

## KI-Trainingsdaten und Urheberrecht

– Chancen und systeminduzierte Grenzen der KI-Verordnung

Boris Paal und Fabian Kieß\*

### I. Einleitung

„Pferdekutscher fanden es auch schlecht, dass Autos automatisiert fahren konnten und man keinen Kutscher mehr brauchte. Ähnlich, wenn man jetzt Illustrator ist.“ Mit dieser Äußerung positionierte sich Richard Socher, Gründer der KI-Suchmaschine *You.com*, zu den Beschwerden und Klagen von Künstlern, Grafikern und Autoren gegen die – aus ihrer Sicht – unangemessenen Entwicklungen betreffend das Training von Technologien Künstlicher Intelligenz (KI).<sup>1</sup> Bereits ein erster Blick lässt die Angreifbarkeit dieser provokanten Einlassung erkennen, da – anders als durch die Automobilindustrie im Verhältnis zur Kutscherbranche – der Output von KI auf der vorangehenden, massenhaften Verwendung von künstlerischen Werken zu Trainingszwecken aufsetzt.<sup>2</sup>

Zahlreiche internationale und nationale Verbände, wie etwa die US-amerikanische *Authors Gild*<sup>3</sup> oder das *Netzwerk Autorenrechte*<sup>4</sup>, haben offene

---

\* Prof. Dr. Boris Paal, M. Jur. (Oxford) ist Inhaber des Lehrstuhls für Law and Regulation of the Digital Transformation an der TU München, TUM School of Social Sciences and Technology, Department of Governance. Dr. Fabian Kieß war wissenschaftlicher Mitarbeiter am vorherigen Lehrstuhl Paal in Leipzig; zurzeit ist er Rechtsreferendar am Hanseatischen Oberlandesgericht.

- 1 J. Edelhoff/E. Özer/T. Soliman, Müssen Künstler und Autoren für KI entschädigt werden?, 23.6.2023, abrufbar unter <https://www.tagesschau.de/wirtschaft/digitales/kuenstliche-intelligenz-grafiker-autoren-urheberrechte-100.html> (zuletzt abgerufen am 30.6.2024).
- 2 N. Maamar, Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen, ZUM 2023, 481 (482); J. Pukas, KI-Trainingsdaten und erweiterte kollektive Lizenzen, GRUR 2023, 614.
- 3 The Authors Guild, Open Letter to Generative AI Leaders, 18.7.2023, abrufbar unter <https://authorsguild.org/app/uploads/2023/10/Authors-Guild-Open-Letter-to-Generative-AI-Leaders.pdf> (zuletzt abgerufen am 30.6.2024).
- 4 Netzwerk Autorenrechte, Offener Brief zur Haltung Deutschlands zur geplanten KI-Grundverordnung, abrufbar unter <https://www.netzwerk-autorenrechte.de/docs/2023-11-21-NAR%20an%20BK%20BMW%20DIG.pdf> (zuletzt abgerufen am 30.6.2024).

Briefe an führende KI-Unternehmen und Regierungen adressiert, um – insbesondere – eine fair(er)e urheberrechtliche Partizipation von Kreativschaffenden an der Verwendung von KI-Trainingsdaten zu erreichen. In dieser Linie stehen auch prominente Gerichtsverfahren: So hat die Bildagentur *Getty Images* in den USA das KI-Unternehmen *Stability AI* verklagt, weil dieses über 12 Millionen Bilder zu Trainingszwecken des eigenen KI-Modells verwendet hatte.<sup>5</sup> In Deutschland sorgt der Fall des Fotoproduzenten *Robert Kneschke* für Aufmerksamkeit, der gegen die Verwendung der von ihm geschaffenen Bilder im Datensatz *Laion-5B* vor dem Landgericht Hamburg vorgeht. Der Datensatz wird vom *LAION e.V.* quelloffen und kostenlos zum Training von KI-Bildgeneratoren zur Verfügung gestellt, worin *Kneschke* eine urheberrechtlich relevante Vervielfältigung i.S.d. § 16 Abs. 1 UrhG sieht.<sup>6</sup> Ende 2023 wurde bekannt, dass die *New York Times* die Unternehmen *Microsoft* und *OpenAI* wegen Urheberrechtsverletzungen im Zusammenhang mit dem Training von KI-Software verklagt hat.<sup>7</sup>

Eine epochale Zäsur für die Wahrnehmung des disruptiven Innovationspotenzials von generativer KI markierte der Markteintritt des *OpenAI* Chatbots *ChatGPT*<sup>8</sup> am 30. November 2022, der innerhalb von nur fünf Tagen bereits eine Million Nutzer aufweisen konnte. Eine Studie der US-Bank *Goldman Sachs* prognostizierte in der Folge, dass die Automatisierung von Arbeitsplätzen weltweit bis zu 300 Millionen Vollzeitarbeitsplätze betreffen könnte, mit dem Potenzial einer vollständigen Verdrängung eines Viertels davon durch KI-Anwendungen.<sup>9</sup>

Vor dem Hintergrund des inhärenten Spannungsfelds von Innovationspotenzialen auf der einen und fairer urheberrechtlicher Partizipation auf

---

5 Getty Images (US), Inc. v. Stability AI, Inc., 1:23-cv-03223 (N.D. Cal.). Hierzu etwa *K. de la Durantaye*, „Garbage in, garbage out“ – Die Regulierung generativer KI durch Urheberrecht, ZUM 2023, 645.

6 *SLD*, Termin verlegt – Klage gegen LAION e.V. vor dem Landgericht Hamburg, abrufbar unter <https://www.sld-ip.com/blog/klage-gegen-laion-e-v/> (zuletzt abgerufen am 30.6.2024).

7 *M. Grynbaum/R. Mac*, The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work, 27.12.2023, abrufbar unter <https://www.nytimes.com/2023/12/27/business/new-york-times-open-ai-microsoft-lawsuit.html>; United States District Court Southern District Of New York: Case 1:23-cv-11195.

8 GPT steht für „generative Pre-trained Transformer“.

9 *Goldman Sachs*, The Potentially Large Effects of Artificial Intelligence on Economic Growth, 26.3.2023, abrufbar unter <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html> (zuletzt abgerufen am 30.6.2024).

der anderen Seite, wird in diesem Beitrag untersucht, ob und inwieweit das Training von KI unter Verwendung von geschützten Werken eine urheberrechtlich relevante Verwertungshandlung darstellt – und welcher legislatorische Handlungs- und Anpassungsbedarf hieraus erwächst.<sup>10</sup> Zu diesem Zweck wird ein besonderes Augenmerk auf der jüngst im Trilog-Verfahren beschlossenen KI-Verordnung<sup>11</sup> (KI-VO) liegen, um urheberrechtliche Chancen und Grenzen des dort verfolgten risikobasierten Ansatzes herauszuarbeiten.

Zunächst erfolgen eine Annäherung an den untersuchungsgegenständlichen KI-Begriff (hierzu II.), eine Darstellung des urheberrechtlichen Schutzes *de lege lata* in Europa und Deutschland (dazu III.) sowie die Bewertung der Implikationen durch die geplante KI-Verordnung (nachfolgend IV.), bevor die Ausführungen mit einem Fazit und Ausblick (dazu V.) schließen.

Dieser Beitrag ist in freundschaftlicher Verbindung *Georgios Gounalakis* zum 65. Geburtstag zugeschrieben; der Jubilar ist mit zahlreichen wegweisenden Veröffentlichungen – auch – zum Urheberrecht hervorgetreten, so insbesondere mit seiner bahnbrechenden Dissertationsschrift aus dem Jahr 1989.<sup>12</sup>

## II. Untersuchungsgegenständlicher KI-Begriff

Bislang hat sich noch kein allgemeingültiger Begriff von KI-Systemen herausgebildet.<sup>13</sup> Gleichwohl sind für die zentralen Kriterien von KI-Systemen bereits belastbare Definitionsansätze vorhanden: Art. 3 Nr. 1 KI-VO definiert KI-Systeme als maschinenbasierte Systeme, die für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt, nach Betriebsaufnahme anpassungsfähig sowie dazu in der Lage sind, aus erhaltenen Eingaben

10 N. Maamar, (Fn. 2), ZUM 2023, 481 (482).

11 Rat der Europäischen Union, Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) – PE-CONS 24/24, 14.5.2024, abrufbar unter <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/de/pdf> (zuletzt abgerufen am 30.6.2024).

12 G. Gounalakis, Kabelfernsehen im Spannungsfeld von Urheberrecht und Verbraucherschutz, 1989.

13 Vgl. etwa S. Bußmann/C. Glasowski/M. Niehaus/S. Stecher, Die Schutzfähigkeit von KI-Trainingsdaten *de lege lata*, RDI 2022, 391.

explizite oder implizite Ergebnisse wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen abzuleiten, die physische oder virtuelle Umgebungen beeinflussen können.<sup>14</sup>

## 1. Generative KI

Die nachstehende Untersuchung legt einen Fokus auf generative KI, welche durch (zumindest partiell) autonome Entscheidungsfindung in der Lage ist, Inhalte wie Texte, Bilder, Videos oder Audios zu erzeugen.<sup>15</sup> Der Entscheidungsablauf generativer KI ist grundsätzlich durch die folgenden vier Prozessschritte gekennzeichnet: Sammlung von Trainingsdaten (Phase 1), Training (Phase 2), Nutzerauftrag (Phase 3) und Output (Phase 4).<sup>16</sup> Die Sammlung von Trainingsdaten erfolgt insbesondere durch Web-Crawling und Web-Scraping.<sup>17</sup> Der sich anschließende Trainingsprozess basiert auf maschinellem Lernen (*machine learning*), wobei maßgeblich künstliche neuronale Netze (KNN) zum Einsatz kommen.<sup>18</sup> Im Übrigen stützt sich das Training auf iterative Vorgänge,<sup>19</sup> hierzu erfolgt insbesondere ein Rückgriff auf die Methoden des überwachten Lernens (*supervised learning*) und des Verstärkungslernens (*reinforcement learning*).<sup>20</sup> Während beim überwachten Lernen durch den iterativen Prozess eine Annäherung der Prognosen an die in den Trainingsdaten enthaltenen richtigen Antworten erfolgt, speist sich der Lernprozess beim Verstärkungslernen aus dem Feedback der Lernumgebung.<sup>21</sup> Den Output-Prozess löst der Nutzerauftrag (vor

14 In diese Richtung insbesondere S. Bußmann/C. Glasowski/M. Niehaus/S. Stecher, (Fn. 13), RDI 2022, 391; H.-U. Detting/S. Krüger, Digitalisierung, Algorithmisierung und Künstliche Intelligenz im Pharmarecht, PharmR 2018, 513 (514). Ausführlich zum KI-Begriff Bitkom, Künstliche Intelligenz, 2017, S. 28 ff.

15 Vgl. K. de la Durantaye, (Fn. 5), ZUM 2023, 645 (646).

16 Treffend N. Maamar, (Fn. 2), ZUM 2023, 481 (481 f.).

17 So D. Bomhard, KI-Training mit fremden Daten – IP-rechtliche Herausforderungen rund um § 44b UrhG, DSRITB 2023, 255 (255 f.); J. Pukas, (Fn. 2), GRUR 2023, 614. Zur Funktionsweise und (datenschutzrechtlichen) Relevanz von Web-Scraping B. Paal, Data Scraping und Art. 82 DS-GVO, ZfDR 2023, 325.

18 Vgl. J. Pukas, (Fn. 2), GRUR 2023, 614.

19 P. J. Pesch/R. Böhme, Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern, GRUR 2023, 997 (999) mit weiterführender und instruktiver Einführung zur technischen Funktionsweise von generativer KI.

20 Vgl. P. Hacker, Immaterialgüterrechtlicher Schutz von KI-Trainingsdaten, GRUR 2020, 1025 (1026).

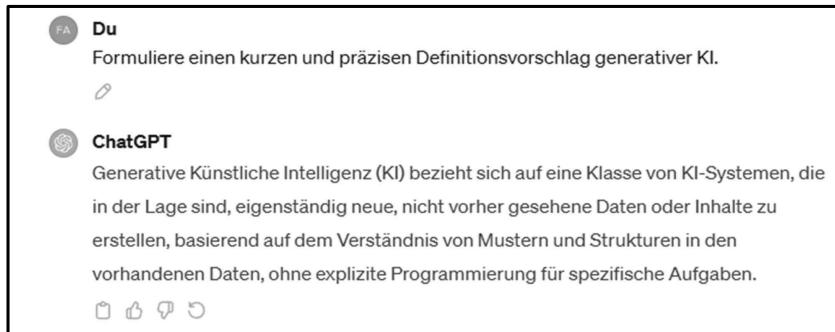
21 P. Hacker, (Fn. 21), GRUR 2020, 1025 (1026) mit weiterführenden Beispielen.

allem in der Form von *prompts*<sup>22)</sup> aus, bevor sodann auf Grundlage einer Wahrscheinlichkeitsrechnung ein Inhalt generiert wird.<sup>23)</sup>

## 2. Anwendungsbeispiel: ChatGPT 3.5

Um den vorbenannten, stark vereinfachten Begriff generativer KI besser einordnen und nachvollziehen zu können, erfolgt nachfolgend die Darstellung eines Anwendungsbeispiels anhand des OpenAI Chatbots *ChatGPT 3.5*. Dieser Chatbot basiert auf dem *Large Language Modell (LLM) GPT 3.5*, welches wiederum eine modifizierte Version des *LLM GPT 3.0* darstellt. *LLM GPT 3.0* ist ein Sprachmodell mit mehr als 175 Milliarden trainierbaren Parametern, für das durch *Open AI* rund 45 Terrabyte an Trainingsdaten bis einschließlich 2021 (Phase 1) insbesondere mittels Verstärkungslernens verwendet wurden (Phase 2).<sup>24)</sup> Im Rahmen der Nutzeranwendung kann mittels Texteingabe in einem *prompt* eine Frage oder eine Aufgabenstellung formuliert werden (Phase 3). Der Output besteht aus der Generierung eines Inhalts in Textform (Phase 4).

Abbildung 1: ChatGPT 3.5 prompt



22 So *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (647); *N. Maamar*, (Fn. 2), ZUM 2023, 481 (482).

23 Vgl. *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (647); *M. von Welser*, ChatGPT und Urheberrecht, GRUR-Prax 2023, 57.

24 Im Ganzen so *M. von Welser*, (Fn. 23), GRUR-Prax 2023, 57.

### III. Urheberrechtlicher Schutz de lege lata

Für die Auseinandersetzung mit dem urheberrechtlichen Schutz von Trainingsdaten ist zu differenzieren zwischen Input- und Output-Vorgang.<sup>25</sup> Der Input-Vorgang besteht aus den Teilschritten der Sammlung der Trainingsdaten, des Trainings und des Nutzerauftrags, während der Output-Vorgang den KI-generierten Inhalt betrifft.<sup>26</sup> Mit Blick auf den untersuchungsgegenständlichen Interessenkonflikt bzw. -ausgleich zwischen Rechtseinhabern und KI-Betreibern erfolgt hier eine Beschränkung auf die urheberrechtliche Ausgangssituation in der EU und Deutschland betreffend die Trainingsphase im Input-Vorgang.<sup>27</sup>

#### 1. Relevante Verwertungshandlung: (temporäre) Speicherung digitaler und digitalisierter Werke

Für die Schaffung funktionsfähiger KI ist die Sammlung massenhafter Trainingsdaten (Big Data) eine zentrale Voraussetzung.<sup>28</sup> Hierzu bedarf es regelmäßig, auch und gerade wegen des Nichtvorhandenseins von quantitativ und qualitativ ausreichenden Daten im Eigenbestand, eines Rückgriffs auf Fremddaten.<sup>29</sup> Während das einzelne Datum *de lege lata* weder unional noch national durch ein Ausschließlichkeitsrecht geschützt ist,<sup>30</sup> bestehen an Daten in Form von digitalen und digitalisierten<sup>31</sup> Werken – wie sich

---

25 Vgl. N. Maamar, (Fn. 2), ZUM 2023, 481 (482).

26 Zu den verschiedenen Phasen s. oben II. I.

27 Sich (zumindest auch) auf die urheberrechtlichen Fragen des Output-Vorgangs konzentrierend S. Bußmann/C. Glasowski/M. Niehaus/S. Stecher, (Fn. 13), RDi 2022, 391; E. Filikhina, Urheberrechtliche Schutz von KI-Systemen, DSRITB 2023, 269; F. Krone, Urheberrechtlicher Schutz von ChatGPT-Texten?, RDi 2023, 117; N. Maamar, (Fn. 2), ZUM 2023, 481 (488 f.).

28 Dies ebenfalls betonend J. Siems/T. Repka, Unrechtmäßige Nutzung von KI-Trainingsdaten als Gefahr für neue Geschäftsmodelle? – Compliance-Anforderungen bei der Entwicklung von KI-Systemen, DSRITB 2021, 517.

29 Vgl. D. Bomhard, (Fn. 17), DSRITB 2023, 255 (256).

30 Vgl. etwa J. Drexel/R. M. Hilty/L. Desaunettes/F. Greiner/D. Kim/H. Richter/G. Surblyte/K. Wiedemann, Ausschließlichkeits- und Zugangsrechte an Daten, GRUR Int. 2016, 914 (914 f.). Hingegen können Ausschließlichkeitsrechte an Datenbank(werk)en bestehen, vgl. RL 96/9/EG iVm §§ 4, 87a ff. UrhG.

31 Digitalisierte Werke unterscheiden sich von digitalen Werken dadurch, dass sie ursprünglich in analoger Form vorlagen und erst durch den Digitalisierungsprozess (oftmals durch Text und Data Mining) in digitaler Form vorliegen, vgl. RegE, BR-Drs. 19/27426, 9.3.2021, S. 88.

etwa aus der ausdrücklichen Formulierung in § 44b Abs. 1 UrhG ergibt – niedrigschwellige Verwertungsrechte im nationalen Recht nach Maßgabe der §§ 15 ff. UrhG. Die urheberrechtlich relevante Handlung beim KI-Training wird grundsätzlich darin zu sehen sein, dass im Rahmen der Sammlung der Daten (Phase 1) eine – zumindest temporäre – Speicherung der digitalen sowie digitalisierten Werke etwa in Form von Texten oder Bildern notwendig ist, was rechtlich als zustimmungspflichtige Vervielfältigung iSd §§ 15 Abs. 1 Nr. 1, 16 UrhG zu qualifizieren ist.<sup>32</sup>

## 2. Hauptanwendungsfall: Web-Scraping und Web-Crawling

Da die massenhafte Sammlung von Daten und damit die Nutzung von urheberrechtlich geschützten Werken erforderlich ist, wird die Einholung von Zustimmungen, so beispielsweise in Form von Lizenzierungen (vgl. § 31 Abs. 1 UrhG), aktuell kaum praxistauglich sein; vielfach dürfte der konkrete Urheber bei digital(isiert)en Werken zudem nur schwer zu ermitteln sein.<sup>33</sup> Der praktische Schwerpunkt bei der Befassung mit dem Training von generativer KI wird somit in der Sammlung von Daten durch Web-Scraping<sup>34</sup> und Web-Crawling<sup>35</sup> zu verorten sein.<sup>36</sup>

### a) Regelmäßig keine Anwendung von § 44a UrhG

Mit Blick auf die Automatisierung von Web-Scraping und Web-Crawling kommt grundsätzlich die Anwendbarkeit von Art. 5 Abs. 1 InfoSoc-RL iVm § 44a UrhG in Betracht. Danach sind vorübergehende Vervielfältigungs-handlungen zulässig, die – lediglich – flüchtig oder begleitend sind. Gegen

<sup>32</sup> D. Bomhard, (Fn. 17), DSRITB 2023, 255 (258) mit weiterführenden Ausführungen zu den relevanten Vervielfältigungshandlungen; N. Maamar, (Fn. 2), ZUM 2023, 481 (483); P. J. Pesch/R. Böhme, (Fn. 19), GRUR 2023, 997 (1004).

<sup>33</sup> Vgl. J. Siems/T. Repka, (Fn. 28), DSRITB 2021, 521 (524 f.).

<sup>34</sup> Web-Scraping umfasst den Vorgang, durch den im Internet frei bzw. öffentlich zugängliche Informationen von einer Webseite automatisiert „ausgekratzt“, sprich extrahiert werden, um die extrahierten Informationen sodann durch Analyse, (Weiter-)Verarbeitung oder Verkauf zu nutzen, vgl. B. Paal, (Fn. 17), ZfDR 2023, 325 (327).

<sup>35</sup> Web-Crawling beschreibt den Vorgang des Durchsuchens des Internets, bei dem ein Webcrawler (auch als Spider bzw. Bot bezeichnet) eine Website aufruft und deren Inhalt analysiert, vgl. D. Bomhard, (Fn. 17), DSRITB 2023, 255.

<sup>36</sup> So auch D. Bomhard, (Fn. 17), DSRITB 2023, 255 (256).

die Anwendbarkeit dieser Bestimmungen im klassischen Szenario des KI-Trainings spricht gleichwohl bereits das Merkmal der „vorübergehenden“ Vervielfältigungshandlung, da das KI-Training zumeist auf iterativen Vorgängen beruht.<sup>37</sup> Davon abgesehen dürfte es auch an der flüchtigen bzw. begleitenden Eigenschaft derartiger Vervielfältigungshandlungen vor dem Hintergrund des Schutzzwecks der InfoSoc-RL und der klaren Regel-Ausnahme-Systematik fehlen. Insbesondere die ErwGr. 9-11 InfoSoc-RL betonen das intendierte hohe Schutzniveau des Urheberrechts; zudem verdeutlicht der Wortlaut von Art. 5 Abs. 5 InfoSoc-RL die Ausnahmestellung der Bestimmung („*nur in bestimmten Sonderfällen*“). Schließlich wird auch unter Berücksichtigung des zu Art. 5 Abs. 1 InfoSoc-RL korrelierenden ErwGr. 33 InfoSoc-RL eine Anwendung der Ausnahmevorschrift vorliegend abzulehnen sein, da der Sammlung von Daten als Vorbereitungshandlung des Trainings ein wirtschaftlicher Wert beizumessen ist.

b) Regelmäßige Anwendung von § 44b Abs. 1 UrhG

Aufgrund des regelmäßigen Rückgriffs auf Web-Scraping und -Crawling kommt die Anwendung der sog. Text und Data Mining-Schranke (TDM-Schranke) aus Art. 4 DSM-RL iVm § 44b UrhG in Betracht.<sup>38</sup>

(1) Anwendbarkeit

§ 44b Abs. 1 UrhG versteht Text und Data Mining als die automatisierte Analyse von digital(isiert)en Werken, „*um daraus Informationen [...] zu gewinnen*“. ErwGr. 8 DSM-RL verdeutlicht die hervorgehobene Relevanz von Text und Data Mining, da dieses als „*vorherrschende Technik in der Digitalwirtschaft*“ zur automatisierten Inhaltsanalyse bezeichnet wird. Teilweise wird die Anwendbarkeit des Text- und Data Mining im Kontext von Trainingsdaten gleichwohl mit Blick auf die Tatbestandsvoraussetzung „*um daraus Informationen [...] zu gewinnen*“ zumindest angezweifelt, da die Sammlung von Daten und der Trainingsvorgang der gescrapten bzw. ge-crawlten Daten nicht unmittelbar auf Informationsgewinnung, sondern auf

---

37 P. J. Pesch/R. Böhme, (Fn. 19), GRUR 2023, 997 (1006).

38 Vgl. K. de la Durantaye, (Fn. 5), ZUM 2023, 645 (651).

die Generierung eines Werks gerichtet sei.<sup>39</sup> Dem ist die (vor allem) in den Erwägungsgründen der DSM-RL hervortretende Intention der Schaffung von Innovationsanreizen (vgl. etwa ErwGr. 18 UAbs. 1 S. 4 DSM-RL) entgegenzuhalten. So weist ErwGr. 18 UAbs. 1 S. 1 DSM-RL ausdrücklich darauf hin, dass Text und Data Mining auch der „*Entwicklung neuer Anwendungen oder Technologien*“ dient.<sup>40</sup> Vor diesem Hintergrund dürfte im Kontext von KI-Trainingsdaten daher auch eine Informationsgewinnung iSd § 44b Abs. 1 UrhG anzunehmen sein, wenn diese durch automatisierte Analyse einen (ggf. auch nur systeminhärenten und damit nicht nachverfolgbaren) Zwischenschritt zur bezweckten Generierung eines Inhalts darstellt.<sup>41</sup>

## (2) Voraussetzungen und Rechtsfolgen

Die Erlaubnis des Text und Data Mining ist durch Art. 4 Abs. 1 DSM-RL iVm § 44b Abs. 2 S. 1 UrhG auf „rechtmäßig zugängliche Werke“ beschränkt, womit § 44b UrhG keinen Zugangsanspruch begründet, sondern diesen vielmehr voraussetzt.<sup>42</sup> Die Gesetzesbegründung zu § 44b UrhG führt hierzu aus, dass beispielsweise im Internet frei zugängliche Werke oder eine Lizenzierung in Form einer Open-Access-Veröffentlichung rechtmäßig zugänglich sind.<sup>43</sup> Ausgeschlossen ist demgegenüber insbesondere das Überwinden von Sicherheitsvorkehrungen, etwa in Form von Hacking.<sup>44</sup> Weiterhin ist für die Erlaubnis des gewerblichen<sup>45</sup> Text und Data Mining nach Art. 4 Abs. 3 DSM-RL iVm § 44b Abs. 3 UrhG erforderlich, dass der Rechteinhaber keinen ausdrücklichen Nutzungsvorbehalt ausgesprochen hat, wobei dies bei online veröffentlichten Inhalten in maschinenlesbarer Form erfolgen muss. Eine nähere Definition des Begriffs der „maschinenlesbaren Form“ erfolgt weder durch unionale noch nationale Gesetzgeber; allerdings beschreibt ErwGr. 21 RL 2013/37/EU (in einem

39 So insbesondere *J. Pukas*, (Fn. 2), GRUR 2023, 614 (615) mit weiterführender Argumentation.

40 *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (651).

41 Im Ergebnis so auch *D. Bomhard*, (Fn. 17), DSRITB 2023, 255 (259 f.); *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (651).

42 So auch *T. Dreier*, in: Dreier/Schulze, UrhG, 7. Auflage 2022, § 44b Rn. 8.

43 Vgl. RegE, BR-Drs. 19/27426, 9.3.2021, S. 88.

44 Zur notwendigen Abgrenzung von Hacking und Scraping *B. Paal*, (Fn. 17), ZfDR 2023, 325 (327).

45 Nach § 60d UrhG ist TDM zur wissenschaftlichen Forschung umfassend möglich und unterliegt nicht dem Nutzungsvorbehalt aus § 44b Abs. 3 UrhG.

anderen Zusammenhang) ein Dokument als maschinenlesbar, „wenn es in einem Dateiformat vorliegt, das so strukturiert ist, dass Softwareanwendungen die konkreten Daten einfach identifizieren, erkennen und extrahieren können“. Die Begriffsbestimmung ist dynamisch ausgestaltet und orientiert sich an dem aktuellen Stand der Technik. Gängig sind somit jedenfalls Formate wie XML, CSV oder txt, wobei zur Verhinderung von Data-Screening insbesondere auf eine *robot.txt*-Anweisung zurückgegriffen wird.<sup>46</sup> Der unionale Gesetzgeber hat sich somit mit Blick auf die intendierte Innovation (vgl. ErwGr. 18 UAbs. 1 S. 4 DSM-RL) für ein *Opt-Out*-Modell entschieden. Schließlich ist die Erlaubnis des Text und Data Mining zeitlich darauf beschränkt, dass nach Art. 4 Abs. 2 DSM-RL iVm § 44b Abs. 2 S. 2 UrhG eine Aufbewahrung nur für die Notwendigkeit des jeweiligen Zweckes eröffnet ist. Aufgrund der iterativen Trainingsvorgänge und der Notwendigkeit sowohl einer technischen Dokumentation als auch einer Kontrolle des Outputs wird die zeitliche Erforderlichkeit im Rahmen des KI-Trainings vielfach gegeben sein, sodass die Löschpflicht „*faktisch immer mehr verwässert*“.<sup>47</sup>

### 3. Chancen und Grenzen des Status quo

Mit Blick auf die vorangegangenen Ausführungen und Ergebnisse ist festzustellen, dass die urheberrechtlichen Instrumentarien *de lege lata* den Regelfall des KI-Trainings abdecken. Zugleich ist deutlich geworden, dass die Ausbalancierung des Interessenkonflikts von Rechteinhabern und KI-Betreibern in der bisherigen Gesetzeslage durch die seinerzeit kaum absehbare disruptive Entwicklung von KI nicht (mehr) dem aktuellen Stand von Technik und praktischen Herausforderungen entspricht. Vor diesem Hintergrund ist zumindest eine Evaluation des Interessenausgleichs veranlasst, was die Erforderlichkeit einer Bewertung der Chancen und Grenzen des Status quo miteinschließt.

---

<sup>46</sup> Dazu weiterführend *D. Bomhard*, (Fn. 17), DSRITB 2023, 255 (267).

<sup>47</sup> Treffend und weiterführend *D. Bomhard*, (Fn. 17), DSRITB 2023, 255 (262 ff.).

### a) Begrüßenswerte Rechtssicherheit

Zu begrüßen ist grundsätzlich, dass durch die klare Regel-Ausnahme-Systematik und die in §§ 15 Abs. 1 Nr. 1, 16 bzw. § 44b UrhG niedergelegten Voraussetzungen ein für Anwender rechtssichere(re)s System besteht, was im Grundsatz als innovationsfördernd anzusehen ist. Im (Rechts-)Vergleich zu der zwar flexiblen, dafür aber im Einzelfall schwer vorhersehbaren „fair-use“-Doktrin aus 17 U.S.C. § 107 dürften insoweit Vorteile des unionalen Regelungsregimes bestehen.<sup>48</sup>

### b) Praktische Nachteile

Das bestehende Modell aus niedrigschwelligen Zustimmungserfordernissen aus §§ 15 Abs. 1 Nr. 1, 16 UrhG auf der einen und die *opt-out*-Regelung aus § 44b UrhG auf der anderen Seite führt allerdings auch zu signifikanten praktischen Nachteilen. Zunächst ist insoweit anzuführen, dass bislang (noch) keine standardisierte maschinenlesbare Form des Nutzungsvorbehalts für digital(isiert)e Werke iSD § 44b Abs. 3 UrhG existiert. Daher greifen betroffene Unternehmen vermehrt auf *robot.txt*-Anweisungen zurück, die aber vielfach nicht nur Text und Data Mining verhindern, sondern darüber hinaus etwa auch das für die Auffindbarkeit und damit die Monetarisierung von digital(isiert)en Werken existenzielle Ranking<sup>49</sup> beeinträchtigen. Denn Suchmaschinen verwenden regelmäßig eigene Crawler und Bots, die dann ebenfalls von der Auswertung der Inhalte ausgeschlossen werden.<sup>50</sup> Ferner resultieren praktische Schwierigkeiten daraus, dass der Nutzungsvorbehalt aus § 44b Abs. 3 UrhG nach gesetzgeberischer Intention nur *ex-nunc*-Wirkung entfaltet,<sup>51</sup> sodass insbesondere bei vorher analog veröffentlichten digitalisierten Werken die flächendeckende Mitteilung des

<sup>48</sup> Näher auf das US-amerikanische Urheberrecht bezüglich KI-Trainingsdaten eingehend und die Rechtsunsicherheit der „fair-use“-Doktrin betonend *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (648 f.).

<sup>49</sup> Zur wettbewerbsrechtlichen Bedeutung etwa *S. Ott*, Ich will hier rein! Suchmaschinen und das Kartellrecht, MMR 2006, 195.

<sup>50</sup> Hierzu im Ganzen *D. Bomhard*, (Fn. 17), DSRITB 2023, 255 (267 f.) mit einem Beispiel, wie zumindest einzelne Bots erlaubt werden können; *N. Maamar*, (Fn. 2), ZUM 2023, 481 (484) ebenfalls mit Beispiel, wie die fehlende Auffindbarkeit verhindert werden kann.

<sup>51</sup> Vgl. RegE, BR-Drs. 19/27426, 9.3.2021, S. 89.

nunmehr erklärten Vorbehalts kaum möglich ist. Weitere praktische Nachteile bestehen dahingehend, dass aus ausgeübten Nutzungsvorbehalten iSd § 44b Abs. 3 UrhG zum einen hohe Hürden für die dann ggf. notwendige massenhafte Lizenzierung resultieren. Denn zu einem effektiven KI-Training werden riesige Datenmengen benötigt, wobei regelmäßig verschiedene Rechteinhaber betroffen sein werden. Da im Einzelfall nur schwer zu ermitteln sein wird, wer Rechteinhaber ist und zudem kein gutgläubiger Rechtserwerb eröffnet ist, dürfte die Lizenzierung oftmals eine erhebliche praktische Hürde darstellen.<sup>52</sup> Zum anderen werden bei ausgeübtem Nutzungsvorbehalt zumeist gravierende Schwierigkeiten in der Beweisführung bezüglich einer Vervielfältigung für Rechteinhaber im Rahmen der Sammlung von Trainingsdaten bestehen,<sup>53</sup> dies gilt vor allem innerhalb des intransparenten Trainingsprozesses.<sup>54</sup>

### c) Begrenzte Wirksamkeit des Territorialitätsprinzips

Schließlich bestehen systeminduzierte Grenzen des Status quo auch aufgrund des Territorialitätsprinzips des europäischen und nationalen Urheberrechts. So hat der BGH im *Staatsgeschenk*-Urteil aus dem Jahr 2007 klargestellt, dass es auf den Ort der Verletzung der maßgeblichen urheberrechtlichen Nutzungsrechte ankommt.<sup>55</sup> Relevant ist im Rahmen des KI-Trainings regelmäßig bereits die vorgelagerte Speicherung als Verwertungshandlung iSd §§ 15 Abs. 1 Nr. 1, 16 UrhG, welche vielfach gerade außerhalb des unionalen Anwendungsbereichs in den USA oder Asien erfolgen wird.<sup>56</sup> Aus der räumlichen Begrenzung leitet sich nicht nur eine eingeschränkte Wirksamkeit des unionalen Urheberrechts ab, sondern darüber

---

52 Vgl. J. Siems/T. Repka, (Fn. 28), DSRITB 2021, 521 (524 f.); D. Bomhard, (Fn. 19), DSRITB 2023, 255 (259) auf die fehlende Möglichkeit des gutgläubigen Erwerbs hinweisend.

53 Insoweit gibt es das nützliche Browser-Tool *HaveIBeenTrained*, abrufbar unter <https://haveibeentrained.com/> (zuletzt abgerufen am 30.6.2024), welches es ermöglicht, zu überprüfen, ob ein Text oder ein Bild etwa in der Datenbank *Laion-5b* ist.

54 Vgl. K. de la Durantaye, (Fn. 5), ZUM 2023, 645 (651 f.); P. J. Pesch/R. Böhme, (Fn. 21), GRUR 2023, 997 (1007).

55 BGH, Urt. v. 24.5.2007 – I ZR 42/04, ZUM 2007, 644 (645) Rn. 22 – *Staatsgeschenk*.

56 Im Ganzen hierzu N. Maamar, (Fn. 2), ZUM 2023, 481 (486).

hinaus droht auch eine innovationsbehindernde Schwächung unionaler KI-Trainings-Märkte gegenüber EU-ausländischen Anbietern.<sup>57</sup>

#### IV. Implikationen durch die KI-Verordnung – Chancen und systeminduzierte Grenzen

Am 8. Dezember 2023 kam es im Trilog nach einer rund 37-stündigen, als „ultramarathon session“ bezeichneten Verhandlung zu einer politischen Einigung über eine KI-Verordnung, die EU-Binnenmarktkommissar *Thierry Breton* als „historisch“ bezeichnete, da die EU als erster Kontinent klare Regeln zur Nutzung von KI aufgestellt habe.<sup>58</sup> Das Europäische Parlament verabschiedete die KI-VO bzw. das „Gesetz über künstliche Intelligenz“ im März 2024, bevor sodann im Mai 2024 die finale Zustimmung des Rats der Europäischen Union erfolgte.

##### 1. Inhalt: Art. 53 Abs. 1 lit. b bis d KI-VO

Die EU-Kommission veröffentlichte den Entwurf der KI-Verordnung im Februar 2022 und damit zu einem Zeitpunkt, an dem die disruptive Entwicklung generativer KI in der jetzigen Ausgestaltung weder quantitativ noch qualitativ absehbar war. Vor diesem Hintergrund adressierte der nunmehr erfolgte, seinerzeitige Entwurf nicht (hinreichend) etwaige Herausforderungen bezüglich KI-Trainingsdaten; vielmehr fokussierte die Transparenzpflicht für Hochrisiko-KI-Systeme nach Art. 13 Abs. 1 des Kommissionsentwurfs<sup>59</sup> (nachfolgend Komm-E) den Ergebnisvorgang.

Demgegenüber griff bereits der Mitte Juni 2023 veröffentlichte Entwurf des EU-Parlaments<sup>60</sup> (nachfolgend Parl-E) das untersuchungsgegenständliche urheberrechtliche Spannungsverhältnis zwischen Rechteinhabern und KI-Betreibern auf. So war in Art. 28b Abs. 4 lit. b Parl-E für sog. Basismodelle, die für generative KI-Systeme verwendet werden, die Verpflichtung vorgesehen, das Basismodell so zu konzipieren, dass – dem Stand der

---

57 Darauf hinweisend *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (652 f.).

58 EU-Kommission, Statement by Commissioner Breton – The European AI Act is here!, 8.12.2023, abrufbar unter [https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT\\_23\\_6471](https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_23_6471) (zuletzt abgerufen am 30.6.2024).

59 EU-Kommission, Vorschlag für eine KI-Verordnung – COM(2021)206 final, 21.4.2021.

60 EU-Parlament, Artificial Intelligence Act – P9\_TA(2023)0236, 14.6.2023.

Technik entsprechende – Maßnahmen zur Verhinderung von Verstößen gegen das EU-Recht und damit auch das Urheberrecht getroffen werden. Ferner enthielt Art. 28b Abs. 4 lit. c Parl-E eine Transparenzverpflichtung, ausreichend detaillierte Zusammenfassungen („*sufficiently detailed summary*“) zu der Nutzung von Inhalten im Trainingsvorgang zu veröffentlichen. Den Begriff der Basismodelle verstand Art. 3 Abs. 1 Nr. 1 lit. c Parl-E als ein „*KI-Systemmodell, das auf einer breiten Datenbasis trainiert wurde, auf eine allgemeine Ausgabe auslegt ist und an eine breite Palette unterschiedlicher Aufgaben angepasst werden kann*“.<sup>61</sup> Bei Verstößen gegen die Verpflichtungen für Basismodelle drohten nach Art. 71 Abs. 4 Parl-E Geldbußen bis zu 10 Mio. EUR oder bis zu 2 % des gesamten Jahresumsatzes des vorangegangenen Geschäftsjahrs.

Im Trilog-Verfahren konnte sich das EU-Parlament sodann politisch weitgehend durchsetzen, gleichwohl erfolgten im Verlauf der sich daran anschließenden technischen Verhandlungen weitere Anpassungen, die auch und gerade urheberrechtliche Relevanz aufweisen: Der Begriff des Basismodells ging funktional und inhaltlich vergleichbar in der Legaldefinition von einem „*KI-Modell mit allgemeinem Verwendungszweck*“ in Art. 3 Nr. 63 KI-VO auf, das als „*KI-Modell – einschließlich der Fälle, in denen ein solches KI-Modell mit einer großen Datenmenge unter umfassender Selbstüberwachung trainiert wird, eine erhebliche allgemeine Verwendbarkeit aufweist und in der Lage ist, unabhängig von der Art und Weise seines Inverkehrbringens ein breites Spektrum unterschiedlicher Aufgaben kompetent zu erfüllen, und das in eine Vielzahl nachgelagerter Systeme oder Anwendungen integriert werden kann, ausgenommen KI-Modelle, die vor ihrem Inverkehrbringen für Forschungs- und Entwicklungstätigkeiten oder die Konzipierung von Prototypen eingesetzt werden*,“ definiert und durch ErwGr. 97 ff. KI-VO konturiert ist. Zudem wurde die Verpflichtung zur EU-rechtskonformen Konzeption aus Art. 28b Abs. 4 lit. c Parl-E inhaltlich weitgehend übereinstimmend in Art. 53 Abs. 1 lit. b KI-VO übertragen und hinsichtlich der technischen Umsetzung konturiert. So müssen nach Art. 53 Abs. 1 lit. b (ii) KI-VO mindestens die in Anhang XII aufgeführten Informationen wie Aufgaben, Architektur, Lizenz oder Modalitäten des KI-Modells Anbietern von KI-Systemen zur Verfügung gestellt werden, die eine Implementierung des betroffenen KI-Modells beabsichtigen. Auch die

---

<sup>61</sup> In der englischen Fassung: „*[A]n AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks*“.

Transparenzverpflichtung des Art. 28 Abs. 4 lit. d Parl-E wurde in Art. 53 Abs. 1 lit. d KI-VO weitgehend überstimmend übertragen, wonach eine „hinreichend detaillierte Zusammenfassung“ des KI-Trainings erforderlich und dem sog. Büro für Künstliche Intelligenz der EU-Kommission vorzulegen ist. Neu hinzugefügt wurde Art. 53 Abs. 1 lit. c KI-VO, der anordnet, dass Anbieter von KI-Modellen mit allgemeinem Verwendungszweck eine Strategie zur Einhaltung des Urheberrechts der Union entwickeln und hierbei insbesondere auf die Einhaltung des Nutzungsvorbehals aus Art. 4 Abs. 3 DSM-RL achten müssen. Eine Abweichung von Art. 71 Abs. 4 Parl-E enthält hingegen Art. 99 KI-VO, der in den in Abs. 3 u. 4 aufgeführten bußgeldbewehrten Vorschriften nicht Art. 53 KI-VO nennt. Vielmehr kann ausschließlich die Europäische Kommission durch das neu zu schaffende Büro für Künstliche Intelligenz nach Art. 93 Abs. 1 KI-VO die Anbieter der KI-Modelle zu geeigneten Umsetzungsmaßnahmen (lit. a) sowie Risikominderungsmaßnahmen verpflichten (lit. b) und sogar die Bereitstellung des Modells auf dem Markt einschränken bzw. zurücknehmen (lit. c). Vor einer Aufforderung zu entsprechenden Maßnahmen hat das Büro für Künstliche Intelligenz ein Ermessen („kann“), ob es einen regulatorischen Dialog mit dem jeweiligen Anbieter einleitet (Art. 93 Abs. 2 KI-VO). Darüber hinaus kann das Büro für Künstliche Intelligenz nach Art. 93 Abs. 3 KI-VO Verpflichtungszusagen mit Bindungswirkung annehmen.

## 2. Bewertung: Ambivalenz durch systemimmanente Grenzen des produkthaftungsrechtlichen Ansatzes

Die KI-Verordnung verspricht begrüßenswerte Vorteile und geht zugleich mit systemimmanenten Grenzen einher.

### a) Territorialer Anwendungsbereich

Zu begrüßen ist zunächst, dass nach Art. 2 Abs. 1 lit. a KI-VO das Regelungswerk unabhängig von der Niederlassung der KI-Anbieter zur Anwendung kommt, sofern die KI-Systeme in der Union verwendet werden. Der Rückgriff auf das sog. „Marktortprinzip“<sup>62</sup> – vergleichbar zu Art. 1 Abs. 2 DMA oder Art. 3 Abs. 2 DS-GVO – führt insoweit zu einer Verbesserung

---

<sup>62</sup> So etwa *B. Paal/L. Kumkar*, Die digitale Zukunft Europas, ZfDR 2021, 97 (110).

der effektiven Anwendung der KI-Verordnung. So müssen auch US-amerikanische oder asiatische Betreiber damit die Strategie zur Einhaltung des Urheberrechts der Union und die Transparenzverpflichtungen aus Art. 53 Abs. 1 lit. c und d KI-VO erfüllen. Allerdings zeigen sich an dieser Stelle bereits systemimmanente Grenzen der KI-Verordnung: Grundsätzlich ist die KI-Verordnung als risikobasierte Produkthaftungsregulierung ausgestaltet, die je nach erwartetem Risiko – vergleichbar zum Digital Services Act<sup>63</sup> – gestuft reguliert und den Adressaten risikobegrenzende Verpflichtungen auferlegt. Hierbei erweisen sich Art. 53 Abs. 1 lit. b und c KI-VO als „*Fremdkörper*“<sup>64</sup>, da die Bestimmung primär nicht Produktrisiken begrenzen soll, sondern im Konflikt zwischen Rechteinhabern und KI-Betreibern zugunsten der Rechteinhaber ein Enforcement ermöglichen soll.<sup>65</sup> Der weitreichende territoriale Anwendungsbereich der Transparenzregel zur Durchsetzung von Rechten hat aber nur begrenzte Wirkung, wenn das durchzusetzende (Urheber-)Recht *de lege lata* auf den Ort der relevanten Verletzungshandlung abstellt, die oftmals nicht in der EU stattfindet.<sup>66</sup> Somit droht durch Art. 53 Abs. 1 lit. b KI-VO ein wettbewerbsrelevante Schwächung unionaler KI-Betreiber, solange nicht hinsichtlich der Sammlung der Trainingsdaten und des Trainingsvorgangs urheberrechtlich ein „Marktortprinzip“ in Bezug auf unional geschützte digital(isiert)e Werke implementiert wird. Dies gilt insbesondere vor dem Hintergrund, dass im Rahmen des Trainingsprozesses im EU-Ausland unverändert kaum die urheberrechtliche Verwertung von unional geschützten digital(isiert)en Werken hinreichend abgesichert werden kann. Das Problem hat der unionale Gesetzgeber zwar erkannt und in ErwGr. 106 KI-VO explizit adressiert, indem er die Sicherstellung gleicher Wettbewerbsbedingungen durch die Anwendung von Art. 53 Abs. 1 lit. b und c KI-VO auch auf EU-ausländische Anbieter von KI-Modellen erstreckt. Gleichwohl verhindert aber auch die Einführung einer verpflichtenden Strategie zur Einhaltung des Urheberrechts der Union jedenfalls nicht, dass bereits fertige Produkte ohne Anwendung des Urheberrechts auf den im Vorfeld abgeschlossenen Trainingsvorgang auf den Markt gelangen können und die KI-VO erst mit ex-nunc-Wirkung durch den Markteintritt iSd Art. 2 Abs. 1 lit. a KI-VO umzusetzen ist.

---

63 Vgl. B. Paal/F. Kieß, Digitale Plattformen im DSA-E, DMA-E und § 19a GWB, ZfDR 2022, 1 (13).

64 Treffend K. de la Durantaye, (Fn. 5), ZUM 2023, 645 (656).

65 In diese Richtung auch K. de la Durantaye, (Fn. 5), ZUM 2023, 645 (655 f.).

66 Vgl. oben III. 3. c.

### b) Verbesserte Durchsetzungsmöglichkeiten

Im (begrenzten) Anwendungsbereich dürfte Art. Art. 53 Abs. 1 lit. a-c KI-VO dazu führen, dass eine Stärkung der bestehenden Durchsetzungsmöglichkeiten bezüglich der Einhaltung bzw. Verletzung der Vorgaben des Urheberrechts erfolgt. Der grundsätzlich niedrigschwellige Anwendungsbereich durch den Rückgriff auf sog. Basismodelle führt insoweit zu einer flächendeckende(re)n sachlichen Anwendung der für das Urheberrecht relevanten Pflichten. Die notwendige Implementierung von Designmaßnahmen nach Art. 53 Abs. 1 lit. b KI-VO zur Einhaltung auch des Urheberrechts und damit etwa auch zur Überprüfung von Nutzungs- vorbehalten iSd § 44b Abs. 3 UrhG dürfte in Kombination mit dem Umsetzungsmechanismus aus Art. 93 KI-VO dazu führen, dahingehend derzeit bestehende Defizite<sup>67</sup> zu reduzieren. Insoweit erscheint es nachvollziehbar, dass aufgrund des (zumindest auch) wettbewerblichen Bezugs des Art. 53 Abs. 1 lit. b-d KI-VO in Art. 93 KI-VO eine Überwachung nur durch die Europäische Kommission mit aus dem Kartellrecht und dem DMA bekannten Instrumentarien erfolgt. So sind die in Art. 93 KI-VO geregelten Instrumentarien der (auch strukturellen) Abhilfemaßnahmen (vgl. Art. 7 VO 1/2003, Art. 18 DMA), des fakultativen „regulatory dialogue“ (vgl. Art. 6 iVm Art. 8 Abs. 3 DMA) und der Verpflichtungszusagen (vgl. Art. 9 VO 1/2003 und Art. 25 DMA) etablierte Instrumente des europäischen Wettbewerbs- bzw. Regulierungsrechts. Schließlich sind Transparenzinstrumente wie die Strategie zur Einhaltung des Urheberrechts der Union und die Veröffentlichung einer Zusammenfassung der verwendeten Trainingsdaten nach Art. 53 Abs. 1 lit. c und d KI-VO grundsätzlich dazu geeignet, die Sanktionierung von Urheberrechtsverletzungen zu verbessern.

### c) Rechtsvergleichender Vorteil der Rechtssicherheit durch Bestimmtheit?

An der Umsetzung des Art. 28b Abs. 4 lit. b und c Parl-E war grundlegend zu kritisieren, dass er auf (zu) unbestimmten und auslegungsbedürftigen Rechtsbegriffen basierte. Dies zeigte sich bereits an dem Begriff der Basismodelle aus Art. 3 Abs. 1 Nr. 1 lit. c Parl-E: Die maßgeblichen Parameter der „breiten Datenbasis“ („broad data“) und „der „breite[n] Palette unterschied-

---

<sup>67</sup> Auf das Defizit hinweisend P.J. Pesch/R. Böhme, (Fn. 19), GRUR 2023, 997 (1003 ff.).

*licher Aufgaben*“ („wide range of distinctive tasks“) wurden im Parl-E nicht näher präzisiert,<sup>68</sup> waren aber gerade grundlegend für die Anwendung der weitreichenden Vorschriften. Insoweit ist es zu begrüßen, dass die Definition von KI-Modellen mit allgemeinem Verwendungszweck in Art. 3 Nr. 63 KI-VO deutlich detaillierter ausfällt und umfangreich durch ErwGr. 97 ff. KI-VO konturiert wird. Insbesondere die Nennung von mindestens einer Milliarde Parameter als Merkmals der allgemeinen Verwendung in ErwGr. 98 KI-VO dient als nützliche Orientierung für potenziell betroffene KI-Modell-Anbieter.

Erhebliche Bestimmtheitsdefizite zeigten sich auch bei der Transparenzverpflichtung aus Art. 28b Abs. 4 lit. c Parl-E auf, da der Begriff der „*hinsichtlich detaillierten Zusammenfassung*“ („sufficiently detailed summary“) auslegungsbedürftig ist und durch den Parl-E nicht näher definiert wurde. Mit Blick auf die Defizite begrüßenswert ist daher die Vorgabe abstrakter Leitlinien der Begriffsauslegung durch ErwGr. 107 KI-VO und der Umstand, dass das Büro für Künstliche Intelligenz in den zu erarbeitenden Praxisleitfäden iSd Art. 56 Abs. 1 KI-VO insbesondere Vorlagen für die angemessene Detailgenauigkeit der Zusammenfassung erstellen soll (vgl. Art. 56 Abs. 2 lit. b iVm ErwGr. 107 KI-VO).

Die im Gesetzgebungsverfahren erhöhte Bestimmtheit kann dazu beitragen, die Vorteile aus dem bestehenden bestimmten und damit rechtssicheren Urheberrecht<sup>69</sup> gegenüber flexibl(er)en Rechtsordnungen – wie insbesondere der USA – zu gewährleisten.

## V. Zusammenfassung und Ausblick

Grundsätzlich ist die KI-Verordnung mit Blick auf ihr Alleinstellungsmerkmal und der möglichen Vorbildfunktion im Allgemeinen sowie überdies auch bezüglich der vereinzelten urheberrechtlichen Regelungen im Besonderen zu begrüßen. Technisch-organisatorische Maßnahmen und Transparenzpflichten, wie konkret in Art. 53 Abs. 1 lit. b-d KI-VO, sind geeignete Instrumente zur Förderung der Einhaltung und Durchsetzung des bestehenden Urheberrechtsschutzes.

---

<sup>68</sup> Vgl. auch *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (654).

<sup>69</sup> Vgl. hierzu III. 3. a.

## 1. Notwendigkeit einer Evaluation

Insgesamt ist der Entwurf der KI-VO – vor allem durch die produktsicherheitsrechtliche Konzeption des Regelungswerkes – systemimmanent nicht geeignet, den vorliegend untersuchungsgegenständlichen urheberrechtlichen Konflikt bei generativer KI als „*Werkzeug der Zukunft*“<sup>70</sup> zwischen Innovationspotenzial und Urheberschutz als Mittel zur „*Erhaltung und Entwicklung kreativer Tätigkeit*“ (vgl. ErwGr. 10 S.1 RL 2001/29/RG) sachgerecht aufzulösen. Da insbesondere die Text und Data Mining-Schranke aus Art. 4 DSM-RL iVm § 44b UrhG als Ausprägung eines Interessenausgleichs tendenziell zugunsten der Innovation (vgl. ErwGr. 18 UAbs. 1 S. 4 DSM-RL) noch vor der disruptiven Entwicklung generativer KI konzipiert und implementiert wurde, sollten unionale und nationale Gesetzgeber die Ausbalancierung der Interessen neu gewichten bzw. jedenfalls evaluieren.

## 2. Interessensausgleich de lege ferenda

Hierbei ist ein behutsames und abgewogenes Vorgehen auch vor dem Hintergrund geboten, dass eine überschießende Regulierung (iSe Over-enforcement) zur Behinderung der Innovation und eine Unterregulierung zum Niedergang ganzer Sektoren führen kann. Ein Königsweg zur gerechten Auflösung der betroffenen Interessen ist bislang zwar nicht ersichtlich. Fest steht aber jedenfalls, dass aus jedem regulierenden Eingriff ein Einfluss auf die Balance zwischen Innovation und Urheberschutz erwächst. Sofern aus der vorzunehmenden Evaluation der Gesetzeslage ein Entschluss zu einem stärker ausgestalteten Urheberrechtsschutz hervorgehen sollte, könnte im Bereich des KI-Trainings zunächst eine Ausnahme von Art. 4 DSM-RL iVm § 44b UrhG statuiert werden, wodurch faktisch ein *opt-in*-Modell entstehen würde. Denkbar ist zudem die Implementierung eines Vergütungsanspruchs kumulativ oder alternativ zum Vorbehalt aus § 44b Abs. 3 UrhG, wobei hier aber zu berücksichtigen ist, dass Urheber unter Umständen ein Interesse daran haben, dass gar keine Nutzung ihrer Inhalte erfolgt.<sup>71</sup> Sollte demgegenüber ein stärkerer Investitionsschutz bezieht werden, kommt vergleichbar zu Art. 3 DSM-RL iVm § 60d UrhG ein Rückgriff auf eine vorbehaltsunabhängige, vergütungsfreie Schrankennutzung

---

<sup>70</sup> So S. Bußmann/C. Glasowski/M. Niehaus/S. Stecher, (Fn. 13), RDI 2022, 391.

<sup>71</sup> Vgl. K. de la Durantaye, (Fn. 5), ZUM 2023, 645 (658).

in Betracht, wodurch aber ein Rückgang kostenfreier Online-Inhalte zu befürchten wäre.<sup>72</sup> Sofern eine Evaluation schließlich dazu führt, dass die aktuelle Interessengewichtung als sachgerecht erachtet wird, sollten neben Transparenzverpflichtungen weitere tradierte Instrumente wie die kollektive Lizenzvergabe (vgl. §§ 51 ff. VGG)<sup>73</sup> zur Verbesserung der Durchsetzungsmöglichkeiten der Rechteinhaber in den Blick genommen werden.

### 3. Fokus auf Wettbewerb der Rechtsordnungen

Unbeschadet dessen sollten auch und gerade mit Blick auf den Wettbewerb der Rechtsordnungen<sup>74</sup> zugunsten des Innovationspotenzials generativer KI angemessene Maßnahmen getroffen werden. Das Alleinstellungsmerkmal einer umfassenden KI-Regulierung kann durch Rechtssicherheit verbesserte Rahmenbedingungen für Innovation gegenüber einer flexiblen und somit schwer vorhersehbaren „rule of reason“ wie die US-amerikanische *fair-use*-Doktrin bewirken. Rechtssicherheit durch klare Regeln könnte insoweit gleichsam ein „unique selling point“ des unionalen Urheberrechts sein. Mit Blick darauf sind die im Gesetzgebungsverfahren erfolgten Konturierungen zu begrüßen, wenngleich etwa die Aufnahme von quantitativen Schwellenwerten mit Vermutungswirkung – vergleichbar zu Art. 3 Abs. 2 DMA oder Art. 33 Abs. 1 DAS – zur Bestimmung von KI-Modellen mit allgemeinem Verwendungszweck über ErwGr. 98 KI-VO hinaus eine noch größere Rechtssicherheit zu bewirken versprechen würde.

Ferner könnte durch die Implementierung eines „Marktortprinzips“ – zumindest im Anwendungsbereich generativer KI – im unionalen Urheberrecht noch stärker darauf hingewirkt werden, dass unionale KI-Betreiber keine Nachteile gegenüber EU-ausländischen KI-Betreibern zu vergegenwärtigen haben. Zu erwägen ist schließlich eine „große Lösung“ durch Schaffung einer digitalen Urheberrechtsverordnung, um das „level playing field“ in der Union zu stärken und eine Kohärenz zu interdependenten

---

72 So *K. de la Durantaye*, (Fn. 5), ZUM 2023, 645 (658).

73 Hierzu insbesondere *J. Pukas*, (Fn. 2), GRUR 2023, 614.

74 Bezüglich des Wettbewerbs von Rechtsordnungen im Datenschutzrecht *B. Paal*, Kollisionsrecht und Wettbewerb der Rechtsordnungen im Datenschutzrecht, in: FS Ebke, Deutsches, Europäisches und Vergleichendes Wirtschaftsrecht, 2022, S. 769 ff.

Gesetzgebungsakten wie DS-GVO<sup>75</sup> oder DMA<sup>76</sup> zu befördern. In der Gesamtwürdigung bleibt also noch viel zu tun und der Jubilar *Georgios Gounalakis* wird sich mit seiner gewichtigen Stimme in die Diskussion einbringen, auch und gerade mit dem Blick auf die Bedeutung der KI für den von ihm (sowohl wissenschaftlich als auch in der praktischen Anwendung) intensiv bearbeiteten Medienbereich – herzliche Geburtstagsglückwünsche und *ad multos annos!*

---

75 Hierzu etwa *L. Franke*, Datenschutzkonformes Training von KI-Systemen mit öffentlich verfügbaren personenbezogenen Daten.

76 Vgl. etwa *P. Hacker*, KI und DMA – Zugang, Transparenz und Fairness für KI-Modelle in der digitalen Wirtschaft, GRUR 2022, 1278.

