im Arbeitsvertrag standardmäßig die Nutzungs- und Verwertungsrechte an ihre im Rahmen des Arbeitsverhältnisses entstandenen kreativen Leistungen an den Arbeitgeber abtreten.

61 Urteil des EuGH vom 2.5.2012, Az. EuGH C-406/10, mit Anmerkung von *Stögmüller* in K&R 2012, 415.

Insbesondere in Bezug auf den sog. Algorithmus ist deshalb genau zu untersuchen, ob es sich dabei um eine allgemein bekannte, schutzrechtsfreie mathematische Rechenregel handelt und damit ein sogenanntes „Banalprogramm“ oder um die an sich schutzfähige Ausgestaltung eines konkreten Programms.⁶²

In der Regel dürften aber solche Codes und Algorithmen, die einer KI zugrunde liegen, dermaßen komplex sein, dass ihnen in ihrer konkreten Ausgestaltung ein urheberrechtlicher Schutz zukommen müsste.⁶³

Fraglich ist aber, wie es um Algorithmen und Software steht, die sich im Rahmen von Machine Learning selbstständig optimieren und autonom weiterentwickeln.

4. Urheberrechtlicher Schutz von durch KI erzeugten Ergebnissen

Wenn die von einer KI erzeugten Ergebnisse nicht nur in Form von ohnehin nicht schutzfähigen Datenansammlungen oder Berechnungen, sondern im äußeren Erscheinungsbild ähnlich einer der im UrhG genannten Werksformen oder anderer Leistungen vorliegen, stellt sich die Frage, ob diese Ergebnisse ihrerseits schutzfähig sind.

4.1. Urheberrechte

Bereits heute gibt es KI-generierte Schöpfungen, die dem äußeren Erscheinungsbild nach einem menschlichen Werk in nichts nachstehen. So können Computer autonom selbstständig Texte zusammenfassen⁶⁴ oder übersetzen, Musik komponieren und produzieren oder „Kunst“ erschaffen, die selbst Experten im Falle eines maschinenerstellten Gemäldes für einen echten Rembrandt hielten.⁶⁵

Fraglich ist daher, ob so entstandene Inhalte Werke i.S.d. § 2 UrhG sein können und falls ja, wer der Urheber ist.

Wie oben bereits ausgeführt wurde, ist der originäre Kern des Urheberrechts die persönliche geistige Schöpfung. Ein von KI hergestelltes Erzeugnis ist also nur dann ein urheberrechtsschutzfähiges Werk, wenn die Gestaltung des Erzeugnisses noch auf einen menschlich-geistigen Schöpfungsakt zurückgeführt werden kann, wenn also die Maschine nur Hilfs- bzw. Ausführungsmittel

62 Dreier in Dreier/Schulze, a.a.O. § 69a, Rn. 20-22.

63 Schaub „Interaktion von Mensch und Maschine“ JZ 2017, 342 (347).

64 So können z.B. durch KI längere Texte zu „Snippets“ zusammengefasst und in einer Suchmaschine angezeigt werden, um das Leistungsschutzrecht der Verleger zu umgehen.

65 Siehe als Beispiel das Projekt „The Next Rembrandt“ von der ING-Group und Microsoft, die dazu alle Gemälde von Rembrandt van Rijn detailliert bis zum einzelnen Pinselstrich durch eine KI analysieren ließen und die daraufhin mit deren typischen Merkmalen ein neues Gemälde errechnete, das am Ende per 3D-Druck auf die Leinwand gebracht wurde. Abrufbar unter <https://www.nextrembrandt.com/> (letzter Zugriff 23.10.2018).

ist.⁶⁶ Nur in diesem Fall ist noch die für den urheberrechtlichen Schöpfungsbegriff unabdingbare Unmittelbarkeit zwischen geistigem Vorgang und Ergebnis gegeben. Wird die Fähigkeit des Computers bewusst in eine Gestaltung mit einbezogen so steht das einem Schöpfungsvorgang i.S.v. § 2 Abs. 2 UrhG nicht entgegen.⁶⁷ Das ist in etwa der Fall, wenn ein Grafikdesigner mittels Bildbearbeitungssoftware bestimmte Effekte erzeugt oder ein Musiker oder Musikproduzent virtuelle Instrumente, Sampler und Sequenzersoftware oder auch Zufallselemente verwendet, um durch Auswahl, Kombination und Anordnung verschiedener Klänge ein Musikstück zu erstellen.⁶⁸ Dabei sind an die sogenannte Schöpfungshöhe keinen allzu großen Anforderungen zu stellen, das Urheberrecht schützt auch die „kleine Münze“.⁶⁹

An dieser Unmittelbarkeit fehlt es aber, wenn das Erzeugnis ausschließlich das Ergebnis einer KI (ähnlich eines Zufallsgenerators) ist. Mag auch die KI oder der Algorithmus selbst auf einem geistigen Schaffensprozess beruhen, so ist das hierdurch entstandene Ergebnis nur mittelbar auf einen menschlichen Schaffensprozess zurückzuführen. Es reicht also für den Urheberrechtsschutz nicht aus, dass der Mensch zwar die Maschine beherrscht, den unmittelbaren Umsetzungsprozess innerhalb und durch die Maschine aber nicht mehr beeinflussen kann.⁷⁰

Nach ganz überwiegender Ansicht sind daher vollständig autonom durch eine KI, deren genauer Funktionsprozess zu komplex ist, um irgendwelche Ergebnisse vorherzusagen, erzeugte Schöpfungen nicht als Werke i.S.v. § 2 UrhG anzusehen und somit nicht als solche schutzfähig.⁷¹ Der Programmierer der KI wird nicht zum Urheber des Werks der KI.⁷² Es gibt in solchen Fällen keinen Urheber.

Im Ergebnis wird es aber auf eine Einzelfallprüfung ankommen, ob der erzeugte Inhalt noch in irgendeiner Form auf eine individuelle unmittelbare geistige Schöpfung zurückgeht oder nicht.

Nichts anderes gilt in dem Fall, dass sich ein autonomes System mittels KI weiterentwickelt und selbst seinen Code verändert. Die Änderungen sind dann nicht mehr vom urheberrechtlichen

66 *Hetmank/Lauber-Rönsberg* „Künstliche Intelligenz – Herausforderungen für das Immaterialgüterrecht“ GRUR, 2018, 574 (577).

67 *Loewenheim* in *Schricker/Loewenheim*, Urheberrecht, 5. Aufl. 2017, Rn. 12-14.

68 *Bullinger* in *Wandtke/Bullinger*, Praxiskommentar zum Urheberrecht, 4. Aufl. 2014, § 2 UrhG, Rn. 72.

69 *Nordemann* in *Loewenheim*, Handbuch des Urheberrechts, 2. Aufl. 2010, § 6, Rn. 16 ff.

70 *Ahlberg* in BeckOK Urheberrecht, § 2 UrhG, Rn. 54-56; *Schöttler* „Künstliche Intelligenz als Schöpfer urheberrechtlicher Werke?“ AnwZert ITR 11/2017 Anm. 2.

71 Anders sieht das *Schulze*, der es für das Entstehen eines Urheberrechtsschutzes ausreichen lassen will, wenn „der Urheber“ aus einer Mehrzahl aus einem Zufallsgenerator (und damit wohl auch einer KI) entstandenen Versionen bewusst eine Auswahl trifft, welche davon als „Werk“ zu bestimmen sei. Siehe *Dreier/Schulze*, Urheberrechtsgesetz, 6. Aufl. 2018, Rn. 8. Das dürfte aber im Ergebnis zu weit gehen.

72 Anders wohl nur *Harte-Bavendamm/Wiebe* in *Kilian/Heussen*, Computerrechts-Handbuch, 33. EL Februar 2017, 1. Abschnitt, Teil 5, Rn. 41, die dem Programmierer in Einzelfällen ein Urheberrecht am Werk der KI zukommen lassen wollen.

Schutz erfasst.⁷³ Die verbleibenden vom Entwickler geschriebenen Elemente dürften aber weiterhin als geschützt gelten.

4.2. Verwandte Schutzrechte

Die verwandten Leistungsschutzrechte schützen Erzeugnisse, die den urheberrechtlich geschützten Werken zwar in ihrer Form und Schutzwirkung ähneln, aber die Schwelle der geistigen Schöpfung nicht erreichen. Geschützt wird hier deshalb vor allem die Investitionsleistung, die zur Erstellung oder Aufführung nötig ist.⁷⁴

4.2.1. Lichtbildschutz

So sind z.B. Fotos, die die erforderliche Schöpfungshöhe nicht erreichen (etwa bloße Schnappschüsse aber auch Fotos von technischen Vorgängen, Satellitenfotos, Werbefotos oder Bildberichterstattung etc.) nicht als Lichtbildwerke gem. § 2 UrhG sondern als Lichtbilder nach § 72 UrhG geschützt. Früher bedeutete das Aufnehmen von Lichtbildern einen gewissen finanziellen und technischen Aufwand wie Aufbau, Einstellen der Kamera, Filmentwicklung und Erstellung von Abzügen, der als schützenswert gilt.⁷⁵ Das Leistungsschutzrecht erwirbt der Fotograf bzw. derjenige, der den Aufnahmeprozess auslöst.⁷⁶

Fraglich ist, ob visuelle Erzeugnisse und Bilddateien, die Fotos oder Lichtbildern ähneln, wie z.B. „The Next Rembrandt“, dem Lichtbildschutz unterfallen. Dafür lässt der Gesetzeswortlaut keinen Raum, wenn er „Lichtbilder sowie Erzeugnisse, die ähnlich wie Lichtbilder hergestellt werden“ (§ 72 Abs. 1 UrhG) als Schutzgegenstand nennt. Konkret heißt das, dass die Bilder anhand eines photochemischen oder photophysikalischen Prozesses durch Strahlung bzw. Licht entstanden sein müssen. Davon sind z.B. auch Aufnahmen eines Computertomographen oder von Infrarotkameras erfasst, vollständig im Computer digital erzeugte Grafiken oder CAD-Zeichnungen oder eben von einer KI berechnete Bilder oder Grafiken jedoch nicht.⁷⁷

4.2.2. Datenbankschutz

Datenbankwerke genießen ebenfalls Urheberrechtsschutz. Darunter sind Sammlungen von Werken, Daten oder anderen unabhängigen Elementen, die aufgrund der Auswahl oder Anordnung der Elemente eine persönliche geistige Schöpfung darstellen und deren Elemente systematisch

73 *Schaub*, a.a.O. S. 347.

74 *Vogel* in *Loewenheim*, Handbuch des Urheberrechts, 2. Aufl. 2010, § 37 Rn. 1.

75 Ob ein auf dem Investitionsaufwand beruhender Schutz von Lichtbildern ohne Einschränkung in Zeiten einer Bilderflut mit jederzeit verfügbaren Digital- und Smartphone-Kameras und der mit einem Klick möglichen massenhaften und unkomplizierten Verbreitung im Internet ohne jeglichen organisatorischen oder finanziellen Aufwand noch angemessen ist, sei dahingestellt. Andererseits bedürfen tatsächlich aufwändige Bilder wie Satellitenfotos aber auch die Erzeugnisse von Berufsfotografen oder Bildjournalisten eines gewissen Schutzes.

76 § 72 Abs 2 UrhG; *Fechner*, a.a.O. Rn. 111.

77 *Vogel* in *Loewenheim*, a.a.O. § 37 Rn. 9.; *Fechner*, a.a.O., Rn. 15.

oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind (§ 4 UrhG). Bei automatisch durch einen Computer gesammelten und ausgewählten oder von Computern eigens erstellten Inhalten fehlt es an der persönlichen geistigen Schöpfung (siehe oben).

In diesem Fall greifen jedoch die §§ 87a ff. UrhG ein, die das Merkmal der Auswahl und Anordnung und damit der persönlichen geistigen Schöpfung nicht erfordern und die daher ein Leistungsschutzrecht „sui generis“ für (automatisch generierte) Datenbanken gewähren (klassisches Beispiel: Telefonbuch).⁷⁸ Damit sollen die Investitionen des Datenbankherstellers, die durch Beschaffung, Sammlung, Aufbereitung, Überprüfung und Darstellung der Daten durchaus erheblich sein können, geschützt werden und durch eine Vergütungsregelung gleichzeitig ein Anreiz geschaffen werden, in den Aufbau und die Pflege von Datenbanken zu investieren.⁷⁹ Den europarechtlichen Rahmen gibt dabei die Datenbank-RL⁸⁰ vor. Durch Art. 7 Abs. 1 Datenbank-RL wird das im UrhG nicht genannte Kriterium der „wesentlichen Investition“ als weitere Voraussetzung des Schutzes vorgegeben.⁸¹

Fraglich ist, welche Investitionen zu berücksichtigen sind. Schutzbegründend sind solche in die Beschaffung, Überprüfung oder Darstellung des Datenbankinhalts, aber auch der Aufwand für die Entwicklung oder Lizenzierung von Computerprogrammen, die für die Erstellung und den Betrieb der Datenbank nötig sind.⁸²

Demgegenüber sind Mittel, die für das Erzeugen der Werke oder der Elemente eingesetzt werden, die sich in der Datenbank befinden, den mit der Beschaffung des Inhalts dieser Datenbank verbundenen Investitionen nicht gleichzustellen. Sie können also bei der Beurteilung, ob die mit der Erstellung dieser Datenbank verbundenen Investitionen wesentlich sind, nicht berücksichtigt werden. Die Datenerzeugung fällt gerade nicht unter den Schutzzweck der §§ 87a ff UrhG, einen Anreiz für die Erstellung von Informationssystemen zu bieten.⁸³ Vom Datenbankhersteller selbst generierte Daten sind nicht erfasst, um eine Monopolisierung von Daten zu vermeiden.⁸⁴

In Kombination mit dem Erfordernis der Unabhängigkeit der einzelnen Elemente der Datenbank folgt daraus, dass Investitionskosten für von einer KI generierte Inhalte wie Bilder, Musikstücke oder Übersetzungen von Romanen nicht von dem Datenbankschutz sui generis erfasst sind.

78 Hubmann/Rehbinder/Peukert, a.a.O. Rn. 728.

79 Ebenda Rn. 729.

80 Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken.

81 Dreier in Dreier/Schulze, a.a.O. § 87a UrhG, Rn. 11.

82 Hetmank/Lauber-Rönsberg, a.a.O. S. 578.

83 OLG Hamburg, Urteil vom 8.6.2017 – 5 U 54/12, BeckRS 2017 Rn. 245, mit Hinweis auf EuGH Urteil vom 9.11.2004, Az. C-444/02, „Fixtures-Fußballspielpläne I“, GRUR Int 2005, 239.

84 Hetmank/Lauber-Rönsberg, a.a.O. S. 578.

Erfasst werden können nur solche Investitionen in die Entwicklung oder Nutzung einer KI, die genau auf die Erstellung einer konkreten Datenbank ausgerichtet sind. Als Beispiel zu nennen wäre hier eine KI, die allgemein zugängliche Daten analysiert und katalogisiert, wie etwa eine automatisierte Analyse von DNA-Sequenzen.⁸⁵

Ein Schutz von KI-generierten Ergebnissen als Datenbank *sui generis* ist also möglich, allerdings nur unter den genannten eng begrenzten Voraussetzungen, die nur ausnahmsweise vorliegen dürften.

4.2.3. Einführung eines Immaterialgüterrechts für KI-produzierte Schöpfungen?

Angesichts des *de lege lata* sehr lückenhaften Schutzes, der Ergebnissen von KI-Prozessen zuteilwird, auf der einen Seite, und der enormen organisatorischen und wirtschaftlichen Investitionen, die die technische Entwicklung (aber auch schon die Beschaffung von geeigneten Korpora) bedeuten, auf der anderen Seite, stellt sich die Frage, ob es nicht der Schaffung eines speziellen immaterialgüterrechtlichen Schutzes bedarf.

Zwar sind die konkrete Ausgestaltung der KI-Software und die Investition an sich in der Regel urheberrechtlich geschützt, wenn aber die immateriellen Resultate ungeschützt von jedermann frei verwendet werden können, kommt der KI kein großer Marktwert mehr zu. Um also sicherzustellen, dass die Investitionen rentabel bleiben und um einen Anreiz zu bieten, solche Investitionen überhaupt zu tätigen, könnte ein entsprechendes Leistungsschutzrecht oder Patentrecht helfen.

Andererseits haben immaterialgüterrechtliche Regelungen immer auch eine wettbewerbsbeschränkende Kehrseite, die Innovationen sogar behindern könnte. Ob ein weitergehender Investitions- und Innovationsschutz gerechtfertigt ist, ist eine rechtliche, ökonomische, gesellschaftliche und nicht zuletzt politische Frage. Zu Konzepten und den Voraussetzungen, die Patent- und Leistungsschutzrecht in Bezug auf KI-generierte Erfindungen und Schöpfungen erfüllen sollten, liefern Sven Hetmank und Anne Lauber-Rönsberg einige lesenswerte Ansätze.⁸⁶

5. Fazit und Ausblick

Die Forschung zu künstlicher Intelligenz hat zuletzt entscheidende Fortschritte gemacht und dürfte unser Leben in vielen Bereichen in den nächsten Jahren und Jahrzehnten bedeutend verändern. Die Entwicklung von KI-Systemen ist ressourcenintensiv und im besonderen Maße abhängig von der Verfügbarkeit einer großen Menge an Trainingsdaten und deren Qualität. Viele dieser Daten sind urheber- oder leistungsschutzrechtlich geschützt und die rechtssichere Erstellung von Korpora stellt Forschung und Wirtschaft oft vor Schwierigkeiten. Durch die Einführung der Text-and-Data-Mining-Schranke in das deutsche Urheberrechtsgesetz hat sich zwar für die

85 Ebenda S. 579.

86 *Hetmank/Lauber-Rönsberg*, a.a.O.; Ebenfalls einen *sui generis*-Schutz für KI-generierte Inhalte ins Spiel bringt *Schöttler* „Künstliche Intelligenz als Schöpfer urheberrechtlicher Werke?“ AnwZert ITR 11/2017 Anm. 2.

nichtkommerzielle Forschung eine Erleichterung bei der Beschaffung von Trainingsdaten ergeben. Die kommerzielle Forschung, die einen wichtigen Innovationstreiber darstellt, profitiert davon jedoch nicht. Kritiker dieser Beschränkung, die befürchten, dass durch diese die zukünftige KI- und damit auch die gesamtwirtschaftliche Entwicklung in Deutschland und der EU einen enormen Standortnachteil gegenüber anderen Rechtsordnungen erleiden, setzen ihre Hoffnungen darauf, dass die Beschränkung auf nichtkommerzielle Forschung mit der kommenden EU-Urheberrechtsreform entfällt.

Gleichzeitig ist der Schutz der Investitionen in KI-Entwicklung lückenhaft. Zwar erfahren die KI-Systeme mit ihren Algorithmen urheberrechtlichen Schutz als Computerprogramme. Die Resultate und Schöpfungen, die die KIs produzieren, sind aber bis auf wenige Ausnahmen ungeschützt und somit nur bedingt wirtschaftlich verwertbar. Es ist aber zu beachten, dass die von KI generierten Daten oft nur ein indirekter Schritt zu oder Bestandteil von Dienstleistungen und Produkten sind, die ihrerseits wirtschaftlich verwertbar sind. Ob darüber hinaus gesetzgeberisch weitere immaterialgüterrechtliche Maßnahmen notwendig sind, um Investitionen in KI wirtschaftlich zu sichern, ist diskutabel und kann nicht eindeutig beantwortet werden.
