

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

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

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

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

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

---

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

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

### I. Anwendbares Recht

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

#### 1. Hinreichender Inlandsbezug der Speicherung zur Anwendung deutschen Urheberechts

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

---

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

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

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

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

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

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

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

---

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

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

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

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

## A. Speicherung von Trainingsdaten aus Internetquellen mit Hilfe von Webcrawlern

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

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

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

---

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

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

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

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

#### *II. Eingriff in das urheberrechtliche Ausschließlichkeitsrecht*

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

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

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

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

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

---

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

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

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

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

### a) Vorübergehender Charakter der Vervielfältigung

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

---

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

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

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

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

318 Auch Käde, CR 2024, 599 Rn. 17.

319 Dazu bereits siehe oben unter Fn. 203.

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

b) Eigene wirtschaftliche Bedeutung der Vervielfältigung

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

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

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

---

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

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

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

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

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

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

---

325 A. A. Leistner, GRUR 2024, 1665 (1666); Schack, NJW 2024, 113 Rn. 6. Ebenfalls keine eigenständige wirtschaftliche Bedeutung allein durch die spätere Anwendung der Analyseergebnisse sehen wohl Spindler, GRUR 2016, II112 (II115); Raue, ZUM 2021, 793 (795).

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

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

328 Wie hier auch Heinze/Wendorf, in: Ebers/Heinze/Krüger/Steinrötter, Künstliche Intelligenz und Robotik, § 9 KI und Urheberrecht Rn. 17; Dregelies, GRUR 2024, 1484 (1485); ähnlich außerdem Spindler, GRUR 2016, II112 (II115).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

Durchführung des Trainingsvorgangs eine darüberhinausgehende, eigene wirtschaftliche Bedeutung zu.<sup>329</sup>

#### c) Schlussfolgerung zur Anwendbarkeit der Schrankenbestimmung

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

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

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

---

329 Vgl. zur Voraussetzung EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 175, 177 – Football Association Premier League; EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 50 – Infopaq II; im Kontext nicht KI-bezogenen Text und Data-Minings halten das auch *Spindler*, GRUR 2016, III12 (III16); *Spindler*, ZGE 2018, 273 (278); *Triaille/Meeùs d'Argenteuil/Francquen*, Study on the legal framework of text and data mining (TDM), S. 47 für denkbar; wie hier zu KI-Trainingsdaten auch *Heinze/Wendorf*, in: *Ebers/Heinze/Krüger/Steinrötter*, Künstliche Intelligenz und Robotik, § 9 KI und Urheberrecht Rn. 17; *Sigmüller/Gassner*, RDI 2023, 124 Rn. 10; im Ergebnis auch *Leistner*, GRUR 2024, 1665 (1666); für das Training generativer KI-Systeme *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 88 ff.; in diese Richtung auch *Chiou*, JIPITEC 2019, 398 Rn. 16; noch offengelassen in *Hofmann*, WRP 2024, 11 Rn. 14; eine eigene wirtschaftliche Bedeutung der Vervielfältigung dann aber ablehnend in *Hofmann*, ZUM 2024, 166 (169); tendenziell ablehnend auch *Maamar*, ZUM 2023, 481 (485); *Vesala*, IIC 2023, 351 (364 f.); *Kögel*, DSRITB 2023, 285 (289).

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

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

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

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

---

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

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

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

a) Ziel der Informationsgewinnung

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

---

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

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

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

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

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

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

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

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

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

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

#### aa) Wortlaut als Ausgangspunkt der Auslegung

Ausgangspunkt der Auslegung ist der Wortsinn des Gewinnungsbegriffs.<sup>345</sup> Das gilt auch im europäischen Recht.<sup>346</sup> Der Wortsinn des Gewinnungsbegriffs

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

---

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

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

Ein besonderer juristischer Sprachgebrauch der Formulierung findet sich, der deutschen Rechtslage entsprechend, nur in der Computerprogramm-RL<sup>356</sup>. Ein Schluss auf ein universell gültiges juristisches Verständnis des Begriffs der Informationsgewinnung ist hieraus allerdings nicht möglich.<sup>357</sup>

### (3) Schlussfolgerung aus dem Wortlaut der Bestimmungen

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

#### bb) Systematische Bezüge

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

---

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

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

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

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

---

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

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

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

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

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

### cc) Technologische Neutralität des Urheberrechts

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

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

---

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

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

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

#### dd) Historische Auslegung

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

##### (1) Nationale Gesetzeshistorie

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

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

---

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

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

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

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

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

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

## (2) Historie der Richtlinienbestimmung

Für das Erfordernis der Wahrnehmbarkeit und damit gegen eine Anwendung des Text und Data Mining-Begriffs auf das Training von KI spricht hingegen, dass KI in den Materialien zum europäischen Legislativprozess<sup>370</sup> kein einziges Mal erwähnt wurde.<sup>371</sup> Weiterhin hat der Richtliniengesetzgeber den Informationsgewinn in ErwGr. 8 S. 1 DSM-RL als „Erkenntnisgewinn“ umschrieben. Damit ein menschlicher Erkenntnisgewinn möglich ist, ist es aber erforderlich, dass die Informationen für Menschen wahrnehmbar gemacht worden sind.

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

---

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

<sup>369</sup> BT-Drs. 19/27426, S. 60.

<sup>370</sup> Rechtssetzungsverfahren 2016/0280/COD.

<sup>371</sup> Hierauf rekurriert auch *Schack*, NJW 2024, II3 Rn. 8.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

Hierin war die Schrankenbestimmung für Text und Data Mining noch auf den Zweck wissenschaftlicher Forschung beschränkt (Art. 3 Nr. 1 DSM-RL-E). Damit bestand im Richtlinienentwurf tatsächlich noch eine engere Bindung der Schrankenbestimmung an den (wissenschaftlich betriebenen) Erkenntnisgewinn. Nach der Erweiterung der Schrankenbestimmung auf allgemeines Text und Data Mining, in deren Zusammenhang auch Formulierungen in den Richtlinientext eingefügt wurden, die insbesondere die Bedeutung der Schrankenbestimmung für Innovationen und die Entwicklung neuer Anwendungen und Technologien hervorheben (ErwGr. 5, 18 DSM-RL, siehe angenommenen Text des Europäischen Parlaments aus erster Lesung, Dokument P8\_TA-PROV(2019)0231),<sup>372</sup> wurde die betreffende Passage in ErwGr. 8 S.1 DSM-RL allerdings nicht mehr angepasst. Vor dem Hintergrund dieser Regelungshistorie ist der Umschreibung der Informationsgewinnung mit dem Begriff des Erkenntnisgewinns in ErwGr. 8 S.1 DSM-RL daher keine besonders hohe Bedeutung beizumessen. Es handelt sich vielmehr um eine redaktionelle Ungenauigkeit.

#### ee) Zweck des zugrundeliegenden Richtlinienrechts

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

---

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

<sup>373</sup> ErwGr. 18 DSM-RL.

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

<sup>375</sup> Vgl. Art. 1 Abs. 1, ErwGr. 5 DSM-RL.

## ff) Schlussfolgerungen

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

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

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

---

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

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

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

b) Bezugspunkt der Informationsgewinnung

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

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

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

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

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

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

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

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

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

---

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

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

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

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

---

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

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

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

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

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

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

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

#### d) Konformität mit dem Dreistufentest

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

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

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

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

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

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

#### aa) Bestimmter Sonderfall

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

#### bb) Beeinträchtigung der normalen Verwertung und ungebührliche Beeinträchtigung der Urheberinteressen

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

---

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

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

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

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

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

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

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

---

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

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

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

### III. Eingreifen einer urheberrechtlichen Schrankenbestimmung

#### e) Zwischenergebnis zur Anwendbarkeit von § 44b UrhG

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

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

#### f) Beschränkung auf rechtmäßig zugängliche Werke

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

---

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

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

399 Dazu im Detail siehe unmittelbar oben.

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

#### aa) Begriff der rechtmäßigen Zugänglichkeit

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

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

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

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

---

400 Vesala, IIC 2023, 351 (359).

401 Vesala, IIC 2023, 351 (359).

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

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

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

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

---

403 Margoni, Kluwer Copyright Blog v. 22.12.2023.

404 Margoni, Kluwer Copyright Blog v. 22.12.2023.

405 Margoni, Kluwer Copyright Blog v. 22.12.2023.

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

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

## (2) Regelungszweck

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

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

## (3) Ergebnis der Auslegung

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

---

11 Rn. 16; *Konertz/Schönhof*, WRP 2024, 289 Rn. 31; *Raue*, GRUR 2017, 11 (14); *Schack*, GRUR 2021, 904 (907); *v. Welser*, GRUR-Prax 2023, 516 Rn. 32; *Schack*, NJW 2024, II 3 Rn. 11. Einschränkend lediglich *Dworschak*, in: *Fromm/Nordemann*, Urheberrecht, UrhG § 44b Rn. 12; *Margoni*, Kluwer Copyright Blog v. 22.12.2023.

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

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

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

nen. Unabhängig davon, ob der Veröffentlichung der Inhalte vom Rechtsinhaber zugestimmt wurde, ist damit ausreichend, dass der Datensammler rechtmäßigen Zugang zum Werk hat.<sup>411</sup>

#### (4) Beispiele für rechtmäßige Zugangskonstellationen

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

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

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

---

411 Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 8; Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 19 (anders noch in der Vorauflage); Hofmann, ZUM 2024, 166 (171); Raue, GRUR 2017, 11 (14); Schack, GRUR 2021, 904 (907); Griffiths/Syndinou/Xalabarder, GRUR Int. 2023, 22 (26); v. Welser, GRUR-Prax 2023, 516 Rn. 32. Einschränkend lediglich Dworschak, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 12; Margoni, Kluwer Copyright Blog v. 22.12.2023.

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

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

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

(6) Abgleich mit den identifizierten Regelungsbedürfnissen und  
Schlussfolgerungen

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

---

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

<sup>415</sup> Das verkennt hingegen *Schack*, NJW 2024, 113 Rn. 11.

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

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

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

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

---

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

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

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

Differenziert werden kann dabei, ob der Datenverarbeiter im eigenen Interesse handelt oder als Dienstleister zugunsten des Zugangsinhabers tätig wird. Ersteres ist insbesondere dann der Fall, wenn der Datenverarbeiter die generierten Inhalte in eigener Verantwortung zu Trainingsdatensätzen aufbereitet. Das umfasst auch die Tätigkeit intermediärer Trainingsdatensammler. Zugunsten des Zugangsinhabers wird hingegen tätig, wer in seinem Interesse KI-Training durchführt, beispielsweise, um in dessen Auftrag anwendungsfallspezifische KI-Systeme zu entwickeln.

#### (1) Zugangsvermittlung im Interesse des Zugangsinhabers

Fraglich ist, ob ein im Interesse des Zugangsinhabers tätig werdender Datenverarbeiter Vervielfältigungen von für ihn rechtmäßig zugänglichen Werken im Sinne des § 44b Abs. 2 S. 1 UrhG anfertigt, wenn ihm der Zu-

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

gang zu den vervielfältigten Werken (beispielsweise Login-Daten zu einer Datenbank) vom Zugangsinhaber vermittelt worden ist.

#### (i) Datenverarbeiter als Vervielfältigungsdienstleister

Vorgelagert klärungsbedürftig ist zunächst aber, ob der Datenverarbeiter in der untersuchten Fallkonstellation überhaupt *eigene* Vervielfältigungshandlungen im Sinne von § 16 UrhG vornimmt. Nur in diesem Fall kommt es auf seinen rechtmäßigen Zugang zu den Werken nach der Zugangsweitergabe an. Eigene Vervielfältigungshandlungen liegen hingegen nicht vor, wenn er als Vervielfältigungsdienstleister<sup>418</sup> für den Zugangsinhaber tätig wird. Dann werden diesem die Vervielfältigungen unmittelbar selbst zugerechnet. Vervielfältigungsdienstleister sind dabei von Hilfspersonen im Sinne des § 53 Abs. 1 S. 2 UrhG oder im Kontext der Wissenschaftsschranken aus §§ 60c, d UrhG zu unterscheiden. Diese nehmen eine eigene Vervielfältigungshandlung vor, welche jedoch von den Schrankenbestimmungen mitprivilegiert wird.<sup>419</sup>

##### (a) Voraussetzungen

Damit der Datenverarbeiter lediglich als Vervielfältigungsdienstleister auftritt, darf er nicht aktiv in die Vervielfältigungshandlung eingreifen.<sup>420</sup> Ein aktiver Eingriff findet jedenfalls auch dann statt, wenn die angefertigten Vervielfältigungen nicht nur mit dem Auftraggeber, sondern auch mit einer unbestimmten Zahl Dritter geteilt werden.<sup>421</sup> Weiterhin muss die Tätigkeit des Datenverarbeiters auf die Organisation der Vervielfältigungen als Hilfsdienstleistung beschränkt bleiben. Hierfür ist zum einen notwendig, dass

---

418 Zum Begriff siehe EuGH, Urt. v. 29.11.2017 - C-265/16, GRUR 2018, 68 Rn. 35 – VCAST; EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 30 – Ocilion.

419 Zur Privatkopieschranke deutlich in BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 25 – Internet-Radiorecorder; Renner, in: Fromm/Nordemann, Urheberrecht, UrhG § 53 Rn. 23; vgl. auch Loewenheim/Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 53 Rn. 27, 30; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 53 Rn. 14; zu den Wissenschaftsschranken außerdem vgl. Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60c Rn. 18; Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60d Rn. 14.

420 Vgl. EuGH, Urt. v. 29.11.2017 - C-265/16, GRUR 2018, 68 Rn. 54 – VCAST.

421 Vgl. EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 45 f. – Ocilion.

die Vervielfältigungen auf die Initiative des Auftraggebers zurückgehen.<sup>422</sup> Der Datenverarbeiter darf weiterhin nicht den Zugang zu den Werken vermitteln, die Gegenstand der Vervielfältigungshandlungen sind.<sup>423</sup> Darüber hinaus darf die Bereitstellung der Vervielfältigungen selbst nicht Hauptgegenstand des Angebots des Datenverarbeiters sein. Er muss sich vielmehr darauf beschränken, *Mittel* für Vervielfältigungshandlungen zur Verfügung zu stellen.<sup>424</sup>

### (β) Aktiver Eingriff des Datenverarbeiters im vorliegenden Fall

In der untersuchten Fallgruppe prozessiert der Datenverarbeiter die Werke im Interesse eines auftraggebenden Zugangsinhabers. Demzufolge ist nicht davon auszugehen, dass die angefertigten Vervielfältigungsstücke mit einer unbestimmten Zahl anderer Interessenten geteilt werden. Außerdem verschafft der Zugangsinhaber und nicht der Datenverarbeiter den Werkzugang.

Fraglich ist allerdings, ob die Vervielfältigungen tatsächlich auf die Initiative des Zugangsinhabers zurückgehen. Dafür spricht zwar, dass er die Vervielfältigungshandlungen auf die vom Zugang umfassten Werke konkretisiert hat. Allerdings steht zwischen der Zugangsverschaffung und der Vervielfältigung noch ein Willensentschluss des Datenverarbeiters. Denn er muss den Zugang erst aktiv für die Anfertigung der Vervielfältigungsstücke nutzen. Der Zugangsverschaffung folgen, abhängig von der konkreten Fallgestaltung, aber jedenfalls nicht unbedingt automatisierte Vervielfältigungshandlungen des Datenverarbeiters nach. Demzufolge löst der Zugangsinhaber den Vervielfältigungsvorgang nicht unmittelbar selbst aus. Das spricht dafür, dass der Datenverarbeiter aktiv in die Vervielfältigungshandlungen eingreift.<sup>425</sup>

Darüber hinaus führt der Datenverarbeiter die Organisation und Durchführung der Vervielfältigungshandlungen durch, um auf Basis der vervielfältigten Werke KI-Training zu betreiben. Dafür müssen dem Zugangsinhaber die Vervielfältigungen nicht überlassen werden. Der Zugangsinhaber hat vielmehr nur ein Interesse am Erhalt oder der Anwendung eines

---

422 Vgl. EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 50 - Ocilion.

423 Vgl. EuGH, Urt. v. 29.11.2017 - C-265/16, GRUR 2018, 68 Rn. 37 f. – VCAST.

424 Vgl. EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 40, 44 – Ocilion.

425 Ein vergleichbares Kriterium zur Bestimmung der Herstellereigenschaft einer Vervielfältigung leitet der BGH in seiner Rechtsprechung auch aus der nationalen Urheberrechtsdogmatik ab, siehe BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 33 – Internet-Radiorecorder.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

trainierten KI-Systems. Ziel der Tätigkeit des Datenverarbeiters ist daher nicht die Bereitstellung technischer Hilfsmittel für Vervielfältigungen des Zugangsinhabers, sondern anstelle des Zugangsinhabers die Vervielfältigungsleistung und das Training von KI durchzuführen. Die Vervielfältigungen und ein sich anschließendes KI-Training sind demzufolge Hauptgegenstand des Angebots des Datenverarbeiters. Insgesamt folgt daraus, dass der Datenverarbeiter im untersuchten Fall aktiv in die Vervielfältigungshandlungen eingreift.

#### (y) Schlussfolgerungen

Aufgrund des aktiven Eingriffs in die Vervielfältigungshandlungen ist der Datenverarbeiter im vorliegen Fall kein Vervielfältigungsdienstleister des Zugangsinhabers. Die auf Basis des vermittelten Zugangs vorgenommenen Vervielfältigungshandlungen zum Training der KI werden direkt dem Datenverarbeiter zugerechnet.

#### (ii) Eigener rechtmäßiger Zugang des Datenverarbeiters

Aufgrund der eigenen Vervielfältigungshandlungen des im Interesse des Zugangsinhabers handelnden Datenverarbeiters im Sinne des § 16 UrhG kommt es für ein Eingreifen der Schrankenbestimmung aus § 44b UrhG zu seinen Gunsten darauf an, ob er aufgrund der Zugangsvermittlung rechtmäßigen Zugang zu den betreffenden Werken im Sinne des § 44b Abs. 2 S. 1 UrhG gehabt hat. Inwieweit der Tatbestand ein solches Verständnis zulässt, hängt auch von der Auslegung des mit § 44b UrhG umgesetzten Art. 4 DSM-RL ab.

#### (a) Strenge Wortlautanwendung

Gegen einen rechtmäßigen Zugang des Datenverarbeiters zu den Werken spricht zunächst, dass er die vom Rechtsinhaber oder einem dritten Zugangsanbieter aufgestellten Zugangsvoraussetzungen nicht eigenhändig erfüllt. Insbesondere hat er, beispielsweise durch den Abschluss eines Abonnements, nicht selbst den Zugang zu einer Werkdatenbank erworben.

(β) Umkehrschluss aus § 53 Abs. 1 S. 2 UrhG

Gegen einen rechtmäßigen Zugang des Datenverarbeiters im vorliegenden Fall spricht weiterhin, dass § 53 Abs. 1 S. 2 UrhG den Einsatz von Hilfspersonen ausdrücklich mitumfasst. Da eine dem entsprechende Privilegierung von Hilfspersonen in § 44b UrhG nicht vorgesehen ist, könnte im Umkehrschluss hieraus gefolgert werden, dass § 44b UrhG den Einsatz von Hilfspersonen für das Text und Data Mining nicht ermöglichen wollte.

Voraussetzung eines Umkehrschlusses ist jedoch, dass das Gesetz die bezeichnete Rechtsfolge nur auf den konkret geregelten Fall beschränken wollte. Dies bestimmt sich nach dem Willen des Gesetzgebers und insbesondere dem Telos der Regelung.<sup>426</sup> In Bezug auf die vorliegende Bestimmung dürfte der Einsatz von Hilfspersonen damit *allein* im Rahmen der Schrankenbestimmung für Privatkopien für freistellungswert gehalten werden.

Grund für die Erweiterung der Schrankenwirkung des § 53 UrhG auf Hilfspersonen war, dass vom Rechtsinhaber nicht effektiv kontrolliert werden kann, wer im privaten Umfeld tatsächlich die Vervielfältigungshandlung vornimmt.<sup>427</sup> Ein solches Kontrolldefizit tritt allerdings immer bei Vervielfältigungshandlungen im nicht-öffentlichen Raum auf. Das kann nicht nur bei der Herstellung von Privatkopien, sondern beispielsweise auch bei Vervielfältigungshandlungen zu Lehr- oder Archivierungszwecken der Fall sein.<sup>428</sup> Das Kontrolldefizit tritt bei der Anfertigung von Privatkopien lediglich besonders zutage. Aus diesem Grund ist die Freistellung von Vervielfältigungen durch Hilfspersonen nur im Rahmen von § 53 UrhG ausdrücklich normiert worden. Hieraus lässt sich aber nicht schlussfolgern, dass der Rückgriff auf Hilfspersonen zur Durchführung von durch Schrankenbestimmungen privilegierten Werknutzungen ausschließlich für die Anfertigung von Privatkopien gestattet sein soll. Ein Umkehrschluss aus § 53 Abs. 1 S. 2 UrhG ist daher nicht möglich.

---

426 Larenz/Canaris, Methodenlehre der Rechtswissenschaft, S. 209.

427 BT-Drs. 15/38, S. 20.

428 Für eine Freistellung der Hilfsdienstleistung hier vgl. Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 60a Rn. 9; Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60c Rn. 18; Stieper, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60d Rn. 14.

(y) Ableitung eines allgemeinen Rechtsgedankens aus § 53 Abs. 1 S. 2 UrhG

Im Gegenteil könnte der in § 53 Abs. 1 S. 2 UrhG zum Ausdruck kommende Rechtsgedanke der Mitprivilegierung von Hilfspersonen als allgemeingültig angesehen werden. Damit wäre er auch auf § 44b UrhG übertragbar. Methodisch ist hierfür kein mehrfaches In-Erscheinung-Treten des Rechtsgedankens im UrhG notwendig.<sup>429</sup> Es reicht aus, dass der einer Regelung zugrundeliegende Normsinn auch auf einen weiteren Kreis von Fällen übertragen werden kann.<sup>430</sup> Wie bereits dargelegt wurde,<sup>431</sup> ist dies für das Kontrolldefizit gegenüber dem Einsatz von Hilfspersonen bei nicht-öffentlichen Vervielfältigungshandlungen der Fall.

Der Normsinn könnte darüber hinaus auch auf Vervielfältigungen von Werken für das Training von KI, die mit Hilfe von Webcrawlern angefertigt worden sind, zutreffen. Denn Maßnahmen wie IP-Adressbeschränkungen, die den Zugang zu einer Werkdatenbank auf ausgewählte Personen beschränken, können selbst mit eingeschränkten technischen Fähigkeiten umgangen werden. Damit lässt sich faktisch nur eine sehr eingeschränkte personelle Zugangs- und demzufolge auch Vervielfältigungskontrolle umsetzen. Wie bei der Schrankenbestimmung für Privatkopien spricht dieses Kontrolldefizit vor dem Einsatz von Hilfspersonen auch beim Training von KI dafür, einen solchen urheberrechtlich mit zu privilegieren. Der in § 53 Abs. 1 S. 2 UrhG zum Ausdruck kommende Normsinn kann damit auch auf die Schrankenbestimmung für Text und Data Mining übertragen werden.

Insgesamt sprechen die vorgenannten teleologischen Gründe dafür, aus § 53 Abs. 1 S. 2 UrhG einen allgemeinen Rechtsgedanken abzuleiten, der insbesondere auch auf die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG anwendbar ist. Folge dessen ist, dass der Einsatz von Hilfspersonen für Vervielfältigungen von Werken als Trainingsdaten im Interesse des Zugangsinhabers von § 44b UrhG umfasst ist. Dies kann bei der Anwendung des Schrankentatbestands erreicht werden, indem auch Datenverarbeiter rechtmäßigen Zugang zu Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben, denen der Werkzugang von einem Zugangsinhaber vermittelt worden ist, damit sie in seinem Interesse die Vervielfältigungen anfertigen.

---

429 Larenz/Canaris, Methodenlehre der Rechtswissenschaft, S. 207.

430 Larenz/Canaris, Methodenlehre der Rechtswissenschaft, S. 207.

431 Dazu siehe unmittelbar oben.

(δ) Teleologische Erwägungen anhand des umgesetzten Art. 4 DSM-RL

Gegen das aufgezeigte Tatbestandsverständnis könnte aus einer unionsrechtlichen jedoch Perspektive sprechen, dass ErwGr. 28 DSM-RL den Kulturerbe-Einrichtungen einen Rückgriff auf Hilfspersonen für die Zwecke der Sammlungserhaltung gestattet. Da sich in der Richtlinie keine entsprechende Aussage in Bezug auf das allgemeine Text und Data Mining findet, ließe sich hieraus ein Umkehrschluss ableiten, nach dem der Einsatz dritter Datenverarbeiter als Hilfspersonen im Interesse des Zugangsinhabers nicht vom Anwendungsbereich des § 44b UrhG umfasst wäre. Ob ein entsprechender Umkehrschluss gezogen werden kann, ist aber auch im Rahmen des ErwGr. 28 DSM-RL davon abhängig, ob die Richtlinie tatsächlich regeln wollte, dass der Einsatz dritter Hilfspersonen ausschließlich bei Vervielfältigungshandlungen zum Zwecke der Erhaltung von Sammlungen der Kulturerbe-Einrichtungen freigestellt sein sollte. Dagegen kann allerdings angeführt werden, dass ErwGr. 28 DSM-RL gemäß seines Wortlauts auf der Annahme beruht, dass Einrichtungen des Kulturerbes regelmäßig nicht über die technischen Mittel und Fähigkeiten verfügen, die für die umfangreichen Vervielfältigungshandlungen zur Sammlungserhaltung notwendig sind. Dasselbe trifft unter Umständen aber auch auf Zugangsinhaber zu, die mangels technischer Fähigkeiten und Ressourcen für umfangreiche Vervielfältigungshandlungen sowie eine automatisierte Analyse de facto nicht von der Schrankenbestimmung aus § 44b UrhG profitieren können. Grund für die besondere Herausstellung in ErwGr. 28 DSM-RL muss vielmehr gewesen sein, dass das Fähigkeiten- und Ressourcendefizit bei Einrichtungen des Kulturerbes besonders deutlich zutage tritt. Wie in Bezug auf § 53 Abs. 1 S. 2 UrhG ergeben sich demzufolge auch aus ErwGr. 28 DSM-RL keine Hinweise darauf, dass der dem Erwägungsgrund zugrundeliegende Rechtsgedanke allein für kulturelle Erhaltungszwecke Anwendung finden sollte. Vielmehr kann der in ErwGr. 28 DSM-RL zum Ausdruck kommende Rechtsgedanke auch hier verallgemeinert und ebenfalls auf die Schrankenbestimmung für Text und Data Mining und den Einsatz von Hilfspersonen für das Training von KI im Allgemeinen übertragen werden.

Für das Tatbestandsverständnis, nach dem die Datenverarbeitung durch Hilfspersonen im Wege eines weiten Verständnisses des Begriffs der rechtmäßigen Zugänglichkeit ermöglicht wird, spricht weiterhin, dass das von der Freistellung durch den in § 44b UrhG umgesetzten Art. 4 DSM-RL (unter anderem) betroffene Vervielfältigungsrecht der Urheber nach Art. 2 InfoSoc-RL dem Zweck dient, dem Urheber die Verwertung seines Wer-

kes zu sichern.<sup>432</sup> Umgekehrt schützt die Tatbestandsvoraussetzung der rechtmäßigen Zugänglichkeit den Urheber beziehungsweise Rechtsinhaber vor einer übergebührlichen Beeinträchtigung seiner Verwertungsinteressen. Hieraus folgt, dass eine Beendigung der Schrankenwirkung des Art. 4 DSM-RL zulasten eines dritten Datenverarbeiters bei der Weitergabe eines Zugangs durch den Schrankenprivilegierten Zugangsinhaber nur dann notwendig ist, wenn es durch die Weitergabe zu einer intensivierten Beeinträchtigung der wirtschaftlichen Verwertungsmöglichkeiten der Urheber kommt. Wird der Datenverarbeiter im Interesse des Zugangsinhabers tätig, heißt das allerdings, dass er die Vervielfältigungen der Werke *anstelle* des Zugangsinhabers anfertigt. Die Werke werden auf Grundlage des zur Verfügung stehenden Werkzugangs also nur einmal für das Text und Data Mining genutzt. Daraus folgt, dass die Zugangsweitergabe an einen Datenverarbeiter in den hier betreffenden Fällen nicht zu einer intensivierten Beeinträchtigung der Verwertungsinteressen betroffener Urheber führt. Insgesamt spricht damit auch eine teleologische Auslegung des Art. 4 DSM-RL vor dem Hintergrund der verwertungsrechtlichen Funktion des Vervielfältigungsrechts dafür, dass auch Datenverarbeiter rechtmäßigen Zugang zu Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben, denen der Zugang als Hilfsperson des Zugangsinhabers vermittelt worden ist.

#### (ε) Auslegungsergebnis

Insgesamt sprechen damit vor allem teleologische Gründe dafür, den Rechtmäßigkeitsbegriff so auszulegen, dass auch solche Datenverarbeiter rechtmäßigen Zugang zu Werken gemäß § 44b Abs. 2 S. 1 UrhG haben, denen der Zugang von einem Zugangsinhaber vermittelt worden ist, damit sie im Interesse des Zugangsinhabers Vervielfältigungen der Werke anfertigen. In diesen Fällen ist also nicht notwendig, dass die Werknutzer die Zugangsvoraussetzung eigenhändig erfüllen und sich beispielsweise den Zugriff auf eine Werkdatenbank durch Abschluss eines Nutzungsvertrags selbst verschafft haben.<sup>433</sup>

---

432 Nordemann, in: Fromm/Nordemann, Urheberrecht, Einl. zum UrhG Rn. 19; v. Ungern-Sternberg, GRUR 2012, 576 (579 f.); v. Ungern-Sternberg, GRUR 2012, II198 (1200); Haberstumpf, ZUM 2020, 810 (816); Grisse/Kaiser, ZUM 2021, 401 (404). Außerdem ErwGr. 10 InfoSoc-RL.

433 Anders wohl Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 8.

(iii) Zugangsweitergabe trotz vertraglicher Weitergabebeschränkung

Fraglich ist allerdings, ob der Zugang des Datenverarbeiters zu den Werken auch dann im Sinne von § 44b Abs. 2 S.1 UrhG rechtmäßig ist, wenn dieser zwar von einem Zugangsinhaber beispielsweise durch Weitergabe von Login-Daten verschafft worden ist, die Weitergabe des Zugangs dem Zugangsinhaber allerdings vertraglich untersagt worden ist.

Dagegen spricht, dass auch die vertragliche Beziehung zwischen dem zugangsgewährenden Rechtsinhaber oder sonstigen Zugangsanbieter und den Werknutzern wie beispielsweise Trainingsdatensammlern zur Bestimmung des rechtmäßigen Zugangs im Rahmen des § 44b Abs. 2 S.1 UrhG herangezogen wird.<sup>434</sup> Demzufolge wäre es konsequent, die vertragliche Ausgestaltung der Zugangsgewährung auch auf einer Sekundärebene zwischen Zugangsinhaber und Datenverarbeiter zu berücksichtigen.

Allerdings vervielfältigen die Datenverarbeiter in der vorliegenden Sachverhaltskonstellation die Werke im Interesse, also anstelle des Zugangsinhabers. Durch die Weitergabe der Zugänge kommt es demzufolge nicht zu einer Intensivierung der wirtschaftlichen Beeinträchtigung der betroffenen Rechtsinhaber beziehungsweise Anbieter von Werkzugängen wie beispielsweise Datenbankbetreibern. Dürften Zugangsinhaber aus einer urheberrechtlichen Perspektive keine Hilfspersonen für ihre Vervielfältigungen einschalten, würde dies weiterhin einen Wertungswiderspruch zu den §§ 53, 60c, 60d UrhG hervorrufen. Im Rahmen dieser Bestimmungen ist der Einsatz von Hilfspersonen zwar ebenfalls an Voraussetzungen gekoppelt, kann durch Rechtsinhaber aber nicht insgesamt untersagt werden.<sup>435</sup>

Aus Perspektive der Datenverarbeiter bestehen zudem strukturelle Ähnlichkeiten zwischen vertraglichen Beschränkungen der Zugangsweitergabe und der Rechtmäßigkeit der öffentlichen Zugänglichmachung der Werke. Weil letztere für Datenverarbeiter (und Zugangsinhaber) keine Bedeutung entfaltet, sondern allein ihr eigener Zugang zum Werk entscheidend ist, kann sich infolgedessen auch eine vertragliche Beschränkung der Zugangsweitergabe im dem Datenverarbeiter vorgelagerten Verhältnis nicht zu seinen Lasten auswirken. Im Gegenteil dient das Abstellen auf den eigenen Zugang und nicht die Rechtmäßigkeit der Zugänglichmachung im Rahmen des § 44b Abs. 2 S.1 UrhG der Gewährleistung von Rechtssicherheit

---

434 Dazu siehe oben unter 3. Teil A. III. 2. f) aa) (4).

435 Vgl. zu den Schrankenbestimmungen m. w. N. zum Einsatz von Hilfspersonen unmittelbar oben.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

für Trainingsdatensammler und andere Datenverarbeiter.<sup>436</sup> Diese Rechts-sicherheit würde für Datenverarbeiter konterkariert, wenn sich die Details des Vertragsverhältnisses zwischen dem Zugangsanbieter und dem Zugangsinhaber auf die Rechtmäßigkeit ihrer Datenverarbeitung im Rahmen des § 44b Abs. 2 S. 1 UrhG auswirken würden. Die Datenverarbeiter müssen vielmehr auf ihre erlangte freie Zugänglichkeit der Werke vertrauen dürfen.

Schließlich würde eine Berücksichtigung vertraglicher Beschränkungen der Zugangsweitergabe im Verhältnis zwischen Zugangsanbieter und Zugangsinhaber außer Acht lassen, dass der Nutzungsvorbehalt mechanismus aus § 44b Abs. 3 UrhG de lege lata zentrales regulatorisches Instrument für die Steuerung der Nutzung von Werken zum Text und Data Mining ist. Es bestünde die Gefahr, dass durch „Zugangsabreden“ zwischen Zugangsanbietern und -inhabern die spezifischen, aus § 44b Abs. 3 UrhG folgenden Voraussetzungen unterlaufen würden.

Alles in allem sprechen damit sowohl systematische als auch teleologische Gründe dafür, dass eine vertragliche Untersagung der Zugangsweitergabe zwischen Zugangsanbieter und -inhaber keine Auswirkungen darauf hat, dass die betreffenden Werke für einen Datenverarbeiter, dem der Zugang zu den Werken vermittelt worden ist, im Sinne des § 44b Abs. 2 S. 1 UrhG rechtmäßig zugänglich sind. Unabhängig davon begeht der Zugangsinhaber in diesen Fällen aber eine sanktionsfähige vertragliche Pflichtverletzung.

#### (2) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzugängen für Datenverarbeitung im Interesse des Zugangsinhabers

Fraglich ist darüber hinaus, ob die Werke für einen Datenverarbeiter auch dann noch rechtmäßig zugänglich im Sinne des § 44b Abs. 2 S. 1 UrhG sind, wenn ihm nicht der Zugang zu den Werken (z. B. durch eine Weitergabe von Login-Daten) vermittelt wird, sondern wenn durch den Zugangsinhaber zuvor angefertigte Vervielfältigungsvorlagen zur Verfügung gestellt werden. Hierfür speichert der Zugangsinhaber die Werke zunächst ab, stellt also selbst Vervielfältigungsstücke her. Diese Werkstücke kann der Datenverarbeiter im Anschluss als Vervielfältigungsvorlage für eigene Vervielfältigungshandlungen nutzen, die er für die automatisierte Analyse der Werke benötigt.

---

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

Um die rechtmäßige Zugänglichkeit der Werke für den Datenverarbeiter zu beurteilen, kommt es insbesondere darauf an, auf welches Zugangsobjekt bei der Bewertung abgestellt wird. Möglich wäre einmal, nur die Quelle in den Blick zu nehmen, die der Datenverarbeiter für die Anfertigung seiner Vervielfältigungsstücke verwendet. Das sind die Werkstücke, die bereits durch den Zugangsinhaber hergestellt worden sind. In der Folge käme es nur darauf an, ob der Datenverarbeiter rechtmäßigen Zugang zu den beim Zugangsinhaber verfügbaren Werkstücken hatte. Da der Zugangsinhaber dem Datenverarbeiter die Werkstücke zur Herstellung von Vervielfältigungen für die automatisierte Analyse aktiv zugänglich macht, ist das der Fall. Infolgedessen bestünde rechtmäßiger Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG. Andererseits könnte aber auch die gesamte Zugangskette mit in den Blick genommen werden. Mittelbar wäre damit auch der Zugang des Zugangsinhabers in die Betrachtung einzubeziehen. Für den Datenverarbeiter käme es, wie im Falle einer Zugangsvermittlung,<sup>437</sup> demzufolge darauf an, ob der Zugangsinhaber rechtmäßigen Zugang zu den Werken im Sinne der Regelung gehabt hat und ob er hiervon als nachgelagerter Datenverarbeiter auch profitieren kann.

Für eine Berücksichtigung der gesamten Zugangskette, mit der auch der Zugang des Herstellers der Vervielfältigungsvorlagen in den Blick nimmt genommen wird, spricht, dass es auch im Rahmen des § 53 Abs. 1 S. 1 UrhG darauf ankommt, dass für die Vervielfältigungshandlungen keine offensichtlich rechtswidrigen Vorlagen genutzt werden. Hieraus ließe sich unter Umständen ein allgemeiner Rechtsgedanke ableiten, in dessen Folge auch im Rahmen des § 44b Abs. 2 S. 1 UrhG der Ursprung der Vervielfältigungsvorlage, also die gesamte Zugangskette, in den Blick genommen werden müsste.

Allerdings beruht die Rechtmäßigkeitsanforderung aus § 53 Abs. 1 S. 1 UrhG auf einem politischen Kompromiss im Vermittlungsausschuss zwischen Deutschem Bundestag und Bundesrat.<sup>438</sup> Diese Regelungshistorie spricht dagegen, ihr eine allgemeingültige Bedeutung auch über den konkreten Regelungsgegenstand hinaus beizumessen. Außerdem konnte bereits gezeigt werden, dass aus § 53 Abs. 1 S. 2 UrhG ein allgemeiner Rechtsgedanke abgeleitet werden kann, wonach Vervielfältigungshandlungen von Hilfspersonen des originär Schrankenberechtigten ebenfalls von der Frei-

---

437 Dazu unmittelbar siehe oben unter 3. Teil A. III. 2. f) bb) (1).

438 BT-Drs. 15/1353.

stellungswirkung der Schrankenbestimmung miterfasst sind.<sup>439</sup> Hierfür kann es keinen Unterschied machen, ob der Hilfsperson der Werkzugang vermittelt wird, oder ob ihm bereits Vervielfältigungen der Werke zum Zwecke der Hilfsdienstleistung verfügbar gemacht werden.

Gegen eine Berücksichtigung der Zugangskette spricht außerdem, dass die Beurteilung des rechtmäßigen Zugangs des Herstellers der Vervielfältigungsvorlage Ähnlichkeiten zu der Fragestellung aufweist, ob eine im Internet zur Verfügung gestellte Vervielfältigungsvorlage rechtmäßig öffentlich zugänglich gemacht worden ist. Die Rechtmäßigkeit der Zugänglichmachung ist für die Anwendung des § 44b Abs. 2 S. 1 UrhG aber ohne Bedeutung.<sup>440</sup> Bei der Berücksichtigung der gesamten Zugangskette droht deswegen die Gefahr eines Wertungswiderspruchs.

Dafür, bei der Beurteilung der rechtmäßigen Zugänglichkeit allein auf die unmittelbare Quelle des Datenverarbeiter abzustellen, kann darüber hinaus angeführt werden, dass der Datenverarbeiter unter Umständen nicht weiß und wissen kann, aus welcher Primärquelle die Vervielfältigungsvorlagen herrühren und welche Zugangsvoraussetzungen hier hätten erfüllt sein müssen. Es erscheint vor diesem Hintergrund unbillig, ihm die außerhalb seines Einflussbereichs liegende Zugangskette zuzurechnen. Für den Datenverarbeiter bedeutete dies nicht zuletzt einen erheblichen Verlust an Rechtssicherheit. Dies zeigt sich insbesondere bei der Nutzung von frei verfügbar gemachten Trainingsdatensätzen. Zwar hätte der Datenverarbeiter selbst zu ihnen rechtmäßigen Zugang. Er könnte aber dennoch nicht beurteilen, ob die Schrankenwirkung des § 44b UrhG eingreift. Dieser Zustand widerspräche dem Ziel der Regelung, eine rechtssichere Grundlage für die Verarbeitung der Werke zu schaffen.<sup>441</sup>

Alles in allem sprechen damit mehrere teleologische Gründe dafür, bei der Beurteilung der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG allein auf den unmittelbaren Zugang des Datenverarbeiters zu seiner Vervielfältigungsvorlage abzustellen. Die Zugangskette, welche der Vervielfältigungsvorlage zugrunde liegt, ist für die Beurteilung seines rechtmäßigen Zugangs damit nicht von Bedeutung.

Da dem Datenverarbeiter der Zugang zu seinen Vervielfältigungsvorlagen im konkreten Fall aktiv vom Zugangsinhaber bereitgestellt wird, damit er in seinem Interesse Analyseleistungen erbringt, hat er im Rahmen der vorliegenden Fallgruppe rechtmäßigen Zugang zu den Werken im Sinne

---

439 Oben unter 3. Teil A. III. 2. f) bb) (1).

440 Dazu siehe oben unter 3. Teil A. III. 2. f) aa).

441 ErwGr. 18 DSM-RL.

von § 44b Abs. 2 S.1 UrhG. Auf das Verhältnis zwischen Zugangs- und Rechtsinhaber sowie den Ursprung der verwendeten Vervielfältigungsvorlagen kommt es nicht an.

(3) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen aus den Werkzeugungen für Datenverarbeitung im Eigeninteresse des Datenverarbeiters

Im Rahmen einer anderen Konstellation ist weiterhin denkbar, dass die vom Zugangsinhaber vervielfältigten Werke einem im eigenen Interesse handelnden Datenverarbeiter als Vervielfältigungsvorlage zur Verfügung gestellt werden. Hierunter fallen insbesondere intermediäre Trainingsdatensammler, die Zugangsinhabern beispielsweise Daten abkaufen, um diese für eine kommerzielle Verwertung in Trainingsdatensätze zu überführen. Denkbar ist aber auch, dass ein Zugangsinhaber die vervielfältigten Werke selbst als Vorlage für weitere Vervielfältigungshandlungen (gegebenenfalls rechtswidrig) öffentlich wiedergibt, damit andere die Inhalte zum eigenen KI-Training vervielfältigen können. Fraglich ist, ob die Datenverarbeiter auch in diesen Fällen rechtmäßigen Zugang den akquirierten Werken im Sinne des § 44b Abs. 2 S.1 UrhG haben, sodass ihre Vervielfältigungshandlungen in den Anwendungsbereich der Schrankenbestimmung für Text und Data Mining fallen können.

(i) Rechtmäßiger Zugang der Datenverarbeiter

Da es für den rechtmäßigen Zugang im Sinne des § 44b Abs. 2 S.1 UrhG nur auf den Zugang des Datenverarbeiters zu den unmittelbaren Vervielfältigungsvorlagen ankommt und ihm die Vorlagen auch in dieser Fallkonstellation aktiv bereitgestellt worden sind, hat der Datenverarbeiter auch hier rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S.1 UrhG.<sup>442</sup> Demzufolge würden die Vervielfältigungshandlungen grundsätzlich in den Anwendungsbereich der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG fallen.

---

442 Dazu ausführlich siehe unmittelbar oben unter 3. Teil A. III. 2. f) bb) (2).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

#### (ii) Begrenzung der Schrankenwirkung durch den Dreistufentest

Unklar ist aber, ob die Freistellung der Nutzungshandlung über § 44b UrhG in diesem Fall mit dem Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL vereinbar ist. Danach darf die Schrankenbestimmung nur in bestimmten Sonderfällen angewandt werden, in denen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt und die berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden.

Weil die Schrankenbestimmung nur Vervielfältigungen für das Text und Data Mining freistellt, betrifft sie nur einen bestimmten Sonderfall, sodass die erste Voraussetzungen des Dreistufentests im vorliegenden Fall gewahrt ist.<sup>443</sup> Ob die normale Verwertung der Werke beeinträchtigt wird oder ob es zu einer ungebührlichen Verletzung der berechtigten Urheberinteressen kommt, richtet sich weiterhin nach einer umfassenden Gesamtabwägung der gegenläufigen Interessen insbesondere im Hinblick auf die Verwertungsmöglichkeiten betroffener Urheber.<sup>444</sup>

#### (a) Argumente für eine ungebührliche Beeinträchtigung der Urheberinteressen

Für eine ungebührliche Beeinträchtigung der Urheberinteressen und in der Folge für eine Verletzung des Dreistufentests spricht, dass der Datenverarbeiter im vorliegenden Fall nicht im Interesse des Zugangsinhabers tätig wird. Infolgedessen ist es möglich, dass neben dem Datenverarbeiter auch der Zugangsinhaber selbst Interesse an der Nutzung der betroffenen Werke zum Text und Data Mining hat. Aus diesem Grund kann es auf Grundlage eines Zugangs des Zugangsinhabers mitunter zu *zwei* wirtschaftlichen relevanten Nutzungen der Werke durch den Zugangsinhaber zum einen und den Datenverarbeiter zum anderen kommen. Damit ist im vorliegenden Fall im Vergleich zur Nutzung der Werke durch den Zugangsinhaber allein eine intensivierte Beeinträchtigung der Verwertungsinteressen der Urheber möglich. Hierauf hat auch eine Vergütung, die der Datenverarbeiter dem Zugangsinhaber für die Bereitstellung der Vervielfältigungsvorlagen zahlt, keinen Einfluss. Denn diese Zahlung ist nicht an den betroffenen Urheber gerichtet.

---

443 Vgl. zur ersten Stufe des Dreistufentests ausführlicher oben unter 3. Teil A. III. 2. d).

444 Dazu bereits siehe oben unter 3. Teil A. III. 2. d).

Weiterhin kann für eine ungebührliche Beeinträchtigung der Urheberinteressen angeführt werden, dass die Bereitstellung von Vervielfältigungsvorlagen zugunsten eines nicht selbst zugangsberechtigten Datenverarbeiters möglicherweise die Entscheidung der betroffenen Urheber unterläuft, einem Datenverarbeiter keinen Zugang zu ihren Werken zu gewähren. Im Gegenteil müssten sie bei Zugrundelegung des obigen Tatbestandsverständnisses hinnehmen, dass ihre Werke für das KI-Training unkontrollierbar in die Welt gelangt sind, sobald sie die Werke in einer zugangsgeschützten Umgebung verfügbar gemacht haben.

Außerdem haben Urheber bei der Nutzung ihrer Werke zum Training von KI ein schutzwürdiges Amortisationsinteresse.<sup>445</sup> Dieses Interesse kann auch durch eine vergütungspflichtige Zugangsbeschränkung geschützt werden. Um dem Amortisationsinteresse bestmögliche Geltung zu verleihen, sollte die Effektivität einer eingesetzten Zugangsbeschränkung weitestgehend aufrechterhalten werden. Vor diesem Hintergrund ist es interessengerecht, die untersuchte Fallkonstellation aus dem Anwendungsbereich der Schrankenbestimmung für Text und Data Mining auszuklammern. Um von der Freistellungswirkung der Schrankenbestimmung profitieren zu können, müsste der Datenverarbeiter infolgedessen einen eigenen Werkzugang erwerben.

#### (β) Argumente gegen eine ungebührliche Beeinträchtigung der Urheberinteressen

Gegen eine ungebührliche Beeinträchtigung der Urheberinteressen spricht allerdings, dass durch die Verwendung der Werke zum Training von KI auch in der hier betreffenden Fallkonstellation keine Kernverwertungsarten der Urheber berührt sind. Weiterhin gegen die Beeinträchtigung der normalen Werkverwertung der Urheber spricht, dass individuelle Lizenzverträge zwischen Urhebern und Datenverarbeitern aufgrund prohibitiv hoher Transaktionskosten im Regelfall nicht zustande kommen.<sup>446</sup>

Gegen eine ungebührliche Beeinträchtigung der Urheberinteressen spricht weiterhin, dass sich aus einer Abwägung von Nutzer- und Urheberinteressen ein umfassendes Freistellungsbedürfnis in Bezug auf das Web Scraping von Werken zum KI-Training ergeben hat.<sup>447</sup> Vor diesem Hin-

---

<sup>445</sup> Dazu siehe oben 2. Teil B. II. 1.

<sup>446</sup> Siehe oben unter 2. Teil B. I. 2. und 3. Teil A. III. 2. d).

<sup>447</sup> Dazu siehe oben 2. Teil B. I. 2.

tergrund ist das Interesse der Urheber an einem bestmöglichen Schutz der von ihnen eingeschalteten Zugangsbeschränkung trotz ihres grundsätzlich anerkennungswürdigen Amortisationsinteresses im Spezifischen nicht schutzbedürftig. Vielmehr sollten die Werke frei zum Training von KI genutzt werden dürfen.<sup>448</sup> Dafür spricht auch, dass aus Perspektive des Datensammlers keine rechtswidrig erlangte Vervielfältigungsvorlage verwendet worden ist. Vielmehr wurden ihm die Werkstücke aktiv zugänglich gemacht. Vor diesem Hintergrund ist auch zu berücksichtigen, dass es Ziel des Normgebers ist, eine rechtssichere Grundlage für das Text und Data Mining zur Verfügung zu stellen.<sup>449</sup> Haben Datenverarbeiter selbst keine Zugangsbeschränkungen umgangen, müssen sie daher darauf vertrauen dürfen, dass sie die Werke rechtmäßig auf Basis des § 44b UrhG verarbeiten dürfen, solange die übrigen Freistellungsvoraussetzungen erfüllt sind.

(γ) Schlussfolgerungen

Grundsätzlich sprechen insbesondere die Nichtbeeinträchtigung der Kernverwertungsmöglichkeiten der Urheber sowie das schutzwürdige Interesse der Datenverarbeiter an einer rechtssicheren Grundlage für Vervielfältigungen zum Training von KI gegen eine ungebührliche Beeinträchtigung der normalen Verwertung von Werken durch die Urheber. Vervielfältigungs-handlungen von Datenverarbeitern, die sie auf Basis von durch Zugangsinhaber zur Verfügung gestellten Vervielfältigungsvorlagen anfertigen, verletzen damit grundsätzlich nicht den Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL. Damit sind sie im Allgemeinen von der Schrankenbestimmung für Text und Data Mining gemäß § 44b UrhG freigestellt.

(δ) Abweichende Beurteilung bei objektiver Erkennbarkeit der zugangsgeschützten Herkunft genutzter Vervielfältigungsvorlagen

Etwas anderes könnte allerdings dann gelten, wenn für den im eigenen Interesse handelnden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlagen einer zugangsgeschützten Quelle entstammen. Denn wenn sich der Datenverarbeiter auf seinen rechtmäßigen Zugang zu

---

448 Dazu siehe oben 2. Teil B. I. 2.

449 ErwGr. 18 DSM-RL.

den Vervielfältigungsvorlagen zurückzieht, obwohl erkennbar ist, dass für den originären Werkzugang Zugangsbeschränkungen überwunden werden müssen oder insbesondere ein Entgelt zu entrichten ist, wohnt seiner Nutzung ein Umgehungsmoment inne. Es ist mit der unmittelbaren Umgehung der zugangshindernden Maßnahmen vergleichbar, die zu einem Nichteingreifen der Schrankenbestimmung aus § 44b UrhG führt. Der Datenverarbeiter hätte beispielsweise selbst ein eigenes Entgelt für den Zugang zu den Werken entrichten müssen, um seine im Eigeninteresse erfolgende Nutzung der Werke angemessen auszugleichen. Außerdem hat der Datenverarbeiter in diesem Fall kein schützenswertes Interesse an einer rechtssicheren Freistellungsregelung durch die Schrankenbestimmung. Dies führt im Ergebnis zu einer ungebührlichen Beeinträchtigung der Urheberinteressen. Infolgedessen ist in diesen Fällen der Dreistufentest aus Art. 5 Abs. 5 Info-Soc-RL verletzt. Ist für den vervielfältigenden Datenverarbeiter objektiv erkennbar, dass die Vervielfältigungsvorlage einer zugangsgeschützten Quelle entstammt, sind seine Vervielfältigungshandlungen abweichend vom zuvor Gesagten demzufolge ausnahmsweise nicht von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 S. 1 UrhG freigestellt.<sup>450</sup> Eine Rückausnahme hiervon muss nur dann gemacht werden, wenn für den Urheber oder Rechteinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, auf Basis des Werkzugangs angefertigte Werkstücke dritten Datenverarbeitern als Vorlagen für Vervielfältigungen zum Text und Data Mining bereitzustellen. Denn dann hätte die Ermöglichung weiterer Vervielfältigungshandlungen zum Text und Data Mining insbesondere in der Höhe des für den Zugangserwerb zu entrichtenden Entgelts berücksichtigt werden können. Zudem liegt kein Umgehungsmoment vor, welches ebenfalls maßgeblich zur Verletzung des Dreistufentests beiträgt.<sup>451</sup>

Dass damit auch die Tätigkeit intermediärer Trainingsdatensammler im Wesentlichen vom Anwendungsbereich des § 44b Abs. 2 S. 1 UrhG erfasst ist, deckt sich auch mit den bereits herausgearbeiteten Regelungsbedürfnissen. Für ihre Tätigkeit ist aber insbesondere noch die Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG in den Blick zu nehmen.<sup>452</sup>

---

450 ähnlich auch Dworschak, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 15.

451 Das wird insbesondere bei Gedächtnisinstitutionen wie Bibliotheken relevant, siehe unten unter 3. Teil A. III. 4. a).

452 Dazu sogleich unter 3. Teil A. III. 2. h).

(4) Zugangsvermittlung für Datenverarbeitung im Interesse des Datenverarbeiters

Möglich ist auch, dass der Zugangsinhaber dem Datenverarbeiter seinen Zugang (beispielsweise Login-Daten) vermittelt, also weiterreicht, damit er auf Basis dieses Werkzugangs im eigenen Interesse Vervielfältigungstücke für das Training von KI oder die Herstellung von Trainingsdatensätzen anfertigen kann. Denkbar ist, dass der Zugangsinhaber hierfür eine Vergütung erhält. Fraglich ist auch in diesem Fall, ob der Datenverarbeiter aufgrund der Zugangsvermittlung rechtmäßigen Zugang zu den betroffenen Werken im Sinne des § 44b Abs. 2 S. 1 UrhG hat. Dann würden seine Vervielfältigungshandlungen in den Anwendungsbereich des Schrankentatbestands fallen.

Wie bereits gezeigt werden konnte, ist es für die Anwendung des Schrankentatbestands aus § 44b UrhG unschädlich, dass der Datenverarbeiter die Zugangsvoraussetzungen nicht eigenhändig erfüllt.<sup>453</sup> Allerdings muss eine Anwendung der Schrankenbestimmung wegen einer Verletzung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL ausscheiden, wenn für den im eigenen Interesse tätig werdenden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlage einer zugangsgeschützten Quelle entstammt. Denn solchen Vervielfältigungshandlungen wohnt ein Umgehungsmoment inne. Sie beeinträchtigen die berechtigten Urheberinteressen daher übergebührlich.<sup>454</sup> Wird dem Datenverarbeiter der Werkzugang beispielsweise in Form von Login-Daten vermittelt, ist für ihn allerdings objektiv erkennbar, dass die Vervielfältigungsvorlagen aus einer geschützten Quelle stammen, deren Zugangsvoraussetzungen er nicht selbst erfüllt. Damit beeinträchtigt eine Anwendung der Schrankenbestimmung aus § 44b UrhG auf diese Fallgruppe die berechtigten Urheberinteressen übergebührlich. Das führt zu einer Verletzung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL. Aus diesem Grund kann die Nutzung in der betreffenden Fallgruppe nicht von der Schrankenbestimmung aus § 44b UrhG freigestellt werden. Eine Ausnahme hiervon muss nur dann gemacht werden, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, den Werkzugang an dritte Datenverarbeiter zum Text und Data Mining weiterzugeben.<sup>455</sup>

---

453 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (1).

454 Dazu ausführlich siehe unmittelbar oben unter 3. Teil A. III. 2. f) bb) (3).

- cc) Fernwirkung einer fehlenden Zugangsberechtigung insbesondere bei der öffentlichen Wiedergabe von Trainingsdatensätzen

In einer umgekehrten Fallkonstellation ist schließlich auch klärungsbedürftig, wie sich eine Bereitstellung von Vervielfältigungsvorlagen an Datenverarbeiter auf die rechtmäßige Zugänglichkeit der Werke für die Datenverarbeiter im Sinne des § 44b Abs. 2 S. 1 UrhG auswirkt, wenn der Vermittler selbst keinen rechtmäßigen Zugang zu den Werken gehabt hat. Das umfasst insbesondere auch Fälle, in denen die Werke als Teil von Trainingsdatensätzen angeboten werden, obwohl der Anbieter des Datensatzes keinen rechtmäßigen Zugang zu den Werken nutzen konnte.

Wie bereits gezeigt wurde, kommt es für die Bestimmung der Rechtmäßigkeit des Zugangs des nachgelagerten Datenverarbeiters allein auf seine Beziehung zu den von ihm genutzten Vervielfältigungsvorlagen an.<sup>456</sup> Sofern der Zugang für den Datenverarbeiter frei möglich ist, beispielsweise, weil die Werke frei im Internet wiedergegeben oder ihm aktiv überlassen werden, hat der Datenverarbeiter rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG. Auf die Urheberrechtskonformität der Zugänglichmachung, also der öffentlichen Wiedergabe der Vervielfältigungsvorlagen, kommt es nicht an.<sup>457</sup> Ebenso wenig ist grundsätzlich die Rechtmäßigkeit des Zugangs des die Vervielfältigungsvorlage zur Verfügung Stellenden von Bedeutung.<sup>458</sup> Das Bedürfnis nach einer rechtssicheren Grundlage für das Text und Data Mining, welches in ErwGr. 18 DSM-RL zum Ausdruck kommt, überwiegt daher grundsätzlich die urheberrechtsverletzende Herkunft der Text und Data Mining-Grundlage.

Allerdings steht auch in diesem Fall der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL einer Anwendung des Schrankentatbestands entgegen, wenn für den vervielfältigenden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlage einer zugangsgeschützten Quelle entstammt.<sup>459</sup> Wann das der Fall ist, ist unter Berücksichtigung aller Umstände des Einzelfalls zu bestimmen. Fraglich ist dabei insbesondere, ob von der Größe des bereitgestellten Datensatzes in Verbindung mit der Art der darin enthaltenen Werke auf eine zugangs geschützte Herkunft der Inhalte geschlossen werden muss. Denkbar wäre

---

455 Vgl. dazu ausführlich oben unter 3. Teil A. III. 2. f) bb) (3).

456 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (2).

457 Dazu siehe oben unter 3. Teil A. III. 2. f) aa).

458 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (2).

459 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (3).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

das beispielsweise beim öffentlich verfügbaren Trainingsdatenset *Books3*. Er besteht aus 197.000 Büchern.<sup>460</sup> Zugunsten einer objektiven Erkennbarkeit kann hier angeführt werden, dass eine unrechtmäßige Herkunft der Inhalte aufgrund der großen Menge an verfügbar gemachten Schriftwerke naheliegen dürfte. Gegen einen solchen Schluss spricht jedoch, dass es auch umfangreiche Trainingsdatensätze gibt, die sich aus urheberrechtlich freien Inhalten zusammensetzen. Hierzu gehört beispielsweise der *Project Gutenberg*-Datensatz. Er besteht ausschließlich aus gemeinfrei gewordenen Schriftwerken.<sup>461</sup> Allein aus der Größe des bereitgestellten Datensatzes und der Art der enthaltenen Werke kann daher nicht ohne weiteres auf eine zugangsgeschützte Herkunft der Inhalte geschlossen werden. Als Ausgangspunkt der Betrachtung muss vielmehr das konkret vervielfältigte Werk dienen. Die zugangsgeschützte Quelle eines Werkstücks ist beispielsweise dann objektiv erkennbar, wenn es sich um ein audiovisuelles Werk handelt, dass als Eigenproduktion eines Streaming-Anbieters erkennbar ist. Diese Inhalte werden vom Rechteinhaber stets nur auf dem zugangsgeschützten Streaming-Portal verfügbar gemacht. Solche Werke könnten aus einer praktischen Perspektive beispielsweise durch einen automatisierten Abgleich mit der Wortbildmarke des Streaming-Anbieters herausgefiltert werden. Bei der Bestimmung der objektiven Erkennbarkeit muss im Übrigen aber stets auch berücksichtigt werden, wie erkennbar die Nutzung des Werkes für den Datenverarbeiter ist. Die zugangsgeschützte Herkunft einer Vervielfältigungsvorlage ist beispielsweise dann nicht objektiv erkennbar, wenn das jeweilige Werk in einem ansonsten unspezifischen, urheberrechtskonform erscheinenden Trainingsdatensatz eingebettet ist. Von Bedeutung ist auch, inwieweit die Datensammlung automatisiert und generalisiert erfolgt, beispielsweise durch ein breit angelegtes Web Scraping von Inhalten für das Training von KI. Für objektive Erkennbarkeit spricht hingegen, wenn das betreffende Werk gezielt vervielfältigt wird oder der vervielfältigte Datensatz insgesamt aus Werkstücken besteht, die einen zugangsgeschützten Ursprung nahelegen.

---

460 Siehe dazu [https://huggingface.co/datasets/the\\_pile\\_books3](https://huggingface.co/datasets/the_pile_books3) (zuletzt angerufen am 22.3.2024); vgl. außerdem *Gao u. a.*, arXiv: 2101.00027 2020, S. 3 f.

461 Dazu <https://www.gutenberg.org> (zuletzt abgerufen am 22.3.2024); [https://data.ncl.ac.uk/articles/dataset/Gutenberg\\_Dataset/24574753](https://data.ncl.ac.uk/articles/dataset/Gutenberg_Dataset/24574753) (zuletzt abgerufen am 22.3.2034); *Gao u. a.*, arXiv: 2101.00027 2020, S. 5.

g) Kreis der Schrankenbegünstigten

§ 44b UrhG spezifiziert den Kreis der Schrankenbegünstigten im Gegensatz zu § 60d UrhG nicht. Aus diesem Grund ist der Anwendungsbereich der Schrankenbestimmung auch dann eröffnet, wenn die Trainingsdatensammler gleichzeitig in den persönlichen Anwendungsbereich der speziellen Schrankenbestimmung für Text und Data Mining zu wissenschaftlichen Zwecken, § 60d UrhG, fallen.<sup>462</sup>

aa) Kommerziell tätige Trainingsdatensammler

Darüber hinaus profitieren von der Schrankenbestimmung sowohl kommerziell auch als nicht-kommerziell tätige Trainingsdatensammler. Etwas anderes ergibt sich auch nicht aus der Anwendung des Dreistufentests.<sup>463</sup> Das Freistellungsbedürfnis, welches sowohl für kommerziell als auch für nicht-kommerziell tätige Trainingsdatensammler besteht, wird in personeller Hinsicht durch § 44b UrhG demzufolge erfüllt.

bb) Intermediäre Trainingsdatensammler

Fraglich ist allerdings, ob der persönliche Anwendungsbereich der Schrankenbestimmung aus § 44b UrhG auch intermediäre Trainingsdatensammler umfasst. Sie speichern ebenfalls Werke mit Hilfe von Webcrawlern zur Verwendung als KI-Trainingsdaten ab. Damit ist auch ihre Tätigkeit von dem herausgearbeiteten Freistellungsinteresse abgedeckt. Im Gegensatz zum KI-Entwickler führen sie aber selbst kein Training von KI-Systemen durch. Infolgedessen findet das Text und Data Mining im Sinne des § 44b Abs. 1 UrhG nicht bei ihnen statt – jedenfalls so lange die intermediären Trainingsdatensammler nicht in besonderen Fallkonstellationen eine vorbereitende Analyse der Werke durchführen, die unter § 44b Abs. 1 UrhG gefasst werden kann.<sup>464</sup>

Für eine Freistellung der intermediären Trainingsdatensammler über § 44b UrhG ist somit maßgeblich, ob sie bei der Speicherung von Wer-

---

462 Vgl. Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 4.

463 Dazu und zu einer abweichenden Auffassung siehe oben unter Teil. A. I. 2. b) cc).

464 So insbesondere in LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

ken für ein späteres KI-Training durch Dritte Vervielfältigungen „für das Text und Data Mining“ beziehungsweise „für die Zwecke des Text und Data Mining“ vornehmen, wie es § 44b Abs. 2 S. 1 UrhG respektive Art. 4 Abs. 1 DSM-RL tatbestandlich voraussetzen. Das hängt davon ab, ob die Schrankenbestimmung mit diesen Formulierungen einen Bezug der Vervielfältigungshandlungen zu einem bereits konkretisierten Text und Data Mining-Projekt notwendig macht. Wäre das der Fall, könnten Vervielfältigungen intermediärer Trainingsdatensammler, die ohne Bezug zu einem spezifischen Text und Data Mining-Projekt stattfinden, nicht unter den Schrankentatbestand aus § 44b Abs. 2 S. 1 UrhG subsumiert werden. Bei der Auslegung der Schrankenbestimmung ist insbesondere auch der unionsrechtliche Hintergrund der Regelung aus Art. 4 DSM-RL zu berücksichtigen.

#### (1) Wortlaut der Regelung

Als Ausgangspunkt der Auslegung ist auf den natürlichen Wortsinn des Tatbestandsmerkmals abzustellen.<sup>465</sup> Im nationalen Umsetzungsrecht wird vorausgesetzt, dass die Vervielfältigungen *für* das Text und Data Mining angefertigt werden müssen. Mit der Präposition „für“ wird ihrem natürlichen Wortsinn nach der Zweck einer Handlung beschrieben.<sup>466</sup> Die gleiche Bedeutung kann weiterhin der Formulierung aus Art. 4 Abs. 1 DSM-RL entnommen werden. Diese bezieht sich ausdrücklich auf die „Zwecke“ der Vervielfältigungshandlungen.

Bezugspunkt der Zweckbindung ist „das Text und Data Mining“. Diese Formulierung ist unbestimmt. Die Zweckbindung wird damit pauschal mit der Analysetechnik als solcher und nicht mit *einem konkreten* Text und Data Mining-Vorhaben verknüpft.<sup>467</sup> Demzufolge sprechen der Wortlaut des § 44b Abs. 2 S. 1 UrhG und des umgesetzten Art. 4 Abs. 1 DSM-RL dafür, dass es ausreichend ist, wenn die in Rede stehenden Vervielfältigungen angefertigt werden, damit zu einem späteren Zeitpunkt die Analysetechnik des Text und Data Mining als solche durchgeführt wird. Hierfür bedarf

---

465 Zur Auslegungsregel für alle *Larenz/Canaris*, Methodenlehre der Rechtswissenschaft, S. 163 f. Im europäischen Recht ebenso unbestritten, siehe EuGH, Urt. v. 05.07.2012 - C-49/11, EuZW 2012, 638 Rn. 32 – Content Services; EuGH, Urt. v. 10.03.2005 - C-336/03, MMR 2005, 364 Rn. 21 – easycar.

466 *Dudenredaktion (Bibliographisches Institut)*, Duden: das Bedeutungswörterbuch, Schlagwort: „für“.

467 Vgl. zur identischen Formulierung im Rahmen der Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG: *Wulf/Löbeth*, GRUR 2024, 737 (740).

es diesem Wortlaut entsprechend noch keiner Konkretisierung eines einzelnen Text und Data Mining-Projekts. Der Wortlaut der Bestimmungen spricht demzufolge dafür, dass auch intermediäre Trainingsdatensammler Vervielfältigungen für das Text und Data Mining im Sinne von § 44b Abs. 2 S. 1 UrhG anfertigen können. Denn auch ihre Vervielfältigungen sind allgemein für das KI-Training, also für die Analysetechnik des Text und Data Mining bestimmt, auch wenn noch kein spezifisches Trainingsprojekt konkretisiert ist.

## (2) Systematische Anknüpfungspunkte

Systematische Bezüge der Schrankenbestimmung aus § 44b UrhG respektive Art. 4 DSM-RL zu anderen urheberrechtlichen Regelungen könnten hingegen für eine enge Auslegung des Tatbestands sprechen. Nach dieser müsste für eine Privilegierung der Vervielfältigungshandlungen bereits ein konkretes Text und Data Mining-Projekt bestehen, dem die Vervielfältigungen gewidmet werden.

Hierfür spricht zum einen, dass im Rahmen von § 60d UrhG a. F., der vorherigen Fassung der Schrankenbestimmung für Text und Data Mining zum Zwecke wissenschaftlicher Forschung, vereinzelt der Bezug der Vervielfältigungshandlungen zu einem konkreten Text und Data Mining-Projekt vorausgesetzt wurde.<sup>468</sup>

Wenn für die Tatbestandserfüllung kein Bezug der Vervielfältigungshandlungen zu einem konkreten Text und Data Mining-Projekt notwendig wäre, könnte außerdem die Gefahr eines Wertungswiderspruchs zwischen Art. 3 und Art. 4 DSM-RL bestehen. Denn Forschungseinrichtungen müssen bei der Aufbewahrung der Werke gemäß Art. 3 Abs. 2 DSM-RL respektive § 60d Abs. 5 UrhG n. F. angemessene Sicherheitsvorkehrungen vorsehen. Bei intermediären Trainingsdatensammlern, die Werke vervielfältigen, ohne ein bereits konkretisiertes Text und Data Mining-Projekt ermöglichen zu wollen, ist jedoch ebenfalls davon auszugehen, dass die Werke zunächst für einen unbestimmten Zeitraum gespeichert werden. Auch hier findet demzufolge eine vorgelagerte Aufbewahrung der Werke statt. Hierfür müssten sie gemäß § 44b UrhG respektive Art. 4 DSM-RL jedoch keine Sicherheitsvorkehrungen einsetzen.

---

468 Vgl. dazu *Stieper*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 60d Rn. 13.

Allerdings ist im Rahmen der Freistellungsregelung aus § 44b UrhG respektive Art. 4 DSM-RL eine Löschungspflicht zu beachten.<sup>469</sup> Diese besteht im Rahmen des Art. 3 DSM-RL beziehungsweise § 60d UrhG nicht. Das rechtfertigt bereits die Verpflichtung zur Einsetzung angemessener Sicherheitsmaßnahmen. Darüber hinaus richtet sich die Schrankenbestimmung für Text und Data Mining zu allgemeinen Zwecken aus § 44b UrhG respektive Art. 4 DSM-RL insbesondere an gewinnwirtschaftlich tätige Trainingsdatensammler. Da die geschäftliche Betätigung eng mit den vervielfältigten Werken verbunden ist, bestehen bereits natürliche Anreize für eine Einsetzung angemessener Sicherheitsmaßnahmen zur Sicherung des wirtschaftlichen Guts. Es bedarf vor diesem Hintergrund keiner besonderen Verpflichtung zur Einhaltung solcher Sicherheitsstandards. Ein Wertungswiderspruch zwischen Art. 3 und Art. 4 DSM-RL bei einem weiten Tatbestandsverständnis der Zweckbestimmung besteht deswegen nicht.

Darüber hinaus geht § 60d UrhG a. F. auf eine ältere unionsrechtliche Grundlage zurück, die durch Art. 3 DSM-RL als neue Spezialregelung hinsichtlich der Vervielfältigung von Werken für wissenschaftlichen Zwecken dienendes Text und Data Mining überlagert worden ist. Ein systematischer Vergleich zwischen § 60d UrhG a. F. und § 44b UrhG ist demzufolge nicht möglich.

Aus den systematischen Bezügen der Schrankenbestimmung für allgemeines Text und Data Mining aus § 44b UrhG respektive Art. 4 DSM-RL zu anderen urheberrechtlichen Regelungen könnten damit keine belastbaren Schlüsse für eine Auslegung des Tatbestands zur Beantwortung der hier aufgeworfenen Fragestellung abgeleitet werden.

### (3) Telos der Schrankenbestimmung

Weitere Anhaltspunkte für die Auslegung der Schrankenbestimmung könnten sich allerdings aus dem Regelungszweck der Schrankenbestimmung ergeben.

Für ein enges Verständnis der von § 44b Abs. 2 S. 1 UrhG vorausgesetzten Zweckbindung der Vervielfältigungshandlungen, nach der für eine Freistellung der Nutzungshandlungen über § 44b UrhG bereits ein konkretes Text und Data Mining-Projekt notwendig ist, spricht, dass im Falle einer Freistellung intermedialer Trainingsdatensammler diese und die KI-Entwickler, also zwei Beteiligte, Vervielfältigungen zum Training von KI anfertigen

---

469 Dazu im Detail siehe unten unter 3. Teil A. III. 2. h).

dürften. Damit käme es zu einer intensivierten Beeinträchtigung der urheberrechtlichen Verwertungsinteressen.

Allerdings würden bei einem weiten Verständnis der von § 44b Abs. 2 S. 1 UrhG vorausgesetzten Zweckbindung trotz alledem die Vervielfältigungen – auch diejenige des intermediären Trainingsdatensammlers – ausschließlich zum Zwecke des Text und Data Mining vorgenommen. Auch die Vervielfältigungen der intermediären Trainingsdatensammler sind demzufolge noch von der gesetzgeberischen Regelungsintention gedeckt, Text und Data Mining auf Grundlage von Werken über § 44b UrhG zu ermöglichen. Außerdem werden auch Vervielfältigungen intermediärer Trainingsdatensammler zum Zwecke des KI-Trainings angefertigt, sodass im Ergebnis dieselbe urheberrechtliche Nutzungsart wie auch bei Vervielfältigungen durch KI-Entwickler betroffen ist.<sup>470</sup> Hierfür ist ohne Belang, ob schon ein konkretes KI-Trainingsprojekt besteht, oder nicht.

Für ein weites Verständnis der von § 44b Abs. 2 S. 1 UrhG vorausgesetzten Zweckbindung spricht darüber hinaus, dass die Schrankenbestimmung gemäß ErwGr. 18 DSM-RL die Entwicklung neuer Technologien ermöglichen soll. Das können insbesondere auch Technologien sein, die auf KI-Systemen beruhen. Die Erstellung von hierfür benötigten Trainingsdatensätzen ist jedoch ressourcen- und zeitintensiv.<sup>471</sup> Sie kann nicht von jedem KI-Entwickler geleistet werden. Intermediäre Trainingsdatensammler, die über die Ressourcen zur Erstellung von Trainingsdatensätzen verfügen, sind für die Entwicklung neuer Technologien im Bereich der KI daher wesentlich. Ein arbeitsteiliges Vorgehen entspricht auch der modernen Wirtschaftsstruktur. Umgekehrt würde eine Zweckbindung der Vervielfältigungen an ein bereits konkretisiertes Text und Data Mining-Projekt kleine KI-Entwickler benachteiligen, die nicht über die Ressourcen zur vollständig eigenhändigen Erstellung der notwendigen Trainingsdatensätze verfügen.

Überdies liegt dem Urheberrecht ein allgemeiner Rechtsgedanke zugrunde, nach dem Vervielfältigungshandlungen von Hilfspersonen originär Schrankenbegünstigter ebenfalls von der Freistellungswirkung einer Schrankenregelung mitumfasst sind.<sup>472</sup> Auch wenn die Tätigkeit intermediärer Trainingsdatensammler noch keinem konkretisierten KI-Trainingsprojekt und keinem individualisierten KI-Entwickler zugeordnet werden kann, handelt es sich dennoch um eine vorgelagerte Hilfstätigkeit für das

---

470 Zur selbstständigen Nutzungsart vgl. unten unter 3. Teil A. III. 2. i) hh) (2) (i).

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

472 Dazu ausführlich siehe oben unter 3. Teil A. III. 2. f) bb) (ii) (γ).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

Speichern von Werken durch KI-Entwickler zur Durchführung konkreter Trainingsprojekte, das auch nach der engen Auslegung des Tatbestands von § 44b Abs. 2 UrhG freigestellt ist. Als eine solche vorgelagerte Hilfstätigkeit sollten demnach auch die Vervielfältigungen intermediärer Trainingsdatensammler vom Schrankentatbestand miterfasst sein.

Vervielfältigungen nur dann über § 44b Abs. 2 UrhG freizustellen, wenn sie für ein bereits konkretisiertes, spezifisches Text und Data Mining-Projekt bestimmt sind, würde schließlich dazu führen, dass die Erstellung von Trainingsdatensätzen nur *on demand* möglich wäre. Insbesondere umfangreiche Trainingsdatensätze, die zum etablierten Standard für die Entwicklung ganzer Systemarten wie beispielsweise Textgenerierungs-KI gehören, könnten nicht im Vorhinein und damit ohne die Dringlichkeit eines spezifischen KI-Entwicklungsprojekts erstellt werden. Es gäbe infolgedessen keine qualitativ bewährte Datenbasis, auf Grundlage derer KI-Systeme diverser Arten entwickelt werden könnten. Das würde die Entwicklung von KI erheblich verlangsamen sowie verteuern und widersprüche damit direkt dem Regelungsziel aus ErwGr. 18 DSM-RL, Innovationen zu fördern.

In diesem Zusammenhang ist auch zu berücksichtigen, dass die Schrankenbestimmung für Text und Data Mining zu wissenschaftlichen Zwecken in § 60d Abs. 1 S. 1 UrhG – abzüglich des semantisch nicht relevanten Artikels „das“ – dieselbe Tatbestandsvoraussetzung wie § 44b Abs. 1 S. 1 UrhG enthält. Ein enges Verständnis, nach dem für jede Vervielfältigungshandlung der Bezug zu einem konkretisierten Text und Data Mining-Projekt erforderlich ist, müsste daher auch im Kontext wissenschaftlicher Datenaggregation der Regelung aus § 60d UrhG zugrunde gelegt werden. Insbesondere für wissenschaftliche Forschung auf Grundlage großer Datensätze bedarf es aber einer Auslagerung der Datensatzerstellung und -verwaltung an hierfür eigens geschaffene Stellen. Denn eben jene Tätigkeiten sind zu ressourcenintensiv, als dass sie von einzelnen Forschern für ihre jeweiligen Forschungsprojekte durchgeführt werden könnten. Insbesondere forschen Personen wie beispielsweise Doktoranden, die nur einen eng begrenzten Zeitraum für ihre Forschungsleistung zur Verfügung haben, wären in ihrer Tätigkeit erheblich beeinträchtigt, wenn sie, anstatt datengetriebene Forschung durchzuführen, ihre Zeit für die Erstellung großer Datensätze aufwenden müssten. Vor diesem Hintergrund ist in der Wissenschaftslandschaft diverser Datenwissenschaften auch praktisch üblich, dass die Datensatzerstellung und -pflege von konkreten Forschungsprojekten ausgelagert erfolgt. Damit wird auch ein besserer Zugang diverser forschender Personen zum Datensatz und damit eine effiziente Ressourcennutzung ermöglicht.

licht. Letztere ist insbesondere auch für öffentliche Forschungseinrichtungen von Bedeutung. Da die Datensammlung im Regelfall außerdem vor einem konkretisierten Forschungsprojekt stattfindet, ist hierfür auch nicht die Aufbewahrungsmöglichkeit aus § 60d Abs. 5 UrhG entscheidend. Aus alldem folgt, dass ein enges Verständnis der Tatbestandsvoraussetzung „für (das) Text und Data Mining“ die datengetriebene Wissenschaft unzumutbar in ihrer Forschungstätigkeit beeinträchtigen würde.

Alles in allem sprechen damit durchgreifende teleologische Gründe insbesondere unter Berücksichtigung des in ErwGr. 18 DSM-RL niedergelegten Zwecks der umgesetzten Richtlinienbestimmung dafür, dass bei der Herstellung von Vervielfältigungen kein spezifisches Text und Data Mining- beziehungsweise KI-Trainings-Projekt gefördert werden muss, damit die Vervielfältigungshandlungen unter § 44b Abs. 2 S. 1 UrhG subsument werden können. Ausreichend ist vielmehr die Zweckbindung der Vervielfältigungshandlung an die Analysetechnik des Text und Data Minings, beispielsweise also die Verwendung der Werke zum KI-Training, im Zeitpunkt der Herstellung des Vervielfältigungsstücks. Die teleologischen Erwägungen sprechen demzufolge auch dafür, dass intermediäre Trainingsdatensammler Vervielfältigungen „für das Text und Data Mining“ im Sinne des § 44b Abs. 2 S. 1 UrhG vornehmen können.

Sobald die Zweckbindung der Vervielfältigungsstücke an das Text und Data Mining aufgehoben wird, ist § 44b Abs. 2 S. 2 UrhG im Weiteren allerdings so auszulegen, dass die Vervielfältigungsstücke ab diesem Zeitpunkt nicht mehr für das Text und Data Mining erforderlich sind. Demzufolge greift im Zeitpunkt der Zweckumwidmung die Löschungspflicht aus § 44b Abs. 2 S. 2 UrhG ein.

#### (4) Zwischenergebnis

Insgesamt führen insbesondere der Zweck der Schrankenbestimmung aus § 44b UrhG respektive Art. 4 DSM-RL sowie die Wortlaute sowohl der unionsrechtlichen Schrankenbestimmung als auch der nationalen Umsetzungsnorm dazu, dass das die Schrankenwirkung begründende Tatbestandsmerkmal der Vervielfältigungen „für das Text und Data Mining“ im Sinne des § 44b Abs. 2 S. 1 UrhG so ausgelegt werden könnte, dass hierfür kein Bezug zu einem bereits konkretisierten Text und Data Mining-Projekt

erforderlich ist.<sup>473</sup> Infolgedessen würden auch intermediäre Trainingsdatensammler in den personellen Anwendungsbereich der Schrankenbestimmung fallen, solange die von ihnen angefertigten Vervielfältigungsstücke im Zeitpunkt ihrer Herstellung allgemein der Ermöglichung des Text und Data Minings und damit insbesondere des KI-Trainings gewidmet sind.<sup>474</sup>

#### (5) Dreistufentest

Fraglich ist allerdings, ob das oben genannte Tatbestandsverständnis und der daraus resultierende weite personelle Anwendungsbereich der Schrankenbestimmung auch mit dem Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL vereinbar wären. Wäre das nicht der Fall, würde dies einer entsprechenden Auslegung des Tatbestandsmerkmals „für das Text und Data Mining“ aus § 44b Abs. 2 S. 1 UrhG entgegenstehen.

Damit die Anwendung der Schrankenbestimmung auch zugunsten intermediärer Trainingsdatensammler erfolgen kann, darf es sich hierbei nur um einen bestimmten Sonderfall (erste Stufe) handeln, im Rahmen dessen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt (zweite Stufe) und die berechtigten Interessen der Rechteinhaber nicht ungebührlich verletzt werden (dritte Stufe). Klärungsbedürftig ist im vorliegenden Fall allein, ob die zweite sowie dritte Stufe der Schutzvorschrift bei der Anwendung der Schrankenbestimmung auf intermediäre Trainingsdatensammler gewahrt wären. Diese münden methodisch in einer umfassenden Gesamtabwägung der betroffenen Urheber- und Nutzerinteressen.<sup>475</sup>

##### (i) Erwägungen zugunsten der betroffenen Urheber

Gegen eine Vereinbarkeit der Freistellung mit dem Dreistufentest spricht zunächst, dass die Vervielfältigungen intermediärer Trainingsdatensammler unter Umständen der Vorbereitung rechtswidriger Nutzungen dienen.

---

<sup>473</sup> In diese Richtung wohl auch *Leistner*, GRUR 2024, 1665 (1667), der eine Privileierung zur Erstellung von Basisdatensammlungen begrüßt. Andere Ansicht zu § 60d UrhG a. F.: *Stieper*, in: *Schricker/Loewenheim*, Urheberrecht, UrhG § 60d Rn. 13.

<sup>474</sup> A. A. wohl *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44b Rn. 7; wie hier wohl implizit *Dworschak*, in: *Fromm/Nordemann*, Urheberrecht, UrhG § 44b Rn. 21.

<sup>475</sup> Dazu methodisch ausführlicher siehe oben unter 3. Teil A. III. 2. d).

Denn die Schrankenbestimmung umfasst weder die Verbreitung noch die öffentliche Wiedergabe der Werke zum Zwecke des Text und Data Minings.

Darüber hinaus intensiviert die Tätigkeit intermediärer Trainingsdatensammler die Nutzung betroffener Werke als KI-Trainingsdaten. Hieraus könnte auch eine intensivierte wirtschaftliche Beeinträchtigung der Urheber folgen. Zudem sind am Sachverhalt beim Hinzutreten eines intermediären Trainingsdatensammlers zwei (zumeist) wirtschaftlich tätige Akteure – der vorgelagert tätige intermediäre Trainingsdatensammler einerseits sowie der KI-Entwickler als dessen Abnehmer andererseits – beteiligt. Aus der Verwendung eines Werks als Trainingsdatum wird in diesen Fallkonstellationen deswegen mehr Wert geschöpft.

## (ii) Erwägungen zugunsten intermediärer Trainingsdatensammler

Intermediären Trainingsdatensammlern muss allerdings zugutegehalten werden, dass ein zwingender Schluss von ihren Vorbereitungshandlungen, also den in Rede stehenden Vervielfältigungen, auf eine nachfolgende, rechtswidrige Nutzung nicht gezogen werden kann. Zwar wirkt die Ausgestaltung des § 44b UrhG in Bezug auf ihr Geschäftsmodell prohibitiv.<sup>476</sup> Das bedeutet aber noch nicht unweigerlich, dass ihre Betätigungen in Urheberrechtsverletzungen münden. Bereits im Rahmen dieser Arbeit konnte ein Raum für urheberrechtskonforme Geschäftsmodelle aufgezeigt werden.<sup>477</sup> Auch erste Gerichtsverfahren zur Anwendung der Schrankenbestimmung aus § 44b UrhG auf die Erstellung von Trainingsdatensätzen liefern Ansatzpunkte hierfür.<sup>478</sup> Letztendlich liegt es in der Verantwortung des intermediären Trainingsdatensammlers, auf Grundlage des urheberrechtlichen Möglichkeitsraums rechtskonforme Geschäftsmodelle zu entwickeln. Es sind keine Gründe ersichtlich, warum ihm dies mittels einer pauschalen Schuldvermutung von vornherein versagt werden sollte.

In diesem Zusammenhang kann auch festgestellt werden, dass die Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a Nr. 1 UrhG die Rechtmäßigkeit der Vervielfältigungshandlung eines Vermittlers von der Rechtmäßigkeit der Werknutzung desjenigen Dritten, der den Vermittlungsdienst nutzt, entkoppelt. Hieran wird deutlich,

---

<sup>476</sup> Siehe dazu unten unter 3. Teil A. III. 2. g) bb) (7).

<sup>477</sup> Siehe dazu unten unter 3. Teil A. III. 2. g) bb) (7).

<sup>478</sup> Siehe LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

dass technische Vorbereitungshandlungen nach dem UrhG nicht unweigerlich das Schicksal der Rechtmäßigkeit nachgelagerter Nutzungshandlungen teilen. Dies muss auch in Ein-Personen-Konstellationen und damit auch für die Tätigkeit intermediärer Trainingsdatensammler gelten.

Gegen eine ungebührliche Verletzung der Urheberinteressen spricht weiterhin, dass das UrhG den wirtschaftlichen Wert des Werkes als Trainingsdatum dem urheberrechtlichen Ausschließlichkeitsrecht durch die Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) entzogen hat. Auch die Tätigkeit intermediärer Trainingsdatensammler beschränkt sich darauf, diesen Wert zu realisieren. Die Wertschöpfung erstreckt sich insbesondere nicht auf andere Verwertungsmöglichkeiten des Werkes. Vor diesem Hintergrund kann es keinen Unterschied machen, ob der Wert allein durch KI-Entwickler oder eben auch durch intermediäre Trainingsdatensammler realisiert wird. Zudem werden die Werke auch nur „für das KI-Training“ angefertigt. Damit ordnet sich die Nutzung intermediärer Trainingsdatensammler als dieselbe urheberrechtliche Nutzungsart ein wie die Nutzung durch KI-Entwickler. Weiterhin sind auch von der Tätigkeit intermediärer Trainingsdatensammler nicht die Kernverwertungsarten der betroffenen Urheber berührt.

Ohnehin können KI-Entwickler selbst in personeller Hinsicht uneingeschränkt von der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) profitieren und eigenhändig Vervielfältigungen von Werken für das Training von KI herstellen. Die Tätigkeit intermediärer Trainingsdatensammler beschränkt sich darauf, diese bereits freigestellten Nutzungen zu ermöglichen. Das kann auch als vorgelagerte Hilfstätigkeit eingeordnet werden, wobei der Einsatz von Hilfspersonen im Urheberrecht grundsätzlich mitprivilegiert ist.<sup>479</sup>

Weiterhin kommt es im Wirtschaftszweig der KI-Entwicklung – wie in vielen anderen Lebensbereichen auch – zur arbeitsteiligen Ausgestaltung von Prozessen. Intermediäre Trainingsdatensammler sind Teil dieser arbeitsteilig organisierten KI-Entwicklung. Sie ermöglichen beispielsweise die Betätigung kleinerer, ressourcenärmerer KI-Entwickler. Sie gehören damit als integraler Bestandteil zum Ökosystem der KI-Entwicklung. Dies bewirkt, dass sie auch vom Innovationsförderungsgedanken aus ErwGr. 18 DSM-RL mitumfasst sind.

Letztlich profitieren von einer Erfassung intermediärer Trainingsdatensammler zudem auch die Urheber mittelbar. Denn eine konsolidierte

---

479 Siehe dazu ausführlich unter 3. Teil A. III. 2. f) bb) (1) (ii) (y).

Marktstruktur erleichtert es ihnen, den de lege lata vorgesehenen Opt-out-Mechanismus aus § 44b Abs. 3 UrhG zu bedienen und die Einhaltung ihrer Nutzungsvorbehaltserklärungen zu überwachen. Intermediäre Trainingsdatensammler, deren Augenmerk auf der Erstellung von Trainingsdatensätzen liegt, könnten zudem mehr Know-How aufbauen und Ressourcen aufwenden, um Nutzungsvorbehalte zu beachten und auch darüber hinaus urheberrechtskonforme Trainingsdatenakquise zu betreiben. Hierzu gehört auch, qualitativ hochwertige Trainingsdatensätze zu erstellen, deren Nutzung nicht zuletzt das Risiko für Urheberrechtsverletzungen im Output eines KI-Systems verringert. Auch eine mögliche gesetzliche Vergütungspflicht könnte de lege ferenda leichter gegenüber wenigen größeren Marktteilnehmern durchgesetzt werden.

#### (6) Schlussfolgerungen

Insgesamt streiten bessere Gründe gegen eine unzumutbare Beeinträchtigung der Urheber. Der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL steht einer Anwendung der Schrankenbestimmung aus § 44b UrhG auf intermediäre Trainingsdatensammler daher nicht entgegen. Im Ergebnis können auch intermediäre Trainingsdatensammler daher von der Schrankenwirkung des § 44b UrhG profitieren, solange die von ihnen angefertigten Vervielfältigungsstücke im Zeitpunkt ihrer Herstellung der Ermöglichung des Text und Data Minings und damit insbesondere des KI-Trainings gewidmet sind.<sup>480</sup>

#### (7) Auswirkungen der Schrankenbestimmung auf die Geschäftsmodelle intermediärer Trainingsdatensammler

Eine vom personellen Anwendungsbereich der Schrankenbestimmung unabhängig zu beantwortende Fragestellung ist, welche urheberrechtskonformen Geschäftsmodelle § 44b UrhG den intermediären Trainingsdatensammlern ermöglicht. Diese der Speicherung von Werken als Trainingsdaten nachgelagerte Problemstellung ist nicht Teil des Forschungsgegenstands der vorliegenden Untersuchung. Anzumerken ist allerdings, dass § 44b UrhG zwar die Vervielfältigung, nicht aber die Verbreitung oder öffent-

---

480 A. A. wohl Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 7; wie hier wohl implizit Dworschak, in: Fromm/Nordemann, Urheberrecht, UrhG § 44b Rn. 21.

liche Wiedergabe der vervielfältigten Werke gestattet. Das schränkt die Möglichkeiten für eine urheberrechtskonforme Betätigung intermediärer Trainingsdatensammler ein. Dennoch verbleiben hier gewisse Spielräume. Denkbar und praktisch verbreitet ist beispielsweise, dass die öffentlich wiedergegebenen Trainingsdatensätze neben Annotationen nur Links zu den erfassten, im Internet abrufbaren Werken enthalten.<sup>481</sup> Diese Linklisten können durch gewinnwirtschaftlich handelnde intermediäre Trainingsdatensammler urheberrechtlich frei wiedergegeben werden, insofern nicht auf rechtswidrig öffentlich zugänglich gemachte Werke verlinkt wird.<sup>482</sup> Die vom intermediären Trainingsdatensammler angefertigten Vervielfältigungsstücke werden in diesem Fall nur zur Herstellung der Annotationen benötigt und können danach wieder gelöscht werden.

Auch das Verbreitungsrecht aus § 17 UrhG wird nicht durch jede Weitergabe hergestellter Trainingsdatensätze berührt. Möglich ist beispielsweise eine konzerninterne Bereitstellung von Trainingsdatensätzen durch konzernzugehörige Dienstleister, die als Intermediäre Trainingsdatensätze für Konzernzwecke zusammenstellen.<sup>483</sup>

In Betracht käme darüber hinaus auch, dass ein Teil des Trainingsprozesses an den intermediären Trainingsdatensammler ausgelagert wird, sodass es nicht zur Verbreitung eines von ihm erstellten, urheberrechtlich relevanten Trainingsdatensatzes im Sinne des § 17 UrhG kommt. Hierfür könnte der auftraggebende KI-Entwickler gegebenenfalls auch Rechenkapazität bereitstellen.

---

481 Ein solcher Sachverhalt lag dem Hamburger Urteil im Fall „Laion“ zugrunde, siehe LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

482 Vgl. zu den Voraussetzungen an eine öffentliche Wiedergabehandlung beim Linking EuGH, Urt. v. 08.06.2016 - C-160/15, GRUR 2016, 1152 Rn. 39 ff. – GS Media; es werden jedoch Überlegungen angestellt, auf die automatisierte Erstellung von Linklisten gewisse Privilegierungen anzuwenden, die für Suchmaschinenbetreiber entwickelt worden sind, siehe Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 97 Rn. 167 ff.

483 Zu konzerninternen Weitergaben und dem Verbreitungsrecht aus § 17 UrhG siehe BGH, Urt. v. 06.05.1981 - I ZR 92/78, GRUR 1982, 100 (101f.) – Schallplattenexport; Dustmann, in: Fromm/Nordemann, Urheberrecht, UrhG § 17 Rn. 19; Raué, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 17 Rn. 35; Loewenheim, in: Schrieker/Loewenheim, Urheberrecht, UrhG § 17 Rn. 41, jeweils alle m. w. N.

h) Löschungspflicht der Trainingsdatensammler

§ 44b Abs. 2 S. 2 UrhG setzt weiterhin voraus, dass die angefertigten Vervielfältigungsstücke gelöscht werden müssen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.

Vor dem Hintergrund dieser Löschungspflicht ist insbesondere klärungsbedürftig, ob auch intermediaire Trainingsdatensammler praktisch von der Schrankenbestimmung des § 44b UrhG profitieren können. Denn sie sammeln die Daten zunächst nur und bereiten sie zu Trainingsdatenkorpora auf. Erst später werden die Daten einem dritten KI-Entwickler zur Verfügung gestellt, wobei Zeitpunkt der Datenweitergabe und Identität des Abnehmers beim Web Scraping noch nicht bekannt sein müssen. Inwieweit intermediaire Trainingsdatensammler von der Schrankenbestimmung profitieren können, hängt davon ab, wann die gesetzlich geregelte Löschungspflicht eingreift.

aa) Zeitpunkt des Eingreifens der Löschungspflicht

Um den Zeitpunkt zu bestimmen, ab dem die Löschungspflicht für die vervielfältigten Werke in Kraft tritt, kommt es auf den in § 44b Abs. 2 S. 2 UrhG angelegten Bezugspunkt der Verpflichtung zur Löschung der Werke an. Denkbar ist zum einen, § 44b Abs. 2 S. 2 UrhG so auszulegen, dass er sich generell auf das Text und Data Mining als Form der Analysetechnik bezieht. Dies legt auch die Formulierung in Art. 4 DSM-RL nahe, die sich nicht auf *ein konkretes* Text und Data Mining-Vorhaben, sondern nur pauschal auf Text und Data Mining als solches bezieht.<sup>484</sup> Folge dieser Auslegung wäre, dass die Löschungspflicht so lange nicht entstehen würde, wie die Daten grundsätzlich noch für etwaige Mining-Aktivitäten beziehungsweise KI-Training erforderlich sind. Jedenfalls unter der Voraussetzung, dass grundsätzlich noch KI-Systeme trainiert werden, sind Trainingsdatensätze noch so lange für ein potenzielles KI-Training notwendig, wie sie auf dem Trainingsdatenmarkt angeboten werden. Für intermediaire Trainingsdatensammler bedeutete das, dass sie Trainingsdatensätze erstellen und für KI-Entwickler auf unbestimmte Zeit bereithalten könnten. Die durch § 60d Abs. 5 UrhG beziehungsweise Art. 3 Abs. 2 DSM-RL gewährte Aufbewahrungserlaubnis zum Zwecke (nicht-kommerzieller) wissenschaftlicher For-

---

484 Wulf/Löbeth, GRUR 2024, 737 (740).

### III. Eingreifen einer urheberrechtlichen Schrankenbestimmung

schung, die in § 44b UrhG beziehungsweise Art. 4 DSM-RL hingegen nicht vorgesehen ist, legt aber nahe, dass im Umkehrschluss hierzu ein strengeres Verständnis der Löschungspflicht angezeigt ist. Danach bezieht sich § 44b Abs. 2 S. 2 UrhG auf einen konkreten Analysevorgang. Die gespeicherten Werke müssen demzufolge bereits gelöscht werden, sobald sie für diesen Analysevorgang nicht mehr benötigt werden. Das ist dann der Fall, wenn die konkrete Datenanalyse beendet ist.<sup>485</sup> Für das KI-Training bedeutet das, dass die Trainingsdaten gemäß § 44b Abs. 2 S. 2 UrhG unmittelbar nach dem Abschluss eines Trainingsvorgangs gelöscht werden müssen. Nach einer zweckorientierten Auslegung, die die Entwicklung neuer Technologien und Innovationen im Allgemeinen fördert<sup>486</sup> und damit auch Nachbesserungen an den Systemen ermöglichen sollte, ist das jedenfalls dann der Fall, wenn das KI-System in die Anwendungsphase übergeht.<sup>487</sup> Insoweit die Werke noch während der Anwendungsphase des Systems verwendet werden müssen, um am selben KI-System durch fortlaufendes Training Nachbesserungen vorzunehmen, greift die Löschungspflicht erst nach Abschluss dieser Nachbesserungen ein.<sup>488</sup> Da Qualitätssicherungs- und Transparenzinitiativen nicht mehr zur jeweiligen automatisierten Analyse, also dem Training des konkreten KI-Systems, gehören, können die Werke auf Basis von § 44b UrhG hierfür allerdings nicht aufbewahrt werden.

Für intermediäre Trainingsdatensammler bedeutet das, dass sie ihren Trainingsdatensatz gemäß § 44b Abs. 2 S. 2 UrhG vernichten müssen, sobald mit ihm das Training eines KI-Systems abgeschlossen worden ist. Da

---

485 Vgl. zu diesem Ansatz in Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 15.

486 ErwGr. I8 DSM-RL.

487 Für ein solch weites Verständnis aufgrund derselben zweckorientierten Erwägungen auch Novelli u. a., Computer Law & Security Review 2024, (10); wohl außerdem Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 9, denn der Zweck des Minings, also des KI-Trainings, ist erreicht, wenn das KI-System in die Anwendungsphase übergeht; insoweit übereinstimmend, als dass eine zweckorientierte, weitere Auslegung des Tatbestandsmerkmals erfolgen muss, auch Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 22, im Detail aber abweichend; ähnlich ohne direkten KI-Bezug auch Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 15, der im Ergebnis von einem projektbezogenen Verständnis ausgeht und innerhalb eines solchen Projekts auch mehrere Analysevorgänge als rechtmäßig erachtet.

488 Folgerichtig Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 22b; Konertz/Schönhof, WRP 2024, 289 Rn. 32; a. A. Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 15; Wulf/Löbeth, GRUR 2024, 737 (740).

dies dem Sinn des Datensammelns zuwiderläuft, wirkt die Löschungspflicht für intermediäre Trainingsdatensammler prohibitiv.<sup>489</sup> Ähnliche Effekte bestehen für Trainingsdatensätze selbst aufbauende KI-Entwickler, die unter erheblichem Ressourceneinsatz vorbereiteten Trainingsdatensätze gemäß § 44b Abs. 2 S. 2 UrhG ebenfalls nur einmalig verwenden dürfen.

#### bb) Regulatorische Defizite des geltenden Urheberrechts

Die Pflicht zur Löschung von mittels Webcrawlern generierter, urheberrechtlich geschützter Trainingsdaten ist, wie im zweiten Teil dieser Untersuchung gezeigt wurde, nicht erforderlich, um einen angemessenen Ausgleich von Urheber- und Nutzerinteressen herbeizuführen. Werden die als Trainingsdaten genutzten Vervielfältigungsstücke nicht gelöscht, wird mit den Daten darüber hinaus weiteres KI-Training ermöglicht. Die Löschungspflicht beeinträchtigt außerdem Qualitäts- und Nachweisanforderungen bei der Entwicklung von KI.<sup>490</sup> Alles in allem wird die in § 44b Abs. 1 S. 2 UrhG vorgesehene Löschungspflicht den identifizierten Regelungsbedürfnissen damit nicht gerecht.<sup>491</sup> Sie beeinträchtigt die Interessen der Trainingsdatensammler und KI-Entwickler unangemessen. Das gilt insbesondere für intermediäre Trainingsdatensammler, deren Geschäftsmodell durch die Löschungspflicht erheblich gestört wird. De lege lata könnte dem sich aus der Löschungspflicht ergebenden regulatorischen Defizit mit Hilfe eines Erwerbs von ECL für das Training von KI abgeholfen werden.<sup>492</sup>

##### i) Opt-out-Mechanismus für Rechtsinhaber

§ 44b Abs. 3 UrhG sieht für Urheber die Möglichkeit vor, die Freistellungswirkung der Schrankenbestimmung durch Erklärung eines Nutzungsvorbe-

---

489 So auch Kögel, InTeR 2023, 179 (181).

490 Margoni/Kretschmer, GRUR Int. 2022, 685 (693); Vgl. auch Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 23, die das Problem aber durch eine de facto Aufhebung der Löschungspflicht lösen wollen, die die Wortlautgrenze der Regelung sprengt und daher nicht mit den gängigen Auslegungsmethoden begründet werden kann. Mit rechtspolitischer Kritik noch in Bomhard, InTeR 2023, 174 (177).

491 Dass die Löschungspflicht für den Umgang mit Trainingsdaten für KI problematisch sei, betont auch de la Durantaye, ZUM 2023, 645 (659). Für eine Streichung plädiert auch Schack, NJW 2024, 113 Rn. 10.

492 Dazu im Detail siehe Pukas, GRUR 2023, 614.

halts zu beenden. Bei online zugänglichen Werken muss die Ausschlusserklärung gemäß § 44b Abs. 3 S. 2 UrhG hierfür in maschinenlesbarer Form erfolgen.<sup>493</sup>

Wie im zweiten Teil der vorliegenden Untersuchung gezeigt werden konnte, ist aus einer regulatorischen Perspektive kein Opt-out-Mechanismus für Urheber vorzusehen. Ein angemessener Ausgleich zwischen Urheber- und Nutzerinteressen kann vielmehr auch ohne Ausschlussmöglichkeit, sondern durch eine gesetzliche Vergütungspflicht sichergestellt werden. Der in § 44b Abs. 3 UrhG geregelte Opt-out-Mechanismus wird den identifizierten Freistellungsbedürfnissen demzufolge nicht gerecht.<sup>494</sup> Wie stark der Widerspruch zwischen den herausgearbeiteten Regelungsbedürfnissen und den geltenden urheberrechtlichen Rahmenbedingungen ausgeprägt ist, hängt im Folgenden von den konkreten Voraussetzungen des Nutzungsvorbehalts und seiner Wirkung ab.

#### aa) Maschinenlesbarkeit des Nutzungsvorbehalts

Damit ein Nutzungsvorbehalt bei online zugänglichen Werken wie den vom Web Scraping urheberrechtlich geschützter Trainingsdaten betroffenen wirksam ist, muss er gemäß § 44b Abs. 3 S. 2 UrhG in maschinenlesbarer Form erfolgen. Im UrhG wird allerdings nicht näher konkretisiert, was hierunter zu verstehen ist. Der Rechtsbegriff ist daher auslegungsbedürftig.

##### (1) Notwendigkeit einer digitalen Hinterlegung der Informationen

Fraglich ist dabei zunächst, ob die mit dem Nutzungsvorbehalt übermittelten Informationen digital in einer Art und Weise hinterlegt sein müssen, die ihre unmittelbare Verarbeitung durch einen Computer ermöglicht, damit sie im Sinne des § 44b Abs. 3 S. 2 UrhG maschinenlesbar sind. Ist das nicht der Fall, genügt eine mittelbare maschinelle Verarbeitbarkeit, die erst durch eine Digitalisierung des Erklärungsinhalts (z. B. Einscannen einer handschriftlichen Notiz) erreicht wird.

---

493 Zu den Anforderungen an den Nutzungsvorbehalt siehe sogleich unter 3. Teil A. III. 2. i) aa).

494 So konkret zu Art. 4 DSM-RL auch *Chiou*, JIPITEC 2019, 398 Rn. 23 f.

Durch erhebliche Fortschritte im Bereich maschinellen Lernens<sup>495</sup> können nahezu alle auch nicht elektronisch dargestellten Informationen wie handgeschriebene Texte digitalisiert und damit mittelbar maschinell verarbeitet werden.<sup>496</sup> Aufgrund leistungsfähiger Texterkennungssoftware gilt das insbesondere für analoge Textdaten. Bilderkennungssystemen versetzen ihre Nutzer zudem in die Lage, Bildinhalte semantisch zu erfassen. Würde eine mittelbare maschinelle Verarbeitbarkeit der Informationen ausreichen, um Maschinenlesbarkeit im Sinne des § 44b Abs. 3 UrhG herbeizuführen, bedeutete dies, dass faktisch nahezu alle Vorbehaltserklärungen unabhängig von ihrer Verkörperungsform maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG wären.<sup>497</sup> In diesem Falle würde das Tatbestandsmerkmal der Maschinenlesbarkeit weitestgehend leerlaufen. Bereits denklogisch muss es aber einen Unterschied zwischen maschinenlesbaren und nicht maschinenlesbaren Erklärungen geben.<sup>498</sup> Darüber hinaus kann vermutet werden, dass der Normgeber dem Tatbestandmerkmal mit seiner Aufnahme in den Gesetzestext eine eigenständige Bedeutung verliehen hat.

Beides spricht dafür, dass ein solch weites Verständnis des Tatbestandsmerkmals der Maschinenlesbarkeit nicht intendiert ist. Insgesamt führen damit teleologische Erwägungen dazu, dass der Inhalt einer Nutzungsvorbehaltserklärung digital in einer Art und Weise hinterlegt werden muss, die eine unmittelbare Verarbeitung durch einen Computer ermöglicht, damit er als maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG gilt.<sup>499</sup> Analoge Erklärungen wie handgeschriebene Texte, die digitalisiert werden können, erfüllen die Voraussetzung der Maschinenlesbarkeit daher nicht.

## (2) Bestehen eines Formalisierungs- und Strukturierungserfordernisses

Klärungsbedürftig ist darüber hinaus aber auch, ob die Nutzungsvorbehaltserklärung in einem formalisierten und strukturierten Format, also in einer Programmier- oder vergleichbar formalisierten Sprache abgegeben werden muss, damit die Voraussetzung der Maschinenlesbarkeit im Sinne des § 44b Abs. 3 S. 1 UrhG erfüllt ist. Sofern das nicht notwendig ist, könnte

---

495 Vgl. *Maslej u. a.*, The AI Index 2023 Annual Report by Stanford University, S. 54 ff.

496 Vgl. zur Sprachverarbeitung *Menzel*, in: Görz/Schmid/Braun, Handbuch der Künstlichen Intelligenz, S. 601.

497 *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; *Bomhard*, InTeR 2023, 174 (178).

498 Ebenso *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332.

499 A. A. *Wulf/Löbeth*, GRUR 2024, 737 (739 f.), wohl bedingt durch ein grammatisches Missverständnis des § 44b Abs. 3 UrhG zugrundeliegenden Art. 4 Abs. 3 DSM-RL und entgegen dem insoweit eindeutigen Wortlaut des § 44b Abs. 3 UrhG.

eine maschinenlesbare Vorbehaltserklärung auch wirksam in natürlicher Sprache abgegeben werden, sofern die Erklärung unmittelbar durch einen Computer verarbeitet werden kann. Hierfür genügt es, wenn sie in elektronischer Form, also insbesondere als elektronischer Text, verkörpert ist.

Auch diese Frage muss im Wege der Auslegung des Tatbestandsmerkmals der Maschinenlesbarkeit aus § 44b Abs. 3 S. 2 UrhG beantwortet werden.<sup>500</sup> Im Rahmen dieser Auslegung ist die Bedeutung des Tatbestandsmerkmals dabei insbesondere im Lichte des Art. 4 Abs. 3 DSM-RL und des hier verwendeten Begriffs der Maschinenlesbarkeit zu bestimmen. Seiner Umsetzung dient § 44b Abs. 3 S. 2 UrhG.<sup>501</sup> Weiterhin zu berücksichtigen ist, dass es für die Erklärung von Nutzungsvorbehalten in natürlicher Sprache keine einheitliche Formulierung gibt. Anders als beispielsweise § 312j Abs. 3 S. 2 BGB gibt das UrhG dem Erklärenden keinen Erklärungswortlaut vor. Aufgrund ihrer Varianz bieten natürlichsprachliche Vorbehaltserklärungen damit keine vollständige Erkennungssicherheit.<sup>502</sup> Folge dessen ist, dass die Auslegung des Maschinenlesbarkeitsbegriffs auch maßgeblich davon mitbeeinflusst wird, wieviel Erkennungssicherheit mit den Nutzungsvorbehaltserklärungen erreicht werden muss.

- (i) Natürlicher Sprachgebrauch des Begriffs „maschinenlesbar“ bzw. „machine-readable“

Ausgangspunkt der Auslegung des Rechtsbegriffs der Maschinenlesbarkeit ist der allgemeine Sprachgebrauch. Da die Computerwissenschaft und -anwendung erheblich durch die englische Sprache geprägt ist, muss hierbei auch das Begriffsäquivalent „machine-readable means“ mit in den Blick genommen werden. Es wird in der englischen Sprachfassung des durch § 44b Abs. 3 UrhG umgesetzten Art. 4 Abs. 3 DSM-RL genutzt. Maschinenlesbar beziehungsweise machine-readable sind Daten nach einem teilweise zugrunde gelegten Begriffsverständnis dann, wenn sie in einer Form gespeichert sind, die von einem Computer verarbeitet werden kann.<sup>503</sup> Volltextsuchen und Suchmaschinen zeigen bereits praktisch, dass eine Verarbeitung von digital hinterlegten, natürlich-sprachlichen Textinhalten durch Computer ohne weiteres möglich ist. Dem entsprechend könnten auch in

---

500 Vgl. hierzu ausführlich auch Hamann, ZGE 2024, II13.

501 BT-Drs. 19/27426, S. 88.

502 Maamar, ZUM 2023, 481 (484).

503 Waite, Oxford English Dictionary, Begriff: machine-readable.

natürlicher Sprache abgegebene Nutzungsverhältniserklärungen unter den Begriff der Maschinenlesbarkeit (beziehungsweise der machine-readability) gefasst werden.

Das aufgezeigte Verständnis wird dem Begriff allerdings nicht einhellig zugrunde gelegt. Im natürlichen Sprachgebrauch ist vielmehr auch eine Interpretation verbreitet, nach der die Daten in strukturierter Form gespeichert sein müssen, sodass sie durch ein einheitliche und gleichbleibende Verarbeitungslogik anwendendes Computerprogramm verarbeitet werden können.<sup>504</sup> Auch wenn sie in einem digitalen Format vorgehalten werden, können hierunter mangels einer Strukturierung der Daten keine Erklärungen in natürlicher Sprache subsumiert werden. Beide Verständnismöglichkeiten zeigen, dass sich im natürlichen Sprachgebrauch kein einheitliches Verständnis des Begriffs der Maschinenlesbarkeit beziehungsweise der machine-readability herausgebildet hat. Die dargelegten Interpretationsmöglichkeiten entscheiden sich insbesondere im Hinblick auf die Notwendigkeit einer Strukturierung der Daten. Gerade hiernach richtet sich aber, ob eine Vorbehaltserklärung in natürlicher Sprache, also unstrukturiert abgegeben werden kann, oder nicht. Aus dem natürlichen Sprachgebrauch können daher keine Schlussfolgerungen für die Auslegung des Rechtsbegriffs der Maschinenlesbarkeit abgeleitet werden.

#### (ii) Begriffsverwendung im nationalen Regelungskontext

Der Begriff der Maschinenlesbarkeit wird allerdings in § 12 Abs. 1 S. 2 EGovG legaldefiniert. Danach sind Informationen maschinenlesbar, wenn sie durch Software automatisiert ausgelesen und verarbeitet werden können. Eine vergleichbare Legaldefinition ist auch in § 3 Nr. 5 DNG normiert. Wie die Möglichkeit zur Volltextsuche und das Suchmaschinenindexing, bei dem nicht strukturierte und formalisierte Webseiteninhalte systematisch und automatisiert ausgewertet werden, praktisch zeigen, bedarf es für eine Verarbeitung digitaler Informationen durch (automatisierte) Software keiner Strukturierung der betroffenen Inhalte. Informationen müssten für eine Maschinenlesbarkeit in diesem Sinne demzufolge digital hinterlegt, nicht aber strukturiert oder formalisiert abrufbar sein. Ein Vergleich zu den gleichlautenden Tatbestandsvoraussetzungen in EGovG und DNG spricht

---

504 Glossar des National Institute of Standards and Technology (U.S. Department of Commerce), Begriff: machine-readable, siehe [https://csrc.nist.gov/glossary/term/Machine\\_Readable](https://csrc.nist.gov/glossary/term/Machine_Readable) (zuletzt abgerufen am 14.4.2024).

deswegen dafür, das Tatbestandsmerkmal der Maschinenlesbarkeit auch in § 44b Abs. 3 S. 2 UrhG so auszulegen, dass zwar eine digitale Hinterlegung, aber keine Strukturierung oder Formalisierung der Vorbehaltserklärungen notwendig ist.<sup>505</sup>

(iii) Einheitliches europäisches Begriffsverständnis

Im Übrigen geht § 44b Abs. 3 UrhG aber auf Art. 4 Abs. 3 DSM-RL zurück und muss daher insbesondere richtlinienkonform ausgelegt werden. Auch in Art. 4 Abs.3 DSM-RL ist der Begriff der Maschinenlesbarkeit zwar nicht ausdrücklich definiert. Es könnte sich aber bereits ein einheitliches europäisches Begriffsverständnis herausgebildet haben,<sup>506</sup> dass auch dem Rechtsbegriff der Maschinenlesbarkeit aus Art. 4 Abs. 3 DSM-RL und damit richtlinienkonform auch § 44b Abs. 2 S. 2 UrhG zugrunde gelegt werden kann. Um dies zu ermitteln, muss herausgearbeitet werden, in welchen Gemeinschaftsrechtsakten der Begriff der Maschinenlesbarkeit verwendet wird und ob den Rechtsbegriffen in den Regelungswerken ein einheitliches Verständnis zugrunde liegt.

(a) Maschinenlesbarkeit in der PSI-Richtlinie

Maschinenlesbarkeit ist zum einen in Art. 2 Nr. 13 RL (EU) 2019/1024 (PSI-RL) definiert. Er liegt dem bereits an anderer Stelle herangezogenen<sup>507</sup> § 3 Nr. 5 DNG zugrunde. Gemäß Art. 2 Nr. 13 PSI-RL ist ein Dateiformat maschinenlesbar, wenn es so strukturiert ist, dass Softwareanwendungen konkrete Daten, einschließlich einzelner Sachverhaltsdarstellungen und deren interner Struktur, leicht identifizieren, erkennen und extrahieren können. Unabhängig davon, welche Voraussetzungen erfüllt werden müssen, damit eine leichte Identifizierung und Extrahierung ermöglicht wird, setzt Art. 2 Nr. 13 PSI-RL eine Datenstrukturierung voraus. Eine solche Strukturierung wird durch Erklärungen in natürlicher Sprache nicht gewährleistet.

---

505 A. A. Hamann, ZGE 2024, 113 (131).

506 Vgl. für eine einheitliche Auslegung von Rechtsbegriffe innerhalb des Gemeinschaftsrechts EuGH, Urt. v. 27.01.2005 - C-188/03, NZA 2005, 213 Rn. 29 – Junk; EuGH, Urt. v. 27.02.2003 - C-373/00, EuZW 2003, 315 Rn. 35.

507 Dazu siehe unmittelbar oben in diesem Abschnitt.

(β) Maschinenlesbarkeit in der Richtlinie zur Nachhaltigkeitsberichterstattung von Unternehmen

Der Begriff der Maschinenlesbarkeit wird darüber hinaus in ErwGr. 55 RL (EU) 2022/2464 (CSRD) (Regelungsziel aufgegriffen durch Art. 29d CSRD) verwendet. Die Richtlinie regelt die Nachhaltigkeitsberichterstattung von Unternehmen. Die Dokumentation der Unternehmen soll danach maschinenlesbar sein. Zur näheren Ausgestaltung wird auf Art. 3 VO (EU) 2019/815 (Delegierte Verordnung) verwiesen. Aus diesem ergibt sich wiederum, dass die Nachhaltigkeitsberichte im xhtml-Format veröffentlicht werden müssen. Das xhtml-Format sieht dabei eine gewisse Formalisierung durch die xhtml-Syntax vor. Eine Strukturierung der Erklärungsinhalte findet hierdurch allerdings nicht statt. Der Erklärungsinhalt wird, eingebettet in die xhtml-Syntax, in natürlicher Sprache wiedergegeben. Der Begriff der Maschinenlesbarkeit, welcher in ErwGr. 55 CSRD verwendet wird, setzt demnach zwar eine formalisierte Einbettung des Erklärungsinhalts voraus. Darüber hinaus lässt er aber ausreichen, wenn der eigentliche Inhalt der Erklärung in natürlicher Sprache vorgehalten. Eine Strukturierung der Erklärung, also eine Übersetzung „in Code“, ist nicht notwendig.

(γ) Maschinenlesbarkeit in der KI-Verordnung

Auch in der KI-VO wird das Tatbestandsmerkmal der Maschinenlesbarkeit verwendet. ErwGr. 133 adressiert die maschinenlesbare Kennzeichnung von KI-generierten Inhalten. Als Beispiel für eine maschinenlesbare Kennzeichnungstechnik werden unter anderem Wasserzeichen genannt. Hierbei handelt es sich um eine Markierungstechnik, bei der selbstgewählte, identifizierbare Zeichen direkt in den Inhalt, beispielsweise ein Lichtbild, eingearbeitet werden. Im Gegensatz zur Einbettung von Informationen in Metadaten werden die Zeichen dabei unmittelbar mit dem Inhalt verflochten. Die Wasserzeichen können, müssen aber nicht mit dem bloßen Auge wahrnehmbar sein. Sie folgen keiner strukturierten oder formalisierten Logik. Sie sind nicht codiert. Vielmehr werden oft verwendertypische Zeichen wie Bildmarken oder zufällige Rauschmuster eingesetzt.

Darüber hinaus verpflichtet Art. 47 KI-VO Anbieter von Hochrisiko-KI-Systemen, maschinenlesbare EU-Konformitätserklärung auszustellen. Diese Erklärungen können physisch oder maschinell unterzeichnet werden (Abs. 1) und müssen in einer Sprache verfügbar sein, die für die jeweils zuständige, nationale Aufsichtsbehörde leicht verständlich ist (Abs. 2). Aus der

gesetzlichen Sprachanforderung sowie der Möglichkeit, die Erklärungen auch physisch zu unterzeichnen, folgt, dass die EU-Konformitätserklärungen nach Art. 47 KI-VO in natürlicher Sprache abgegeben werden müssen. Maschinenlesbarkeit im Sinne von Art. 47 KI-VO setzt demzufolge ebenso wenig eine Strukturierung wie eine Formalisierung der Erklärungsinhalte voraus. Vielmehr ist eine Erklärung in natürlicher Sprache ausreichend. Insgesamt setzt der Begriff der Maschinenlesbarkeit innerhalb der KI-VO keine Codierung der jeweiligen Inhalte voraus.

#### (δ) Maschinenlesbarkeit in der Datenschutz-Grundverordnung

Das Tatbestandsmerkmal der Maschinenlesbarkeit wird außerdem in der DS-GVO aufgegriffen. Gemäß Art. 12 Abs. 7 DS-GVO können bei der Datenerhebung zu übermittelnde Informationen auch in Form standardisierter Bildsymbole bereitgestellt werden. In elektronischer Form müssen diese Bildsymbole gemäß Art. 12 Abs. 7 S. 2 DS-GVO maschinenlesbar sein. Dem Regelungszweck der Herstellung von Barrierefreiheit zugunsten sehbehinderter Menschen<sup>508</sup> entsprechend ist hierunter zu verstehen, dass die Symbole computergestützt vorlesbar sein müssen.<sup>509</sup> Hierfür bedarf es zwar einer gewissen Formalisierung, aber keiner Strukturierung der vorzulesenden Textinhalte. Vielmehr genügt auch eine elektronisch auslesbare Erklärung in natürlicher Sprache.

Nach Art. 20 DS-GVO haben Personen außerdem das Recht, die sie betreffenden personenbezogenen Daten in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten. Die Norm erfasst die Maschinenlesbarkeit und die Strukturierung der Daten als eigenständige Merkmale. Umgekehrt folgt hieraus, dass die Voraussetzung der Maschinenlesbarkeit aus Art. 20 DS-GVO selbst ebenfalls keine Strukturierung der Daten notwendig macht.

---

508 Quaas, in: Wolff/Brink/v. Ungern-Sternberg, Beck'scher Online-Kommentar Datenschutzrecht, DS-GVO Art. 12 Rn. 54; Heckmann/Paschke, in: Ehmann/Selmayr, Datenschutz-Grundverordnung, DS-GVO Art. 12 Rn. 55.

509 Vgl. Heckmann/Paschke, in: Ehmann/Selmayr, Datenschutz-Grundverordnung, DS-GVO Art. 12 Rn. 55.

(€) Schlussfolgerung zur Existenz eines einheitlichen europäischen Begriffsverständnisses

Insgesamt wird der Rechtsbegriff der Maschinenlesbarkeit innerhalb der europäischen Rechtsordnung an verschiedenen Stellen aufgegriffen. Teilweise setzt er dabei eine Strukturierung und Formalisierung der übermittelten Informationen, teilweise aber auch weder eine Strukturierung noch eine Formalisierung der Inhalte voraus. Mitunter wird vom Normgeber sogar eine Erklärung in natürlicher Sprache implizit vorausgesetzt. Hieraus folgt, dass sich kein einheitliches europäisches Verständnis des Rechtsbegriffs der Maschinenlesbarkeit herausgebildet hat, welches Art. 4 Abs. 3 DSM-RL und im Wege richtlinienkonformer Auslegung auch § 44b Abs. 3 UrhG zugrunde gelegt werden kann.<sup>510</sup> Insbesondere kann damit auch keine Parallele zur Legaldefinition der Maschinenlesbarkeit aus Art. 2 Nr. 13 PSI-RL gezogen werden.<sup>511</sup> Eine Parallele gerade zu dieser Regelung zu ziehen, wäre willkürlich. Außerdem dient sie der Erfüllung eines anderen Zweckes als Art. 4 DSM-RL.<sup>512</sup> Mangels eines einheitlichen europäischen Begriffsverständnisses ist die Bedeutung des Rechtsbegriffs der Maschinenlesbarkeit in Art. 4 Abs. 3 DSM-RL vielmehr rechtsaktsautonom zu bestimmen. Hierfür kann sich insbesondere auch am Telos der Regelung orientiert werden.

(iv) Telos der Regelung

Zweck der Voraussetzung der Maschinenlesbarkeit aus Art. 4 Abs. 3 DSM-RL respektive § 44b Abs. 3 S. 2 UrhG ist, eine automatisierte Vervielfältigung von Werken zu ermöglichen.<sup>513</sup> Dafür müssen die Nutzungsvorbehalte in einer Art und Weise erklärt werden, die es automatisierten Webcrawlern ermöglicht, den Erklärungsinhalt zu erfassen und zu verarbeiten. Dass ein automatisiertes Auffinden und infolgedessen eine Verarbeitung von Textinhalten in natürlicher Sprache technisch möglich ist, zeigen bereits

---

510 A. A. Griffiths/Synodinou/Xalabarder, GRUR Int. 2023, 22 (29); Bomhard, David/Siglmüller, RDi 2024, 45 (50) Fn. 45.

511 Wie hier auch LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 69; a. A. Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 31.

512 Vgl. dazu ausführlicher LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 69.

513 Auch Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 10; BT-Drs. 19/27426.

der Betrieb von Suchmaschinen und die Suchfunktion in Dokumenten. Auch HTML-Code kann mit Hilfe simpler Skripte automatisiert ausgewertet und nach bestimmten Sprachformulierungen durchsucht werden. Solche Suchalgorithmen können auch für die Verarbeitung von Inhalten genutzt werden, die mit Hilfe automatisierter Webcrawler für das Training von KI vervielfältigt werden sollen. Eine automatisierte Verarbeitung ist somit möglich, solange die Daten digital hinterlegt und insofern durch Maschinen verarbeitbar sind. In natürlicher Sprache abgegebene Nutzungs- vorbehaltserklärungen stehen damit einer automatisierten Vervielfältigung aus einer technischen Perspektive nicht per se entgegen.<sup>514</sup> Dass die Informationen strukturiert und formalisiert vorgehalten werden, ist für eine automatisierte Verarbeitung aus einer technischen Perspektive demzufolge nicht notwendig.<sup>515</sup> Die Frage nach der Angemessenheit der konkreten Nutzungsvorbehaltserklärung im weiteren Sinne ist dabei von der Frage der Maschinenlesbarkeit der Erklärung zu trennen.<sup>516</sup>

Nach ErwGr. 18 UAbs. 2 S. 2 DSM-RL sollen darüber hinaus in Allgemeine Geschäftsbedingungen einer Webseite eingebettete Nutzungsvorbehaltserklärungen die Voraussetzungen des Art. 4 Abs. 3 DSM-RL erfüllen können. Zwar bezieht sich der Erwägungsgrund ausweislich seiner Formulierung nur auf in den Allgemeinen Geschäftsbedingungen einer Webseite eingebettete Vorbehaltserklärungen, *sofern* diese maschinenlesbar sind. Daraus kann aber nicht geschlussfolgert werden, dass im Umkehrschluss aus ErwGr. 18 UAbs. 2 S. 2 DSM-RL Vorbehaltserklärungen in natürlicher Sprache grundsätzlich nicht maschinenlesbar seien.<sup>517</sup> Denn es folgt aus der allgemeinen Lebenserfahrung, dass Erklärungen innerhalb von Allgemeinen Geschäftsbedingungen einer Webseite so wie die Vertragsbestimmungen selbst in natürlicher Sprache abgegeben werden. Das muss auch dem Richtliniengeber bei der Formulierung des ErwGr. 18 DSM-RL bewusst gewesen sein. Dass nach ErwGr. 18 UAbs. 2 S. 2 DSM-RL Vorbehaltserklärungen in Allgemeinen Geschäftsbedingungen einer Webseite grundsätzlich geeignet sein können, die Voraussetzungen aus Art. 4 Abs. 3 DSM-RL zu erfüllen, spricht demzufolge dafür, dass die Abgabe der Vorbehaltserklärung in natürlicher Sprache der Maschinenlesbarkeit im Sinne des Art. 4 Abs. 3

---

514 Auch Leistner verweist auf die Leistungsfähigkeit moderner Sprachverarbeitungssysteme, siehe *Leistner*, GRUR 2024, 1665 (1673).

515 A. A. Maamar, ZUM 2023, 481 (484); Jacobsen/Hartmann, MMR-Aktuell 2021, 441332; wohl auch Hamann, ZGE 2024, 113 (131).

516 Zur Angemessenheit sogleich unter 3. Teil A. III. 2. i) bb).

517 So aber Hamann, ZGE 2024, 113 (147).

DSM-RL jedenfalls nicht entgegenstehen kann. Ansonsten wäre ErwGr. 18 UAbs. 2 S. 2 DSM-RL gegenstandslos. Umgekehrt kann damit auch aus ErwGr. 18 UAbs. 2 S. 2 DSM-RL abgeleitet werden, dass eine digitale Hinterlegung des Nutzungsvorbehalts ausreichend sein muss. Eine Strukturierung oder Formalisierung der Informationen wäre demnach nicht erforderlich.

Für ein Formalisierungs- und Strukturierungserfordernis der Vorbehalts erklärun gen könnte jedoch sprechen, dass eine formalisierte und strukturierte Informationsübermittlung, beispielsweise in Form von HTML-Metadaten,<sup>518</sup> zur größtmöglichen Erkennungs- und damit Rechtssicherheit für den Erklärungsempfänger führt. Allerdings bezweckt Art. 4 Abs. 3 DSM-RL respektive § 44b Abs. 3 UrhG nicht, die größtmögliche Rechtssicherheit für Trainingsdatensammler herbeizuführen. Um Abgrenzungsschwierigkeiten im Anwendungsbereich der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen, Art. 5 Abs. 1 InfoSoc-RL respektive § 44a UrhG zu beenden,<sup>519</sup> sollte Rechtssicherheit lediglich durch das Ob der Freistellung von Vervielfältigungshandlungen zum Text und Data Mining, nicht aber den Opt-out-Mechanismus im Speziellen herbeigeführt werden.

In diesem Zusammenhang spricht außerdem gegen die Notwendigkeit einer Formalisierung oder Strukturierung der Vorbehalts erklärun gen, dass ohnehin nicht jede Software alle formalisierten Erklärungsmuster auslesen und verarbeiten kann. Ein Trainingsdatensammler, der HTML-Metadaten auswertet, nimmt beispielsweise keine Vorbehalts erklärun zur Kenntnis, die mittels des Robots Exclusion Standards abgegeben worden ist. Ohne gesetzlich normierten Erklärungsstandard besteht selbst bei einer Strukturierung und Formalisierung der Erklärungsmuster damit immer das Risiko, dass Nutzungsvorbehalts erklärun gen nicht ausgelesen und verarbeitet werden können.<sup>520</sup> Unabhängig davon, ob eine Strukturierung und Formalisierung der Vorbehalts erklärun gen erfolgt, oder ob der Nutzungsvorbehalt in natürlicher Sprache erklärt wird, gibt es demzufolge keine vollständige Erkennungssicherheit für Trainingsdatensammler. Ein Grundrisiko ist damit immer vom Datensammler zu tragen. Eine Formalisierung oder Strukturierung führt nicht zwangsläufig zur Eliminierung dieses Risikos.

Darüber hinaus müssen Anbieter von KI-Modellen mit allgemeinem Verwendungszweck nach Art. 53 Abs. 1 lit. c KI-VO eine Strategie auf den

---

518 Mit einem Überblick über diese Technologie Hamann, ZGE 2024, II3 (157 ff.).

519 Vgl. ErwGr. 18 UAbs. 1 DSM-RL.

520 Wulf/Löbeth, GRUR 2024, 737 (738).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

Weg bringen, auf Grundlage derer Nutzungsverbehalte im Sinne des § 44b Abs. 3 UrhG ausgelesen und beachtet werden sollen. Hierfür muss gemäß der gesetzlichen Anforderung auch auf „modernste Technik“ zurückgegriffen werden. Diese wird aber nur dann benötigt, wenn in natürlicher Sprache erklärte Vorbehaltserklärungen digital erfasst und semantisch verarbeitet werden sollen.<sup>521</sup> Auch das spricht im Ergebnis gegen eine Formalisierungs- und Strukturierungsnotwendigkeit der Vorbehaltserklärungen.

Zweck des Opt-out-Mechanismus aus Art. 4 Abs. 3 DSM-RL ist außerdem nicht, größtmögliche Sicherheit für Trainingsdatensammler herzustellen. Der Opt-out-Mechanismus dient vorrangig der Wahrung der Interessen der Rechtsinhaber. Dem Interesse der Rechtsinhaber wird umgekehrt aber eine niederschwellige Möglichkeit zur Erklärung des Nutzungsverbehalts gerecht. Eine solche wird durch die Erklärbarkeit von Nutzungsverbehalten in natürlicher Sprache gewährleistet.

Gegen das Erfordernis der Formalisierung und Strukturierung von Vorbehaltserklärungen spricht zudem, dass dies zu einem Ausschluss solcher Rechtsinhaber führen würde, die nicht in der Lage sind, die Strukturierungs- und Formalisierungsmechanismen zu bedienen. Das ist insbesondere dann der Fall, wenn die Rechtsinhaber nicht über die notwendigen technischen Fähigkeiten verfügen. Da die Schrankenbestimmung für Text und Data Mining einen Opt-out-Mechanismus vorsieht, sollte jedoch allen Rechtsinhabern unabhängig von ihrer technischen Vorbildung ermöglicht werden, diesen Mechanismus auch auszuüben. Das gilt im Weiteren auch für solche Rechtsinhaber, die auf die technische Infrastruktur eines Dritten, beispielsweise eines Content-Management-Systemanbieters, zurückgreifen. Ihre Möglichkeit zur Erklärung des Nutzungsverbehalts darf nicht davon abhängen, inwieweit der Dienstleister dem Rechtsinhaber Einfluss auf die technische Infrastruktur gewährt, also ob er ihm beispielsweise die Möglichkeit eröffnet, HTML-Metadaten zu bearbeiten.

Für eine Erklärbarkeit des Nutzungsverbehalts in natürlicher Sprache spricht schließlich auch, dass auch der nationale Umsetzungsgesetzgeber davon ausging, dass der Nutzungsverbehalt im Impressum oder den Nutzungsbedingungen einer Webseite erklärt werden kann.<sup>522</sup> Auch dafür ist notwendig, dass die Vorbehaltserklärung in natürlicher Sprache abgegeben werden können muss.

---

521 LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 67.

522 BT-Drs. 19/27426, S. 89.

(v) Systematik innerhalb des nationalen Regelungsgefüges

Weiterhin möglich wäre, einen systematischen Schluss aus § 49 UrhG zu ziehen. Denn die Schrankenbestimmung ermöglicht gemäß § 49 Abs. 1 S. 1 UrhG die Nutzung einzelner Rundfunkkommentare sowie Artikel aus Zeitungen und Informationsblättern zu privilegierten Zwecken, wenn die geschützten Inhalte nicht mit einem Vorbehalt der Rechte versehen sind. Allerdings ist der in § 49 UrhG angelegte Vorbehaltsmechanismus zwar funktional mit dem Nutzungsvorbehalt aus § 44b Abs. 3 UrhG vergleichbar. Nichtsdestotrotz sind beide Tatbestände nicht identisch formuliert. Insbesondere sieht § 49 UrhG keine Voraussetzung der Maschinenlesbarkeit vor. Die Schrankenbestimmung adressiert zudem vorwiegend analoge Nutzungen, während § 44b Abs. 3 S. 2 UrhG auf Online-Nutzungen von Werken zugeschnitten ist. Beide Sachverhaltskonstellationen unterscheiden sich demzufolge voneinander. Ein systematischer Schluss aus § 49 UrhG kann deswegen nicht gezogen werden.

(vi) Schlussfolgerung

Insgesamt sprechen damit die Legaldefinitionen des Rechtsbegriffs der Maschinenlesbarkeit im EGovG und DNG, vor allem aber durchgreifende teleologische Gründe auch auf Ebene des umgesetzten Unionsrechts dafür, das Tatbestandsmerkmal der Maschinenlesbarkeit im Sinne des § 44b Abs. 3 S. 2 UrhG so auszulegen, dass keine Formalisierung oder Strukturierung des Erklärungsinhalts notwendig sind. Ein einheitliches europäisches Begriffsverständnis steht dem nicht entgegen. Ausreichend ist eine digitale Hinterlegung der Vorbehaltsklärung, mittels derer die Erklärung aus einer technischen Perspektive unmittelbar durch den Computer verarbeitet werden kann. Um maschinenlesbar und damit wirksam zu sein, müssen Nutzungsvorbehalte im Sinne des § 44b Abs. 3 S. 2 UrhG daher nicht „in codierter Form“ erklärt werden. Die Vorbehaltsklärung kann vielmehr auch in natürlicher Sprache abgegeben werden.<sup>523</sup>

---

523 Wie hier auch Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 16; Steinrötter/Schauer, in: Barudi, Das neue Urheberrecht, § 14 Text und Data Mining Rn. 14; Akinci/Heidrich, IPRB 2023, 270 (272); Leistner, GRUR 2024, 1665 (1673); wohl ebenso Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 9; Schippan, ZUM 2024, 670 (675), wobei es hier etwas unklar bleibt; BT-Drs. 19/27426, S. 89; dazu tendierend auch LG Hamburg, Urt. v. 27.09.2024 - 310

bb) Angemessenheit des Nutzungsvorbehalts

Gemäß Art. 4 Abs. 3 DSM-RL müssen die Werke zudem in angemessener Weise mit dem Nutzungsvorbehalt versehen werden.

(1) Verhältnis zwischen Angemessenheits- und  
Maschinenlesbarkeitsvoraussetzung

Im Richtlinientext wird beispielhaft konkretisiert („etwa“), dass die Angemessenheitsvoraussetzung bei online veröffentlichten Inhalten zur Folge hat, dass die Nutzungsvorbehalte in maschinenlesbarer Form abgegeben werden müssen. Die Voraussetzung der Maschinenlesbarkeit, welche in § 44b Abs. 3 S. 2 UrhG umgesetzt worden ist, leitet sich im Richtlinientext damit aus der Grundvoraussetzung der Angemessenheit ab. Da in der Richtlinie zwar eine beispiel-, aber keine regelgebende Formulierung („in der Regel angemessen, wenn“; „insbesondere dann angemessen, wenn“; „Angemessenheit liegt in der Regel vor, wenn“) verwendet wird, kommt der Voraussetzung der Maschinenlesbarkeit damit keine Indizwirkung für die Erfüllung der Angemessenheitsvoraussetzung als Regelbeispiel zu. Hieraus folgt, dass Vorbehaltserklärungen gegen die Nutzung von online veröffentlichten Werken gemäß Art. 4 Abs. 3 DSM-RL aufgrund des Angemessenheitserfordernisses zwar maschinenlesbar sein müssen. Aus der Maschinenlesbarkeit der Vorbehaltserklärung kann umgekehrt allerdings noch nicht geschlussfolgert werden, dass die Erklärung aufgrund ihrer Maschinenlesbarkeit bereits angemessen im Sinne des Art. 4 Abs. 3 DSM-RL ist. Die Angemessenheit der Vorbehaltserklärung bleibt vielmehr als eigene Tatbestandsvoraussetzung bestehen.

---

O 227/23, GRUR-RS 2024, 25458 Rn. 66; a. A. hingegen *Hamann*, ZGE 2024, 113 (144); im Ergebnis auch *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; *Gauling*, CR 2021, 609 Rn. 11; *Müller-ter Jung/Rexin*, CR 2023, 169 (174); sich hieran anschließend *Bomhard*, in: *Ahlberg/Götting/Lauber-Rönsberg*, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 31, solange es keine einheitliche Standardformulierung in natürlicher Sprache gibt; auch schon in *Bomhard*, InTeR 2023, 174 (278); *Maamar*, ZUM 2023, 481 (484); *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (30).

(2) Richtlinienkonforme Auslegung des § 44b UrhG zur nationalen Anknüpfung der Angemessenheitsvoraussetzung

Trotz des Wortlauts von Art. 4 Abs. 3 DSM-RL ist die Angemessenheitsvoraussetzung nicht in § 44b Abs. 3 UrhG umgesetzt worden. Sie wird lediglich in der Gesetzesbegründung aufgegriffen.<sup>524</sup> Es bedarf deswegen einer diese Diskrepanz korrigierenden, richtlinienkonformen Auslegung des nationalen Schrankentatbestands.<sup>525</sup> Fraglich ist jedoch, an welches Tatbestandsmerkmal in § 44b UrhG die richtlinienkonforme Auslegung angeknüpft werden kann. Denkbar ist hierfür zum einen, die Auslegung im unbestimmten Rechtsbegriff der Maschinenlesbarkeit aus § 44b Abs. 3 S. 2 UrhG zu verankern. Hiergegen spricht aber, dass jede Nutzungsvorbehaltserklärung nach Art. 4 Abs. 3 DSM-RL angemessen sein muss, unabhängig davon, für welche Art von Inhalten der Vorbehalt erklärt wird. Die Voraussetzung der Maschinenlesbarkeit muss hingegen nur bei online veröffentlichten Inhalten erfüllt werden. Eine richtlinienkonforme Anknüpfung an dieses Tatbestandsmerkmal ist damit nicht möglich.<sup>526</sup>

Die Angemessenheitsvoraussetzung ist deswegen im Wege einer richtlinienkonformen Auslegung des Tatbestandsmerkmals „vorbehalten“ in § 44b Abs. 3 UrhG hineinzulesen. Damit haben sich Rechteinhaber Nutzungen ihrer Werke zum Text und Data Mining nur dann gemäß § 44b Abs. 3 S. 1 Hs. 2 UrhG überhaupt vorbehalten, wenn sie ihre Erklärung auch in einer angemessenen Art und Weise abgegeben haben.

(3) Konkrete Anforderungen an die Vorbehaltserklärungen

Die Vorbehaltserklärung muss so ausgestaltet sein, dass sie mit üblichen Mitteln und zumutbarem Aufwand erkannt werden kann. Zur Beurteilung dessen kommt es unter anderem auf die Verkehrssitte und bei online zugänglichen Werken auf den Stand der Technik an.<sup>527</sup>

---

524 Siehe in BT-Drs. 19/27426, S. 89.

525 Insoweit auch *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332. Auch *Bullinger*, in: *Wandtke/Bullinger*, Urheberrecht, UrhG § 44b Rn. 15; *Raue*, ZUM 2021, 793 (796 f.) beziehen sich auf die Angemessenheitsvoraussetzung, ohne jedoch eine dogmatische Verankerung im nationalen Recht vorzunehmen.

526 A. A. *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332.

527 Vgl. *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; für eine Berücksichtigung der Verkehrssitte auch *Raue*, ZUM 2021, 793 (797).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

#### (i) Ort der Vorbehaltserklärung bei online veröffentlichten Inhalten

Bei der Erklärung von Vorbehalten gegen die Nutzung online verfügbarer Inhalte in natürlicher Sprache ist in diesem Zusammenhang fraglich, an welchem Ort der Nutzungsvorbehalt erklärt werden muss, damit er mit zumutbarem Aufwand erkannt werden kann. Hierfür ist maßgeblich, an welchen Orten der durchschnittliche Werknutzer mit einer Vorbehaltserklärung rechnet. Diese Verkehrserwartung wird durch die Nennung des Impressums sowie der Allgemeinen Geschäfts- beziehungsweise Nutzungsbedingungen einer Webseite in den Gesetzesbegründungen geprägt.<sup>528</sup> Grund hierfür ist, dass die sich aus den Gesetzesmaterialien ergebenden Anforderungen erheblich sowohl im fach- und populärwissenschaftlichen als auch im anwendungsbezogenen Diskurs aufgegriffen wurden.<sup>529</sup> Dies hat zum Herausbilden einer entsprechenden Nutzererwartung geführt. Infolgedessen ist es für Trainingsdatensammler zumutbar, diese Orte einer Webseite nach Nutzungsvorbehaltserklärungen gegen die Verwendung von Werken zum Text und Data Mining abzusuchen.

Werden auf einer Webseite mehrere Werke wiedergegeben, so würden Urheber und Rechtsinhaber insbesondere bei der Verwendung umfangreicher Werkdatenbanken darüber hinaus erheblich belastet, wenn sie für jede ihrer Werkwiedergaben eine individuelle Vorbehaltserklärung abgeben müssten. Gleichzeitig besteht kein schutzwürdiges Interesse der Trainingsdatensammler an einer Vielzahl individueller Erklärungen, sofern die abgegebene Vorbehaltserklärung auch in generalisierter, also mehrere Inhalte gleichzeitig erfassender Form zweifelsfrei den jeweils betroffenen Inhalten zugeordnet werden kann. Die Angemessenheitsvoraussetzung bringt die Interessen von Urhebern und Rechtsinhabern sowie Trainingsdatensammern in einen Ausgleich, sodass beides zugunsten der Urheber und Rechtsinhaber zu berücksichtigen ist. Infolgedessen kann eine Vorbehaltserklärung in den Nutzungsbedingungen oder dem Impressum einer Webseite auch für sämtliche der auf einer Webseite abrufbaren Werkwiedergaben oder eine

---

528 ErwGr. 18 UAbs. 2 DSM-RL; BT-Drs. 19/27426, S. 89.

529 Siehe beispielsweise <https://bff.de/news/die-bff-justiziarin-dorothe-lanc-informiert-zum-thema>; <https://www.zdf.de/nachrichten/wirtschaft/urheberrecht-kuenstliche-intelligenz-ki-internet-100.html>; <https://www.boersenverein.de/beratung-service/recht/kuenstliche-intelligenz>, <https://bvpa.org/text-und-data-mining-durch-nutzung-svorbehalt-untersagen>; <https://www.golem.de/news/chatgpt-stable-diffusion-dall-e-die-crux-mit-dem-opt-out-aus-text-und-data-mining-2303-172254-2.html> (alle zuletzt abgerufen am 4.7.2025).

ausgewählte Gruppe hiervon erklärt werden.<sup>530</sup> Demnach ist ausreichend, dass sich die abgegebene Vorbehaltserklärung zweifelsfrei auf eine oder mehrere Werkwiedergaben bezieht.<sup>531</sup> Ebenso zumutbar ist der Aufwand für eine Überprüfung der jeweiligen Subpage, also Einzelseite, auf der ein konkretes, online zugängliches Werk verfügbar gemacht worden ist.<sup>532</sup> Grund hierfür ist, dass der Ort des Nutzungsvorbehalts durch den Fundort des Werkes bereits hinreichend konkretisiert ist.

Von Werknutzern, die für das Text und Data Mining gezielt und in wesentlichem Umfang in einem spezifischen Dateiformat verkörperte Werke vervielfältigen, kann außerdem erwartet werden, dass sie sich in Bezug auf das konkrete Dateiformat vertiefte technische Kenntnisse aneignen. Denn an die Vervielfältigung schließt sich eine Verarbeitung der Werke an, die einen Umgang mit den Besonderheiten des jeweiligen Dateiformats notwendig macht. Infolgedessen ist es für Werknutzer, die gezielt und in wesentlichem Umfang in einem spezifischen Dateiformat verkörperte Werke vervielfältigen, zumutbar, zusätzlich auch die Metadaten der Werke im Hinblick auf hierin erklärte Nutzungsvorbehalte auszulesen.<sup>533</sup> Im Ergebnis ist damit grundsätzlich sowohl die Abgabe einer „asset-“ im Sinne einer dateibezoogenen als auch einer „location-based“ Vorbehaltserklärung möglich.<sup>534</sup>

Eine automatisierte Vervielfältigung und Verarbeitung von Werken wäre hingegen nicht gewährleistet, wenn eine umfassende, über die jeweilige Quelle des Inhalts hinausgehende Suche nach Vorbehaltserklärungen des betroffenen Rechtsinhabers beispielsweise auf einer persönlichen Webseite oder auf Nachrichtenportalen durchgeführt werden müsste.<sup>535</sup> Hinzu kommt, dass dem Datensammler ein wesentlicher Teil der vom Text und Data Mining betroffenen Rechtsinhaber online verfügbarer Werke mangels hinreichender Urheberennennungen nicht bekannt ist. Daher folgt aus der

---

530 LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 63; im Anschluss hieran auch *Leistner*, GRUR 2024, 1665 (1667, 1673); a. A. *Hamann*, ZGE 2024, 113 (148).

531 LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 62.

532 *Jacobsen/Hartmann*, MMR-Aktuell 2021, 441332; vgl. außerdem *Raue*, ZUM 2021, 793 (797); im Anschluss hieran *Hamann*, ZGE 2024, 113 (135); *Kleinkopf*, Text- und Data-Mining, S. 249.

533 Vgl. ErwGr. 18 UAbs. 2 DSM-RL.

534 Vgl. zur Terminologie *Bäcker*, Der Text und Data Mining Vorbehalt: Technische Umsetzung der Maschinenlesbarkeit; außerdem in *Stieper/Denga*, GRUR 2024, 1473 (1476) aufgegriffen.

535 Ähnlich auch *Leistner*, GRUR 2024, 1665 (1673).

Angemessenheitsvoraussetzung, dass nur im Umfeld der als Vervielfältigungsvorlage genutzten Quelle nach Vorbehaltserklärungen gesucht werden muss. Für online veröffentlichte Werke bedeutet dies, dass sich eine Suche nach Vorbehaltserklärungen auf die Webseite beschränken kann, auf der die jeweilige Vervielfältigungsvorlage des Werkes öffentlich wiedergegeben wird.<sup>536</sup> Nicht angemessen sind demzufolge Pauschalerklärungen, beispielsweise auf der Webseite einer Verwertungsgesellschaft „für ihr gesamtes Repertoire“<sup>537</sup>.

(ii) Sprache der Vorbehaltserklärung

Vor dem Hintergrund der Angemessenheitsvoraussetzung ist darüber hinaus fraglich, in welcher Sprache Vorbehaltserklärungen abgegeben werden dürfen, die die in natürlicher Sprache verfasst sind. Auch hierfür kommt es insbesondere auf die Verkehrsgewohnheiten an. Da es sich um die am weitesten verbreitete Verkehrssprache handelt, ist das Suchen nach Nutzungs vorbehalten in englischer Sprache zumutbar.<sup>538</sup> Wegen der leichten Identifizierbarkeit müssen Werknutzer außerdem mit Nutzungs vorbehalten in der Hauptsprache einer jeweiligen Webseite rechnen, auf der die betreffenden Werke veröffentlicht worden sind. Da die betroffenen Rechtsinhaber für Werknutzer hingegen oftmals nicht bekannt sind,<sup>539</sup> ist es für sie nicht zumutbar, die Sprache der jeweiligen Rechtsinhaber mit zu berücksichtigen.<sup>540</sup>

(iii) Mindestmaß an Struktur und Klarheit

Neben alldem müssen die Erklärungen ein Mindestmaß an Struktur und Klarheit aufweisen. Es soll ein digitales Verständnis der Vorbehaltserklärungen ermöglichen. Berücksichtigt werden müssen demnach nur übliche, klar verständliche Formulierungen und Wortbausteine sowie Keywords wie

---

536 Heine, GRUR-Prax 2024, 87 Rn. 17.

537 So zum Beispiel aber die Verwertungsgesellschaften GEMA (<https://www.gema.de/documents/d/guest/tdm-vorbehaltserklaerung-2023-12-pdf>) sowie VG Bild-Kunst (<https://www.bildkunst.de/service/reservation-of-rights-for-text-and-data-mining>).

538 Bomhard, DSRITB 2023, 255 (266); Bomhard, InTeR 2023, 174 (178).

539 Vgl. dazu oben unter 2. Teil B. I. 2.

540 A. A. wohl Bomhard, DSRITB 2023, 255 (266); auch in Bomhard, InTeR 2023, 174 (178): „sämtliche in Betracht kommende Sprachen.“

beispielsweise „TDM reserved“, „Text und Data Mining vorbehalten“, „Art. 4 Abs. 3 DSM-RL“, „§ 44b Abs. 3 UrhG“ sowie „Nutzungsvorbehalt“<sup>541</sup> Diese Aufzählung ist allerdings nicht abschließend.

cc) Ausdrücklichkeit des Nutzungsvorbehalts

Darüber hinaus müssen die Nutzungsvorbehaltserklärungen gemäß Art. 4 Abs. 3 DSM-RL ausdrücklich erklärt werden. Die Voraussetzung wurde ebenfalls nicht mit § 44b UrhG in nationales Recht umgesetzt. Auch zur Behebung dieses Umsetzungsdefizits bedarf es deswegen einer richtlinienkonformen Auslegung des in § 44b Abs. 3 S. 1 UrhG verwendeten Rechtsbegriffs des sich *Vorbehaltens*.<sup>542</sup> Mithin wurde ein Nutzungsvorbehalt nur dann gemäß § 44b Abs. 3 UrhG wirksam abgegeben, wenn der Vorbehalt auch ausdrücklich erklärt worden ist. Das ist der Fall, wenn der Inhalt der Erklärung expressis verbis transportiert wird.<sup>543</sup> Entscheidend ist dabei, ob die Handlung des Vorbehaltserklärenden unmittelbar auf die Abgabe des Erklärungsinhalts, also des Vorbehalts, gerichtet ist. Nicht ausreichend sind demnach Handlungen, mit denen der Vorbehaltserklärende vorrangig einen anderen, selbstständigen Zweck verfolgt und dabei den Vorbehaltswillen indirekt miterklärt.<sup>544</sup> Eine ausdrückliche Erklärung kann nach diesen Maßstäben auch in formalisierter und standardisierter Form abgegeben werden, sofern der verwendete Befehl gerade dazu dient, den Vorbehaltswillen des Erklärenden auszudrücken. Eine Erklärung in Codeform steht dem Ausdrücklichkeitserfordernis daher nicht entgegen.

dd) Notwendigkeit einer Beschränkung der Vorbehaltserklärung auf Nutzung zum Zwecke des Text und Data Minings

Fraglich ist weiterhin aber, ob aus dem Ausdrücklichkeitskriterium des Art. 4 Abs. 3 DSM-RL, welches richtlinienkonform in § 44b Abs. 3 S. 1

---

541 So auch *Bomhard*, InTeR 2023, 174 (178).

542 Vgl. dazu unmittelbar oben unter 3. Teil A. III. 2. i) bb); wie hier ebenfalls LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 62.

543 Vgl. *Armbrüster*, in: Münchener Kommentar zum BGB, BGB Vor § 116 Rn. 6.

544 Vgl. v. *Savigny*, System des heutigen römischen Rechts, S. 245; auch *Armbrüster*, in: Münchener Kommentar zum BGB, BGB Vor § 116.

UrhG hineingelesen werden muss,<sup>545</sup> abgeleitet werden kann, dass nur solche Vorbehaltserklärungen ausdrücklich und damit wirksam sind, die „konkret-individuell“<sup>546</sup> auf ein Verbot der Nutzung des Werkes für das Text und Data Mining beschränkt sind. Dies hätte zur Folge, dass durch zu weit gefasste Vorbehaltserklärungen nicht die Rechtsfolge des § 44b Abs. 3 UrhG ausgelöst werden könnte.

Dafür, dass Vorbehaltserklärungen für ein Wirksamwerden auf Nutzungen zum Zwecke des Text und Data Mining beschränkt werden müssen spricht, dass gemäß ErwGr. 18 UAbs. 2 S. 3 DSM-RL andere Nutzungen als die zum Text und Data Mining Vorgenommenen nicht von der Vorbehaltserklärung betroffen sein sollen.<sup>547</sup> Allerdings wird aus ErwGr. 18 DSM-RL nicht deutlich, ob dies als Voraussetzungen für die Wirksamkeit des Nutzungsvorbehalts zu verstehen ist. Denn der Erwägungsgrund könnte auch als rechtsfolgenbezogene Klarstellung verstanden werden, nach der die quasi-dingliche Wirkung des in Art. 4 Abs. 3 DSM-RL verankerten Nutzungsvorbehalts allein die Freistellung der Nutzungshandlung über die Schrankenbestimmung für Text und Data Mining beendet. Für letzteres spricht, dass gemäß ErwGr. 45 InfoSoc-RL die unionsrechtlichen Schrankenbestimmungen einer vertraglichen Abweichung im Grundsatz nicht entgegenstehen.<sup>548</sup> Es wäre vor diesem Hintergrund wertungswidersprüchlich, dürfte ein Rechtsinhaber beispielsweise die Privatkopiefreiheit gegenüber einem Werknutzer wirksam vertraglich einschränken, in Verbindung mit einer Erklärung des Nutzungsvorbehalts gemäß Art. 4 Abs. 3 DSM-RL eine solche Einschränkung aber faktisch nicht mehr vornehmen, ohne die Unwirksamkeit der Vorbehaltserklärung in Bezug auf das Text und Data Mining in Kauf nehmen zu müssen. Eine Auslegung des ErwGr. 18 UAbs. 2 S. 3 DSM-RL, nach der der Erwägungsgrund eine Wirksamkeitsvoraussetzung für Nutzungsvorbehaltserklärungen beinhaltet, würde demzufolge zu einer allgemeinen Schlechterstellung derjenigen Rechtsinhaber führen, deren Werke potenziell von Text und Data Mining-Aktivitäten betroffen sind. Eine solche Schlechterstellung wird von Art. 4 Abs. 3 DSM-RL aber nicht bezweckt. Im Ergebnis spricht dies dafür, dass ErwGr. 18 UAbs. 2 S. 3 DSM-

---

545 Dazu siehe oben unter 3. Teil A. III. 2. i) cc).

546 Hamann, ZGE 2024, 113 (135).

547 Hamann, ZGE 2024, 113 (134 f.).

548 Zur vertraglichen Abweichung von Schrankenbestimmungen, die nach hier vertretenen Auffassung im Grundsatz möglich ist, umfassend auch mit weiteren Nachweisen in: Stieper, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 263 f.

RL keine Wirksamkeitsvoraussetzung, sondern eine rechtsfolgenbezogene Klarstellung in Bezug auf die quasi-dingliche Wirkung des Nutzungsvorbehalts enthält.<sup>549</sup> Der Richtliniengeber wollte demnach nur einem Schluss vom Kleinen aufs Große aus Art. 4 Abs. 3 DSM-RL zulasten der Anwendung anderer Schrankenbestimmungen vorbeugen. Das bedeutet, dass keine Notwendigkeit besteht, eine Vorbehaltserklärung auf Nutzungen zum Zwecke des Text und Data Mining zu beschränken. Solange die Vorbehaltserklärung nicht implizit, sondern ausdrücklich abgegeben worden ist, kann sie sich neben dem Text und Data Mining demzufolge auch andere Nutzungszwecke einbeziehen.<sup>550</sup>

#### ee) Weitere formalisierte Einzelfälle

Untersucht werden kann in diesem Zusammenhang im Weiteren, ob die im Verkehr üblichen, strukturiert-formalisierten Mechanismen zur Erklärung von Nutzungsvorbehalten den aus § 44b Abs. 3 S. 2 UrhG folgenden Anforderungen genügen. Nur wenn die zuvor spezifizierten Voraussetzungen erfüllt sind, wird mit einer Verwendung einer dieser Mechanismen die Freistellungswirkung des § 44b Abs. 2 S. UrhG beendet.

##### (1) Robots Exclusion Standard (robots.txt)

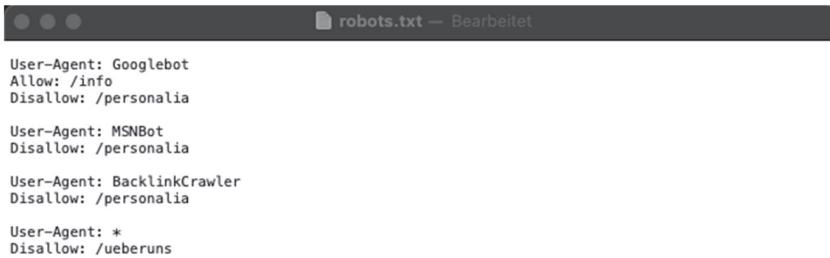
Mithilfe des Robots Exclusion Protocols können vom Webseitenbetreiber Zugangsregeln für automatisierte Webcrawler aufgestellt werden. Diese Webcrawler können grundsätzlich auch Werke für das KI-Training generieren. Möglich ist, alle Webcrawler allgemein (als „\*“ symbolisiert) oder ausgewählte Webcrawler namentlich anzusprechen. Das Auslesen der Webseite oder spezifizierter Seitenbereiche kann gestattet oder untersagt werden.

---

549 A. A. Hamann, ZGE 2024, 113 (134 f.).

550 A. A. Hamann, ZGE 2024, 113 (134 f.).

### III. Eingreifen einer urheberrechtlichen Schrankenbestimmung



```
User-Agent: Googlebot
Allow: /info
Disallow: /personalia

User-Agent: MSNBot
Disallow: /personalia

User-Agent: BacklinkCrawler
Disallow: /personalia

User-Agent: *
Disallow: /ueberuns
```

Um mit Hilfe des Robots Exclusion Protocols einen im Sinne des § 44b Abs. 3 UrhG wirksamen Nutzungsvorbehalt zu erklären, müsste die Verwendung der robots.txt-Datei den Vorbehaltswillen angemessen und ausdrücklich sowie gemäß § 44b Abs. 3 S. 2 UrhG maschinenlesbar übermitteln.

Für eine Angemessenheit des Erklärungsmusters spricht zunächst, dass die Verwendung des Robots Exclusion Protocols von der Internet Engineering Task Force im September 2022 als Internetstandard vorgeschlagen worden ist.<sup>551</sup> Das Robots Exclusion Protocol gehört damit zum Stand der Technik, der für die Übermittlung von Vorbehaltserklärungen eingesetzt wird. Von Datensammlern kann deswegen erwartet werden, dass sie mit dem Robots Exclusion Protocol übermittelte Vorbehaltserklärungen mit üblichen Mitteln und zumutbarem Aufwand erkennen können.

Darüber hinaus wenden sich Vorbehaltserklärende mit einem Einsatz des Robots Exclusion Protocols nicht spezifisch gegen die Verwendung ihrer Werke für das Text und Data Mining. Vielmehr gilt der Vorbehalt zweckunabhängig gegen jede automatisierte Vervielfältigung durch Webcrawler. Eine solche überschießende Vorbehaltserklärung ist für die Wirksamkeit im Sinne des § 44b Abs. 3 UrhG jedoch unschädlich.<sup>552</sup>

Fraglich ist jedoch, ob die Vorbehaltserklärungen mit Hilfe des Robots Exclusion Protocols auch ausdrücklich abgegeben werden. Dies ist nach richtlinienkonformer Auslegung des § 44b Abs. 3 S. 1 UrhG notwendig.<sup>553</sup> Im Gegensatz dazu erfüllen konkludente Vorbehaltserklärungen nicht die gesetzlichen Anforderungen.<sup>554</sup> Konkludent sind Vorbehaltserklärungen, die vordergründig einem anderen Zweck dienen, implizit aber auch den

---

551 <https://datatracker.ietf.org/doc/html/rfc9309> (zuletzt abgerufen am 11.7.2024).

552 Dazu siehe unmittelbar oben unter 3. Teil A. III. 2. i) dd).

553 Dazu ebenfalls siehe oben unter 3. Teil A. III. 2. i) cc).

554 Raue, ZUM 2021, 793 (796).

Erklärungsinhalt übermitteln.<sup>555</sup> Auch wenn das Robots Exclusion Protocol nutzungszweckunabhängig jede automatisierte Vervielfältigung der betroffenen Werke durch Webcrawler untersagt, dient es nichtsdestotrotz keinem anderen Zweck als der Übermittlung der (umfassenden) Vorbehaltserklärung. Die mit Hilfe des Robots Exclusion Protocols übermittelte Erklärung ist trotz ihrer Formalisierung und Strukturierung sowie ungeachtet der überschießenden Vorbehaltswirkung demzufolge ausdrücklich. Im Ergebnis erfüllt sie damit die Voraussetzungen, die an eine Vorbehaltserklärung im Sinne des § 44b Abs. 3 S. 1 UrhG gestellt werden. Da das Robots Exclusion Protocol auch in einer Art und Weise digital hinterlegt ist, die eine unmittelbare Verarbeitung durch Maschinen ermöglicht, sind die mit Hilfe dieses Protokolls abgegebenen Vorbehaltserklärungen auch maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG. Insgesamt sind Nutzungsvorbehalte, die mit Hilfe des Robots Exclusion Protocols erklärt werden, demzufolge wirksam.<sup>556</sup> Sie lösen die in § 44b Abs. 3 UrhG angelegte Beendigungswirkung aus.

## (2) Spawning-Protokoll (ai.txt)

Neben dem Robots Exclusion Protocol wird auch das Spawning-Protokoll zur Abgabe von Vorbehaltserklärungen genutzt. Das Spawning-Protokoll ermöglicht Webseitenbetreibern, sich spezifisch gegen Datensammler zu wenden, die die auf der Webseite wiedergegebenen Werke für das Training von KI vervielfältigen wollen. Die Syntax entspricht dabei grundsätzlich derjenigen des Robots Exclusion Protocols, bei dem die Vorbehaltserklärung über eine im Wurzelverzeichnis der Webseite abgelegte Textdatei übermittelt werden.<sup>557</sup> Allerdings wird die ai.txt-Datei nicht von den Datensammlern selbst, sondern vom Intermediär *Spawning* ausgelesen. Dieser

---

555 Vgl. v. Savigny, System des heutigen römischen Rechts, S. 245; auch Armbriüster, in: Münchener Kommentar zum BGB, BGB Vor § 116.

556 Wie hier auch Baumann, NJW 2023, 3673 Rn. 19; de la Durantaye, ZUM 2023, 645 (651); Gausling, CR 2021, 609 Rn. 5; Heine, GRUR-Prax 2024, 87 Rn. 16; Hofmann, WRP 2024, 11 Rn. 18; Konertz/Schönhof, WRP 2024, 289 Rn. 33; Maamar, ZUM 2023, 481 (484); Müller-ter Jung/Rexin, CR 2023, 169 Rn. 29; Schack, NJW 2024, 113 Rn. 20; Schwarz/Söbbing, RDi 2023, 415 Rn. 33; a. A. hingegen Bomhard, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 33; Hamann, ZGE 2024, 113 (153).

557 Hamann, ZGE 2024, 113 (155). Außerdem ausdrücklich referiert in: <https://spawning.substack.com/p/aitxt-a-new-way-for-websites-to-set> (zuletzt aufgerufen am 2.5.2024).

gibt die Vorbehaltserklärungen über eine Schnittstelle gebündelt an die Trainingsdatensammlern weiter.<sup>558</sup>

Da auch die mit Hilfe des Spawning-Protokolls abgegebenen Vorbehaltserklärungen zwar formalisiert und strukturiert, aber ausschließlich zum Zwecke der Vorbehaltsumvermittlung erklärt worden sind, erfüllen sie trotz ihrer Formalisierung und Strukturierung die Ausdrücklichkeitsanforderungen, die sich im Wege einer richtlinienkonformen Auslegung des § 44b Abs. 3 UrhG ergeben.

Über die bereitgestellte Anwendungsschnittstelle können Datensammler die mit Hilfe des Spawning-Protokolls hinterlegten Vorbehalte darüber hinaus automatisiert und ohne Informationsverlust abrufen, sodass der Vorbehaltszweck auch mit Hilfe des Spawning-Protokolls sicher erreicht wird. Das Spawning-Protokoll gehört damit ebenfalls zum Stand der Technik.<sup>559</sup> Da die Nutzung der Anwendungsschnittstelle im Übrigen niederschwellig und kostenlos möglich ist, ist auch die Berücksichtigung des Spawning-Protokolls für Datensammler zumutbar. Im Ergebnis sind deswegen auch Vorbehaltserklärungen mit Hilfe des Spawning-Protokolls angemessen. Infolgedessen qualifizieren sich die auf Basis des Spawning-Protokolls abgegebenen Erklärungen auch als Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG. Aufgrund ihrer unmittelbaren digitalen Hinterlegung, die sogar eine Standardisierung und Formalisierung der Erklärungsinhalte mit sich bringt, sind auch die Anforderungen aus der Maschinenlesbarkeitsvoraussetzung aus § 44b Abs. 3 S. 2 UrhG erfüllt. Alles in allem sind mit Hilfe des Spawning-Protokolls erklärte Nutzungsvorbehalte insgesamt wirksam im Sinne des § 44b Abs. 3 UrhG. Sie beenden daher die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining.<sup>560</sup>

### (3) World Wide Web Consortium (W3C) TDM Reservation Protocol

Möglich ist darüber hinaus auch, Vorbehaltserklärungen auf Basis des vom *World Wide Web Consortium* entwickelten TDM Reservation Protocol abzugeben. Hierin sind verschiedene Kommunikationsmethoden zusammen-

---

558 Vgl. <https://spawning.substack.com/p/aitxt-a-new-way-for-websites-to-set> (zuletzt abgerufen am 30.4.2024).

559 Vgl. zum Begriff des Stands der Technik: *Martini*, in: Paal/Pauly, Kommentar zu Datenschutz-Grundverordnung und Bundesdatenschutzgesetz, DS-GVO Art. 25 Rn. 39a; *Mayen*, in: Scheurle/Mayen, Telekommunikationsgesetz, TKG § 113d Rn. 6.

560 In diesem Sinne wohl auch *Hamann*, ZGE 2024, 113 (155).

gefasst.<sup>561</sup> Da auch das W3C TDM Reservation Protocol ausschließlich den Zweck der Übermittlung von Vorbehaltserklärungen erfüllt und der Vorbehaltswille mit ihm ebenfalls ohne Informationsverlust sowie unentgeltlich und automatisiert übermittelt werden kann, erfüllen auch Erklärungen auf Basis des W3C TDM Reservation Protocols die Voraussetzungen der Ausdrücklichkeit und Angemessenheit aus § 44b Abs. 3 UrhG. Eine digitale Hinterlegung der Vorbehaltserklärungen beispielsweise als HTML-Metadaten-Tags ermöglicht zudem das unmittelbare computergestützte Auslesen der Vorbehalte. Demzufolge sind die Vorbehaltserklärungen auch maschinenlesbar im Sinne des § 44b Abs. 3 S. 2 UrhG. Infolgedessen erfüllen auch Nutzungsvorbehalte auf Basis des W3C TDM Reservation Protocols die Voraussetzungen des § 44b Abs. 3 UrhG. Auch durch sie wird die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining damit beendet.<sup>562</sup>

#### (4) Technische Schutzmaßnahmen gemäß § 95a UrhG

Unklar ist demgegenüber, ob Nutzungsvorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG durch den Einsatz technischer Schutzmaßnahmen gemäß § 95a UrhG wirksam miterklärt werden.

Nutzungsvorbehalte dienen dem Zweck, den Vorbehaltswillen der Rechtsinhaber kommunikativ zum Ausdruck zu bringen. Im Unterschied hierzu werden technische Schutzmaßnahmen, die über § 95a UrhG vor Umgehung geschützt sind, eingesetzt, um auf einer faktischen Ebene den Zugriff auf oder die Vervielfältigung von Werken zu verhindern. Sie wirken also tatsächlich gegen die Nutzungshandlung.<sup>563</sup> Technische Schutzmaßnahmen können den Vorbehaltszweck neben der faktischen Nutzungsbeschränkung demzufolge nur mittelbar neben dem Hauptzweck der Schutzmaßnahmen miterreichen. Läge im Einsatz technischer Schutzmaßnahmen eine Vorbehaltserklärung, würde diese damit nur konkludent abgegeben.<sup>564</sup> Schon deswegen wäre das Ausdrücklichkeitserfordernis nicht erfüllt, welches richtlinienkonform in das Tatbestandsmerkmal des Vorbehaltens aus

---

561 Siehe zum Inhalt des TDM Reservation Protocol <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202> (zuletzt abgerufen am 2.5.2024).

562 So auch Hamann, ZGE 2024, 113 (160); „gerade visionär“; Schippan, ZUM 2024, 670 (676); außerdem wohl ebenso de la Durantaye, ZUM 2023, 645 (651); Hofmann, WRP 2024, 11 Rn. 18; Konertz/Schönhof, WRP 2024, 289 Rn. 33; Maamar, ZUM 2023, 481 (484); Schwarz/Söbbing, RDi 2023, 415 Rn. 15.

563 Vgl. auch Conrad/Schubert, GRUR 2018, 350 (355).

564 Ebenso Hamann, ZGE 2024, 113 (163).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

§ 44b Abs. 3 S. 1 UrhG hineingelesen werden muss. Demzufolge werden beim Einsatz technischer Schutzmaßnahmen im Sinne des § 95a Abs. 2 S. 1 UrhG keine Nutzungsvorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG miterklärt.<sup>565</sup>

Darüber hinaus beenden zugangsbeschränkende Maßnahmen bereits die rechtmäßige Zugänglichkeit der betreffenden Werke im Sinne des § 44b Abs. 2 S. 1 UrhG. In diesen Fällen kann die Schrankenbestimmung für Text und Data Mining daher schon deswegen nicht angewendet werden.

#### ff) Werk- oder wiedergabebezogene Wirkung des Nutzungsvorbehalts

Gemäß § 44b Abs. 3 S. 1 UrhG beendet eine wirksame Erklärung des Nutzungsvorbehalts die Freistellungswirkung, die von der Schrankenbestimmung für Text und Data Mining, § 44b UrhG, ausgeht. Im Detail regelt die Schrankenbestimmung aber nicht die Wirkung und Reichweite des Nutzungsvorbehalts. Das betrifft unter anderem die Fragestellung, ob die Vorbehaltswirkung nur die Nutzung einer konkreten Werkwiedergabe beziehungsweise eines spezifischen Vervielfältigungsstücks als Vervielfältigungsvorlage erfasst, in deren beziehungsweise dessen Zusammenhang der Nutzungsvorbehalt erklärt worden ist (wiedergabebezogene Wirkung). Nach einem weiteren Verständnis könnte die Vorbehaltswirkung aber auch werkbezogen eintreten, also über die konkrete Werkwiedergabe oder das spezifische Vervielfältigungsstück hinaus die Freistellung der Nutzung des betreffenden Werkes insgesamt beendet (werkbezogene Wirkung).

Für eine wiedergabebezogene Wirkung des Nutzungsvorbehalts spricht, dass Nutzungsvorbehaltserklärungen aufgrund der Angemessenheitsvoraussetzung mit zumutbarem Aufwand erkannt werden können müssen. Nur dann sind sie wirksam und vom Datensammler zu berücksichtigen. Würden Nutzungsvorbehaltserklärungen werkbezogen wirken, müsste ein Datensammler bei der Vervielfältigung eines Werkes sämtliche der jemals verfügbar gemachten Repräsentationen eines Werkes, also Vervielfältigungsstücke oder öffentliche Wiedergaben, daraufhin absuchen, ob an anderer Stelle und zu irgendeinem Zeitpunkt bereits eine Vorbehaltserklärung für das Werk abgegeben worden ist. Gerade bei online verfügbaren Werken wäre dieser Aufwand kaum zu erbringen. Vorbehaltserklärungen an anderen Repräsentationen des Werkes aufzufinden ist für Datensammler

---

565 Wie hier auch Hamann, ZGE 2024, II3 (163); a. A. Meys, GRUR Int. 2020, 457 (467).

infolgedessen nicht zumutbar. Im Sinne des § 44b Abs. 3 S. 1 UrhG sind aber nur in zumutbarer Weise erkennbare Nutzungsverbehaltsklärungen überhaupt wirksam.<sup>566</sup> Die Wiedergabebezogene Wirkung des Nutzungsverbehalts folgt damit als Kehrseite aus der in § 44b Abs. 3 S. 1 UrhG verankerten Angemessenheitsvoraussetzung.

Darüber hinaus könnten Datensammler im Falle einer werkbezogenen Wirkung des Nutzungsverbehalts nie sicher sein, ob nicht doch ein nicht identifizierter Nutzungsverbehalt gegen die Verwendung des Werkes erklärt worden ist. Die hieraus resultierende Unsicherheit würde zu prohibitiven Effekten und damit zu einer Einschränkung der effektiven Wirksamkeit der Schrankenbestimmung führen. Dies wäre auch mit dem Grundsatz der praktischen Wirksamkeit des Unionsrechts nicht vereinbar.

Im Ergebnis sprechen beide Argumente für eine Wiedergabebezogene Wirkung des Nutzungsverbehalts. Die von der Vorbehaltsklärung ausgehende Beendigungswirkung beschränkt sich damit stets auf die Nutzung der jeweiligen öffentlichen Wiedergabe beziehungsweise des spezifischen Vervielfältigungsstücks als Vervielfältigungsvorlage für die Zwecke des Text und Data Mining, an der beziehungsweise dem der Nutzungsverbehalts erklärt worden ist. Die Nutzung anderer Werkrepräsentationen ist hiervon nicht betroffen.<sup>567</sup>

#### gg) Zeitliche Wirkung des Nutzungsverbehalts

§ 44b Abs. 3 UrhG regelt weiterhin nicht ausdrücklich, welche zeitliche Wirkung einem Nutzungsverbehalt zukommt. Einerseits könnte er Wirkung lediglich für die Zukunft entfalten (*ex nunc*), also sich nur auf solche Vervielfältigungshandlungen beziehen, die ab dem Zeitpunkt der Vorbehaltsklärung vorgenommen werden. Auf Grundlage eines strengen Verständnisses, nach dem ein Nutzungsverbehalt auch rückwirkend (*ex tunc*) eingreift, müssten auch vor der Vorbehaltsklärung für das Text und Data Mining hergestellte Vervielfältigungsstücke mit Wirksamwerden des Vorbehalts beseitigt werden. Für eine *ex nunc* Wirkung der Vorbehaltsklärung spricht, dass der Begriff des *sich Vorbehalte*ns dem natürlichen Sprachverständnis nach voraussetzt, dass sich die vorbehaltende Person

---

566 Siehe oben unter 3. Teil A. III. 2. i) bb).

567 So auch *Leistner*, GRUR 2024, 1665 (1673); wohl außerdem implizit *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 140 f.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

eine Möglichkeit oder Entscheidung für die Zukunft offenhalten will.<sup>568</sup> Das Vorbehalten bezieht sich damit auf ein zukünftiges Tun oder Unterlassen. Bereits abgeschlossene Vervielfältigungshandlungen werden deswegen schon begrifflich nicht erfasst. Es gibt einen terminologischen Unterschied zwischen dem *Löschen* einer Vervielfältigung und dem *Vorbehalten* einer Vervielfältigungshandlung.<sup>569</sup>

Für ein auf die Zukunft bezogenes Verständnis der Vorbehaltswirkung spricht zudem, dass der Nutzungsvorbehalt nach Art. 4 Abs. 3 DSM-RL die „Anwendung der Ausnahme und Beschränkung (für Text und Data Mining)“ regelt. Das betroffene Vervielfältigungsrecht bezieht sich gemäß Art. 2 InfoSoc-RL auf die Vervielfältigung als *Handlung*, also die *Herstellung* des Vervielfältigungsstücks.<sup>570</sup> Dies führt dazu, dass es einer Anwendung der Schrankenbestimmung nur so lange bedarf, wie die Herstellung des Vervielfältigungsstücks als *Handlung* des Werknutzers andauert. Ist das Vervielfältigungsstück bereits hergestellt worden, kommt es nicht mehr auf die „Anwendung der Ausnahme und Beschränkung“ an. Die hergestellten Vervielfältigungsstücke können vielmehr weiterhin bestehen. Die in Art. 4 Abs. 3 DSM-RL terminologisch angelegte Wirkung des Nutzungsvorbehalts deutet demzufolge ebenfalls auf eine ex nunc-Wirkung der Vorbehaltserklärung hin. Von einer solchen Rechtswirkung ist darüber hinaus auch der Umsetzungsgesetzgeber ausgegangen.<sup>571</sup>

Im Ergebnis sprechen durchgreifende Gründen dafür, dass der Nutzungsvorbehalt aus § 44b Abs. 3 UrhG Wirkung ex nunc entfaltet.<sup>572</sup> Vervielfältigungsstücke, die vor Wirksamwerden einer Vorbehaltserklärung für das Text und Data Mining und insbesondere für das KI-Training hergestellt worden sind, werden vom Nutzungsvorbehalt demzufolge nicht erfasst. Eine Löschung der Werke aus bereits angefertigten Trainingsdatensets muss demzufolge nicht erfolgen. Weiterhin müssen Werke nicht aus trainierten KI-Systemen „herausgerechnet“ werden. Hiervon zu unterscheiden

---

568 Dudenredaktion (*Bibliographisches Institut*), Duden: das Bedeutungswörterbuch, Begriff: vorbehalten.

569 Heesen/Jüngels, RuZ 2021, 45 (55).

570 Dies wird auch deutlich in der deutschen Umsetzung, § 16 Abs. 1 UrhG.

571 BT-Drs. 19/27426, S. 89.

572 Allgemeine Meinung, siehe unter anderem: *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 27; *Dreier*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 44b Rn. 9; *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 44b Rn. 10; *Steinrötter/Schauer*, in: Barudi, Das neue Urheberrecht, § 14 Text und Data Mining Rn. 14; Heesen/Jüngels, RuZ 2021, 45 (55); BT-Drs. 19/27426, S. 89.

ist allerdings die Zulässigkeit der Herstellung weiterer, mitunter für das KI-Training notwendiger Vervielfältigungen. Eine solche ist in Folge des Nutzungsvorbehalts nicht mehr über § 44b UrhG freigestellt.

#### hh) Der Vorbehaltserklärende

Werknutzungen werden nur dann durch § 44b Abs. 2 S. 1 UrhG erlaubnisfrei gestellt, wenn sich der Rechtsinhaber die Nutzungen nicht vorbehalten hat (§ 44b Abs. 3 S. 1 UrhG). § 44b UrhG weist die Vorbehaltsbefugnis damit „dem Rechtsinhaber“ zu. Gleichzeitig wird im Regelungstext nicht spezifiziert, welche Anforderungen an die Vorbehaltsbefugnis im Sinne des § 44b Abs. 3 UrhG im Einzelnen zu stellen sind.

Auch wenn sowohl das UrhG als auch die unionsrechtlichen Bestimmungen des Urheberrechts begrifflich zwischen *Urhebern* und *Rechtsinhabern* differenzieren, folgt aus der gesetzlich gewährten Vorbehaltsbefugnis der Rechtsinhaber, dass erst recht auch Urheber Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG abgeben können.

##### (1) Einfache und ausschließliche Rechtsinhaber

Mit Blick auf die Vorbehaltsbefugnis der Rechtsinhaber ist darüber hinaus jedoch unklar, ob § 44b Abs. 3 UrhG nur *ausschließlichen* Rechtsinhabern die Möglichkeit eröffnet, einen Nutzungsvorbehalt zu erklären, oder ob auch Inhaber eines *einfachen* Nutzungsrechts nach § 44b Abs. 3 UrhG vorbehaltsbefugt sind.

Für eine Vorbehaltsbefugnis auch einfacher Rechtsinhaber spricht, dass auch sie, unabhängig von Umfang und Qualität ihrer Rechtsposition, *Rechtsinhaber* im Sinne des § 44b Abs. 3 S. 1 UrhG sind. Damit fallen sie nach dem Wortlaut des § 44b Abs. 3 S. 1 UrhG in den persönlichen Anwendungsbereich der Regelung. Darüber hinaus haben auch einfache Rechtsinhaber mit dem Erwerb des Nutzungsrechts eine Investition getätigt, an der ein Schutzinteresse besteht. Dieses ließe sich durch die Vorbehaltserklärung realisieren.

Allerdings bezieht sich der Begriff des „Rechtsinhabers“ innerhalb des UrhG auch ohne spezifizierende Zusätze („*ausschließlich*“ oder „*einfach*“) entweder auf ausschließliche (beispielsweise in § 10 Abs. 3 UrhG) oder auf einfache (beispielsweise in § 60d Abs. 6 UrhG oder § 95a Abs. 1 UrhG) Rechtsinhaber. Die konkrete Bedeutung ist vom Kontext der jeweiligen

Regelung abhängig. Es besteht damit kein einheitliches Verständnis des Rechtsinhaberbegriffs, das auch § 44b Abs. 3 S. 1 UrhG zugrunde gelegt werden könne. Ein Schluss auf die notwendige Qualität der Rechtsposition ist aus dem Wortlaut des § 44b Abs. 3 S. 1 UrhG demzufolge nicht möglich.

Allerdings versetzt ein einfaches Nutzungsrecht einen Rechtsinhaber nicht in die Lage, anderen Rechtsinhabern die Nutzung desselben Werkes zu verbieten (§ 31 Abs. 2 UrhG). Es wäre vor diesem Hintergrund wertungs-widersprüchlich, wäre der einfache Rechtsinhaber im Anwendungsbereich der Schrankenbestimmung für Text und Data Mining über § 44b Abs. 3 UrhG befugt, anderen, von der Schrankenbestimmung privilegierten Nutzern die rechtmäßige Nutzung des Werkes zum Zwecke des Text und Data Mining zu verbieten. Die Nichtexklusivität des einfachen Nutzungsrechts hat außerdem zur Folge, dass sich auch das Interesse des einfachen Rechtsinhabers an einem Schutz seiner Investition in die Ermöglichung der Werknutzung nicht darauf erstreckt, dass die Werknutzung unter Ausschluss anderer, also exklusiv erfolgt. Ein Schutz seiner Investition über § 44b Abs. 3 UrhG ist demzufolge nicht nötig.

Insgesamt sprechen damit bessere Gründe dafür, dass neben Urhebern nur ausschließliche Rechtsinhaber Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG abgeben können. Inhaber einfacher Nutzungsrechte sind hingegen nicht vorbehaltsbefugt.<sup>573</sup>

Ungeachtet dessen zielt die Abgabe der Nutzungsvorbehaltserklärung auf die Herbeiführung der von § 44b Abs. 3 UrhG vorgeschriebenen Rechtsfolge ab. Damit wird zwar keine vom Vorbehaltserklärenden selbstgewählte Rechtsfolge spezifiziert. Es fehlt also an den Voraussetzungen einer Willenserklärung im bürgerlichrechtlichen Sinne.<sup>574</sup> Dennoch ist die Erklärung des Nutzungsvorbehalts eine geschäftsähnliche Handlung.<sup>575</sup> Damit finden auch die Regelungen zur Stellvertretung bei der Abgabe von Willenserklärungen, §§ 164 ff. BGB, entsprechende Anwendung.<sup>576</sup> Der Urheber oder ausschließliche Rechtsinhaber kann demzufolge auch einfachen Rechtsin-

---

573 A. A. LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 Rn. 58; im Anschluss hieran Leistner, GRUR 2024, 1665 (1667, 1671); weiterhin auch Brockmeyer, Text und Data Mining, S. 79, nach dem alle Rechtsinhaber berechtigt sind; offengelassen von Vesala, IIC 2023, 351 (357); Hamann, ZGE 2024, 113 (138 ff.).

574 Zu den Voraussetzungen einer Willenserklärung vgl. Köhler, BGB Allgemeiner Teil § 6 Rn. 1f.

575 Hamann, ZGE 2024, 113 (138); allgemein zur geschäftsähnlichen Handlung und ihren Voraussetzungen siehe Köhler, BGB Allgemeiner Teil § 5 Rn. 7.

576 Mansel, in: Jauernig, Bürgerliches Gesetzbuch, BGB § 164 Rn. 2.

habern Vertretungsmacht für die Erklärung eines Nutzungsvorbehalts in Bezug auf die jeweilige Werkwiedergabe erteilen oder eine Vorbehaltsserklärung nachträglich genehmigen (§ 184 BGB analog). Hierfür spricht auch der in Art. 4 Abs. 3 EUV verankerte Grundsatz der effektiven Anwendung des § 44b Abs. 3 UrhG zugrundeliegenden Unionsrechts.<sup>577</sup> Eine solche Genehmigung kann auch konkludent abgegeben werden, beispielsweise durch die (gerichtliche) Geltendmachung der Rechtsverletzung im betreffenden Fall.<sup>578</sup>

## (2) Inhaltliche, räumliche und zeitliche Beschränkungen des Nutzungsrechts

Fraglich ist des Weiteren aber auch, wie das Ausschließlichkeitsrecht eines Rechtsinhabers konkret ausgestaltet sein muss, damit er vorbehaltsbefugt ist. Problematisch wird dies zwar nicht bei Inhabern unbeschränkt ausschließlicher Nutzungsrechte. Deren umfassende Rechtsposition beinhaltet auch die Befugnis, einen Nutzungsvorbehalt nach § 44b Abs. 3 UrhG zu erklären. Im Gegensatz hierzu ist allerdings unklar, ob und wann Inhaber von inhaltlich, räumlich oder zeitlich beschränkten Nutzungsrechten, wie sie durch § 31 Abs. 1 S. 1 UrhG ermöglicht werden, zur Abgabe von Nutzungsvorbehaltsserklärungen befugt sind.

### (i) Inhaltliche Beschränkungen und die Vorbehaltsbefugnis

Inhaltliche Beschränkungen der eingeräumten Rechtsposition können insoweit stattfinden, als dass die Aufspaltung selbstständig verkehrsfähige Rechte herausbildet. Der Nutzungsvorbehalt entfaltet seine Wirkung vor diesem Hintergrund gegen die Erlaubnisfreiheit des Rechts der Vervielfältigung von Werken zum Text und Data Mining. Dies umfasst auch Vervielfältigungen zum Training von KI.<sup>579</sup> Da ein Markt für Trainingsdatensätze vorhanden ist und das Training selbst auch von anderen Sachverhaltskonstellationen wie der Archivierung oder der allgemeinen wissenschaftlichen Verwendung von Werken abgegrenzt werden kann, ist diese Vervielfältigung von Werken für das Training von KI eine eigenständige

---

577 Leistner, GRUR 2024, 1665 (1671).

578 Leistner, GRUR 2024, 1665 (1671).

579 Dazu siehe oben unter 3. Teil A. III. 2. a) bis e).

Nutzungsart.<sup>580</sup> Die Vorbehaltswirkung des § 44b Abs. 3 UrhG erstreckt sich damit auf diese selbstständige Nutzungsart. Nutzungsvertrag und Nutzungsart sind über § 44b Abs. 3 UrhG also miteinander verbunden. Aus diesem Grund sind nur solche Inhaber inhaltlich beschränkter Nutzungsrechte vorbehaltbefugt, die das ausschließliche Nutzungsrecht zur Vervielfältigung von Werken für das Training von KI innehaben, insoweit die Vervielfältigung nicht von § 44b Abs. 2 UrhG freigestellt ist.<sup>581</sup> Umgekehrt hat dies zur Folge, dass beispielsweise solche Rechtsinhaber keinen wirksamen Nutzungsvertrag im Sinne des § 44b Abs. 3 UrhG erklären können, die das ausschließliche, aber beschränkte Recht innehaben, ein Werk zu verfilmen oder auszustellen.

Ob den Rechtsinhabern das Nutzungsrecht für nicht von § 44b UrhG freigestellte Vervielfältigungshandlungen zum KI-Training eingeräumt wurde, ist im Weiteren von der Ausgestaltung und (insbesondere) Auslegung der jeweiligen Nutzungsverträge abhängig. Insbesondere bei pauschalen Rechtseinräumungen kommt es dabei vor allem auf den Zweck der Rechtsübertragung an (§ 31 Abs. 5 UrhG). Von Bedeutung kann auch eine Einräumung von Rechten an unbekannten Nutzungsarten gemäß § 31a UrhG sein. Unbekannt ist eine Nutzungsart, wenn sie dem durchschnittlichen Urheber zum Zeitpunkt des Abschlusses des Lizenzvertrags nicht als praktisch durchführbare eigenständige Nutzungsart geläufig ist.<sup>582</sup> Hierzu genügt schon, dass zwar die technische Möglichkeit, nicht aber die wirtschaftliche Bedeutung und urheberrechtliche Verwertbarkeit der Nutzungsart für den durchschnittlichen Urheber bekannt ist.<sup>583</sup> Unschädlich ist ebenso Kenntnis in (technischen) Fachkreisen.<sup>584</sup> Dass Werke in großem Umfang verwertet werden können, um insbesondere generative KI zu trainieren, und in der Verwertung von Werken als Trainingsdaten daher auch eine erhebliche wirtschaftliche Auswertungsmöglichkeit liegt, ist dem durchschnittlichen Urheber wohl seit der von einer gesellschaftlichen Debatte begleiteten Veröffentlichung des KI-gestützten Chatbots

---

580 Pukas, GRUR 2023, 614 (617); im Anschluss hieran Schack, NJW 2024, 113 Rn. 16; Hamann, ZGE 2024, 113 (138). Ebenso Kögel, InTeR 2023, 179 (181).

581 Heine, GRUR-Prax 2024, 87 Rn. 19.

582 Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 31a Rn. 44.

583 BGH, Urt. v. 11.10.1990 - I ZR 59/89, GRUR 1991, 133 (136) – Videozweitauswertung; BGH, Urt. v. 26.01.1995 - I ZR 63/93, GRUR 1995, 212 (213) – Videozweitauswertung III.

584 Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 31a Rn. 44; Strobl, ZUM 2023, 492 (498).

*ChatGPT* für breite Verbraucherkreise am 30. November 2022 bekannt. Vor diesem Veröffentlichungsdatum ist die Nutzung (also Vervielfältigung) von Werken für das Training von KI im Umkehrschluss eine unbekannte Nutzungsart im Sinne des § 31a UrhG.<sup>585</sup> Rechtsinhabern, denen vor diesem Datum Rechte an unbekannten Nutzungsarten im Sinne des § 31a UrhG eingeräumt worden sind, kommt deswegen auch das Nutzungsrecht zur Vervielfältigung von Werken für das Training von KI zu, insoweit die Vervielfältigungen nicht von § 44b Abs. 2 UrhG freigestellt sind. Sie sind daher vorbehaltbefugt, solange kein Widerruf der Rechtseinräumung im Sinne des § 31a Abs. 1 S. 3 UrhG erfolgt ist.

Verwertungsgesellschaftsrechtliche Wahrnehmungsverträge, von denen pauschal das Vervielfältigungsrecht umfasst wird, sind so auszulegen, dass hierunter auch das Recht zur Vervielfältigung von Werken für das Training von KI fällt.<sup>586</sup> Weiterhin ausreichend ist eine Einräumung des Vervielfältigungsrechts „für die Entwicklung neuer Software“ oder die „Auswertung zugrundeliegender Informationen wie Muster, Trends oder Korrelationen“. Denn beides umfasst unmittelbar auch die Nutzung der Werke zum Training neuer KI-Systeme.

Nicht vorbehaltbefugten Rechtsinhabern kann trotz alledem auch Vertretungsmacht für die Erklärung des Nutzungsvorbehalts in Bezug auf ihre jeweilige Werkwiedergabe eingeräumt werden. Dies richtet sich nach den Regelungen der Stellvertretung, §§ 164 ff. BGB, die analog angewendet werden können.<sup>587</sup>

## (ii) Zeitliche sowie räumliche Beschränkungen und die Vorbehaltbefugnis

Der Nutzungsvorbehalt gegen die Freistellung von Vervielfältigungshandlungen für das Text und Data Mining wirkt darüber hinaus sowohl zeitlich als auch räumlich uneingeschränkt. Das wirft die Frage auf, ob Inhaber räumlich oder zeitlich beschränkter nutzungsrechte ebenfalls vorbehaltbefugt sind.

Gegen eine Vorbehaltbefugnis des Inhabers eines zeitlich oder räumlich beschränkten Nutzungsrechts spricht, dass es widersprüchlich erscheint, ihm ein Vorbehaltrecht zuzubilligen, welches umfassender ist als das ihm eingeräumte Verwertungsrecht. In Bezug auf zeitlich beschränkte Rechts-

---

585 So auch *Nordemann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 31a Rn. 49.

586 *Pukas*, GRUR 2023, 614 (618).

587 Dazu vgl. bereits unmittelbar oben.

einräumungen muss allerdings berücksichtigt werden, dass der Nutzungs- vorbehalt nur wiedergabebezogene Wirkung entfaltet.<sup>588</sup> Damit wird die Freistellungswirkung der Schrankenbestimmung stets nur in Bezug auf die Nutzung einer konkreten öffentlichen Wiedergabe oder eines spezi- fischen Vervielfältigungsstücks als Vervielfältigungsvorlage beendet. Der Nutzungs- vorbehalt beendet die Freistellungswirkung zwar zeitlich unbeschränkt. Aufgrund der wiedergabebezogenen Wirkung des Nutzungs- vorbehals kann er von einem neuen Rechtsinhaber oder dem Urheber nach Erlöschen des ersten, zeitlich beschränkten Nutzungsrechts in Bezug auf seine eigene öffentliche Wiedergabe des Werkes oder neu angefertigten Ver- vielfältigungsstücke aber erneut ausgeübt werden. Aufgrund der wiedergabe- bezogene Wirkung des Nutzungs- vorbehals kommt es demzufolge nicht zu einem Widerspruch zwischen der unbeschränkten Vorbehalsfolge und einer zeitlich beschränkten Nutzungsrechtseinräumung.

Die wiedergabebezogene Wirkung des Nutzungs- vorbehals führt weiter- hin dazu, dass die absolut wirkende Vorbehaltserklärung auch bei einer räumlichen Beschränkung des Nutzungsrechts de facto nur Sperrwirkung in Bezug auf die Nutzung des jeweiligen Vervielfältigungsstücks oder der jeweiligen öffentlichen Wiedergabe entfaltet, die von der räumlich be- schränkten Nutzungsrechtseinräumung erfasst ist. Ein Nutzungs- vorbehalt, der beispielsweise an der deutschen Sprachfassung eines Spielfilms erklärt worden ist, entfaltet seine Vorbehaltswirkung nur gegen die Nutzung dieser Sprachfassung. Auch wenn dem Vorbehalt de jure eine absolute Wirkung zukommt, berührt die Erklärung damit nicht die Vervielfältigung von Werkexemplaren anderer Sprachfassungen. Die wiedergabebezogene Wirkung des Nutzungs- vorbehals führt daher auch bei einer räumlichen Be- schränkung des Nutzungsrechts dazu, dass es nicht zu einem Widerspruch zwischen der unbeschränkten Wirkung des Nutzungs- vorbehals und der räumlichen Beschränkung des Nutzungsrechts kommt. Nicht zuletzt ver- mittelt auch ein ausschließliches, aber räumlich oder zeitlich beschränktes Nutzungsrecht den Rechtsinhabern eine Rechtsposition, Kraft derer sie in die Lage versetzt sind, anderen die Nutzung des betreffenden Werkes zu untersagen. Damit ist auch die Vorbehalsbefugnis des Inhabers eines räumlich oder zeitlich beschränkten, ausschließlichen Nutzungsrechts in- teressengerecht.

Alles in allem ergeben sich keine Wertungswidersprüche zwischen der räumlichen oder zeitlichen Beschränkung eines ausschließlichen Nutzungs-

---

588 Dazu siehe oben unter 3. Teil A. III. 2. i) ff).

rechts und der Möglichkeit zur Abgabe eines de jure unbeschränkten Nutzungsvertrags, wie § 44b Abs. 3 UrhG ihn ermöglicht. Urheber oder ausschließliche Rechtsinhaber werden von der Vorbehaltserklärung des Inhabers eines anderen ausschließlich, aber räumlich oder zeitlich beschränkten Nutzungsrechts nicht beeinträchtigt. Maßgeblich hierfür ist die Wiedergabebezogene Wirkung der Vorbehaltserklärung. Insgesamt sprechen die vorgenannten Gründe dafür, dass auch Inhaber eines ausschließlich, aber räumlich oder zeitlich beschränkten Nutzungsrechts uneingeschränkt vorbehaltbefugt sind. Auch sie können damit Nutzungsverträge im Sinne des § 44b Abs. 3 UrhG abgeben.

(3) Vertragliche Schutzpflicht beim Auseinanderfallen von  
Wiedergabebeherrschung und Vorbehaltbefugnis

In der praktischen Umsetzung der durch § 44b Abs. 3 UrhG gewährten Vorbehaltsmöglichkeit kann es unter Umständen zu einem Auseinanderfallen der Herrschaftsgewalt über die Wiedergabe des Werkes und der Befugnis zur Erklärung des Nutzungsvertrags im Sinne des § 44b Abs. 3 UrhG kommen. Dies ist insbesondere bei der öffentlichen Wiedergabe von Werken durch einfache Rechtsinhaber der Fall. Sie sind kraft ihres Nutzungsrechts zwar wiedergabe-, nicht aber vorbehaltbefugt. Der in diesen Fällen betroffene Urheber oder ausschließliche Rechtsinhaber ist de facto nicht in der Lage, in der Sphäre des nutzenden Rechtsinhabers die Vorbehaltserklärung umzusetzen.<sup>589</sup> Auch die Möglichkeit, den Nutzungsvertrag als Stellvertreter des Vorbehaltbefugten analog §§ 164 ff. BGB erklären zu können,<sup>590</sup> hilft hier nur bedingt weiter. Denn aus dem rechtlichen Können folgt nicht unbedingt, dass der nutzende Rechtsinhaber diese Befugnis auch tatsächlich ausübt. Prohibitive Effekte können zum Beispiel bestehen, wenn Kosten mit der Umsetzung der Vorbehaltserklärungen verbunden sind und der nutzende Rechtsinhaber im Übrigen kein eigenes Interesse daran hat, die Verwendung der Werke zum KI-Training zu unterbinden. Aus dem Abhängigkeitsverhältnis zwischen wiedergebendem Nutzer und vorbehaltbefugtem Urheber beziehungsweise ausschließlichem Rechtsinhaber leitet sich jedoch eine vertragliche Schutzpflicht (§ 241 Abs. 2 BGB) in Form einer Mitwirkungspflicht des Nutzers ab. Sie beinhaltet, dass auf Verlangen des Vorbehaltbefugten diejenigen erforderlichen Maßnahmen

---

589 Heine, GRUR-Prax 2024, 87 Rn. 20.

590 Dazu siehe oben unter 3. Teil A. III. 2. i) hh) (1).

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

zu ergreifen sind, damit der Urheber oder ausschließliche Rechtsinhaber die Vorbehaltserklärung in angemessener Weise abgeben kann.<sup>591</sup> Hiervon betroffen sind unter anderem auch Verwertungsgesellschaften, die Nutzern einfache Lizenzen zur Nutzung des Werkes einräumen, ohne aber in die Wiedergabehandlung der Lizenznehmer eingebunden zu sein.

Die Mitwirkungspflicht, die sich grundsätzlich bereits im Wege ergänzender Vertragsauslegung aus dem Lizenzvertrag zwischen den Parteien ableitet, kann auch durch eine ausdrückliche Bestimmung in den Lizenzbedingungen abgesichert werden. Das ist beispielsweise in Abschnitt I. Nr. 3 Abs. 3 GEMA Tarif VR-OD15 der Fall.<sup>592</sup>

#### ii) Fernwirkung des Nutzungsvorbehalts bei Bereitstellung von Vervielfältigungsvorlagen oder einer Zugangsvermittlung an Dritte

Online verfügbare Werke können insbesondere durch Inhaber von Datenbankzugängen, über die die Werke abrufbar sind, Dritten Datenverarbeitern für Text und Data Mining-Aktivitäten und damit auch für das Training von KI weitervermittelt werden. Das gilt auch dann, wenn in diesen Datenbanken Nutzungsvorbehalte im Sinne des § 44b Abs. 3 UrhG erklärt worden sind. Vor dem Hintergrund dieses Dreiecksverhältnisses zwischen Vorbehaltserklärendem, Zugangsinhaber und Datenverarbeiter ist klärungsbedürftig, ob und inwiefern für den Zugangsinhaber wahrnehmbare Nutzungsvorbehalte, beispielsweise in den Nutzungsbedingungen der Online-Datenbank, zulasten des außenstehenden Datenverarbeiters weiterwirken, auch wenn dieser unter Umständen keine Kenntnis von der Vorbehaltserklärung hat.<sup>593</sup>

Die Weitervermittlung ist auf zwei Arten umsetzbar. Zum einen kann einem Datenverarbeiter der Zugang zu im Internet verfügbaren Werken direkt weitergegeben werden. Das ist beispielsweise bei der Überlassung von Login-Daten zu einer geschützten Datenbank der Fall, aus der der Datenverarbeiter die zum Training benötigten Werke selbstständig erstmals vervielfältigt. Zum anderen können dem Datenverarbeiter auch Werkstücke als Vervielfältigungsvorlagen bereitgestellt werden, die zuvor vom Zu-

---

591 Zur Angemessenheit des Nutzungsvorbehalts siehe oben unter 3. Teil A. III. 2. i) bb).

592 Siehe den Tarif unter <https://ogy.de/0rf7> (zuletzt abgerufen am 30.7.2024).

593 Zu einer vergleichbaren Konstellation bei der Weitergabe des Zugangs zu Werken im Kontext der Tatbestandsvoraussetzung der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG bereits oben unter 3. Teil A. III. 2. f) bb).

gangsinhabern angefertigt worden sind. In diesem Fall fertigt der Datenverarbeiter in seinem eigenen Herrschaftsbereich Vervielfältigungsstücke auf Basis der zunächst vom Zugangsinhaber hergestellten Vervielfältigungsvorlagen an. Dies ist insbesondere bei Endkunden intermediärer Trainingsdatensammler der Fall, insofern die Intermediären nicht nur Linklisten bereitstellen, sondern die Werke selbst abspeichern und vorverarbeiten, bevor sie auf deren Grundlage fertige Trainingsdatensätze auf dem Markt anbieten.

#### (1) Bereitstellung bereits angefertigter Vervielfältigungsvorlagen

Ob der durch den Zugangsinhaber wahrnehmbare Nutzungsvorbehalt auch zulasten des nachgelagerten Datenverarbeiters wirkt, der vom Zugangsinhaber angefertigte Werkstücke als Vervielfältigungsvorlage nutzt, lässt sich mit Hilfe der wiedergabebezogenen Wirkung des Nutzungsvorbehalts beantworten.<sup>594</sup> Denn der Zugangsinhaber bietet dem Datenverarbeiter seine eigenen, zuvor auf Grundlage des Zugangs angefertigten Werkstücke als neue Vervielfältigungsvorlagen an. Demzufolge kann eine mit den Ausgangswiedergaben, nicht aber mit den neuen Vervielfältigungsvorlagen verknüpfte Vorbehaltserklärung keine Beendigungswirkung für die Nutzung dieser neuen Vervielfältigungsvorlagen entfalten. Die Vorbehaltswirkung bleibt vielmehr auf die Nutzung der Ausgangswiedergabe beschränkt. Die Nutzungshandlungen des nachgelagerten Datenverarbeiters sind dementsprechend von § 44b Abs. 2 UrhG gedeckt und erlaubnisfrei gestellt, auch wenn an der ursprünglichen Werkwiedergabe ein Nutzungsvorbehalt erklärt worden ist. Unabhängig hiervon sind die Vervielfältigungshandlungen des Zugangsinhabers urheberrechtswidrig, wenn er bei ihnen wirksame Nutzungsvorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG nicht beachtet. Zudem könnte durch den Zugangsinhaber eine rechtswidrige Verbreitung (§ 17 UrhG) oder öffentliche Wiedergabe (§ 15 Abs. 2 UrhG) der vervielfältigten Werke stattfinden. Das ist im Detail aber einzelfallabhängig.

Fraglich ist jedoch, ob der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL einer Anwendung der Schrankenbestimmung zugunsten des Datenverarbeiters in der untersuchten Fallgruppe dann entgegensteht, wenn die Bereitstellung der Vervielfältigungsvorlagen der Werke im Interesse des Datenverarbeiters, also nicht im Rahmen einer Hilfstätigkeit für den Zugangsinhaber erfolgt, und der Datenverarbei-

---

<sup>594</sup> Zur wiedergabebezogenen Wirkung siehe oben unter 3. Teil A. III. 2. i) ff).

ter objektiv erkennen kann, dass die ursprünglich genutzten Inhalte mit einem Nutzungsvorbehalt versehen waren.<sup>595</sup> Gegen eine Unvereinbarkeit mit dem Dreistufentest spricht aber, dass die Bereitstellung der Vervielfältigungsvorlagen der Werke im vorliegenden Fall nicht generell zu einer urheberrechtskonformen Nutzung der Werke auf Basis von § 44b UrhG führt, wie es bei der Bereitstellung von Vervielfältigungsvorlagen *zugangsbeschränkter* Werke an Dritte der Fall ist. Vielmehr verletzen Zugangsinhaber, die die Werke trotz des Nutzungsvorbehalts zunächst vervielfältigen, bereits bei der Herstellung der Vervielfältigungsvorlagen mangels Anwendung der Schrankenbestimmung aus § 44b Abs. 2 UrhG die Rechtsposition betroffener Urheber beziehungsweise Rechtsinhaber. Es kommt damit nicht zu einer vollständigen Umgehung der Wirkung des Nutzungsvorbehalts. Die Urheberrechtsverletzung wird nur beim Zugangsinhaber lokalisiert. Der vorliegenden Sachverhaltskonstellation wohnt demzufolge kein Umgehungsmoment inne. Damit besteht ein Unterschied zur Bereitstellung von Vervielfältigungsvorlagen zugangsbeschränkter Werke an Dritte, die de jure zu einer Umgehung der eingesetzten Zugangsbeschränkung führt.<sup>596</sup> Infolgedessen kommt es nicht zu einer ungebührlichen Verletzung der Urheberinteressen. Eine Einschränkung der Anwendung von § 44b Abs. 2 UrhG für den Fall, dass der Datenverarbeiter im eigenen Interesse tätig wird und objektiv erkennen kann, dass die ursprünglichen Werkwiedergaben mit einem Nutzungsvorbehalt versehen waren, findet über den Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL demzufolge nicht statt.

## (2) Zugangsvermittlung

Auch die Auswirkung einer Vorbehaltserklärung auf Vervielfältigungshandlungen, die auf Grundlage eines dem Datenverarbeiter vom Zugangsinhaber weitervermittelten Werkzugangs vorgenommen worden sind, richtet sich nach der wiedergabebezogenen Wirkung des Nutzungsvorbehalts.<sup>597</sup> Aufgrund dessen, dass dem Datenverarbeiter lediglich der Werkzugang vermittelt, also beispielsweise Login-Daten weitergegeben worden sind, nutzt er als Vervielfältigungsvorlage weiterhin die ursprüngliche Wiedergabe beziehungsweise das ursprüngliche Werkstück, an dem der Nutzungs-

---

595 Vgl. zu einem Parallelproblem bei der Weiterwirkung von Zugangsbeschränkungen im Rahmen des § 44b Abs. 2 S. 1 UrhG oben unter 3. Teil A. III. 2. f) bb).

596 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (3).

597 Zur wiedergabebezogenen Wirkung siehe oben unter 3. Teil A. III. 2. i) ff).

vorbehalt erklärt worden ist. Damit bleibt der wiedergabebezogen wirkende Nutzungsvorbehalt ausschlaggebend für die Freistellungswirkung des § 44b Abs. 2 UrhG. Die Zugangsweitergabe hat auf die Wirkung des Nutzungs-vorbehalts demzufolge keinen Einfluss. Auch wenn Datenverarbeitern der Werkzugang von einem Zugangsinhaber vermittelt wird, müssen sie beim Vervielfältigen der Inhalte Nutzungsvorhalte im Sinne des § 44b Abs. 3 UrhG auslesen. Dies setzt unter Umständen voraus, dass sich die Daten-verarbeiter Zugriff auf die Nutzungsbedingungen einer Werkdatenbank ver-schaffen, bevor sie Vervielfältigungen für die Datenanalyse anfertigen.

jj) Zwischenergebnis zum Nutzungsvorbehalt

Alles in allem müssen Nutzungsvorbehaltserklärungen ausdrücklich und angemessen sein, um die in § 44b Abs. 3 UrhG angelegte Beendigungswir-kung auszulösen. Beides wird zwar nicht ausdrücklich in § 44b Abs. 3 UrhG vorausgesetzt. Dennoch sind die Anforderungen im Wege einer richtlinien-konformen Auslegung des „sich Vorbehaltens“ im Sinne des § 44b Abs. 3 S. 1 UrhG mitzuberücksichtigen. Trotz der Ausdrücklichkeitsanforderung dürfen die Nutzungsvorhalte allerdings mit anderen, nicht auf § 44b Abs. 3 UrhG bezogenen Vorbehaltserklärungen verbunden werden. Vor-behaltserklärungen gegen die Nutzung online zugänglicher Werke müssen darüber hinaus gemäß § 44b Abs. 3 S. 2 UrhG maschinenlesbar sein. Dies setzt zwar voraus, dass die Vorbehaltserklärungen digital hinterlegt sind, sodass sie unmittelbar durch einen Computer verarbeitet werden können. Im Übrigen ist aber weder eine Formalisierung noch eine Strukturierung der Vorbehaltserklärungen notwendig. Damit können Nutzungsvorhalte bei online verfügbaren Werken grundsätzlich auch in natürlicher Sprache abgegeben werden. Dies wird aber durch die richtlinienkonform zu berück-sichtigende Angemessenheitsvoraussetzung eingegrenzt. Hiernach müssen die Erklärungen insbesondere gewisse Klarheitsanforderungen erfüllen. Es bestehen weiterhin gewisse Einschränkungen hinsichtlich der verwendba-ren Sprache und des Ortes, an dem die digitalen Vorbehaltserklärungen hinterlegt werden dürfen. Insgesamt wird durch die Erklärbarkeit des Nut-zungsvorbehalts in natürlicher Sprache eine niedrige Hürde für Rechtsin-haber zur Abgabe der Vorbehaltserklärung aufgestellt. Praktisch können die Anforderungen unter anderem auch durch die Verwendung des Ro-bots Exclusion Standards, des Spawning-Protokolls und des W3C TDM Reservation Protocols erfüllt werden. Wegen der niederschweligen Wirk-

samkeitsvoraussetzungen und der leichten Erklärbarkeit des Vorbehalts ist aus einer praktischen Perspektive zu befürchten, dass viele Rechtsinhaber die Freistellungswirkung der Schrankenbestimmung aus § 44b UrhG für die Nutzung ihrer Werke beenden werden. Folge dessen ist eine erhebliche Unterminierung des im zweiten Teil der vorliegenden Untersuchung identifizierten Freistellungsbedürfnisses für die Speicherung von urheberrechtlich geschützten Trainingsdaten aus Internetquellen. Nur die wiedergabebezogene Wirkung des Nutzungsvorbehalts trägt dem Freistellungsbedürfnis insoweit Rechnung, als das eine weitreichende, werkbezogene Ausschlusswirkung der Vorbehaltserklärungen vermieden wird.

j) Fazit zur Schrankenbestimmung für Text und Data Mining

Alles in allem kann das KI-Training nach einer richtlinienkonformen Auslegung des Tatbestands unter den Text und Data Mining-Begriff des § 44b Abs. 1 UrhG gefasst werden. Damit wird das Web Scraping von urheberrechtlich geschützten Trainingsdaten grundsätzlich von § 44b Abs. 2 UrhG freigestellt. Das gilt sowohl für kommerziell als auch für nicht-kommerziell tätige Trainingsdatensammler. Allerdings bleibt die Schrankenbestimmung aufgrund der in § 44b Abs. 2 S. 2 UrhG geregelten Löschungspflicht und dem in § 44b Abs. 3 UrhG vorgesehenen Opt-out-Mechanismus dennoch hinter den im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnissen zurück.

Bei der Erfassung und Speicherung von Werken zum Training von KI haben Datenverarbeiter auch dann rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG, wenn ihnen Vervielfältigungsvorlagen von einem Zugangsinhaber zur Verfügung gestellt worden sind, ohne dass die Datenverarbeiter selbst die Zugangsvoraussetzungen erfüllen. Aufgrund des Dreistufentests muss hiervon aber eine Ausnahme gemacht werden, wenn für den im eigenen Interesse handelnden Datenverarbeiter objektiv erkennbar ist, dass die Vervielfältigungsvorlagen auf Grundlage von Werkstücken aus einer zugangsgeschützten Quelle hergestellt wurden. Das gilt jedoch nicht, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, auf Basis des Werkzugangs angefertigte Werkstücke dritten Datenverarbeitern als Vorlagen für Vervielfältigungen zum Text und Data Mining bereitzustellen. Nach diesen Maßstäben ist zwar auch eine direkte Zugangsweitergabe eines vermittelnden Zugangsinhabers,

beispielsweise durch die Weitergabe von Login-Daten an einen Datenverarbeiter möglich. Hier führt eine Anwendung des Dreistufentests aber dazu, dass eine Anwendung der Schrankenbestimmung aus § 44b UrhG zugunsten des Datenverarbeiters im Regelfall dann ausscheidet, wenn er nicht als Hilfsperson des Zugangsinhabers, sondern im eigenen Interesse tätig wird.

Vervielfältigungen müssen gemäß § 44b Abs. 2 S. 1 UrhG gelöscht werden, wenn sie für das Text und Data Mining nicht mehr erforderlich sind. Dies bezieht sich auf ein Training eines konkreten KI-Systems. Die Vervielfältigungen der Werke dürfen aber aufbewahrt werden, solange noch Nachbesserungen am jeweiligen KI-System vorgenommen werden, für die die Vervielfältigungen notwendig sind.

Vervielfältigungen von im Internet abrufbaren Werken können jedoch nicht durch die Schrankenbestimmung für Text und Data Mining freigegeben werden, wenn ein maschinenlesbarer Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist. Hierfür genügt grundsätzlich eine Vorbehaltserklärung in natürlicher Sprache. Es müssen aber die Angemessenheits- und Ausdrücklichkeitsvoraussetzungen erfüllt sein, die sich aus einer richtlinienkonformen Auslegung des § 44b Abs. 3 UrhG ergeben. Die Möglichkeit, einen Nutzungsvorbehalt zu erklären, konterkariert erheblich die im zweiten Teil der vorliegenden Untersuchung identifizierten Freistellungsbedürfnisse.

### 3. Andere Schrankenbestimmungen

Klärungsbedürftig bleibt damit, ob sich Trainingsdatensammler zur Verwirklichung der identifizierten Freistellungsinteressen auf andere urheberrechtliche Schrankenbestimmungen berufen können.

#### a) Schrankenbestimmung für Zitate (§ 51 UrhG)

Da trainierte KI-Systeme bei ihrer Anwendung mittelbar auf die Trainingsdaten Bezug nehmen, könnte zum einen die Schrankenbestimmung für Zitate (§ 51 UrhG) angewendet werden. Nach § 51 Abs. 1 S. 1 UrhG ist hierfür notwendig, dass von den Trainingsdatensammlern ein Zitatzweck verfolgt wird. Beim Web Scraping von Werken für das KI-Training könnte das nur der Fall sein, um eine geistige Auseinandersetzung zwischen den für das Training vervielfältigten Werken und den Aussagen des Trainings-

datensammlers oder KI-Nutzers zu ermöglichen.<sup>598</sup> Die Erstellung eines Trainingsdatenkorpus beinhaltet jedoch regelmäßig keinen Aussagegehalt, sodass eine geistige Auseinandersetzung mit den vervielfältigten Werken bereits deswegen nicht stattfindet.<sup>599</sup> Auch bei der Datenannotation wird ein für das Training genutztes Werk lediglich beschrieben. Einen eigenständigen Aussagegehalt enthält die Annotation hingegen nicht. Selbst wenn die KI-Nutzer mit der auf das Training folgenden Anwendung der KI einen Aussagegehalt transportieren, mangelt es ihnen außerdem typischerweise an der Kenntnis über die konkret verwendeten Trainingsdaten. Eine geistige Auseinandersetzung mit den Trainingswerken ist dann ebenfalls nicht möglich. Insgesamt wird mit dem Web Scraping von Werken zum Training von KI keine geistige Auseinandersetzung angestrebt, sodass die Datensammlung mangels Verfolgung eines Zitatzwecks nicht über § 51 UrhG freigestellt werden kann.<sup>600</sup>

b) Schrankenbestimmung für die Benutzung eines Datenbankwerks (§ 55a UrhG)

Zumindest auf das Web Scraping von Datenbankwerken könnte darüber hinaus § 55a UrhG anwendbar sein. Da die dauerhafte Vervielfältigung, so wie sie bei der Erfassung und Speicherung von Werken für das Training von KI stattfindet, das für die übliche Benutzung der Datenbankwerke erforderliche Maß im Regelfall aber überschreitet,<sup>601</sup> kann § 55a UrhG im Ergebnis nicht das Web Scraping von Datenbankwerken für das Training von KI freistellen.<sup>602</sup> Wenn es sich bei den zum Training vervielfältigten Datenbankwerken hingegen um öffentlich geteilte Trainingsdatensätze handelt, liegt im Upload der Datenbankwerke bereits eine schlichte Einwilligung bezogen auf ihre Weiterverwendung für das Training von KI. In diesen Fällen ist eine Freistellung demnach auch ohne Eingreifen der Schrankenbestimmung gewährleistet.<sup>603</sup>

---

598 Vgl. zur Voraussetzung EuGH, Urt. v. 29.07.2019 - C-516/17, GRUR 2019, 940 Rn. 78 – Reformistischer Aufbruch.

599 So auch *Spindler*, GRUR 2016, II112 (II116); *Spindler*, ZGE 2018, 273 (278).

600 Etwas weniger streng, größtenteils aber ablehnend *Kleinkopf*, Text- und Data-Mining, S. 183 f.

601 *Lüft/Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 55a Rn. 75; sehr kritisch auch *Spindler*, in: Schrieker/Loewenheim, Urheberrecht, UrhG § 55a.

602 *Kleinkopf*, Text- und Data-Mining, S. 187.

c) Schrankenbestimmung für die Nutzung als unwesentliches Beiwerk (§ 57 UrhG)

Eine Freistellung des Web Scrapings von Werken für das Training von KI ließe sich im Einzelfall auch über die Schrankenbestimmung für unwesentliches Beiwerk, § 57 UrhG, erreichen.

aa) Irrtümlich nicht erfasste Nutzungsvorbehalte (§ 44b Abs. 3 UrhG)

Denkbar ist zum einen, dass Trainingsdatensammler trotz einer Implementierung technischer Sicherungsmechanismen zur Berücksichtigung von Nutzungsvorbehalten unbeabsichtigerweise Inhalte vervielfältigen, für deren Nutzung Vorbehaltserklärungen im Sinne des § 44b Abs. 3 UrhG abgegeben worden sind. Diese unbeabsichtigten Vervielfältigungen könnten in den Anwendungsbereich der Schrankenbestimmung für die Nutzung von Werken als unwesentliches Beiwerk (§ 57 UrhG) fallen.<sup>604</sup>

Entscheidend ist dabei zum einen, welches Objekt in diesen Fällen Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG ist. Dafür, dass sich der anvisierte Trainingsdatenkorpus bestehend aus den über § 44b UrhG rechtmäßig nutzbaren Werken in seiner Gesamtheit als Hauptgegenstand der Nutzungshandlung qualifiziert, spricht, dass Trainingsdaten vom relevanten Verkehrskreis schon deswegen als Datensätze und nicht als einzelne Datenpunkte betrachtet werden, weil das Training von KI in der Regel einer Vielzahl an Datenpunkten bedarf und daher stets nur der Datensatz als Ganzes technische und wirtschaftliche Bedeutung entfaltet.<sup>605</sup> Eine Gruppierung von Vervielfältigungsobjekten kann im Rahmen des § 57 UrhG darüber hinaus trotz dessen, dass § 57 UrhG lediglich im Singular von *dem Gegenstand* der Vervielfältigung spricht, als Hauptgegenstand im Sinne des § 57 UrhG angesehen werden.<sup>606</sup> Denn Zweck des § 57 UrhG ist, eine Werknutzung freizustellen, die die urheberrechtlichen Verwertungsinteressen nur unerheblich beeinträchtigen. Dafür ist nur von untergeordneter Bedeutung, im Verhältnis zu welchem Objekt oder welchen Objekten

---

603 Anders *Kleinkopf*, Text- und Data-Mining, S. 187, die in diesen Fällen § 55a UrhG zur Anwendung bringen will.

604 Dazu im Kontext erweiterter kollektiver Lizzenzen ähnlich auch in *Pukas*, GRUR 2023, 614 (619 f.).

605 Vgl. zu den quantitativen Anforderungen für die Trainingsdatensätze im Überblick oben I. Teil B. II. 3.

606 *Jacobs*, in: FS Ahrens, S. 225 (227).

die Werknutzung als unwesentliches Beiwerk erscheint. Dafür, dass sich dieser *anvisierte* Trainingsdatenkorpus bestehend aus den über § 44b UrhG rechtmäßig nutzbaren Werken als Hauptgegenstand der Nutzungshandlung im Sinne des § 57 UrhG qualifiziert, spricht weiterhin, dass Trainingsdatensammler aufgrund der Einsetzung technischer Sicherungsmechanismen zur Berücksichtigung von Nutzungsvorbehalten im vorliegenden Fall darauf abzielen, nur solche Inhalte zu erfassen, die nicht mit einem Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG versehen worden sind. Deswegen ist eine Abgrenzung zu den fälschlicherweise trotz eines erklärten Nutzungsvorbehalts vervielfältigten Werken möglich. Beide Werkgruppen sind in der vorliegenden Fallkonstellation demzufolge nicht gleichrangig. Insgesamt sprechen damit mehrere Gründe dafür, dass der Trainingsdatenkorpus bestehend aus den Inhalten, die nicht mit einem Nutzungsvorbehalt versehen worden sind, Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG ist.<sup>607</sup>

Einzelne Werke, die unbeabsichtigt trotz eines wirksamen Nutzungsvorbehalts vervielfältigt worden sind, könnten im Verhältnis zu diesem Hauptgegenstand als unwesentliches Beiwerk erscheinen.<sup>608</sup> Voraussetzung dessen ist, dass sie nur beiläufig einbezogen und vom Durchschnittsbetrachter nicht wahrgenommen werden, sodass sie beliebig weggelassen oder ausgetauscht werden können und die Gesamtwirkung des Hauptgegenstands der Vervielfältigung nicht beeinflussen.<sup>609</sup> Dies ist unter Berücksichtigung aller Umstände des Einzelfalles zu beurteilen.<sup>610</sup> In Relation zum gesamten Trainingsdatenkorpus aus den aufgrund von § 44b UrhG rechtmäßig erfassten Werken machen einzelne, unbeabsichtigerweise trotz Nutzungsvorbehaltserklärung vervielfältigte Werke einen nur untergeordneten Teil des Datenkorpus aus. Die einzelnen Werke aus dem Trainingsdatensatz herauszunehmen, würde seine Gesamtwirkung nicht beeinflussen. Die betroffenen Werke sind daher beliebig austauschbar. Ob darüber hinaus eine nur beiläufige Einbeziehung der Werke stattfindet, die von § 57 UrhG freigestellt ist, kann bei der Berücksichtigung sämtlicher Umstände des Einzelfalls davon abhängig gemacht werden, ob der Trainingsdatensammler wirksame Maßnahmen zur Beachtung erklärter Nutzungsvorbehalte gemäß § 44b

---

607 Ähnlich zum Widerspruch bei ECL-Lizenzlösungen schon in *Pukas*, GRUR 2023, 614 (619 f.).

608 Vgl. *Pukas*, GRUR 2023, 614 (619 f.).

609 BGH, Urt. v. 17.11.2014 - I ZR 177/13, GRUR 2015, 667 Rn. 27, 31 – Möbelkatalog; *Vogel*, in: Schrieker/Loewenheim, Urheberrecht, UrhG § 57 Rn. 10 m. w. N. Kritisch zur Austauschbarkeit *Jacobs*, in: FS Ahrens, S. 225 (229 f.).

610 BGH, Urt. v. 17.11.2014 - I ZR 177/13, GRUR 2015, 667 Rn. 21 – Möbelkatalog.

Abs. 3 UrhG ergriffen und damit bestmögliche Anstrengungen unternommen hat, dass es nicht zur Vervielfältigung der betreffenden Werke kommt. Dann ist die fehlerhafte, unbeabsichtigte Vervielfältigung einzelner, über § 44b Abs. 3 UrhG zur Nutzung durch den Urheber vorbehaltener Werke als unwesentliches Beiwerk des im Übrigen urheberrechtskonform akquirierten Trainingsdatensatzes von § 57 UrhG gedeckt. Die Vervielfältigung der Werke ist in diesem Fall demzufolge urheberrechtlich frei möglich. Mit dieser einzelfallbezogenen Anwendung von § 57 UrhG werden auch nicht strukturell die Voraussetzungen des § 44b UrhG unterlaufen.

Die Anwendung des § 57 UrhG auf diesen Einzelfall, ändert aber nichts daran, dass die im Sinne des § 44b Abs. 3 UrhG erklärten Nutzungsvorbehalte bestmöglich berücksichtigt werden müssen. Selbst die einzelfallbezogene Anwendung der Schrankenbestimmung für die Nutzung von Werken als unwesentliches Beiwerk führt daher nicht zu einer allgemeinen Freistellung des Web Scrapings von im Internet abrufbaren Werken für das Training von KI. Auch die Anwendung von § 57 UrhG kann die im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnisse daher nicht erfüllen.

bb) Trainingsdatensätze aus primär nicht urheberrechtlich geschützten Inhalten

Möglich ist darüber hinaus auch, dass Trainingsdatensammler, deren Ziel die Erfassung und Verarbeitung urheberrechtlich nicht geschützter Inhalte ist, für die Vervielfältigung einzelner Werke, die trotz dessen irrtümlich gespeichert werden, von der Schrankenbestimmung aus § 57 UrhG profitieren könnten. Denkbar ist das beispielsweise, wenn der Trainingsdatensammler amtliche und damit gemäß § 5 UrhG nicht geschützte Werke vervielfältigt, in einem dieser Werke aber urheberrechtlich geschützte Inhalte eingebettet sind, beispielsweise ein Kartenausschnitt.<sup>611</sup> Damit § 57 UrhG für die Vervielfältigung dieser Werke eingreifen kann, müssen die unbeabsichtigt vervielfältigten Werke auch hier gemäß § 57 UrhG neben dem eigentlichen Gegenstand der Vervielfältigung, dem Trainingsdatensatz aus urheberrechtsfreien Inhalten, als unwesentliches Beiwerk erscheinen.

---

<sup>611</sup> Vgl. zu einer ähnlichen Sachverhaltskonstellation BGH, Urt. v. 21.01.2021 - I ZR 59/19, ZUM 2021, 527 – Kastellaun.

Für eine Einordnung des vom Trainingsdatensammler zielgerichtet akquirierten Trainingsdatensatzes aus urheberrechtsfreien Inhalten als Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG ist zunächst unschädlich, dass der Trainingsdatensatz als Objektgruppe aus mehreren Datenpunkten besteht.<sup>612</sup> Weiterhin dient § 57 UrhG dem Zweck, der Allgemeinheit wirtschaftlich unerhebliche Werknutzungen freizugeben. Dabei kann es nicht darauf ankommen, im Verhältnis zu welchem Objekt die wirtschaftlich unerhebliche Werknutzung stattfindet. Aus diesen Gründen ist ebenfalls nicht notwendig, dass der Hauptgegenstand der Vervielfältigung im Sinne des § 57 UrhG selbst urheberrechtlich geschützt ist.<sup>613</sup> Darüber hinaus besteht aufgrund der Zweckrichtung der Datensammlung keine Gleichwertigkeit zwischen den unbeabsichtigt vervielfältigten Werken als Nebenfolge der Datensammlung und den zielgerichtet gespeicherten urheberrechtsfreien Inhalten.<sup>614</sup> Letztere verkörpern im Verhältnis zu den vervielfältigten Werken demzufolge einen eigenständigen Gegenstand. Auch im vorliegenden Fall kann der Trainingsdatensatz aus urheberrechtsfreien Inhalten in seiner Gesamtheit damit als Hauptgegenstand der Vervielfältigungshandlung im Sinne des § 57 UrhG angesehen werden.

Für die Anwendung des § 57 UrhG kommt es infolgedessen darauf an, ob die versehentlich erfassten Werke im Verhältnis zum Trainingsdatensatz aus urheberrechtsfreien Inhalten als unwesentliches Beiwerk im Sinne des § 57 UrhG erscheinen.<sup>615</sup> Auch im vorliegenden Fall machen die vervielfältigten Werke nur einen unwesentlichen Teil des gesamten Trainingsdatensatzes aus. Sie könnten demzufolge weggelassen oder ausgetauscht werden, ohne dass die Qualität und damit die Gesamtwirkung des Datensatzes beeinflusst wird. In die anschließend vorzunehmende Beurteilung der Beiläufigkeit der Vervielfältigung, die in einer Gesamtabwägung mündet, kann eingestellt werden, dass die Vervielfältigungsstücke nicht zielgerichtet hergestellt werden. Ob dies tatsächlich der Fall ist, lässt sich unter anderem anhand der für den Webcrawler definierten Regeln bewerten. Von Bedeutung ist auch, auf welche Quellen der Trainingsdatensammler das Web Scraping ausgerichtet hat. Im Detail ist das einzelfallabhängig. Sind die Vervielfältigungen

---

<sup>612</sup> Dazu siehe unmittelbar oben unter 3. Teil A. III. 3. c) aa).

<sup>613</sup> Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 57 Rn. 1; Jacobs, in: FS Ahrens, S. 225 (227); vgl. z. B. auch OLG Köln, Urt. v. 23.08.2013 - 6 U 17/13, GRUR-RR 2014, 58 (59), das als Hauptgegenstand einen „Verkaufskatalog“ und eine „Internetseite“ ansieht.

<sup>614</sup> Zu einer vergleichbaren Fallkonstellation etwas ausführlicher siehe unmittelbar oben unter 3. Teil A. III. 3. c) aa).

<sup>615</sup> Zu den Voraussetzungen siehe unmittelbar oben unter 3. Teil A. III. 3. c) aa).

beiläufig, ist davon auszugehen, dass sie unwesentliches Beiwerk des Trainingsdatensatzes aus primär urheberrechtsfreien Inhalten im Sinne des § 57 UrhG sind. Infolgedessen werden die Vervielfältigungshandlungen in diesem Fall durch die Schrankenbestimmung aus § 57 UrhG erlaubnisfrei gestellt. Auch das führt jedoch nicht zu einer umfassenden Freistellung des Web Scrapings von Werken als KI-Trainingsdaten. Durch die Anwendung des § 57 UrhG in diesem Einzelfall kann das im zweiten Teil der vorliegenden Untersuchung identifizierte Freistellungsbedürfnis daher ebenfalls nicht vollständig erfüllt werden.

d) Schrankenbestimmung für Karikatur, Parodie und Pastiche (§ 51a UrhG)

Im Einzelfall könnte die Vervielfältigung von Werken als Trainingsdaten auch durch die Schrankenbestimmung für Karikatur, Parodie und Pastiche aus § 51a UrhG erlaubt sein.<sup>616</sup> Denkbar ist, dass bereits die Vervielfältigung konkret ausgewählter Werke, die zum Training einer künstlerisch eingesetzten KI genutzt werden sollen, als vorgelagerte Nutzung von der Schrankenbestimmung privilegiert wird, wenn mit der KI anschließend Karikaturen, Parodien oder Pastiche erzeugt werden sollen.<sup>617</sup> Denn dann könnte bereits die (Auswahl und) Vervielfältigung der als Trainingsdaten genutzten Werke eine künstlerische Betätigung darstellen, die für einen effektiven Grundrechtsschutz und eine wirksame Anwendung des § 51a UrhG von der Schrankennutzung mit zu umfassen ist.<sup>618</sup> Notwendig wäre hierfür aber, dass bereits zum Zeitpunkt der Speicherung der Trainingsdaten hinreichend konkretisiert ist, dass der Output des trainierten, künstlerisch genutzten KI-Systems in den Anwendungsbereich der Schrankenbestimmung für Karikatur, Parodie und Pastiche aus § 51a UrhG fallen wird. Zudem darf nicht *auch*, sondern *nur* die Erzeugung von Karikaturen, Parodien und Pastiche bezeichnet sein.<sup>619</sup> Auch die Abwägung im Rahmen des Dreistufen-

---

616 Insgesamt dazu im Detail *Lennartz/Kraetzig*, GRUR 2024, 1258, wobei die AutorInnen von einem weiten Pastichebegriff ausgehen, der jede abgeleitete künstlerische Betätigung erfasst.

617 *Lennartz/Kraetzig*, GRUR 2024, 1258 (1261, 1262).

618 *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260 f.).

619 Großzügiger hingegen *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 158 f., nach denen es ausreicht, wenn die Erfüllung der Voraussetzung des § 51a UrhG „Hauptzweck“ des KI-Systems ist.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

tests, die bei der Anwendung des § 51a UrhG eine wesentliche Bedeutung entfaltet,<sup>620</sup> müsste für den Output damit vorweggenommen werden. Ohne Kenntnis vom konkreten Output des KI-Systems ist das jedoch kaum prognostizierbar. Insgesamt ist eine Freistellung der Vervielfältigungen über § 51a UrhG damit nur in begrenzten Einzelfällen denkbar.

#### e) Intermediäre Trainingsdatensammler als Bibliotheken (§ 60d UrhG)

Von den regulatorischen Freistellungsbedürfnissen für das Web Scraping urheberrechtlich geschützter Trainingsdaten ist auch die Tätigkeit intermediärer Trainingsdatensammler umfasst. Sie nehmen eine Vermittlerrolle zwischen Urhebern und KI-Entwicklern ein. Vor diesem Hintergrund könnte ihre Datensammlung nicht nur durch § 44b UrhG, sondern auch durch die erweiterte Schrankenbestimmung aus § 60d UrhG freigestellt sein. Diese privilegiert unter anderem Forschungseinrichtungen sowie Vermittler wie Archive oder Bibliotheken. Da es für die Erzielung einer generellen Freistellungswirkung zugunsten intermediärer Trainingsdatensammler nicht auf eine Forschungstätigkeit ankommen kann, müssen sie unter den Begriff der Kulturerbe-Einrichtung im Sinne des § 60d Abs. 3 Nr. 1 UrhG gefasst werden können, um in den persönlichen Anwendungsbereich der erweiterten Schrankenbestimmung zu fallen. In Betracht kommt lediglich, dass sie sich als Bibliotheken im Sinne der Regelung qualifizieren.

Der Bibliotheksbumpf ist weder in § 60d UrhG noch in Art. 2 Nr. 3, Art. 3 DSM-RL legaldefiniert, dessen Umsetzung § 60d UrhG dient. Auch in der InfoSoc-RL (Art. 5 Abs. 2 lit. c) sowie der verwaiste Werke-RL<sup>621</sup> (Art. 1) wird er ohne eine nähere Begriffsbestimmung aufgegriffen. Seine Bedeutung muss deswegen mit Hilfe der üblichen Auslegungsmethoden bestimmt werden. Die Basis der Auslegung bildet dabei auch im europäischen Recht der natürliche Wortsinn des verwendeten Rechtsbegriffs.<sup>622</sup> Im allgemeinen deutschen Sprachgebrauch wird unter einer Bibliothek ein Aufbewahrungsort für eine systematisch geordnete Sammlung von Büchern verstanden.<sup>623</sup>

---

620 Stieper, GRUR 2020, 699 (703).

621 Richlinie 2012/28/EU des Europäischen Parlaments und des Rates vom 25. Oktober 2012 über bestimmte zulässige Formen der Nutzung verwaister Werke.

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

Selbst wenn davon abgesehen wird, dass intermediäre Trainingsdatensammler nicht nur Schriftwerke, also „Bücher“, vervielfältigen, ist darüber hinaus fraglich, ob sie eine Systematisierung ihrer Inhalte vornehmen. Dagegen spricht, dass die Werke innerhalb der Trainingsdatensätze nicht in Kataloge aufgenommen werden. Hierfür ursächlich sind unter Umständen ein Fehlen der für eine Systematisierung notwendigen Metadaten sowie die Menge der verarbeiteten Inhalte. Eine Systematisierung ist zudem auch nicht für das Training notwendig. Zwar kann auf Plattformen intermediärer Trainingsdatensammler mitunter eine Katalogisierung, also systematische Anordnung verschiedener Trainingsdatensätze stattfinden.<sup>624</sup> Vom Bibliotheks begriff wird allerdings eine systematische Anordnung der gesammelten Inhalte selbst vorausgesetzt. Das sind im vorliegenden Fall die für das Training bestimmten Werkstücke.

Mangels Systematisierung der vervielfältigten Werke fallen intermediäre Trainingsdatensammler demzufolge nicht unter den in § 60d Abs. 3 Nr. 1 UrhG respektive Art. 2 Nr. 3 DSM-RL normierten Bibliotheks begriff. Es kommt damit bereits nicht darauf an, ob als Bibliothek eine räumliche Einrichtung erforderlich ist.<sup>625</sup> Intermediäre Trainingsdatensammler können damit jedenfalls so lange nicht von der erweiterten Schrankenwirkung des § 60d UrhG profitieren, wie sie nicht aus anderen Gründen in den persönlichen Anwendungsbereich der Schrankenbestimmung fallen.<sup>626</sup> Die regulatorischen Defizite im Hinblick auf die Löschungspflicht (vgl. § 44b Abs. 2 S. 2 UrhG)<sup>627</sup> sowie den Nutzungsvorbehalt (vgl. § 44b Abs. 3 UrhG)<sup>628</sup> können durch die Anwendung der Schrankenbestimmung aus § 60d UrhG damit nicht zugunsten intermediärer Trainingsdatensammler geheilt werden.

---

623 *Dudenredaktion (Bibliographisches Institut)*, Duden: das Fremdwörterbuch, Eintrag: „Bibliothek“.

624 Siehe zum Beispiel die Plattform „huggingface.co“, auf der unter anderem eine große Zahl an Trainingsdatensätzen angeboten wird.

625 Vgl. dazu Henke, E-Books im Urheberrecht, S. 118 ff.

626 Zu einem entsprechenden Fall, bei dem eine wissenschaftliche Analysen vorbereitende Datensammlung als ausreichend erachtet wurde, siehe LG Hamburg, Urt. v. 27.09.2024 - 310 O 227/23, GRUR-RS 2024, 25458 – Laion.

627 Siehe dazu oben unter 3. Teil A. III. 2. h).

628 Siehe dazu oben unter 3. Teil A. III. 2. i).

#### 4. Rolle der Gedächtnisinstitutionen für das Training von KI

Kulturerbe-Einrichtungen im Sinne des § 60d Abs. 3 Nr. 1 UrhG, zu denen insbesondere Bibliotheken und Archive gehören, könnten eine wesentliche Rolle bei der Speicherung von KI-Trainingsdaten einnehmen. Denn sie halten einen erheblichen Teil des Kulturerbes und damit auch eine große Menge an Werken vor. Diese können zum Training von KI-Systemen genutzt werden. Für Trainingsdatensammler ist die Verwendung dieser Werkbestände mitunter vorteilhaft, weil eine große Zahl an Werken spezifischer Gattungen aus einer Quelle akquiriert werden kann. Die Werke und ihre Werkrepräsentationen erfüllen zudem gewisse Qualitätsanforderungen. Nutzer, die Trainingsdaten für die Entwicklung von KI-Systemen benötigen, könnten die Sammlungen der Gedächtnisinstitutionen vor diesem Hintergrund auf Basis von §§ 44b, 60d UrhG umfassend zur Erstellung von Trainingsdatenkorpora vervielfältigen. Dies umfasst auch digitale Sammlungsbestandteile wie elektronische Zeitschriftenbibliotheken oder anderen Datenbanken, zu denen jedenfalls größere Gedächtnisinstitutionen üblicherweise Zugänge bereithalten.

##### a) Rechtmäßige Zugänglichkeit der Werke für Nutzer der Gedächtniseinrichtungen

Damit Nutzer solcher Gedächtnisinstitutionen von der Freistellungswirkung der Schrankenbestimmungen aus den §§ 44b, 60d UrhG profitieren können, ist zunächst notwendig, dass sie gemäß § 44b Abs. 2 S. 1 UrhG rechtmäßigen Zugang zu den Werken haben, die von den Gedächtnisinstitutionen bereitgestellt werden. Zu unterscheiden ist dabei einmal zwischen Werken, die sich wie beispielsweise Buchbestände oder eigene Digitalisate physisch oder elektronisch im unmittelbaren Herrschaftsbereich der Gedächtnisinstitutionen befinden, und solchen Werken, die über Zugänge abrufbar sind, die eine Gedächtnisinstitution bei Dritten einlizenziert hat. Hierzu zählen beispielsweise Inhalte einer elektronischen Zeitschriftenbibliothek.

Zu Werken, die sich physisch oder elektronisch im unmittelbaren Herrschaftsbereich der Gedächtnisinstitutionen befinden, haben Nutzer aufgrund dessen rechtmäßigen Zugang im Sinne des § 44b Abs. 2 S. 1 UrhG, dass sie mit Abschluss ihres Nutzungsvertrags ordnungsgemäß die Zugangsbeschränkungen der Gedächtnisinstitution (z. B. Login-Beschrän-

kung mit Nutzernummer und Passwort) überwinden.<sup>629</sup> Hier bestehen keine Unterschiede zu Nutzern von Online-Datenbanken, die mit Abschluss eines Abonnements rechtmäßigen Zugang zu den Inhalten der Datenbanken im Sinne des § 44b Abs. 2 S. 1 UrhG erhalten.

Etwas anderes könnte allerdings für Werke gelten, die über Zugänge abrufbar sind, die die Gedächtnisinstitutionen bei Dritten einlizenziert haben. In diesem Fall vermitteln die Gedächtnisinstitutionen den Nutzern den Zugang weiter. Die Nutzer handeln dabei nicht als Hilfspersonen der Gedächtnisinstitution, sondern im eigenen Interesse. Zwar ist für die Anwendung des Schrankentatbestands aus § 44b UrhG grundsätzlich unschädlich, dass die Nutzer die vom Dritten aufgestellten Zugangsvoraussetzungen nicht selbst erfüllen, also insbesondere nicht selbst Zugangsdaten erworben haben.<sup>630</sup> Allerdings können die Nutzer im vorliegenden Fall zumeist objektiv erkennen, dass die Werke einer zugangsgeschützten Quelle entstammen, die nicht zum unmittelbaren Herrschaftsbereich der Gedächtnisinstitution, also der eigenen Sammlung im engeren Sinne, gehört. Vervielfältigungen der durch diese Drittzugänge akquirierten Werke wohnt damit ein Umgehungsmoment inne. Denn der Nutzer spart sich eigene Aufwendungen, die für den Erwerb des Zugangs zu den Drittinhälften notwendig gewesen wären. Diese Umgehung beeinträchtigt die berechtigten Urheberinteressen übergehörlich. Darüber hinaus stellen Dritte (Urheber beziehungsweise Rechtsinhaber) durch die Zugangskontrolle den Schutz ihres Amortisationsinteresse sicher. Die Wahrung dieses Interesses würde durch eine Zugangsweitergabe an Nutzer von Gedächtnisinstitutionen zur Speicherung von Trainingsdaten vereitelt. Infolgedessen führt die Anwendung der Schrankenbestimmung aus § 44b UrhG in diesem Fall zu einer Verletzung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL.<sup>631</sup> Infolgedessen fällt die Vervielfältigung zugangsgeschützter Drittinhälte, die nicht im engeren Sinne zur Sammlung der Gedächtnisinstitutionen gehören und zu denen diese lediglich den Zugang an Nutzer weitervermitteln, um die Werke etwa für Trainingsdatensätze zu nutzen, nicht unter die Schrankenregelung des § 44b UrhG. Etwas anderes könnte aber dann gelten, wenn es für den Urheber oder sekundären Rechtsinhaber beim Zugangserwerb der Gedächtnisinstitution erkennbar gewesen ist, dass es sich um eine Gedächtnisinstitution handelt. Denn dann hätte er, bedingt durch die Funktion dieser Einrichtungen, wissen müssen,

---

629 Vgl. zur Voraussetzung siehe oben unter 3. Teil A. III. 2. f) aa).

630 Dazu siehe oben unter 3. Teil A. III. 2. f) bb) (1).

631 Hierzu ausführlich unter 3. Teil A. III. 2. f) bb) (3).

dass eine Zugangsvermittlung an Einrichtungsnutzer naheliegt. Dies hätte auch bei der Bestimmung der Gegenleistung für die Zugangsgewährung berücksichtigt werden können. Ein Schutz der Amortisationsinteressen hätte demzufolge sichergestellt werden können. Dem Sachverhalt haftet in diesem Fall außerdem kein Umgehungsmoment an.

Darüber hinaus findet die Zugangsvermittlung an Nutzer der Gedächtnisinstitutionen nur innerhalb der Infrastruktur der Einrichtungen statt. Eine unkontrollierte Weitergabe des Zugangs an beliebige Dritte, die faktisch zu einem Leerlaufen der Zugangsbeschränkung führen könnte, ist nicht zu befürchten.

Alles in allem führen die vorgenannten Gründe dazu, dass der Dreistufentest aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL nicht verletzt ist, wenn für den Urheber oder sekundären Rechtsinhaber im Rahmen der Zugangsverschaffung erkennbar gewesen ist, dass es sich um eine Gedächtnisinstitution handelt.<sup>632</sup> Alles in allem haben Nutzer von Gedächtnisinstitutionen zu von der Einrichtung bereitgestellten Werken rechtmäßigen Zugang im Sinne von § 44b Abs. 2 S. 1 UrhG, insofern die betreffenden Werke nicht in Datenbanken Dritter wie beispielsweise Verlagsdatenbanken wiedergegeben werden oder die Gedächtnisinstitutionen ihren Nutzern den Zugang zu Werken über Drittdatenbanken weitervermitteln, wobei beim Zugangserwerb durch die Gedächtnisinstitution für den jeweiligen Datenbankbetreiber erkennbar gewesen sein muss, dass es sich um eine Gedächtnisinstitution handelt.

b) Freistellung über § 60d UrhG zugunsten der Nutzer von  
Gedächtnisinstitutionen

Klärungsbedürftig ist weiterhin auch, ob Nutzer von Gedächtnisinstitutionen wie die Einrichtungen selbst von der erweiterten Freistellungswirkung des § 60d Abs. 3 S. 1 UrhG profitieren können. Hiervon könnten insbesondere kommerziell Forschende profitieren. Denn sie sind zwar nicht vom persönlichen Anwendungsbereich der Schrankenbestimmung aus § 60d

---

<sup>632</sup> Ein etwas weiteres Verständnis legt unter Umständen der Regierungsentwurf zugrunde. Danach haben die Nutzer von Gedächtniseinrichtungen stets rechtmäßigen Zugang zu solchen Inhalten, eine teilweise Begrenzung über den Dreistufentest findet nicht statt, siehe BT-Drs. 19/27426, S. 97. Es bleibt allerdings unklar, ob in der Gesetzesbegründung tatsächlich Fälle der Zugangsvermittlung adressiert werden, dazu vgl. unten unter 3. Teil A. III. 5. b).

UrhG erfasst. Dennoch führt ihre wissenschaftliche Forschungstätigkeit dazu, dass die erweiterte Freistellungswirkung des § 60d UrhG sachlichen anwendbar wäre.

Damit Nutzer wie kommerziell Forschende von der erweiterten Freistellungswirkung profitieren können, müssen sie als Teil der Gedächtnisinstitutionen verstanden werden. Aus einer wortlautbezogenen Perspektive spricht zunächst dafür, dass der persönliche Anwendungsbereich nach § 60 Abs. 3 Nr. 1 UrhG nur solche Bibliotheken und Museen umfasst, die öffentlich zugänglich sind. Um allein angestelltes Forschungspersonal von Museen und Bibliotheken zu privilegieren, bedürfte es keiner Beschränkung der Privilegierungswirkung auf *öffentliche zugängliche Einrichtungen*. Vielmehr kommt der Einschränkung dann ein eigenständiger Bedeutungsgehalt zu, wenn auch die Nutzer der Einrichtungen in den persönlichen Anwendungsbereich der Schrankenbestimmung mit einbezogen würden. Gegen ein solches Tatbestandsverständnis spricht aus einer systematischen Perspektive allerdings, dass eben diese Nutzer als Einzelforscher nach § 60d Abs. 3 Nr. 2 UrhG vom Schrankentatbestand privilegiert werden, so lange sie keine kommerziellen Zwecke verfolgen. Die Nutzer als Teil der Gedächtnisinstitutionen aufzufassen und ihnen damit über § 60d Abs. 3 Nr. 1 UrhG die Schrankenwirkung zu eröffnen, unterläuft die besondere Voraussetzung des § 60d Abs. 3 Nr. 2 UrhG. Nicht zuletzt bestehen Parallelen zur in § 60e Abs. 4 S. 2 UrhG geregelten Anschlussvervielfältigung, die über Terminals in Bibliotheken ermöglicht wird. Auch diese Regelung umfasst ungeachtet ihres Wortlauts, der auf den ersten Blick eine Mitprivilegierung der Bibliotheksnutzer nahelegt, nur die öffentliche Wiedergabe der Bibliotheken an den Terminals. Die Nutzervervielfältigungen richten sich hingegen nach gesondert anzuwendenden Schrankenbestimmungen, insbesondere § 53 UrhG.<sup>633</sup> Eine Mitprivilegierung der Nutzer erfolgt demzufolge nicht.

Für eine Mitprivilegierung der Nutzer von Gedächtnisinstitutionen spricht aus einer gesetzeshistorischen Perspektive jedoch, dass der nationale Umsetzungsgesetzgeber davon ausging, dass „individuelle Nutzer“, die der Gedächtnisinstitution „angehören“, in den Anwendungsbereich der Schrankenbestimmung fallen sollen.<sup>634</sup> Es ist allerdings bereits unklar, welche Personengruppe der Umsetzungsgesetzgeber hiermit adressiert hat.

---

633 BGH, Urt. v. 16.04.2015 - I ZR 69/11, GRUR 2015, II01 Rn. 41 – Elektronische Leseplätze II.

634 BT-Drs. 19/27426, S. 97.

### *III. Eingreifen einer urheberrechtlichen Schrankenbestimmung*

Denn der Nutzerbegriff könnte sich zum einen auf die Nutzer von Gedächtnisinstitutionen im allgemeinsprachlichen Sinn, also insbesondere Nutzer von Bibliotheken, beziehen. Nach einem anderen, engeren Verständnis könnten hingegen individualisierte *Werknutzer* innerhalb der privilegierten Organisation adressiert worden sein. Bei Zugrundelegung des zweiten Begriffsverständnisses hätte der Umsetzungsgesetzgeber nur klar gestellt, dass die rechtstechnische Privilegierung von Organisationsstrukturen zu einer Privilegierung der individuellen Personen führt, die in der Organisationsstruktur zusammengefasst werden. Demnach könnte beispielsweise die Adressierung von Hochschulen im Schrankentatbestand zu einer Privilegierung der individuellen, an den Hochschulen forschenden Mitglieder wie Professoren oder wissenschaftlichen Mitarbeitern führen. Für das zweite Begriffsverständnis spricht, dass der nationale Umsetzungsgesetzgeber mit seiner Erläuterung zu § 60d Abs. 3 UrhG an ErwGr. 14 DSM-RL anknüpft.<sup>635</sup> Danach sollen auch Forschungsorganisationen oder Einrichtungen kulturellen Erbes „angehörende Personen“ von der erweiterten Schrankenbestimmung erfasst sein. Der Nutzerbegriff, welcher seinem allgemeinen Begriffsverständnis nach eine weite Auslegung des Tatbestands nahelegt,<sup>636</sup> wird in ErwGr. 14 DSM-RL nicht verwendet.

Vielmehr könnte die in ErwGr. 14 DSM-RL gewählte Formulierung nach einer richtlinienkonformen Auslegung gegen eine Erstreckung der Schrankenwirkung auf Nutzer von Gedächtnisinstitutionen über § 60d Abs. 3 Nr. 1 UrhG sprechen. Denn es besteht ein qualitativer Unterschied zwischen einer Institution *angehörenden* Personen, also eigenem Forschungspersonal, und *Nutznießern* einer Institution. Letztere profitieren als Außenstehende von der (Forschungs-)Leistung der Gedächtnisinstitution. Hierunter fallen beispielsweise Nutzer einer Bibliothek. Sie machen sich die Sammlungsleistung und eine gegebenenfalls stattfindende wissenschaftliche Aufbereitung der Sammlung durch die Bibliothek zunutze, ohne der Einrichtung unmittelbar zugehörig zu sein.

Gegen ein Verständnis des § 60d Abs. 3 Nr. 1 UrhG, nach dem auch Nutzer der Gedächtnisinstitutionen von der erweiterten Schrankenwirkung des § 60d UrhG profitieren, spricht darüber hinaus auch, dass Zweck der Erweiterung der Schrankenwirkung auf Gedächtnisinstitutionen in § 60d Abs. 3 Nr. 1 UrhG die Ermöglichung von Forschung ist, die die Einrichtung

---

635 BT-Drs. 19/27426, S. 96.

636 Dazu siehe unmittelbar oben in diesem Abschnitt.

gen selbst im Zusammenhang mit ihrer Haupttätigkeit betreiben.<sup>637</sup> Eine Erweiterung der Schrankenwirkung auf individuelle Nutzer von Gedächtnisinstitutionen ist in diesem Regelungszweck nicht angelegt.

Alles in allem sprechen damit sowohl systematische als auch gesetzgebungshistorische sowie teleologische Erwägungen dafür, § 60d Abs. 3 Nr. 1 UrhG so auszulegen, dass sich die Schrankenwirkung nicht auf Nutzer von Gedächtnisinstitutionen erstreckt. Nach § 60d Abs. 3 Nr. 1 UrhG privilegiert sind demnach nur die Gedächtnisinstitutionen selbst. Die urheberrechtsrelevanten und freigestellten Nutzungshandlungen werden dabei von den Gedächtnisinstitutionen unmittelbar angehörendem Forschungspersonal vorgenommen. Nutzer von Gedächtnisinstitutionen, denen über die Sammlungen und Datenbankzugänge der Institutionen rechtmäßiger Zugang zu Werken vermittelt wird,<sup>638</sup> können Vervielfältigungen zum Training von KI daher nur auf Basis der allgemeinen Schrankenbestimmung für Text und Data Mining, § 44b UrhG, vornehmen.<sup>639</sup>

## 5. Zwischenergebnis zum Eingreifen einer urheberrechtlichen Schrankenbestimmung

Alles in allem kann der eigentliche KI-Trainingsvorgang unter die in § 44b Abs. 1 UrhG normierte Legaldefinition des Text und Data Mining subsumiert werden. Damit sind die beim Web Scraping vorgenommenen Vervielfältigungshandlungen grundsätzlich von der Schrankenbestimmung des § 44b Abs. 2 S. 1 UrhG gedeckt. Löschungspflicht (§ 44b Abs. 2 S. 2 UrhG) und Opt-out-Mechanismus für Urheber (§ 44b Abs. 3 UrhG) führen allerdings dazu, dass die Reichweite der Schrankenbestimmung für Text und Data Mining dennoch hinter dem im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnis zurückbleibt. Es kann auch nicht durch die Anwendung anderer Schrankenbestimmungen, insbesondere § 44a UrhG und § 57 UrhG, befriedigt werden, auch wenn es im Einzelfall insbesondere zu einer Freistellung der Vervielfältigungen von Werken zum Training von KI über § 57 UrhG kommen kann. Das ist der Fall, wenn die Werke unbeabsichtigt bei der Speicherung urheberrechtsfreier Trainingsdaten miterfasst, oder wenn trotz wirksamer technischer Sicherungsmaßnahmen zur Berücksichtigung von Nutzungsvorbehaltserklärungen im Sinne des

---

637 ErwGr. 8 DSM-RL.

638 Dazu siehe unmittelbar oben unter 3. Teil A. III. 5. a).

639 Zu den Voraussetzungen siehe oben unter 3. Teil A. III. 2.

§ 44b Abs. 3 UrhG Vorbehaltserklärungen im Einzelfall nicht erkannt und berücksichtigt wurden. Von der erweiterten Schrankenbestimmung aus § 60d UrhG können außerdem keine Nutzer von Gedächtnisinstitutionen wie Bibliotheken oder Archiven profitieren. Wollen sie über die Gedächtnisinstitutionen erlangte Werke zum Training von KI vervielfältigen, sind sie demzufolge insbesondere auf § 44b UrhG verwiesen. Von der erweiterten Freistellungswirkung des § 60d UrhG profitieren darüber hinaus auch intermediäre Trainingsdatensammler nicht grundsätzlich. Sie sind keine Bibliotheken und damit Kulturerbe-Einrichtungen im Sinne des § 60d Abs. 3 Nr. 1 UrhG. Demzufolge eröffnen sie nicht den persönlichen Anwendungsbereich der Schrankenbestimmung. Sie müssen daher ebenfalls insbesondere auf § 44b UrhG zurückgreifen.

#### *IV. Gesetzliche Vergütung für Urheber*

Das im zweiten Teil dieser Untersuchung identifizierte Regelungsbedürfnis sieht neben der Freistellung des Web Scrapings von Werken zum Training von KI vor, dass den Urhebern für die anfallenden Vervielfältigungen ihrer Werke eine angemessene Vergütung gezahlt werden sollte.<sup>640</sup> Die Schrankenbestimmung für Text und Data Mining (§ 44b UrhG), die grundsätzlich auf Vervielfältigungen von Werken zum Training von KI angewendet werden kann, gewährt den betroffenen Urhebern eine solche Vergütung allerdings nicht. Gleiches gilt gemäß § 60h Abs. 2 Nr. 3 UrhG für die spezielle Schrankenbestimmung des Text und Data Minings zum Zwecke wissenschaftlicher Forschung (§ 60d UrhG).

#### *V. Zwischenfazit zum Web Scraping von Werken als Trainingsdaten für KI*

Insgesamt hält das Urheberrecht mit der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) eine Schrankenbestimmung bereit, von deren Freistellungswirkung das Web Scraping urheberrechtlich geschützter Werke als Trainingsdaten umfasst ist. Mit den im zweiten Teil der vorliegenden Untersuchung identifizierten Freistellungserfordernissen übereinstimmend ist § 44b UrhG auch zugunsten kommerziell tätiger Trainingsdatensammler anwendbar. Allerdings führen die Löschungspflicht der

---

<sup>640</sup> Dazu ausführlich siehe oben unter 2. Teil B. II. 1.

## *B. Durchführung des Trainings auf Grundlage von Werken*

Trainingsdatensammler (§ 44b Abs. 2 S. 2 UrhG), der selbst durch eine Vorbehaltserklärung in natürlicher Sprache auslösbarer Opt-out-Mechanismus (§ 44b Abs. 3 UrhG) und die Vergütungsfreiheit der Vervielfältigungs-handlungen dazu, dass die Schrankenbestimmung die identifizierten Rege-lungsbedürfnisse für ein interessengerechtes urheberrechtliches Regelungs-regime dennoch nicht vollständig erfüllt. De lege ferenda sind deswegen Anpassungen an der Schrankenbestimmung für Text und Data Mining insbesondere direkt in § 44b UrhG notwendig.

## *B. Durchführung des Trainings auf Grundlage urheberrechtlich geschützter Trainingsdaten*

Nach der Speicherung der für das Training bestimmten Werke mit Hilfe von Webcrawlern wird auf Basis der erfassten Werke das KI-Training durchgeführt. Für die Durchführung des gesamten Trainingsprozess konn-te dabei ein regulatorisches Freistellungsinteresse identifiziert werden. Die Freistellung sollte im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte erfolgen.

Damit das geltende Urheberrecht das im zweiten Teil der Untersuchung identifizierte Freistellungsbedürfnis erfüllt, ist es erforderlich, dass alle Teil-aspekte des Trainingsvorgangs<sup>641</sup> durch eine unmittelbar begrenzende Aus-gestaltung der urheberrechtlichen Verwertungsrechte aus den §§ 15 ff. UrhG vom Ausschließlichkeitsrecht ausgenommen sind.

### *I. Vorverarbeitung der Trainingsdaten (Normalisierung und Feature Encoding)*

Unter dem Ausdruck der Vorverarbeitung von Trainingsdaten werden meh-rere Verarbeitungsschritte zusammengefasst. Die erfassten Trainingsdaten müssen zunächst normalisiert werden. Als letztes wird das Feature En-coding der Inhalte durchgeführt.<sup>642</sup> Erst dann eignen sich die urheberrecht-

---

641 Dazu siehe unter 2. Teil C. am Anfang.

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

lich geschützten Trainingsdaten als Berechnungsgrundlage für den Algorithmus.

## 1. Normalisierung der Trainingsdaten

Da bei der Normalisierung digitale Kopien der als Trainingsdaten erfassten Werke angefertigt werden, ist ein Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG denkbar. Teil des Normalisierungsverfahrens, also der Transponierung der Daten auf eine vergleichbare Skala, ist darüber hinaus regelmäßig eine Veränderung der geschützten Inhalte.<sup>643</sup> Berücksichtigt werden muss damit außerdem, inwieweit solche Veränderungen Einfluss auf den möglichen Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG und eine mögliche urheberrechtliche Freistellung der Normalisierungshandlungen haben.

### a) Normalisierung als Vervielfältigung im Sinne des § 16 UrhG: Anwendung eines normativierten Vervielfältigungsbegriffs

Ob bei der Normalisierung von Werken für das Training von KI Vervielfältigungsstücke im urheberrechtlichen Sinne angefertigt werden, hängt davon ab, was unter einem Vervielfältigungsstück im Sinne des § 16 UrhG zu verstehen ist. Denn auch wenn der Vervielfältigung im urheberrechtlichen Sinne ein tatsächlicher Speichervorgang zugrunde liegt, handelt es dennoch um einen normativen Rechtsbegriff, der infolgedessen auch der Auslegung zugänglich ist.<sup>644</sup> Daraus folgt, dass nicht jede tatsächliche Speicherung auch eine Vervielfältigung im Rechtssinne sein muss.<sup>645</sup> Eine normative Aufladung des Vervielfältigungsbegriffs ergibt sich auch aus der Rechtspre-

---

643 Zu möglichen Veränderungsschritten im Detail unten unter 3. Teil B. I. 1. a) bb) (2).

644 Hugenholtz, EIPR 2000, 482 (485); sich hieran anschließend Schönberger, ZGE 2018, 35 (53); a. A. wohl Heerma, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; Chiou, JIPITEC 2019, 398 Rn. 6; außerdem Jager, Artificial Creativity?, S. 332.

645 So auch Hugenholtz, EIPR 2000, 482 (485); sich hieran anschließend Schönberger, ZGE 2018, 35 (53); de lege ferenda auch Becker, ZGE 2016, 239 (267); Hofmann, ZGE 2016, 482 (502); a. A. BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 25 – Internet-Radiorecorder; Sucker, Der digitale Werkgenuss im Urheberrecht, S. 89; außerdem wohl Heerma, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; Chiou, JIPITEC 2019, 398 Rn. 6; außerdem Jager, Artificial Creativity?,

## B. Durchführung des Trainings auf Grundlage von Werken

chung des EuGH, nach der das Kopieren eines Schutzgegenstands nicht in das Vervielfältigungsrecht des Rechteinhabers eingreift, wenn der Inhalt nach der tatsächlichen Speicherung so verändert wird, dass der Schutzgegenstand nicht mehr wiedererkennbar ist.<sup>646</sup>

### aa) Der Begriff des Vervielfältigungsstücks nach dem UrhG

Was unter einem Vervielfältigungsstück zu verstehen ist, ist im UrhG nicht ausdrücklich definiert. Dennoch hat sich ein allgemeines Begriffsverständnis herausgebildet. Unter einer Vervielfältigung im Sinne des UrhG wird dabei jede in beliebiger Weise vorgenommene körperliche Festlegung eines Werkes verstanden, die unmittelbar oder mittelbar zur menschlichen Wahrnehmung der geistigen Schöpfung geeignet ist.<sup>647</sup> Dabei wird ausschließlich eine technische Perspektive eingenommen.<sup>648</sup> Bei der Normalisierung urheberrechtlich geschützter Trainingsdaten zur Entwicklung von KI-Systemen werden auf eine einheitliche Skala transformierte, digitale Kopien von Werken, also technisch gesehen Repräsentationen der Werke angefertigt, sodass eine tatbestandsmäßige Festlegung der Werke erfolgt. Sie ist jedenfalls mittelbar auch zur menschlichen Wahrnehmung der geistigen Schöpfung geeignet. Nach dem tradierten, national geprägten Begriffsverständnis würden bei der Normalisierung von zum Training erfassten Werken daher Vervielfältigungsstücke im Sinne des § 16 UrhG hergestellt.

### bb) Funktionsbezogene Auslegung des Vervielfältigungsbegriffs

Das Vervielfältigungsrecht ist durch Art. 2 InfoSoc-RL allerdings unionsrechtlich harmonisiert. Welche Bedeutung der Begriff des Vervielfältigungsstücks (§ 16 Abs. 1 UrhG) beziehungsweise der Vervielfältigung (Art. 2 Info-

---

S. 333, die allesamt ein streng technisches Verständnis des Vervielfältigungsbegriffs zugrunde legen.

646 EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 31 – Pelham.

647 BGH, Urt. v. 04.10.1990 - I ZR 139/89, GRUR 1991, 449 (453) – Betriebssystem; BGH, Urt. v. 06.10.2016, GRUR 2017, 266 Rn. 37 – World of Warcraft I; BT-Drs. IV/270, S. 47.

648 BGH, Urt. v. 05.03.2020 - I ZR 32/19, GRUR 2020, 738 Rn. 25 – Internet-Radiorecorder; *Heerma*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; *Sucker*, Der digitale Werkgenuss im Urheberrecht, S. 89; *Chiou*, JIPITEC 2019, 398 Rn. 6; *Jager*, Artificial Creativity?, S. 333.

Soc-RL) hat, muss sich daher nach einer richtlinienkonformen Auslegung des Vervielfältigungsbegriffs richten. Das betrifft auch die Frage, ob Werkfestlegungen zum Zwecke rein technischer Verwendung, wie sie bei der Normalisierung urheberrechtlich geschützter Trainingsdaten angefertigt werden, in den Anwendungsbereich des Vervielfältigungsrechts fallen.

Auch in der InfoSoc-RL wird der Vervielfältigungsbegriff nicht ausdrücklich definiert. Das führt dazu, dass eine richtlinienkonforme Auslegung daher insbesondere anhand des Sinn und Zwecks des Vervielfältigungsrechts aus Art. 2 InfoSoc-RL erfolgen muss.<sup>649</sup> Diese teleologische Perspektive kann unter anderem auch zu einer funktionsbezogenen, einschränkenden Auslegung des Vervielfältigungsrechts führen.<sup>650</sup> Die dem Sachverhalt zugrundeliegende technische Ausgangslage ist hierfür nicht allein maßgeblich. Die Sozialbedeutung der Nutzungshandlung kann in diesem Zusammenhang jedoch nur dann zur Konkretisierung des Vervielfältigungsbegriffs genutzt werden, insofern sie sich mit Hilfe einer der Auslegungsmethoden an den Tatbestand aus Art. 2 InfoSoc-RL anknüpfen lässt.<sup>651</sup>

---

649 v. *Ungern-Sternberg*, GRUR 2012, II98 (1200); v. *Ungern-Sternberg*, GRUR 2015, 533 (534); insbesondere im Kontext des urheberrechtlichen Vervielfältigungsrechts nun auch an erster Stelle betont in EuGH, Urt. v. 13.07.2023 - C-426/21, GRUR 2023, 1284 Rn. 26 – Ocilion; außerdem bereits in EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 28 – Pelham, wobei die hier zum Tonträgerherstellerrecht getätigten Ausführungen ebenfalls auf das in Art. 2 InfoSoc-RL harmonisierte Urheberrecht übertragen werden können. In diesem Sinne auch *Grisse/Kaiser*, ZUM 2021, 401 (404); *Haberstumpf*, ZUM 2020, 810 (817). Eine teleologisch geprägte Auslegung des Vervielfältigungsrechts deutete sich bereits in der EuGH-Entscheidung *Allposters* an. Entscheidend ist für den EuGH hier, dass die Übertragung eines Kunstwerkes von einem Poster auf eine Leinwand die Reproduktion näher an das Original heranrückt, siehe EuGH, Urt. v. 22.01.2015 - C-419/13, GRUR 2015, 256 Rn. 42 – *Allposters*. Damit geht, was für den EuGH entscheidend gewesen sein muss, eine Vergrößerung der vom Werkstück ausgehenden Substitutionsgefahr, also eine Intensivierung der Beeinträchtigung der Verwertungsinteressen betroffener Urheber, einher.

650 v. *Ungern-Sternberg*, GRUR 2012, II98 (1204 f.); im Anschluss hieran auch *Antoine*, ZGE 2023, 387 (399); im Kontext digitalen Werkgenusses ähnlich auch *Raue*, ZGE 2017, 514 (530); vgl. zu entsprechenden Überlegungen de lege ferenda *Becker*, ZGE 2016, 239 (267); *Hofmann*, ZGE 2016, 482 (502); ähnlich außerdem *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260), die die Abkehr vom technisch-zentrierten zu einem auf die Sozialbedeutung bezogenen Verständnis des Vervielfältigungsrechts fordern.

651 A. A. wohl *Lennartz/Kraetzig*, GRUR 2024, 1258 (1260), die losgelöst von den anerkannten Auslegungsmethoden eine Konkretisierung anhand der Sozialbedeutung der Nutzungshandlung vornehmen wollen.

## B. Durchführung des Trainings auf Grundlage von Werken

- (1) Herleitung einer Schutzbereichsbegrenzung aus der verwertungsrechtlichen Funktion des Vervielfältigungsrechts

Grundsätzlich strebt die InfoSoc-RL an, ein hohes Schutzniveau für Urheber zu gewährleisten.<sup>652</sup> Allerdings ist Sinn und Zweck der urheberrechtlichen Verwertungsrechte und damit des Vervielfältigungsrechts im Speziellen, dem Urheber die Verwertung seines Werkes, also die Monetarisierung dessen Nutzung, zu sichern.<sup>653</sup> Daraus folgt, dass nur solche Werkverwendungen auch erlaubnispflichtige Werknutzungen und damit Vervielfältigungen im urheberrechtlichen Sinne sein können, die die Verwertungsmöglichkeit der Urheber zumindest potenziell beeinträchtigen.<sup>654</sup> Hierfür genügt, dass aus der Werkverwendung ein unmittelbares oder mittelbares Vermarktungspotenzial entsteht und damit eine Substitutionsgefahr für die Verwertung des Werkes durch den Urheber hervorgerufen wird.<sup>655</sup> Eine solche kann grundsätzlich auch durch veränderte Erzeugnisse begründet werden, sodass auch diese im Allgemeinen unter das Vervielfältigungsrecht fallen.<sup>656</sup> Geht von der Werkverwendung hingegen keine Substitutionsge-

---

652 ErwGr. 4, 9 InfoSoc-RL.

653 Nordemann, in: Fromm/Nordemann, Urheberrecht, Einl. zum UrhG Rn. 19; v. Ungern-Sternberg, GRUR 2012, 576 (579 f.); v. Ungern-Sternberg, GRUR 2012, II98 (1200); Haberstumpf, ZUM 2020, 810 (816); Grisse/Kaiser, ZUM 2021, 401 (404). Außerdem ErwGr. 10 InfoSoc-RL.

654 öst. OGH, Beschl. v. 23.02.2010 - 4 Ob 208/09 f, ZUM 2010, 629 (630); v. Ungern-Sternberg, in: Schricker/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23; v. Ungern-Sternberg, GRUR 2012, 576 (580); Schönberger, ZGE 2018, 35 (53); Antoine, Verändernde Werknutzungen, S. 179 f.; in diese Richtung auch Grisse/Kaiser, ZUM 2021, 401 (404); grundsätzlich wie hier auch Antoine, ZGE 2023, 387 (398), wobei sie eine Einschränkung der funktionsbezogenen Auslegung auf die „Nutzungsfälle an den Rändern des urheberrechtlichen Ausschließlichkeitsrechts“ befürwortet, dazu sogleich. Zum Verhältnis zu den urheberpersönlichkeitsrechtlichen Bestimmungen siehe Antoine, ZGE 2023, 387 (401 ff.).

655 v. Ungern-Sternberg, in: Schricker/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23. Die Intensivierung der Substitutionsgefahr durch eine weitere Annäherung des Werkstücks an das Original scheint auch für den EuGH in der *Allposters*-Entscheidung maßgebend dafür gewesen zu sein, beim Medientransfer eines Bildes von Plakat auf Leinwand eine neue Vervielfältigung anzunehmen, vgl. EuGH, Urt. v. 22.01.2015 - C-419/13, GRUR 2015, 256 Rn. 42 – Allposters.

656 Hiervon ging der EuGH implizit bereits im Jahr 2012 aus, als er die Herstellung veränderter Erzeugnisse beiläufig unter das Vervielfältigungsrecht aus Art. 2 InfoSoc-RL fasste und auf Grundlage dessen ein Eingreifen der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus Art. 5 Abs. 1 InfoSoc-RL untersuchte, vgl. EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 53 – Infopaq II.

fahr für die Verwertung des Werkes durch den Urheber hervor, besteht auch kein Grund für eine Beteiligung des Urhebers an von Dritten erwirtschafteten Einnahmen. Anderes folgt auch nicht aus dem Grundsatz einer möglichst umfassenden Urheberbeteiligung. Denn eine Beteiligung der Urheber ohne Beeinträchtigung ihrer Verwertungsinteressen stellt keinen angemessenen Ausgleich, sondern eine ungerechtfertigte Überkompensation dar. Sie ist nicht durch das Urheberrecht bezweckt.<sup>657</sup>

Keine Substitutionsgefahr wird durch beiläufige oder zufällige, also nicht zielgerichtete und insgesamt untergeordnete Werkverwendungen begründet.<sup>658</sup> Andererseits genügt aber, dass die Kopie des Werkes eine Werkrezeption ermöglichen soll. Das ist beispielsweise im Arbeitsspeicher von Computern der Fall. Eine Substitutionsgefahr für die Verwertung des genutzten Werks durch den Urheber wird im Kontext der Anfertigung veränderter Werkkopien weiterhin dann nicht ausgelöst, wenn die eigenschöpferischen Züge eines Werkes nach der Bearbeitung in einer Werkkopie nicht mehr wiedererkennbar sind.<sup>659</sup> Dass die Werkverwendung von

---

657 Anders Jager, *Artificial Creativity?*, S. 333, der sein rein-technisches Verständnis gerade auf den Grundsatz der möglichst umfassenden Urheberbeteiligung stützt.

658 v. Ungern-Sternberg, GRUR 2012, 576 (580).

659 Vgl. zum Tonträgerherstellerrecht EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 38 – Pelham. Da auch der Harmonisierung des originären Urheberrechts ein verwertungsschützender Charakter zugutekommt und sowohl das Tonträgerherstellerrecht als auch das Urheberrecht gemeinsamen auf Art. 2 InfoSoc-RL beruhen, ist eine Übertragung des Wiedererkennbarkeitskriteriums aus der Rechtsprechung des EuGH auf das Urheberrecht möglich. So auch Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 23 Rn. 41; Lauber-Rönsberg/Ahlberg, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 23 Rn. 35; Loewenheim/Leistner, in: Loewenheim, Handbuch des Urheberrechts, § 8 Schutzzumfang Rn. 13; Nordemann/Czychowski, in: Hasselblatt, Münchener Anwalts-Handbuch Gewerblicher Rechtsschutz, § 40 Urheberrecht und verwandte Schutzrechte Rn. 560; Paul, in: Borges/Hilber, Beck'scher Online-Kommentar IT-Recht, UrhG § 23 Rn. 6; Raue, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 30; BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 49 – Porsche 911; BGH, Urt. v. 15.12.2022 - I ZR 173/21, GRUR 2023, 571 Rn. 29 – Vitrinenleuchte; Grisse/Kaiser, ZUM 2021, 401 (404); Grünberger, ZUM 2023, 309 (318); v. Ungern-Sternberg, GRUR 2022, 1777 (1779); Haberstumpf, ZUM 2020, 810 (416); Lauber-Rönsberg, ZUM 2020, 733 (736); Wagner, MMR 2020, 728 (730); kritisch allerdings Ohly, ZUM 2021, 745 (746); Schack, ZGE 2023, 263 (272 ff.); Stieper, GRUR 2020, 699 (706); ablehnend allerdings Antoine, Verändernde Werknutzungen, S. 181 ff. Antoine, die ihre Skepsis vor allem auf dem investitionsschützenden Charakter des Tonträgerherstellerrecht begründet, berücksichtigt dabei allerdings nicht, dass auch dem harmonisierten Urheberrecht ein Amortisationsgedanke zu grunde liegt (ErwGr. 10 InfoSoc-RL, er adressiert sowohl das Urheberrecht als auch

der Kunstfreiheit umfasst ist, ist vor dem Hintergrund der teleologischen Auslegung demnach nicht notwendig.<sup>660</sup> Darüber hinaus betreffen die urheberrechtlichen Schrankenbestimmungen Sachverhaltskonstellationen, in denen die Verwertungsinteressen der Urheber jedenfalls potenziell beeinträchtigt sind, also zumindest eine mittelbare Verwertungsmöglichkeit für Urheber besteht.<sup>661</sup> Daher besteht auch kein Konflikt zwischen der funktionsbezogenen Auslegung des Vervielfältigungsrechts, die sich danach richtet, ob das Verwertungsinteresse eines Urhebers *überhaupt* beeinträchtigt ist, und dem Ausgleich gegenläufiger Interessen, den das Urheberrecht mit Hilfe der Schrankenbestimmungen vornimmt.<sup>662</sup> Die funktionsbezogene Auslegung auf den Randbereich des Urheberrechts zu beschränken, ist daher nicht notwendig.<sup>663</sup>

---

die verwandten, in der InfoSoc-RL geregelten Schutzrechte). Die für den EuGH maßgebliche Erwägung, dass Rechtsinhaber in die Lage versetzt werden müssen, ihre in die Schaffung des Schutzgegenstands fließenden ökonomischen Ressourcen durch die Verwertung des Werkes ausgleichen zu können, um auch das Schaffen aus einer wirtschaftlichen Perspektive zu ermöglichen, greift deswegen ebenso für den originären urheberrechtlichen Schutz. Vor diesem Hintergrund ist eine Anwendung des Wiedererkennbarkeitskriteriums auf das Urheberrecht sogar angezeigt. Eine Umsetzung der Wiedererkennbarkeitsschwelle im Vervielfältigungsrecht ist dabei dogmatisch konsequenter als eine Berücksichtigung in § 23 UrhG, in diese Richtung auch Grünberger, ZUM 2023, 309 (320 ff.). Etwas widersprüchlich hier der BGH, der die Wiedererkennbarkeitsformel zwar i. R. d. § 23 UrhG prüft (BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 47 ff. – Porsche 911.), letztlich aber zutreffend konstatiert, dass es bei mangelnder Wiedererkennbarkeit bereits an einer Vervielfältigungshandlung im Sinne des § 16 UrhG fehlt, sodass es auf das Vorliegen einer freien Benutzung nach § 24 UrhG a. F. schon nicht mehr ankommt, siehe BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 63 – Porsche 911.

660 Grisse/Kaiser, ZUM 2021, 401 (414); Haberstumpf, ZUM 2020, 810 (817).

661 Das gilt selbst für die Schrankenbestimmung zur Nutzung von Werken als unwesentliches Beiwerk (§ 57 UrhG). Denn auch sie geht auf eine unionsrechtliche Grundlage zurück, siehe Vogel, in: Schricker/Loewenheim, Urheberrecht, UrhG § 57 Rn. 5. Damit kann auch diese Schrankenbestimmung nur dann eingreifen, wenn im Ergebnis der dargelegten funktionsbezogenen Auslegung überhaupt eine Vervielfältigung angefertigt worden ist. Nur dann liegt eine urheberrechtsrelevante und erlaubnisbedürftige Werknutzung vor. Das setzt jedenfalls eine potenzielle Beeinträchtigung der urheberrechtlichen Verwertungsinteressen voraussetzt. Dieses kann beispielsweise auch bei der hintergründigen Nutzung eines Werkes gegeben sein. Es ist daher von der Frage zu trennen, ob die Werknutzung für den Nutzer un wesentlich im Sinne der Bestimmung ist.

662 Anders Antoine, ZGE 2023, 387 (400), die eine Kollision mit den urheberrechtlichen Schrankenbestimmungen annimmt und den in den Schrankenbestimmungen verhandelten Fällen des Interessenausgleichs eine Begrenzung für die funktionsbezogene Auslegung des Vervielfältigungsrechts entnimmt.

(2) Wiedererkennbarkeit eigenschöpferischer Züge nach der Normalisierung

Bei der Normalisierung von urheberrechtlich geschützten Trainingsdaten, die mit Veränderungen an den Werken einhergeht, ist insbesondere klärungsbedürftig, ob die eigenschöpferischen Züge eines Werkes nach der Herstellung normalisierter Werkkopien noch wiedererkennbar sind. Das hängt davon ab, welche konkreten Verarbeitungsschritte Teil des Normalisierungsprozesses sind. Dies ist wiederum von der Art der verwendeten Inhalten sowie der Art des zu trainierenden KI-Systems abhängig.

Praktisch verbreitet ist beispielsweise eine Expansion von Textinhalten. Diese umfasst unter anderem eine Auflösung von Abkürzungen.<sup>664</sup> Der Text bleibt dabei aber noch im Wesentlichen erkennbar. Weitere mögliche Normalisierungsschritte sind unter anderem eine Entfernung der Interpunktionszeichen oder eine Reduktion der Wörter auf ihren Wortstamm.<sup>665</sup> Am Ende kommt es regelmäßig zu einer Tokenisierung der Inhalte, bei der der Text in einzelne Token aufgeteilt wird.<sup>666</sup> Sie können aus ganzen Sätzen, Wörtern oder Silben bestehen. Ob die schöpferischen Züge eines Sprachwerks in normalisierter Form noch im urheberrechtlichen Sinne wiedererkennbar sind, hängt vor diesem Hintergrund von den im Einzelfall angewendeten Normalisierungsschritten sowie den schutzbegründenden Elementen des jeweiligen Werks ab. Für die Zwecke dieser Untersuchung ist daher davon auszugehen, dass die schutzbegründenden Elemente eines Schriftwerks auch nach seiner Normalisierung noch wiedererkennbar sind.

Auch für die Normalisierung von visuellen Werken kann auf unterschiedliche Normalisierungsverfahren zurückgegriffen werden. Teilweise bleiben dabei wesentliche schöpferische Merkmale der Inhalte wie Motiv, Bildausschnitt, Perspektive, Bildschärfe oder Kontrastgebung für einen Beobachter identifizierbar.<sup>667</sup> Im Detail kommt es auch hierbei auf die im Einzelfall angewendeten Normalisierungsverfahren und die schutzbegrün-

---

663 So aber *Antoine*, ZGE 2023, 387 (400); *Antoine*, Verändernde Werknutzungen, S. 176 f.

664 *Yse*, Text Normalization for Natural Language Processing (NLP).

665 *Yse*, Text Normalization for Natural Language Processing (NLP).

666 Mit Beispielen *Menzli*, Tokenization in NLP; *Yse*, Text Normalization for Natural Language Processing (NLP).

667 Zu den schöpferischen Elementen eines Lichtbildwerkes siehe *Loewenheim/Leister*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 2 Rn. 213. Zur zwei beispielhaften Normalisierungsreihen von Bildinhalten siehe *NVIDIA Corporation*, Normalize Operator — NVIDIA DALI 1.29.0 documentation.

## B. Durchführung des Trainings auf Grundlage von Werken

denden Elemente der normalisierten Werke an. Daher ist betreffend der Normalisierung von visuellen Werken für die Zwecke dieser Untersuchung ebenfalls davon auszugehen, dass schutzbegründende Elemente der Werke im urheberrechtlichen Sinne wiedererkennbar bleiben.

An dieser Stelle kann demzufolge dahinstehen, wo genau die Wiedererkennbarkeitsgrenze des Schutzbereichs zu ziehen ist und wie sich dies auf einzelne Normalisierungsverfahren auswirkt. Vielmehr führen Veränderungen bei der Normalisierung von Werken nicht grundsätzlich zur mangelnden Wiedererkennbarkeit eigenschöpferischer Elemente in den urheberrechtlich geschützten Trainingsdaten. Im Einzelfall kann das aber durchaus anders zu beurteilen sein.

### (3) Weitere funktionsorientierte Erwägungen zur Normalisierung von Werken

Weiterhin können die generierten Trainingsdaten auch in normalisierter Form noch auf dem Trainingsdatenmarkt verwertet werden. Damit kommt den Werkrepräsentationen auch nach der Normalisierung noch eine wirtschaftliche Bedeutung zu. Infolgedessen beeinträchtigt die Herstellung normalisierter Kopien von Werken auch aus dieser Perspektive die Verwertungsinteressen der Urheber.

#### cc) Schlussfolgerungen

Alles in allem können auch normalisierte Kopien von Werken Vervielfältigungsstücke im urheberrechtlichen Sinne sein.<sup>668</sup> Solange die schutzbegründenden Elemente im normalisierten Erzeugnis noch wiedererkennbar sind, hat hierauf grundsätzlich auch die Veränderung der Werke bei der Normalisierung keinen Einfluss. Die Normalisierung urheberrechtlich geschützter Trainingsdaten greift damit – je nach angewendetem Normalisierungsverfahren – in das Vervielfältigungsrecht betroffener Urheber im Sinne des § 16 UrhG ein.<sup>669</sup> Vervielfältigungen, die im Zuge der Normalisierung möglicherweise hergestellt werden, sind demzufolge erlaubnis-

---

668 Wie hier auch Käde, Kreative Maschinen und Urheberrecht, S. 70; Pesch/Böhme, GRUR 2023, 997 (1004); wohl außerdem Konertz/Schönhof, WRP 2024, 289 Rn. 6.

669 Weitergehend Dornis/Stober, Urheberrecht und Training generativer KI-Modelle, S. 70; Dregelies, GRUR 2024, 1484 (1484), die pauschal eine Vervielfältigung annehmen.

pflichtig, wenn nicht eine urheberrechtliche Schrankenbestimmung die Werknutzung gesetzlich gestattet. Bereits hieraus folgt, dass das UrhG das Freistellungsbedürfnis für die als Teil des Trainingsprozesses durchgeführte Normalisierung der urheberrechtlich geschützten Trainingsdaten nicht erfüllt. Denn nach der durchgeführten regulatorischen Bedarfsanalyse ist eine Freistellung des gesamten Trainingsprozesses im Wege einer unmittelbar begrenzenden Ausgestaltung der Verwertungsrechte interessengerecht.

Nichtsdestotrotz könnte die Freistellung von im Rahmen der Normalisierung vorgenommenen Vervielfältigungshandlungen durch die Anwendung einer urheberrechtlichen Schrankenbestimmung erreicht werden. Damit ließe sich zumindest das Ob des regulatorischen Freistellungsbedürfnisses erfüllen. Ob eine Schrankenbestimmung die Normalisierung der Trainingsdaten freistellt, hängt aber auch davon ab, ob sie die Herstellung veränderter Vervielfältigungsstücke umfasst.

b) Anwendung urheberrechtlicher Schrankenbestimmungen zur Erlaubnis der Vervielfältigungshandlungen

Erforderlich ist also, dass die im Rahmen der Normalisierung von urheberrechtlich geschützten Trainingsdaten möglicherweise vorgenommenen Vervielfältigungshandlungen durch die Anwendung einer urheberrechtlichen Schrankenbestimmung erlaubnisfrei gestellt sind.

aa) Schrankenbestimmung für vorübergehende Vervielfältigungs-handlungen (§ 44a UrhG)

Vervielfältigungen von Werken im Rahmen der Normalisierung urheberrechtlich geschützter Trainingsdaten könnten von der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG gedeckt sein. Wie auch bei der Speicherung der Trainingsdaten mit Hilfe von Webcrawlern scheitert die Anwendung von § 44a UrhG allerdings zum einen an der eigenen wirtschaftlichen Bedeutung der Vervielfältigungen, die neben dem Einsatz für das Training einer konkreten KI auch im normalisierten Zustand auf dem Trainingsdatenmarkt veräußert werden können. Darüber hinaus ist auch davon auszugehen, dass dauerhafte Ver-

## B. Durchführung des Trainings auf Grundlage von Werken

vielfältigungen der Inhalte stattfinden.<sup>670</sup> Eine Freistellung von im Rahmen der Normalisierung stattfindenden Vervielfältigungshandlungen über die Schrankenbestimmung aus § 44a UrhG findet demzufolge nicht statt.

### bb) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)

Die Vervielfältigungen von Werken, die gegebenenfalls im Rahmen der Normalisierung urheberrechtlich geschützter Trainingsdaten angefertigt werden, könnten wie die Speicherung der Trainingsdaten jedoch von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 UrhG freigestellt sein. Denn das KI-Training auf Basis von Werken fällt unter den Begriff des Text und Data Mining im Sinne des § 44b Abs. 1 UrhG.<sup>671</sup>

#### (1) Grundsätzliche Anwendbarkeit von § 44b UrhG

Von der Schrankenbestimmung aus § 44b Abs. 2 UrhG gedeckt sind alle Vervielfältigungshandlungen, die *für* das KI-Training durchgeführt werden, hierfür also notwendig sind. Da das KI-Training auf Basis normalisierter Daten stattfindet, ist der vorgelagerte Normalisierungsschritt für das Training von KI notwendig. Infolgedessen werden die bei der Normalisierung urheberrechtlich geschützter Trainingsdaten gegebenenfalls stattfindenden Vervielfältigungshandlungen ebenfalls grundsätzlich vom Schrankentatbestand aus § 44b Abs. 2 UrhG umfasst.<sup>672</sup>

#### (2) Auswirkungen von Veränderungen der Werke auf die Anwendbarkeit der Schrankenbestimmung

Fraglich ist allerdings, ob die Schrankenbestimmung aus § 44b Abs. 2 UrhG auch die Herstellung *veränderter* Vervielfältigungsstücke der Werke

---

<sup>670</sup> Zu dieser Voraussetzung ausführlich oben unter 3. Teil A. III. 1.

<sup>671</sup> Dazu bereits ausführlich siehe oben unter 3. Teil A. III. 2. i) ee) (3) aa) bis dd); wie hier auch *Jager, Artificial Creativity?*, S. 351, 157 ff.

<sup>672</sup> So für im Kontext des allgemeinen Text und Data Mining auch ErwGr. 8 S. 6 DSM-RL; zudem auch *Bullinger*, in: *Wandtke/Bullinger, Urheberrecht, UrhG § 60d Rn. 11; Bomhard, InTeR 2023, 174 (176); Kögel, InTeR 2023, 179 (181); Raua, ZUM 2021, 793 (795, 797)*; für die alte deutsche Schrankenbestimmung für Text und Data Mining, § 60d UrhG a. F., außerdem *Geiger/Frosio/Bulayenko*, IIC 2018, 814 (831). Ebenso in ErwGr. 8 S. 6 DSM-RL.

freistellt, wie sie im Rahmen des Normalisierungsverfahrens angefertigt werden.

Grundsätzlich bestimmt § 62 Abs. 1 S. 1 UrhG, dass die aus den Schrankenbestimmungen folgenden Nutzungsberechtigungen nicht zu einer Erlaubnis von Änderungen an den betreffenden Werken führen. Allerdings sind Änderungen abweichend hiervon gemäß § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG zulässig, wenn der Urheber seine Einwilligung nach Treu und Glauben nicht versagen kann. Dieser unbestimmte Rechtsbegriff führt über den Rechtsgrundverweis aus § 62 Abs. 1 S. 2 UrhG auch im Anwendungsbereich des Änderungsverbots zu einer umfassenden Abwägung von Urheber- und Nutzerinteressen.<sup>673</sup> Verändernde Werknutzungen sind danach zulässig, wenn sie zur Erfüllung des jeweiligen Schrankenzwecks erforderlich und zumutbar sind.<sup>674</sup> Darüber hinaus müssen verändernde Werknutzungen über §§ 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG im Wege richtlinienkonformer Auslegung des Tatbestandsmerkmals der Erlaubnis nach „Treu und Glauben“ aber auch dann zulässig sein, wenn die mit einer Schrankenregelung umgesetzte Richtlinienbestimmung des Unionsurheberrechts eine Veränderung des Werks gestattet.

Dafür, dass Art. 4 DSM-RL, der mit § 44b UrhG umgesetzt worden ist,<sup>675</sup> verändernde Vervielfältigungen von Werken zum Zwecke des Text und Data Minings freistellt, spricht zum einen, dass das Umarbeitungsrecht an Computerprogrammen (Art. 4 Abs. 1 lit. b Computerprogramm-RL) vom Schrankentatbestand mitumfasst ist. Nach dieser Regelungslogik müssten auch verändernde Vervielfältigungen sonstiger Werke von der Schrankenbestimmung freigestellt sein. Vor allem zielt die DSM-RL mit Art. 4 weiterhin ausdrücklich darauf ab, unter anderem Vervielfältigungen von Werken freizustellen, die während der Normalisierung der gesammelten Daten für das Text und Data Mining angefertigt werden.<sup>676</sup> Wie dargelegt wurde, führt die Normalisierung in vielen Fällen aber zu einer Veränderung der Werke. Demzufolge entspricht es dem niedergelegten Zweck des Art. 4

---

673 Peukert, in: Schricker/Loewenheim, Urheberrecht, UrhG § 62 Rn. 9; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 62 Rn. 6; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 62 Rn. 10; a. A. Nordemann, in: Fromm/Nordemann, Urheberrecht, UrhG § 62 Rn. 7.

674 OLG Hamburg, Urt. v. 05.06.1969 - 3 U 21/69, GRUR 1970, 38 (39) – Heintje; Engels, in: Götting/Lauber-Rönsberg/Ahlberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 62 Rn. 10 f.; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 62 Rn. 6, 12; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 62 Rn. 10.

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

676 ErwGr. 8 DSM-RL.

## B. Durchführung des Trainings auf Grundlage von Werken

DSM-RL, auch Veränderungen der Werke jedenfalls im Rahmen des Normalisierungsverfahrens urheberrechtlich freizustellen. Insgesamt ist deswegen davon auszugehen, dass die Schrankenbestimmung aus Art. 4 DSM-RL auch die Herstellung veränderter Vervielfältigungen jedenfalls für die Normalisierung von Werken umfasst.<sup>677</sup> Dies ist richtlinienkonform über das Tatbestandsmerkmal der Erlaubnis nach „Treu und Glauben“ in § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG auch bei der Bestimmung der Reichweite des Änderungsverbots zu berücksichtigen. Änderungen an Werken, die im Rahmen des Normalisierungsverfahrens stattfinden und zur Herstellung veränderter Vervielfältigungsstücke führen, sind demzufolge nicht vom Änderungsverbot aus § 62 Abs. 1 S. 1 UrhG betroffen. Sie werden vielmehr durch § 44b Abs. 2 UrhG freigestellt.

### c) Zwischenergebnis zur Normalisierung urheberrechtlich geschützter Trainingsdaten

Alles in allem können bei der Normalisierung urheberrechtlich geschützter Trainingsdaten, die regelmäßig vor dem eigentlichen Trainingsvorgang durchgeführt wird, Vervielfältigungshandlungen gemäß § 16 UrhG stattfinden. Diese Vervielfältigungen sind jedoch von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG umfasst. Das gilt auch für die Herstellung von im Rahmen des Normalisierungsverfahrens veränderten Vervielfältigungsstücken. Einer individuellen Erlaubnis durch die betroffenen Urheber bedarf es demzufolge jedenfalls dann nicht, wenn kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist.

Für den gesamten Trainingsprozess, zu dem auch der Normalisierungsschritt gehört, konnte aus einer regulatorischen Perspektive ein Freistellungsbedürfnis identifiziert werden. Danach sollte das Training von KI auf Basis von Werken durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte aus dem urheberrechtlichen Schutzbereich ausgeklammert werden. Vor diesem Hintergrund folgt aus dem möglichen Eingriff in das Vervielfältigungsrecht der von der Normalisierung betroffenen Urheber, dass das geltende Urheberrecht die im zweiten Teil der Untersuchung herausgearbeiteten Freistellungsbedürfnisse für die Durchführung des KI-Trainings nicht vollständig erfüllt. Aufgrund des Eingreifens der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG wird

---

<sup>677</sup> A. A. Novelli *u. a.*, Computer Law & Security Review 2024, S. 10.

das geltende Urheberrecht in Bezug auf die Normalisierungsprozesse aber jedenfalls dem Ob des regulatorischen Freistellungsbedürfnisses gerecht.

## 2. Feature Encoding

Im Anschluss an die Normalisierung wird das Feature Encoding der als Trainingsdaten gespeicherten Werke durchgeführt. Dabei werden die Werke in numerische Repräsentationen, sogenannte Tensoren, umgewandelt. Nur sie können im KNN, einem statistischen Modell, weiterverarbeitet werden.

Auch die Umwandlung der urheberrechtlich geschützten Trainingsdaten in Tensoren könnte einen Eingriff in das Vervielfältigungsrecht aus § 16 UrhG begründen. Wie bei der Normalisierung urheberrechtlich geschützter Trainingsdaten würden dies dazu führen, dass das im zweiten Teil der Untersuchung identifizierte Freistellungsbedürfnis vom geltenden Urheberrecht nicht verwirklicht wird.<sup>678</sup>

### a) Umwandlung in Tensoren als Vervielfältigung im Sinne des § 16 UrhG

Fraglich ist daher zum einen, ob mit dem Feature Encoding eine weitere Vervielfältigungshandlung verbunden ist, die gemäß § 16 UrhG dem Urheber vorbehalten wird.

Unter einer Vervielfältigung im Sinne des UrhG wird im Allgemeinen jede in beliebiger Weise vorgenommene körperliche Festlegung eines Werkes verstanden, die unmittelbar oder mittelbar zur menschlichen Wahrnehmung der geistigen Schöpfung geeignet ist.<sup>679</sup> Allerdings sind nach einer richtlinienkonformen Auslegung des Vervielfältigungsbegriffs nur solche Werkverwendungen auch erlaubnispflichtige Vervielfältigungen, die die Verwertungsmöglichkeit der Urheber zumindest potenziell beeinträchtigen.<sup>680</sup> Das ist insbesondere dann nicht der Fall, wenn die eigenschöpfen-

---

678 Zum regulatorischen Freistellungsbedürfnis siehe 2. Teil C.

679 BGH, Urt. v. 04.10.1990 - I ZR 139/89, GRUR 1991, 449 (453) – Betriebssystem; BGH, Urt. v. 06.10.2016, GRUR 2017, 266 Rn. 37 – World of Warcraft I; BT-Drs. IV/270, S. 47.

680 öst. OGH, Beschl. v. 23.02.2010 - 4 Ob 208/09 f, ZUM 2010, 629 (630); v. *Ungern-Sternberg*, in: Schricker/Loewenheim, Urheberrecht, UrhG §15 Rn. 23; v. *Ungern-Sternberg*, GRUR 2012, 576 (580); *Schönberger*, ZGE 2018, 35 (53); *Antoine*, Verändernde Werknutzungen, S. 179 f.; in diese Richtung auch *Grisse/Kaiser*, ZUM 2021, 401 (404); grundsätzlich wie hier auch *Antoine*, ZGE 2023, 387 (398), wobei sie

## B. Durchführung des Trainings auf Grundlage von Werken

rischen Züge eines Werkes im umgewandelten Erzeugnis nicht wiedererkennbar sind.<sup>681</sup> Die Verwertungsmöglichkeiten der Urheber werden darüber hinaus auch dann nicht beeinträchtigt, wenn die Werkverwendungen beiläufig oder zufällig und insgesamt untergeordnet sind.<sup>682</sup>

### aa) Wiedererkennbarkeitsgrenze

Beim Feature Encoding werden die erfassten Werke in einem numerischen Repräsentationsformat abgespeichert, sodass digitale Festlegungen der Werke entstehen. Fraglich ist aber, ob in diesen numerischen Festlegungen noch die eigenschöpferischen Züge der umgewandelten Werke wiedererkennbar sind. Dafür kommt es darauf an, wo genau die Wiedererkennbarkeitsgrenze verläuft, wie der Begriff also zu verstehen ist.

#### (1) Faktische oder normative Wiedererkennbarkeit

Zum einen ist möglich, die aus der Wiedererkennbarkeit folgende Schutzbereichsbegrenzung rein-faktisch zu verstehen. Dann wäre darauf abzustellen, ob übernommene Inhalte von Betrachtern rein tatsächlich noch als Werke der Urheber identifiziert werden können.<sup>683</sup> Dabei ist stets auf die konkreten schutzbegründenden Merkmale eines Werkes und nicht auf andere Faktoren wie insbesondere Urheberennnungen, identifizierende Informationsgehalte oder andere nicht schutzfähige Elemente abzustellen.<sup>684</sup> Ist

---

eine Einschränkung der funktionsbezogenen Auslegung auf die „Nutzungsfälle an den Rändern des urheberrechtlichen Ausschließlichkeitsrechts“ befürwortet, dazu sogleich. Zum Verhältnis zu den urheberpersönlichkeitsrechtlichen Bestimmungen siehe *Antoine*, ZGE 2023, 387 (401 ff.). Ausführlich dazu und zur funktionsbezogenen Auslegung des Vervielfältigungsrechts allgemein siehe oben unter 3. Teil B. I. 1. a) aa) und bb).

681 Vgl. EuGH, Urt. v. 29.07.2019 - C-476/17, ZUM 2019, 738 Rn. 38 – Pelham; BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 49, 56 – Porsche 911.

682 v. *Ungern-Sternberg*, GRUR 2012, 576 (580).

683 In diesem Sinne LG Hamburg, Urt. v. 09.12.2020 - 308 O 431/17, ZUM-RD, 164 (174) – Hey, Pippi Langstrumpf; *Grisse/Kaiser*, ZUM 2021, 401 (413); *Lauber-Rönsberg*, ZUM 2020, 733 (736); wohl ebenfalls *Homar*, ZUM 2019, 731 (736); *Wagner*, MMR 2020, 728 (730); in Bezug auf die Wiedererkennbarkeitsgrenze auch *Antoine*, Verändernde Werknutzungen, S. 187 ff., wobei sie im Allgemeinen ebenfalls eine funktional-beschränkende Auslegung des Vervielfältigungsrechts fordert, siehe S. 193.

684 BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 49 – Porsche 911; deutlich auch *Grisse/Kaiser*, ZUM 2021, 401 (411).

ein Werk tatsächlich nicht identifizierbar, sind die übernommenen eigenschöpferischen Elemente also vollständig unkenntlich gemacht worden, scheidet eine Wiedererkennbarkeit im urheberrechtlichen Sinne aus.

Zum anderen könnte die Wiedererkennbarkeitsgrenze aber auch normativ verstanden werden. In der Folge bestünden verschiedene Möglichkeiten, die Schutzbereichsbegrenzung auszustalten. Denkbar ist zum einen, über den Begriff der Wiedererkennbarkeit weitere teleologische Erwägungen zu berücksichtigen, die in einer Abwägung der im konkreten Fall widerstrebenden Grundrechte münden.<sup>685</sup> Darüber hinaus könnte auch darauf abgestellt werden, ob dem Ursprungswerk in der veränderten Gestaltung noch eine wesentliche Bedeutung zukommt.<sup>686</sup> Schließlich ließen sich auch alle schöpferischen Züge in einer Gesamtschau berücksichtigen. Dabei käme es darauf an, ob der Gesamteindruck der veränderten Gestaltung noch mit dem Ursprungswerk übereinstimmt.<sup>687</sup>

Unabhängig davon, ob es zu einem – jedenfalls denkbaren – normativen Verständnis des Wiedererkennbarkeitsbegriffs kommt, muss das natürliche Sprachverständnis des Begriffs aber Grundlage der Auslegung bleiben. Es bildet die Basis, aber auch die Grenze der Rechtsauslegung.<sup>688</sup> Um in den Schutzbereich des Vervielfältigungsrechts zu fallen, muss das Werk im veränderten Erzeugnis daher stets tatsächlich identifizierbar sein. Hierfür spricht auch, dass aus einer funktionsbezogenen Perspektive auf das Vervielfältigungsrecht<sup>689</sup> keine Substitutionsgefahr für die Verwertung des veränderten Werks durch den Urheber besteht, wenn es im Erzeugnis tatsächlich nicht identifizierbar ist. Die mit dem Vervielfältigungsrecht geschützten Verwertungsinteressen des Urhebers sind in diesem Fall daher nicht betroffen. Eine Normativierung des Wiedererkennbarkeitskriteriums kann demzufolge nur bei bestehender Identifizierbarkeit schutzbereichsbegrenzend, nicht aber in einer Art und Weise stattfinden, die die Wiedererkennbarkeit des Werkes allein auf Basis normativer Kriterien begründet.<sup>690</sup> Voraussetzung jeder Wiedererkennbarkeit ist demzufolge zunächst, dass

---

685 So insbesondere *Stieper*, ZUM 2019, 713 (719); *Schulze*, GRUR 2020, 128 (130).

686 Vgl. *Finke*, Werk und Veränderung, S. 204.

687 So vgl. BGH, Urt. v. 15.12.2022 - I ZR 173/21, GRUR 2023, 571 Rn. 28, 32 – Vitrinenleuchte; BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 56 – Porsche 911.

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

689 Dazu ausführlich siehe oben unter 3. Teil B. I. 1. a) bb).

690 So auch *Grisse/Kaiser*, ZUM 2021, 401 (405, 413); *Haberstumpf*, ZUM 2020, 810 (817); in diese Richtung auch *Stieper*, ZUM 2019, 713 (719).

die schutzbegründenden Elemente des Werkes in der veränderten Gestaltung noch objektiv erkennbar und nicht vollständig unkenntlich gemacht worden sind.<sup>691</sup> Vollständig unkenntlich gemacht worden sind die eigenschöpferischen Züge des Ursprungswerkes, wenn die veränderte Gestaltung dem Ursprungswerk bei seiner Rezeption nicht mehr zugeordnet werden kann.<sup>692</sup> Nur wenn im tatsächlichen Sinne Identifizierbarkeit im veränderten Erzeugnis besteht, kann in einem zweiten Schritt untersucht werden, ob die Wiedererkennbarkeit im Sinne der EuGH-Rechtsprechung aufgrund normativ-einschränkender Kriterien dennoch abzulehnen ist.<sup>693</sup> Die Herstellung des Erzeugnisses wäre nach richtlinienkonformer Auslegung damit keine Vervielfältigungshandlung gemäß § 16 UrhG.

(2) Beurteilungsperspektive für die faktische Wiedererkennbarkeit schutzbegründender Elemente

Für die Bestimmung der faktischen Wiedererkennbarkeit ist dem natürlichen Wortsinn nach die menschliche Wahrnehmung entscheidend.<sup>694</sup> Unklar ist aber, welche Betrachtungsperspektive bei Bestimmung der tatsächlichen Identifizierbarkeit genau eingenommen wird.

Einerseits könnte auf die Wahrnehmungsfähigkeit eines durchschnittlichen Rezipienten abgestellt werden.<sup>695</sup> Damit müsste die tatsächliche Wiedererkennbarkeit schöpferischer Elemente aus der Perspektive eines durchschnittlich geschulten und aufmerksamen Betrachters mitunter abgelehnt werden, obwohl bei einem gezielten und aufmerksamen Vergleich ein überdurchschnittlich vorgeprägter Betrachter wie beispielsweise ein Fachrichter

691 *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 42; *Grisse/Kaiser*, ZUM 2021, 401 (405); *Haberstumpf*, ZUM 2020, 810 (817); *Stieper*, ZUM 2019, 713 (719); vgl. außerdem BGH, Urt. v. 15.12.2022 - I ZR 173/21, GRUR 2023, 571 Rn. 29 – Vitrinenleuchte, woraus deutlich wird, dass der BGH jedenfalls davon ausgeht, dass vor einer Gesamtschau übernommene eigenschöpferische Züge identifiziert werden müssen.

692 *Grisse/Kaiser*, ZUM 2021, 401 (410).

693 Das Zweistufenverhältnis betont auch *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 42, 45.

694 *Grisse/Kaiser*, ZUM 2021, 401 (414).

695 BGH, Urt. v. 30.04.2020 - I ZR 115/16, GRUR 2020, 843 Rn. 29 – Metall auf Metall IV; *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 43; *Apel*, MMR 2019, S.601 (602); *Haberstumpf*, ZUM 2020, 810 (817); *Finke*, Werk und Veränderung, S. 203; vgl. mit der gleichen Beurteilungsperspektive zur Rechtslage vor Umsetzung der InfoSoc-RL LG Hamburg, Urt. v. 09.12.2020 - 308 O 431/17, ZUM-RD, 164 Rn. 37 – Hey Pippi Langstrumpf.

noch eigenschöpferische Elemente des Ursprungswerkes im veränderten Erzeugnis identifizieren könnte. Nach einem anderen, den Schutzbereich des Urheberrechts weiter fassenden Verständnisses könnte es aber auch allein auf eine objektive Identifizierbarkeit ankommen, selbst wenn nur überdurchschnittlich geprägtes Fachpublikum zur Identifizierung in der Lage wäre. Gegen ein so weites Verständnis der tatsächlichen Wiedererkennbarkeitsgrenze spricht allerdings, dass die Perspektive des durchschnittlichen Betrachters auch bei der Beurteilung der für den urheberrechtlichen Schutz notwendigen Individualität nach § 2 UrhG herangezogen wird.<sup>696</sup> Es ist damit konsequent, wenn eben jene Beurteilungsperspektive auch für die Wiedererkennbarkeit eingenommen wird und damit auch aus dem urheberrechtlichen Schutzbereich herausführt. Die Wiedererkennbarkeitsgrenze ließe bei einem weiten Verständnis, nach dem sogar die Identifizierbarkeit durch überdurchschnittlich geschultes Fachpublikum wie Fachgerichte ausreichend wäre, in den meisten Fällen außerdem faktisch leer. Insbesondere im prozessualen Kontext, der denklogisch eine Identifizierung durch den Urheber oder Rechteinhaber voraussetzt, erschöpfte sie sich vielmehr in der Feststellung einer Selbstverständlichkeit.<sup>697</sup> Um zu erfassen, ob die Verwertungsinteressen der Urheber durch die Verwendung eines Werkes beeinträchtigt sind, kommt es schließlich maßgeblich auf die Perspektive der Konsumenten an. Denn sie entscheiden darüber, ob das Erzeugnis eines Dritten Substitutionsgefahr für die Verwertung des Ursprungswerks durch den Urheber auslöst, oder nicht. Die Wiedererkennbarkeitsgrenze leitet sich indes aus einer an dem Schutz der Verwertungsinteressen orientierten Auslegung des Vervielfältigungsrechts ab. Demzufolge ist die Konsumentensicht auch für die Bestimmung der Wiedererkennbarkeit heranzuziehen.<sup>698</sup>

Alles in allem sprechen damit durchgreifende Erwägungen dafür, die Wiedererkennbarkeit auf Grundlage der Wahrnehmungsfähigkeit eines durchschnittlichen Rezipienten zu beurteilen. Auch bei der Beurteilung der

---

696 OLG Düsseldorf, Urt. v. 21.01.1997 - 20 U 13/96, ZUM 1998, 65 (67); OLG Karlsruhe, Urt. v. 10.11.1999 - 6 U 220/97, ZUM 2000, 327 (329); OLG Köln, Urt. v. 12.06.2009 - 6 U 215/08, GRUR-RR 2010, 182 (184) – Pferdeskulptur; *Bullinger*, in: Wandtke/Bullinger, Urheberrecht, UrhG § 2 Rn. 8; *Loewenheim/Leistner*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 2 Rn. 162; vgl. *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 2 Rn. 54.

697 *Antoine*, Verändernde Werknutzungen, S. 185.

698 So auch BGH, Urt. v. 30.04.2020 - I ZR 115/16, GRUR 2020, 843 Rn. 29 – Metall auf Metall IV.

## B. Durchführung des Trainings auf Grundlage von Werken

tatsächlichen Wiedererkennbarkeit im Sinne einer Identifizierbarkeit der eigenschöpferischen Züge eines Werkes wird damit eine in gewissem Grad normative Betrachtungsperspektive zugrunde gelegt. Wie auch bei Beurteilungen aus Sicht des Durchschnittsverbrauchers, die im Lauterkeitsrecht verbreitet sind, können Gerichte die Perspektive des durchschnittlichen Rezipienten kraft ihres Erfahrungswissens selbst einnehmen.<sup>699</sup> Verkehrsbefragungen in Rezipientenkreisen können dementsprechend zwar eine gewisse Indizwirkung entfalten. Erforderlich sind sie aber nicht.<sup>700</sup>

Eine Wiedererkennung ist weiterhin ohne Kenntnis über das im Ausgangspunkt wiederzuerkennende Werk bereits begrifflich nicht möglich. Deswegen ist bei der Beurteilung der faktischen Wiedererkennbarkeit davon auszugehen, dass die maßgeblichen Durchschnittsrezipienten auch Kenntnis vom jeweiligen Ursprungswerk haben.<sup>701</sup>

### (3) Technische Rekonstruierbarkeit

Um einem Missbrauch der Schutzbereichsbegrenzung über die Wiedererkennbarkeitsvoraussetzung vorzubeugen, müssen auch solche Festlegungen eines Werkes vom urheberrechtlichen Vervielfältigungsrecht mitumfasst sein, im Rahmen derer die eigenschöpferischen Elemente des Werkes zwar unkenntlich gemacht wurden, diese allerdings mit Hilfe technischer Mittel wieder rekonstruiert werden können.<sup>702</sup> Sofern für die Rekonstruktion allerdings nicht-übliche, über die durchschnittlichen technischen Fähigkeiten der Rezipienten hinausgehende Mittel eingesetzt werden müssen, ist eine Rekonstruktion der Werke jedenfalls von einem durchschnittlichen Rezipienten tatsächlich nicht zu erwarten. Es ist damit auch kein Multiplikations- bezie-

---

699 So auch *Leistner*, GRUR 2019, 1008 (1010); vgl. zum Durchschnittsverbraucher im Lauterkeitsrecht EuGH, Urt. v. 18.10.2012 - C-428/11, GRUR 2012, 1269 Rn. 53 – *Purely Creative* u. a.; außerdem ErwGr. 18 RL 2005/29/EG.

700 Gleichermaßen gilt für Sachverständigungsgutachten, siehe *Leistner*, GRUR 2019, 1008 (1010); vgl. zum Lauterkeitsrecht *Dreyer*, in: Harte-Bavendamm/Henning-Bodewig, Gesetz gegen den unlauteren Wettbewerb (UWG): mit Preisangabenverordnung und Geschäftsgeheimnisgesetz: Kommentar, UWG § 5 Rn. 157; a. A. für die Wiedererkennbarkeit aber *Grisse/Kaiser*, ZUM 2021, 401 (414).

701 So auch *Apel*, MMR 2019, S.601 (602).

702 Insoweit übereinstimmend *Grisse/Kaiser*, ZUM 2021, 401 (414); *Pesch/Böhme*, GRUR 2023, 997 (1005); *Käde*, Kreative Maschinen und Urheberrecht, S. 69 f.; *Jager*, Artificial Creativity?, S. 332; im allgemeineren Kontext auch *Dustmann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 16 Rn. 16.

hungsweise Substitutionseffekt auf dem relevanten Markt zu erwarten.<sup>703</sup> Damit fehlt es in diesen Fällen an einer Beeinträchtigung der Verwertungsinteressen betroffener Urheber. Die technisch umgewandelten, unkenntlich gemachten Werkrepräsentationen sind in diesem Fall infolgedessen keine Vervielfältigungsstücke der Werke im urheberrechtlichen Sinne.

Zur Bestimmung des nötigen Rekonstruktionsaufwands ist nicht auf subjektive Merkmale wie konkrete technische Fähigkeiten eines Nutzers oder gegen eine Nutzergruppe gerichtete technische Zugangsbeschränkungen, sondern auf den objektiv, also für jedermann erforderlichen Aufwand für eine Wiederherstellung der Wahrnehmbarkeit der eigenschöpferischen Elemente eines Werkes abzustellen.<sup>704</sup> Im Ergebnis sind die eigenschöpferischen Züge in technischen Kopien dann wiedererkennbar, wenn die Rekonstruktion vom durchschnittlichen Rezipienten ohne besondere technische Fähigkeiten oder mit üblicherweise bekannten technischen Hilfsmitteln und Programmen vorgenommen werden kann.<sup>705</sup> Das ist zum Beispiel bei der Konvertierung eines Werkes in eine RAR-Datei der Fall. Sie ermöglicht zwar unmittelbar keine Rezeption der schöpferischen Eigenschaften eines Werkes. Von durchschnittlichen Computernutzern kann das Werk mit leicht öffentlich verfügbaren Programmen aber ohne Weiteres wieder extrahiert werden. Hierdurch werden die schöpferischen Elemente des Werkes wieder wahrnehmbar. Gleiches gilt für eine Zerlegung in eine Vielzahl kleinster Werksegmente, die sich unter Zuhilfenahme öffentlich und kostenlos verfügbarer Computerprogramme für einen technisch auch nur durchschnittlich versierten Werkrezipienten ohne besondere Schwierigkeiten wieder zu einer vollständigen, wahrnehmbaren Festlegung des Werkes zusammenfügen lassen.<sup>706</sup>

---

703 A. A. zu einem etwas anders gelagerten Problem *Jager, Artificial Creativity?*, S. 333, wobei die Erwägungen auf einer sehr theoretischen Betrachtungsweise beruhen; außerdem *Dornis*, CR 2024, 765 Rn. 26 f., der die teleologische Perspektive auf die Reichweite des Schutzbereichs jedoch unberücksichtigt lässt.

704 Lediglich von subjektiven Schranken geht allerdings *Sesing-Wagenpfeil* aus, siehe *Sesing-Wagenpfeil*, ZGE 2024, 212 (243 ff.). Die dort zugrunde gelegte Beurteilungsgrundlage trägt deswegen nicht vollständig.

705 *Grisse/Kaiser*, ZUM 2021, 401 (414); a. A. aber *Dustmann*, in: *Fromm/Nordemann*, Urheberrecht, UrhG § 16; *Pesch/Böhme*, GRUR 2023, 997 (1005); *Käde*, Kreative Maschinen und Urheberrecht, S. 69 f.; *Sesing-Wagenpfeil*, ZGE 2024, 212 (243 ff.); in einem etwas anderen Kontext auch *Jager, Artificial Creativity?*, S. 332, nach denen es nicht auf den Aufwand für die Rekonstruktion der Inhalte ankommt, was aus den o. g. Gründen allerdings zu weitreichend ist.

## B. Durchführung des Trainings auf Grundlage von Werken

- (4) Wiedererkennbarkeitsgrenze bei ausschließlich technisch bedingten Erzeugnissen ohne eigene Werkqualität

Die beim Feature Encoding von Trainingsdaten erzeugten Werkrepräsentationen werden als Teil eines technischen Verfahrens angefertigt. Klärungsbedürftig ist in diesem Zusammenhang, ob die Wiedererkennbarkeitsgrenze das Vervielfältigungsrecht nach einer richtlinienkonformen, funktionsbezogenen Auslegung auch dann beschränkt, wenn das veränderte Erzeugnis ausschließlich technisch bedingt ist. Ist das nicht der Fall, so wären ausschließlich technisch bedingte Werkrepräsentationen wie die Produkte des Feature Encodings auch dann Vervielfältigungsstücke im Sinne des § 16 UrhG, wenn die Werke in der Repräsentation nicht wiedererkennbar sind.

Dafür, die Wiedererkennbarkeitsgrenze des urheberrechtlichen Schutzgegenstands nicht auf ausschließlich technisch bedingte Produkte anzuwenden, spricht, dass sie von der überwiegenden rechtswissenschaftlichen Literatur in § 23 Abs. 1 S. 2 UrhG verortet wird.<sup>707</sup> Nach dem Wortlaut des § 23 Abs. 1 S. 2 UrhG muss zwischen dem Ausgangs- und einem neugeschaffenen Werk ein hinreichender Abstand (unter anderem im Sinne fehlender Wiedererkennbarkeit) bestehen. Neue Werke mit eigenen schöpferischen Merkmalen werden bei ausschließlich technisch bedingten Prozessen hingegen nicht angefertigt.

Allerdings muss die die richtlinienkonforme Auslegung im nationalen Recht nicht an § 23 Abs. 1 S. 2 UrhG angeknüpft werden. Denn sie geht auf den Sinn und Zweck von Art. 2 InfoSoc-RL zurück. Die funktional-einschränkende Auslegung des Vervielfältigungsrechts und damit auch die Wiedererkennbarkeitsgrenze kann demzufolge unmittelbar in einer richtlinienkonformen Auslegung des § 16 UrhG berücksichtigt werden.<sup>708</sup> Infolgedessen bedarf es auch keiner Überwindung des Wortlauthindernisses aus § 23 Abs. 1 S. 2 UrhG, nach dem ein neues Werk geschaffen worden sein muss.<sup>709</sup> Dies macht deutlich, dass die Frage nach der Werkqualität des

---

706 Vgl. zu einem ähnlichen Fall, aber im Kontext der öffentlichen Wiedergabe nach Art. 3 InfoSoc-RL: EuGH, Urt. v. 17.06.2021 - C-597/19, GRUR 2021, 1067 Rn. 59 – Mircom/Telenet.

707 So bisher die weit überwiegende Ansicht, beispielsweise *Raue*, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 23 Rn. 42; *Lauber-Rönsberg/Ahlberg*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 23 Rn. 3 f.; *Grünberger*, ZUM 2023, 309 (317 ff.). Vgl. auch BT-Drs. 19/27426, S. 79.

708 Siehe oben unter 3. Teil B. I. a) bb).

709 In diese Richtung auch BGH, Urt. v. 07.04.2022 - I ZR 222/20, GRUR 2022, 899 Rn. 52, 55, 58 – Porsche 911; zustimmend *Grünberger*, ZUM 2023, 309 (318).

veränderten Erzeugnisses keine Frage der richtlinienkonformen Einschränkung des Vervielfältigungsrechts, sondern eine Frage der nationalen Anknüpfung im UrhG ist. Für die funktionsbezogene Auslegung ist nur maßgeblich, ob ein abgeleitetes Erzeugnis die Verwertungsinteressen betroffener Urheber beeinträchtigt.<sup>710</sup> Die Werkqualität des generierten Erzeugnisses ist hierfür nicht von Bedeutung.

Insgesamt sprechen damit bessere Gründe dafür, dass die § 16 UrhG nach richtlinienkonformer Auslegung immanente Schutzbereichsgrenze der Wiedererkennbarkeit nicht nur bei der Schöpfung abgeleiteter Werke überschritten werden kann.<sup>711</sup> Damit können auch ausschließlich technisch bedingte Verfahrenserzeugnisse Werke in nicht wiedererkennbarer und demzufolge urheberrechtlich nicht relevanter Art und Weise enthalten. Werkrepräsentationen, die beim Feature Encoding urheberrechtlich geschützter Trainingsdaten hergestellt werden, qualifizieren sich nach der funktionsbezogenen Auslegung des Vervielfältigungsrechts damit ebenfalls dann nicht als Vervielfältigungsstück im Sinne des § 16 UrhG, wenn die eigenschöpferischen Merkmale der umgewandelten Werke nach Abschluss des technischen Verfahrens nicht mehr wiedererkennbar sind.

bb) Anwendung des Maßstabs der Wiedererkennbarkeit auf den konkreten Fall

Beim Feature Encoding werden die digital erfassten und normalisierten Werke in numerische Repräsentationen umgewandelt. Die Werke werden nach Abschluss des Umwandlungsprozesses also nur noch als Anordnung von Zahlen anhand von Achsen dargestellt.

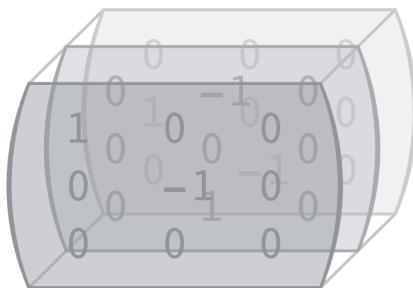
---

710 Dazu im Detail oben unter 3. Teil B. I. 1. a) bb).

711 Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 23 Rn. 28; Grünberger, ZUM 2023, 309 (318); Siglmüller/Gassner, RDi 2023, 124 Rn. 15; Maamar, ZUM 2023, 481 (489 f.); wohl auch Pesch/Böhme, GRUR 2023, 997 (1006); Schwarz/Söbbing, RDi 2023, 415 Rn. 13; a. A. Lauber-Rönsberg/Ahlberg, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 23 Rn. 36; Wandtke/Hauck, GRUR Prax 2020, 542 (543), die allesamt Werkqualität voraussetzen. Wohl ebenso, aber noch zu § 24 UrhG a. F.: Ory/Sorge, NJW 2019, 710 (713).

## B. Durchführung des Trainings auf Grundlage von Werken

Diese können beispielsweise wie folgt aussehen:



Für einen Rezipienten besteht bei diesen rein-numerischen Darstellungen keine Möglichkeit mehr, den eigentlichen Inhalt oder die ursprüngliche Form eines Werkes, also die jeweiligen schutzbegründenden Elemente, wahrzunehmen und zu verarbeiten. Die schutzbegründenden Merkmale eines Werkes aus diesen Zahlenmustern zu erkennen, ist für einen durchschnittlichen Rezipienten nicht mehr möglich. Die durchschnittlichen Rezipienten sind auch nicht in der Lage, die numerischen Darstellungen noch den Ursprungswerken zuzuordnen. Selbst wenn Tensor und zugehöriges Werk nebeneinander abgebildet sind, kann ein durchschnittlicher Rezipient keine inhaltliche Verwandtschaft beider Darstellungen erkennen.

Zwar können die Tensoren mit Hilfe technischer Mittel in wahrnehmbare Darstellungsformate retransformiert werden. Eine Rückumwandlung der Tensor-Repräsentation setzt aber vertiefte technische Kenntnisse und die Nutzung von Software voraus, die außerhalb des vom durchschnittlichen Rezipienten üblicherweise Beherrschten liegt.<sup>712</sup> Im Gegensatz zu anderen technischen Maßnahmen wie dem Entpacken von RAR-Dateien ist die Retransformation von Tensoren für durchschnittlich technisch gebildete Werkrezipienten daher nicht möglich. Auch wenn sie technisch möglich ist, kann sie von einem durchschnittlichen Rezipienten demzufolge nicht erwartet werden.

Insgesamt folgt damit auf Grundlage der herausgearbeiteten Maßstäbe, dass die umgewandelten Werke in den rein-numerischen Tensoren tatsächlich nicht mehr identifizierbar sind. Dies führt zu einem Ausbleiben der Wiedererkennbarkeit im urheberrechtlichen Sinn. Auf ein normatives Ver-

---

712 Das verkennt Käde, Kreative Maschinen und Urheberrecht, S. 69 f.

ständnis des Wiedererkennbarkeitsbegriffs kommt es an dieser Stelle demzufolge nicht an.

Nach richtlinienkonformer Auslegung des Vervielfältigungsbegriffs findet beim Feature Encoding urheberrechtlich geschützter Trainingsdaten demzufolge keine Vervielfältigung der Inhalte gemäß § 16 Abs. 1 UrhG statt.<sup>713</sup> Die Festlegung der Werke in Form numerischer Tensoren genügt nicht, um das Werk auch nur mittelbar menschlich wahrnehmbar zu machen.<sup>714</sup>

#### b) Schlussfolgerung

Im Rahmen des Feature Encodings von urheberrechtlich geschützten Trainingsdaten werden urheberrechtlich nicht relevante Werkrepräsentationen angefertigt. Dies gilt auch für weitere Kopien der Tensoren, die während des KI-Trainings angefertigt werden. Es bestehen systemische Ähnlichkeiten zu anderen Arten sekundärer Werkwiedergaben wie beispielsweise Inhaltsbeschreibungen, die ebenfalls kein Vervielfältigungsstück im Sinne des § 16 UrhG sind.

Der Vorverarbeitungsschritt des Feature Encodings urheberrechtlich geschützter Trainingsdaten fällt demzufolge bereits aufgrund einer unmittelbar begrenzenden, teleologisch geprägten Auslegung des Vervielfältigungsrechts nicht in den Schutzbereich des Ausschließlichkeitsrechts der Urheber. Damit werden die im zweiten Teil der vorliegenden Untersuchung für den Trainingsprozess identifizierten Regelungsbedürfnisse insoweit vollständig erfüllt.

Letztendlich folgt hieraus aber nicht, dass die weitergehende Verarbeitung des Tensors keine urheberrechtlich relevante Nutzungshandlung mehr begründen könnte. Denn sobald die Retransformation eines Tensor in ein menschlich wahrnehmbares Format stattfindet, wird das Werk wieder tatsächlich identifizierbar. Dies kann auch zur Wiedererkennbarkeit im urheberrechtlichen Sinn führen. Aufgrund dessen ist möglich, dass im Moment der Retransformation ein Vervielfältigungsstück im Sinne des § 16 UrhG

---

713 A. A. Käde, Kreative Maschinen und Urheberrecht, S. 69 f.; Spindler, GRUR 2016, 1112 (1113); im Anschluss hieran auch Wiss. Dienst des Deutschen Bundestages, KI und ML, eine urheberrechtliche Betrachtung (BT WD 10 – 3000 – 67/18), S. 7; wohl auch Dornis/Stober, Urheberrecht und Training generativer KI-Modelle, S. 70; Dregelies, GRUR 2024, 1484 (1484 f.).

714 A. A. Käde, Kreative Maschinen und Urheberrecht, S. 69 f.

## *B. Durchführung des Trainings auf Grundlage von Werken*

hergestellt wird. Diese Umwandlung und die Weiterverarbeitung des umgewandelten Erzeugnisses bedürften dann der Erlaubnis durch den Urheber oder eine Schrankenbestimmung.

### **3. Zwischenfazit zur Vorverarbeitung urheberrechtlicher geschützter Trainingsdaten**

Alles in allem können die Freistellungsbedürfnisse, die für die Vorverarbeitung urheberrechtlich geschützter Trainingsdaten als Teil des Trainingsprozesses ermittelt wurden, durch das geltende Urheberrecht nur teilweise erfüllt werden. Im Hinblick auf die Normalisierung von Trainingsdaten hängt dies damit zusammen, dass die bei der Normalisierung angefertigten Kopien der Werke Vervielfältigungsstücke des § 16 UrhG sein können. Damit bleibt das geltende Urheberrecht hinter den für die Normalisierung als Teil des Trainingsprozesses identifizierten Freistellungsbedürfnissen zurück. Allerdings gilt das nur für das Wie, nicht aber das Ob der Freistellung. Denn über eine Anwendung der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG kann dennoch eine Freistellung der Normalisierung erreicht werden. Voraussetzung hierfür ist jedoch, dass der Rechtsinhaber keinen Nutzungsvorbehalt gemäß § 44b Abs. 3 UrhG erklärt hat.

Im Gegensatz hierzu werden die hinsichtlich des Trainingsprozesses identifizierten Regelungsbedürfnisse beim anschließenden Feature Encoding der urheberrechtlich geschützten Trainingsdaten vollständig erfüllt. Die Herstellung der Tensoren, also der numerischen Repräsentationen der Werke, greift nicht in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG ein.

## *II. Trainingsanalyse der Werke selbst*

Auch die eigentliche Trainingsanalyse der Werke selbst muss durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte urheberrechtlich freigestellt sein. Die Trainingsanalyse der Werke ist mit dem (automatisierten) Werkgenuss von Werken vergleichbar.<sup>715</sup> Der eigentliche Werkgenuss wird von den urheberrechtlichen Verwertungsrechten nicht

---

<sup>715</sup> Zur ausführlichen Begründung siehe oben unter 2. Teil C. II. 1.

erfasst.<sup>716</sup> Das gilt auch für den automatisierten Werkgenuss, also die automatisierte, informationsbezogene Auswertung von Werken, das Data Mining.<sup>717</sup> Dies spricht dafür, dass auch das mit dem automatisierten Werkgenuss wesensverwandte, eigentliche Training einer KI auf Basis von Werken nicht Teil des urheberrechtlichen Ausschließlichkeitsrechts ist. Wenn die in einem Werk enthaltenen Informationen nicht Teil des urheberrechtlichen Schutzbereichs sind und ihre Extraktion damit frei möglich ist, kann es außerdem keinen Unterschied machen, wie viele dieser Informationen aus dem Werk gewonnen worden sind oder aus wie vielen Werken eine Informationsextraktion stattfindet. Gegen die Gleichstellung des KI-Trainings mit dem Werkgenuss kann daher auch nicht eingewendet werden, dass das KI-Training quantitativ als auch qualitativ über den menschlichen Werkgenuss hinausgeht.<sup>718</sup> Im Ergebnis ist die eigentliche Analyse der Werke beim Training von KI damit – wie auch der automatisierte Werkgenuss – nicht Teil der urheberrechtlichen Verwertungsrechte.<sup>719</sup> Die Trainingsanalyse selbst wird demzufolge durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte aus dem Schutzbereich des Urheberrechts ausgeklammert. Das Freistellungsbedürfnis für das Training von KI wird hinsichtlich des eigentlichen Analysevorgangs damit erfüllt. Das „*right to read*“ ist insofern nicht nur das „*right to mine*“ im datenwissen-

716 Heerma, in: Wandtke/Bullinger, Urheberrecht, UrhG § 15 Rn. 10 m. w. N. außerdem bereits BT-Drs. IV/270, S. 28. Nach Sucker gilt das jedenfalls für den analogen Werkgenuss, während der digitale Werkgenuss mittelbar durch die zwangsläufig stattfindenden Vervielfältigungshandlungen vom urheberrechtlichen Ausschließlichkeitsrecht erfasst wird, siehe Sucker, Der digitale Werkgenuss im Urheberrecht, S. 47 ff., 89. Diese – zutreffende – folgenorientierte Betrachtungsweise hat aber keinen Einfluss darauf, dass der eigentliche menschliche Analysevorgang, also der Werkgenuss im engeren Sinne, auch im digitalen Kontext nicht dem Ausschließlichkeitsrecht der Urheber zugewiesen ist.

717 Das zeigt schon die Regelung des § 44b Abs. 2 UrhG. Siehe außerdem Stieper, in: Schrieker/Loewenheim, Urheberrecht, UrhG § 60d Rn. 3; Raue, ZUM 2019, 684 (685); de la Durantaye, Allgemeine Bildungs- und Wissenschaftsschranke, S. 240 f.; Kleinkopf, Text- und Data-Mining, S. 181; Cocoru/Boehm, An analytical review of text and data mining practices and approaches in Europe - Policy recommendations in view of the upcoming copyright legislative proposal, S. 8; European Copyright Society, General Opinion on the EU Copyright Reform Package, S. 5; BT-Drs. 18/12329, S. 40; Hilty/Moscon, Modernisation of the EU Copyright Rules, S. 127 f.

718 So aber Schack, NJW 2024, 113 Rn. 9.

719 Vgl. auch Raue, ZUM 2019, 684 (684 f.); Wulf/Löbeth, GRUR 2024, 737 (738); ebenso Kögel, DSRITB 2023, 285 (287).

## B. Durchführung des Trainings auf Grundlage von Werken

senschaftlichen Sinne,<sup>720</sup> sondern auch das „*right to train*“ beim Training von KI auf Basis von Werken.<sup>721</sup>

## III. Vervielfältigungen während der Trainingsanalyse

Während der Trainingsanalyse kommt es außerdem zu weiteren Vervielfältigungen der Trainingsdaten. Sie werden beispielsweise vorübergehend in den Arbeitsspeicher der für das Training genutzten Rechenarchitektur geladen. Vervielfältigt werden allerdings nicht die Werke selbst, sondern die Tensoren, also die numerischen Repräsentationen der Trainingsdaten, die im Rahmen der Datenvorverarbeitung angefertigt worden sind.<sup>722</sup> Diese sind keine Vervielfältigungsstücke der Werke. Ihre Vervielfältigung im Rahmen der Trainingsanalyse berührt demzufolge auch nicht das urheberrechtliche Ausschließlichkeitsrecht.<sup>723</sup> Die Vervielfältigungen während der Trainingsanalyse sind vielmehr frei möglich. Dies steht im Einklang mit den identifizierten regulatorischen Freistellungsbedürfnissen.

## IV. Gewichtungswerte und andere Parameter des trainierten Modells

Ziel des Trainingsvorgangs ist eine Konfiguration der im KNN enthaltenen, numerischen Parameter. Sie soll in einer Art und Weise erfolgen, die eine bestmögliche Qualität des vom KI-System generierten Outputs gewährleistet. Auch die Konfiguration der Parameter innerhalb des KNN sollte als Teil des Trainingsprozesses aus einer regulatorischen Perspektive durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte freigestellt sein.<sup>724</sup>

Ob die identifizierten Freistellungsbedürfnisse im vorliegenden Fall vom geltenden Urheberrecht erfüllt werden, hängt davon ab, ob die konfigurierten Parameter des KNN Vervielfältigungsstücke der zum Training verwen-

---

720 Zur Formulierung *Spindler*, ZGE 2018, 273 (283); im Kontext wissenschaftlicher Forschung *Raeu*, ZUM 2019, 684 (686).

721 A. A. *Schack*, NJW 2024, 113 Rn. 9; im Kontext generativer KI-Systeme *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 103.

722 Siehe zum Feature Encoding unmittelbar oben unter 3. Teil B. I. 2.

723 A. A. wohl *Jager*, Artificial Creativity?, S. 331 f., dessen Standpunkt auf der Annahme beruht, dass nicht nur numerische Repräsentationen, sondern stets die Werke selbst verarbeitet werden. Ebenfalls a. A. *Schack*, NJW 2024, 113 Rn. 4.

724 Vgl. oben unter 2. Teil C. I.

deten Werke im Sinne des § 16 UrhG sind. Der Vervielfältigungsbegriff wird dabei nicht rein technisch, sondern normativ anhand der vom Vervielfältigungsrecht geschützten Verwertungsinteressen der Urheber bestimmt.<sup>725</sup> Eine Vervielfältigung liegt im konkreten Fall danach insbesondere dann nicht vor, wenn die eigenschöpferischen Züge des Werkes in den numerischen Parametern des trainierten KNN nicht mehr wiedererkennbar sind.<sup>726</sup> Die objektiv zu bestimmende Wiedererkennbarkeit eines Werkes in einem Erzeugnis ist dabei von subjektiven Zugangsbeschränkungen zu unterscheiden.<sup>727</sup> Für das Fehlen der Wiedererkennbarkeit ist nicht erforderlich, dass der neuen, veränderten Gestaltung, also den Parametern des trainierten KNN, selbst als Werk urheberrechtlicher Schutz zukommt.<sup>728</sup> Die Schutzbereichsgrenze der Wiedererkennbarkeit besteht auch bei rein technischen Verarbeitungen wie der Konfiguration von Parametern innerhalb eines KNN.

Damit die schutzbegründenden Elemente der zum Training verwendeten Werke in den Parametern eines KNN noch wiedererkennbar sind, müssen sie dort zunächst noch tatsächlich und objektiv identifiziert werden können.<sup>729</sup> Bei einer vollständigen Unkenntlichmachung der eigenschöpferischen Züge scheidet eine Wiedererkennbarkeit hingegen aus. Bei der Beurteilung kommt es auf die Wahrnehmung eines durchschnittlichen Betrachters an.<sup>730</sup> Die Notwendigkeit einer technischen Rekonstruktion der Inhalte ist für die tatsächliche Wiedererkennbarkeit der schutzbegründenden Merkmale dabei unschädlich, solange die Wiederherstellung vom durchschnittlichen Rezipienten ohne besondere technische Fähigkeiten oder mit üblicherweise bekannten technischen Hilfsmitteln und Programmen vorgenommen werden kann.<sup>731</sup>

Erst wenn ein durchschnittlicher Betrachter schutzbegründende Elemente in den Parametern eines KNN objektiv identifizieren kann, ist in einem zweiten Schritt zu untersuchen, ob es trotz einer objektiv-tatsächlichen Er-

---

725 Dazu m. w. N. auch zu abweichenden Ansichten siehe oben unter 3. Teil B. I. 1. a) bb).

726 Zur Herleitung genauer oben unter 3. Teil B. I. 1. a) bb).

727 Das erkennen *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 77 f.

728 Ausführlicher hierzu mit abweichenden Ansichten oben unter 3. Teil B. I. 2. a) aa) (4).

729 Dazu ausführlich m. w. N. oben unter 3. Teil B. I. 2. a) aa) (1).

730 Hierzu bereits siehe 3. Teil B. I. 2. a) aa) (2).

731 Str., ausführlich hierzu mit abweichenden Ansichten oben unter 3. Teil B. I. 2. a) aa) (3).

## B. Durchführung des Trainings auf Grundlage von Werken

kennbarkeit der zum Training verwendeten Werke in den Parametern des KNN dennoch zu einer normativ-einschränkenden Auslegung der Wiedererkennbarkeitsgrenze mit der Folge kommen kann, dass die Parameter des KNN trotz der Identifizierbarkeit schutzbegründender Elemente dennoch keine Vervielfältigungsstücke im Sinne des § 16 UrhG beinhalten.

### 1. Objektive Identifizierbarkeit schutzbegründender Elemente in Parametern eines KNN

Ein durchschnittlicher Betrachter verfügt typischerweise nicht über die technischen Fähigkeiten, um die Architektur und Parameter eines KNN einzusehen. Prohibitiv wirkt bereits, dass die hierfür notwendigen Programme nicht üblicherweise bekannt sind. Hierin besteht auch ein Unterschied zur Speicherung von Werken auf Computerfestplatten, die über marktübliche, in der Regel auf einem Computer vorinstallierte und dem objektiven Durchschnittsbetrachter zur Verfügung stehende Programme typischerweise leicht wahrnehmbar gemacht werden können.<sup>732</sup> Selbst wenn der durchschnittliche Betrachter die Architektur und Parameter eines KNN einsehen könnte, sind im KNN selbst nur numerische Darstellungen, also einzelne Zahlenwerte sichtbar. Sie wurden auf Basis der zum Training genutzten Werke ermittelt. Ein durchschnittlicher Betrachter kann aus Angaben wie „ $w_1 = 0,5$  ;  $w_2 = 1,8$  ;  $w_3 = 21,9$ “ allerdings keine Schlüsse auf die schutzbegründenden Merkmale eines Werkes ableiten. Die Zahlenwerte vermitteln an den durchschnittlichen Betrachter also nicht die schöpferischen Merkmale eines zum Training genutzten Werkes, also beispielsweise einen Pinselstrich bestimmter Technik und Führung oder ein Sujet, in der die Persönlichkeit eines Urhebers zum Ausdruck kommen.<sup>733</sup> Aufgrund dieser numerischen Repräsentation besteht auch keine Vergleichbarkeit zur Speicherung von Werken im Arbeitsspeicher eines Computers. Die Werke werden dort – vom nativen Maschinencode abgesehen – nicht als numerische Repräsentationen, sondern in Klarform vorübergehend abgelegt.<sup>734</sup>

---

732 Anders jedoch *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 77, die eine Gleichstellung beider Vorgänge vornehmen.

733 Darauf stellen auch *Käde*, ZUM 2024, 174 (177); *Konertz/Schönhof*, WRP 2024, 534 Rn. 8; *Kögel*, DSRITB 2023, 285 (288) ab.

734 A. A. *Jager*, Artificial Creativity?, S. 332.

## 2. Möglichkeiten der Invertierung und die Wiedererkennbarkeit

Zwar ist in der Computerwissenschaft noch nicht abschließend geklärt, ob und wenn ja zu welchem Grad eine Memorisierung von Werken in den Parametern eines trainierten KNN stattfindet und mit welcher Ergebnisqualität eine Invertierung, also eine Extraktion der zum Training genutzten Werke aus dem trainierten KNN, erfolgen kann.<sup>735</sup> Selbst wenn dies möglich wäre, ist eine Invertierung allerdings technisch komplex.<sup>736</sup> Die Invertierung könnte nicht mit für technische Laien bekannten Mitteln erfolgen, sondern setzte vielmehr besondere Fähigkeiten und Ressourcen voraus. Sie ist computerwissenschaftlichen Fachkreisen vorbehalten. Die Rekonstruktion von Werken aus den Parametern eines KNN ist für den durchschnittlichen Betrachter demzufolge nicht möglich. Infolgedessen ist unabhängig davon, ob eine Invertierung der KNN theoretisch möglich ist, nicht von einer technischen Rekonstruierbarkeit der zum Training genutzten Werke auszugehen, die zu einer tatsächlichen Identifizierbarkeit und damit zu einer Wiedererkennbarkeit im urheberrechtlichen Sinne führt.<sup>737</sup>

---

735 Dazu vgl. beispielhaft *Feldman*, arXiv:1906.05271 2019; *Nagarajan/Raffel/Goodfellow*, Paper zur NeurIPS Conference 2018; *Carlini u. a.*, arXiv: 2301.13188 2023, wobei es hier, wie der Auffand der Rekonstruktion zeigt, weniger um strukturelle, sondern um Einzelfallprobleme geht; aus der juristischen Literatur *Pesch/Böhme*, GRUR 2023, 997 (1001); *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 59 ff.

736 Insoweit zustimmend auch *Jager*, Artificial Creativity?, S. 331. Eindrucksvoll insbesondere das Beispiel von *Carlini u. a.*, arXiv: 2301.13188 2023: Die Autoren untersuchten 160 Millionen verwendete Trainingsdaten eines generativen KI-Systems und nutzten 350.000 Datenpaare, die im Trainingsdatensatz doppelt hinterlegt worden sind. Die hieraus extrahierten Bildbeschreibungen wurden jeweils 500 Mal als Prompt in das generative KI-System eingegeben. Die 175 Millionen hieraus generierten Inhalte wurden nunmehr mit den 350.000 zuvor ermittelten Datenpaaren verglichen. Hieraus ergaben sich 109 Übereinstimmungen, was einer Übereinstimmungsquote zum gesamten Trainingsdatensatz von 0,00000062 % entspricht. Für diese Berechnungen mussten alle Bildinhalte in Vektoren umgewandelt werden. Für das Experiment waren also fortgeschrittene technische Fähigkeiten und erhebliche Rechenkapazitäten notwendig.

737 Ein vergleichbares Konzept verfolgt auch *Leistner*, GRUR 2024, 1665 (1669). Andere Ansicht insbesondere *Jager*, der den Erfolg einiger Forschungsversuche zur Invertierung als ausreichend für eine generelle Wahrnehmbarmachung der zum Training verwendeten Werke erachtet, siehe *Jager*, Artificial Creativity?, S. 331. A. A. außerdem *Pesch/Böhme*, GRUR 2023, 997 (1005) sowie jüngst *Dornis*, CR 2024, 765 Rn. 22 ff. und insbesondere Rn. 26, der sich allerdings ebenfalls nicht mit der Funktion des urheberrechtlichen Vervielfältigungsrechts und dessen Folgen auseinandersetzt.

### 3. Identifizierbarkeit schutzbegründender Elemente im Output von KI und die Wiedererkennbarkeit

Sollten schutzbegründende Merkmale von zum Training genutzten Werken ausnahmsweise im Output eines KI-Systems enthalten sein, werden die schutzbegründenden Eigenschaften erst mit der Generierung dieses konkreten Outputs wieder tatsächlich, objektiv identifizierbar. Die Generierung des Outputs ist sowohl zeitlich als auch wirtschaftlich vom Training eines KI-Systems abgrenzbar.<sup>738</sup> Sie geht außerdem auf einen eigenen Willensentschluss zurück. Die Entwicklung eines KI-Systems inklusive der Konfiguration der Parameter innerhalb des KNN und die Generierung konkreter, gegebenenfalls urheberrechtsverletzender KI-Erzeugnisse sind demzufolge voneinander unabhängig. Gegebenenfalls liegen auch zwei unabhängige Nutzungshandlungen vor.<sup>739</sup> Hieraus folgt, dass nicht von der Generierung des *Outputs* auf eine Werknutzung beim *Training* der KI geschlossen werden kann.<sup>740</sup> Eine Betrachtung hat vielmehr jeweils getrennt voneinander zu erfolgen. Die Identifizierbarkeit schutzbegründender Elemente von zum Training verwendeten Werken im Output eines KI-Systems spricht demzufolge nicht automatisch für eine urheberrechtsrelevante, erlaubnisbedürftige *Vervielfältigung* der zum Training verwendeten Werke in den Parametern des KNN.<sup>741</sup> Bei der Erzeugung von Output durch eine KI wird auch aus einer technischen Perspektive außerdem nicht die Gesamtheit der Parameter, sondern das Rechenergebnis des KNN, also der Aktivierungszustand der letzten Neuronenschicht, sichtbar gemacht. Dieser wurde *auf Basis* der Parameter des KNN errechnet.<sup>742</sup> Damit kann das Verhältnis von Parametern und Output eines KNN auch nicht mit klassischen Werkwiedergaben, beispielsweise durch einen Schallplattenspieler, verglichen werden. In diesem Fällen wird das untersuchte Vervielfältigungsstück (Schallplatte) und kein auf Basis des Vervielfältigungsstücks (Schalplatte) abgeleitetes Erzeugnis wahrnehmbar gemacht. Vielmehr ist die Generierung des Outputs eine eigene urheberrechtlich relevante Vervielfältigungs-

---

738 So auch *Leistner*, GRUR 2024, 1665 (1669).

739 So auch *Käde*, ZUM 2024, 174 (177); *Schwarz/Söbbing*, RDi 2023, 415 Rn. 13.

740 A. A. *Dornis*, CR 2024, 765 Rn. 9; *Sesing-Wagenpfeil*, ZGE 2024, 212 (229 ff., 233 ff.).

741 Anders jedoch *Pesch/Böhme*, GRUR 2023, 997 (1005) mittels „geschickten Eingabeprompts“; *Konertz/Schönhof*, WRP 2024, 534 Rn. 8 am Ende; insbesondere auch *Sesing-Wagenpfeil*, ZGE 2024, 212 (229 f.), dessen Untersuchung maßgeblich auf der gegenteiligen Annahme gründet.

742 Insoweit technisch unzutreffende Grundannahme in *Sesing-Wagenpfeil*, ZGE 2024, 212 (229 ff.).

handlung im Sinne des § 16 UrhG, wenn in dem Output schutzbegründende Elemente wiedererkennbar sind. Diese Generierung ist für sich genommen sodann erlaubnispflichtig.<sup>743</sup>

Darüber hinaus erscheint ein zwingender Schluss von der Wahrnehmbarmachung eines Werkes auf das Vorliegen eines Vervielfältigungsstücks desselben Werkes jedenfalls zweifelhaft.<sup>744</sup> Beispielsweise macht eine Textanleitung zur Reproduktion eines Werks der bildenden Kunst die schöpferischen Züge des Werkes nicht urheberrechtlich relevant wiedererkennbar. Somit qualifiziert sie sich nicht als Vervielfältigungsstück des Werkes. Dennoch ist die Anleitung geeignet, das urheberrechtlich geschützte Werk durch ihre Ausführung wahrnehmbar zu machen. Die Wahrnehmbarmachung eines Werkes kann, muss damit aber nicht zwingend auf Basis eines Vervielfältigungsstücks erfolgen.

Sollen die *Parameter* eines KNN als Vervielfältigungen von zum Training genutzten Werken im Sinne des § 16 UrhG eingeordnet werden, müssen ihre schutzbegründenden Merkmale vielmehr direkt bei einer Analyse des KNN und nicht des Outputs der KI wiedererkennbar sein. Dies ist, wie dargelegt wurde, bereits aufgrund der numerischen Darstellungsform und einer komplexen technischen Rekonstruktion nicht der Fall. Nach dem Stand der Technik existiert kein Lesegerät, also kein „Parameterspieler“, mit dessen Hilfe die in Parametern eines KNN präsentierten Werke sichtbar gemacht werden können. Es kommt deswegen auch nicht darauf an, ob tatsächlich schutzbegründende Eigenschaften der Werke in den numerischen Parametern abgelegt werden.

#### 4. Schlussfolgerungen

Alles in allem sind die schutzbegründenden Merkmale der zum Training genutzten Werke aus den vorgenannten Gründen damit nicht in den Parametern des trainierten KNN im urheberrechtlichen Sinne wiedererkennbar.<sup>745</sup> Dementsprechend werden die Werke beim Trainingsprozess nicht in den Parametern des KNN im Sinne des § 16 UrhG vervielfältigt.<sup>746</sup> Falls

---

743 So richtigerweise auch *Leistner*, GRUR 2024, 1665 (1669).

744 So jedoch *Dornis*, CR 2024, 765 Rn. 9, 11 ff.; *Sesing-Wagenpfeil*, ZGE 2024, 212 (229); *Kögel*, InTeR 2023, 179 (181).

745 Auf die mangelnde Erkennbarkeit stellt auch *Käde*, ZUM 2024, 174 (177) ab.

746 Tendenziell auch *Hofmann*, WRP 2024, 11 Rn. 12; zustimmend auch *Maamar*, ZUM 2023, 481 (483); *Lennartz/Kraetzig*, GRUR 2024, 1258 (1259); *Kraetzig*, GRUR 2022,

## B. Durchführung des Trainings auf Grundlage von Werken

nach dem Training des KNN von einem technischen Fachexperten eine Invertierung durchgeführt, in deren Folge schutzbegründende Elemente zum Training verwendeter Werke objektiv identifizierbar gemacht werden, wird erst mit dieser Invertierung eine neues Vervielfältigungsstück im Sinne des § 16 UrhG hergestellt. Erst die Invertierung stellt infolgedessen eine erlaubnisbedürftige urheberrechtlich relevante Nutzungshandlung dar.

Im Ergebnis ist die Konfiguration der Parameter des KNN, letzter Schritt des Trainingsprozesses, damit durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte urheberrechtlich freigestellt. Die für diesen Teilaспект des KI-Trainings im zweiten Teil der Untersuchung identifizierten Regelungsbedürfnisse werden durch das geltende Urheberrecht also erfüllt.

## V. Fazit zur Durchführung des Trainings auf Basis von Werken

Weil die vier Teilaaspkte des Trainingsvorgangs im Wesentlichen nicht in den Anwendungsbereich des urheberrechtlichen Vervielfältigungsrechts aus § 16 UrhG fallen, erfüllt das geltende Urheberrecht die für die Durchführung des KI-Trainings identifizierten Regelungsbedürfnisse weitestgehend. Etwas anderes gilt nur für die Normalisierung der urheberrechtlich geschützten Trainingsdaten. Sie ist zwar durch die Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erlaubnisfrei gestellt. Dieser Regelungsmechanismus erweist sich aus einer regulatorischen Perspektive aber nicht als zweckmäßig für die Freistellung der Normalisierungshandlung. Das ist de lege ferenda durch eine Anpassung der Reichweite des urheberrechtlichen Vervielfältigungsrechts aus § 16 UrhG zu lösen.

---

1707 Rn. 7; ebenso *Käde*, ZUM 2024, 174 (177 f., 182); *Käde/v. Maltzan*, InTeR 2020, 201 (207); *Kögel*, DSRITB 2023, 285 (288); *Schwarz/Söbbing*, RDi 2023, 415 Rn. 13; a. A. *Dornis*, CR 2024, 765 Rn. 6; *Sesing-Wagenpfeil*, ZGE 2024, 212 (242 f.); *v. Weller*, GRUR-Prax 2023, 516 Rn. 21; *Jager*, Artificial Creativity?, S. 331; *Kleinkopf*, Text- und Data-Mining, S. 166 f.; dazwischen *Pesch/Böhme*, GRUR 2023, 997 (1004 f.), nach denen eine Vervielfältigung in den Parametern grundsätzlich nicht, im Einzelfall und im frühen Trainingsstadium allerdings stattfinden kann; *Konertz/Schönhof*, WRP 2024, 534 Rn. 8, die außer im Ausnahmefall keine Vervielfältigung annehmen.

## C. Speicherung urheberrechtlich geschützter Inputdaten

Ein Freistellungsbedürfnis konnte im zweiten Teil der vorliegenden Untersuchung weiterhin für die Speicherung von urheberrechtlich geschützten Inputdaten ermittelt werden.<sup>747</sup> Diese können beispielsweise mittels Sensor-technik wie Kameras oder Mikrofonen oder durch Webcrawler erfasst, aber auch manuell in ein KI-System eingegeben werden. Hierzu gehört auch die Verwendung von geschützten Werken als Prompt eines generativen KI-Systems. Von der Speicherung urheberrechtlich geschützter Trainingsdaten unterscheidet sie sich dadurch, dass in der Anwendungsphase des KI-Systems – in dieser werden die Inputdaten verarbeitet – nur vorübergehende Speicherungen der Werke notwendig sind.

Die Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten sollte aus einer regulatorischen Perspektive im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte erfolgen. Damit das Freistellungsbedürfnis erfüllt wird, darf die Speicherung von Werken als Inputdaten also nicht in den Anwendungsbereich des urheberrechtlichen Vervielfältigungsrechts (§ 16 UrhG) fallen.

### I. Eingriff in das Vervielfältigungsrecht

Bei der Speicherung urheberrechtlich geschützter Inputdaten werden vorübergehende, digitale Kopien der betroffenen Werke hergestellt. Diese dienen im Anschluss als Berechnungsgrundlage des KI-Systems. Die Erfassung und Speicherung der Werke als Inputdaten fällt dabei nicht in den Anwendungsbereich des urheberrechtlichen Vervielfältigungsrechts aus § 16 UrhG, wenn trotz des digitalen Kopierens keine Vervielfältigungsstücke der Werke im Sinne des § 16 Abs. 1 UrhG hergestellt werden.

Wie bereits gezeigt werden konnte,<sup>748</sup> ist der Vervielfältigungsbegriff als Rechtsbegriff nicht rein technischer, sondern normativer Natur.<sup>749</sup> Seine

---

747 Dazu ausführlich siehe oben unter 2. Teil D.

748 Ausführlich zur Begründung und Ausgestaltung eines normativen, funktionsbezogenen Vervielfältigungsbegriffs siehe oben unter 3. Teil B. I. 1. a) bb).

749 Hugenholz, EIPR 2000, 482 (485); sich hieran anschließend Schönberger, ZGE 2018, 35 (53); a. A. wohl Heerma, in: Wandtke/Bullinger, Urheberrecht, UrhG § 16 Rn. 16; Chiou, JIPITEC 2019, 398 Rn. 6; außerdem Jager, Artificial Creativity?, S. 332.

### C. Speicherung urheberrechtlich geschützter Inputdaten

Bedeutung muss daher insbesondere auch unter Berücksichtigung des Zwecks des in § 16 UrhG umgesetzten Art. 2 InfoSoc-RL bestimmt werden. Hieraus folgt, dass nur solche Werkverwendungen auch erlaubnispflichtige Werknutzungen im Sinne von Vervielfältigungen sind, die die Verwertungsmöglichkeit der Urheber zumindest potenziell beeinträchtigen.<sup>750</sup> Für eine Beeinträchtigung der Verwertungsinteressen genügt, dass ein unmittelbares oder mittelbares Vermarktungspotential aus der Werkverwendung entsteht und damit Substitutionsgefahr für das verwendete Werk hervorgerufen wird.<sup>751</sup>

Im vorliegenden Fall spricht gegen eine Beeinträchtigung des Vermarktungspotentials der betroffenen Werke, dass die Werke nicht zum Genuss der schöpferischen Merkmale, sondern nur als Berechnungsgrundlage für ein KI-System und dementsprechend zur Erfüllung einer technischen Funktion abgespeichert werden. Je nachdem, welche Art von KI-System konkret eingesetzt wird, kann die Erfassung von Werken als Inputdaten auch beiläufig und zufällig sein. Außerdem müssen die Inputdaten eines KI-Systems mitunter aus einem konkreten Anwendungsszenario heraus erhoben und sofort verarbeitet werden können. Auch die Beziehung des jeweiligen Werks zu seiner Umwelt kann von Bedeutung sein. Lizenzverträge zur Nutzung von als Inputdaten erfassten Werken können demzufolge aus tatsächlichen Gründen in einigen Fällen nicht abgeschlossen werden.

Allerdings können generierte Inputdaten losgelöst von der Anwendung eines konkreten KI-Systems jedenfalls im Nachhinein dazu verwendet werden, um KI-Systeme zu trainieren. Für viele KI-Entwickler sind die von Systemnutzern generierten Inputdaten hierfür auch notwendig. Denn unter Umständen kann im Entwicklungsstadium des KI-Systems keine ausreichende Menge an Trainingsdaten hinreichender Diversität akquiriert werden. Die Verwendung von aus einem realen Anwendungsumfeld abgespeisten Inputdaten für die Optimierung oder das Training eines Systems

---

750 öst. OGH, Beschl. v. 23.02.2010 - 4 Ob 208/09 f, ZUM 2010, 629 (630); v. *Ungern-Sternberg*, in: Schricker/Loewenheim, Urheberrecht, UrhG § 15 Rn. 23; v. *Ungern-Sternberg*, GRUR 2012, 576 (580); *Schönberger*, ZGE 2018, 35 (53); *Antoine*, Verändernde Werknutzungen, S. 179 f.; in diese Richtung auch *Grisse/Kaiser*, ZUM 2021, 401 (404); grundsätzlich wie hier auch *Antoine*, ZGE 2023, 387 (398), wobei sie eine Einschränkung der funktionsbezogenen Auslegung auf die „Nutzungsfälle an den Rändern des urheberrechtlichen Ausschließlichkeitsrechts“ befürwortet, dazu sogleich. Zum Verhältnis zu den urheberpersönlichkeitsrechtlichen Bestimmungen siehe *Antoine*, ZGE 2023, 387 (401 ff.).

751 Siehe oben für weitere Details zur Schutzbereichsgrenze aus der funktionsbezogenen Auslegung des Vervielfältigungsbegriffs unter 3. Teil B. I. 1. a) bb).

ist auch verbreitet.<sup>752</sup> Daraus folgt, dass als Inputdaten generierte Werke zwar mitunter nicht als solche monetarisiert werden können. Dennoch besteht jedenfalls ein Vermarktungspotenzial der als Inputdaten generierten Werke zur Verwendung als Trainingsdaten. In dieser Form besteht auch Substitutionsgefahr für die Verwertungsmöglichkeiten des Urhebers. Alles in allem beeinträchtigt die Speicherung urheberrechtlich geschützter Inputdaten das Verwertungsinteresse betroffener Urheber damit zumindest potenziell. Hieraus folgt, dass nach einer funktionsbezogenen Auslegung des Vervielfältigungsbegriffs bei der Speicherung urheberrechtlich geschützter Inputdaten für KI Vervielfältigungsstücke erfassster Werke im Sinne des § 16 UrhG hergestellt werden. Damit findet, beispielsweise bei der Aufnahme von urheberrechtlich geschützten Werken der Baukunst durch die Kameras autonomer Fahrzeuge, ein Eingriff in das urheberrechtliche Ausschließlichkeitsrecht in Form des Vervielfältigungsrechts statt. Die Vervielfältigungen bei der Speicherung von Werken als Inputdaten sind demzufolge im Grundsatz erlaubnispflichtig. Diese Erlaubnis könnte zwar durch eine urheberrechtliche Schrankenbestimmung gewahrt werden. Aus alledem folgt allerdings bereits, dass das UrhG das für die Speicherung von urheberrechtlich geschützten Inputdaten identifizierte Regelungsbedürfnis nach einer Freistellung im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte in den §§ 15 ff. UrhG nicht erfüllt.

*II. Anwendung urheberrechtlicher Schrankenbestimmungen*

Nichtsdestotrotz ließe sich durch die Anwendung einer urheberrechtlichen Schrankenbestimmung zumindest das Ob des regulatorischen Freistellungsbedürfnisses erfüllen.

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

Solange die Werke nur als Berechnungsgrundlage des KI-Systems in dessen Anwendungsphase dienen sollen, ist eine vorübergehende Speicherung der Inhalte ausreichend. Diese könnte von der Schrankenbestimmung aus § 44a UrhG abgedeckt sein. Dafür muss die Vervielfältigung der Werke als

---

<sup>752</sup> Vgl. als Beispiele *Hoffmann/Vogt*, MMR-Aktuell 2020, 431466; *Shah*, AIwire v. 8.3.2023; *Stecklow/Cunningham/Jin*, Reuters v. 6.4.2023.

### *C. Speicherung urheberrechtlich geschützter Inputdaten*

Inputdaten vorübergehend und flüchtig oder begleitend im Sinne des § 44a UrhG sowie integraler und wesentlicher Teil eines technischen Verfahrens sein. Ihr alleiniger Zweck muss sich außerdem darin erschöpfen, eine Übertragung in einem Netz zwischen Dritten durch einen Vermittler (§ 44a Nr. 1 UrhG) oder eine rechtmäßige Nutzung (§ 44b Nr. 2 UrhG) zu ermöglichen. Den Vervielfältigungen darf dabei selbst keine eigene wirtschaftliche Bedeutung zukommen.

#### a) Vorübergehende Vervielfältigung

Vorübergehend ist eine Vervielfältigung dann, wenn sie nach der Zweckerfüllung innerhalb des technischen Verfahrens zur Löschung vorgesehen ist.<sup>753</sup> Zweck der Vervielfältigung der Werke ist, sie als Input eines KI-Systems zu verwenden. Die Vervielfältigungsstücke der erfassten Werke werden hierfür nach der Umwandlung in numerische Tensoren und spätestens nach der Eingabe dieser Tensoren in den Algorithmus nicht mehr benötigt. Denn der Algorithmus läuft im Weiteren eigenständig ab, eine Bezugnahme auf die Inputdaten ist nach der ersten Berechnungsstufe nicht mehr erforderlich. Insofern mit den gespeicherten Werken keine Sekundärzwecke wie beispielsweise das KI-Training verfolgt werden und sich der Datenlebenszyklus infolgedessen ausschließlich an der Verwendung der Werke als Inputdaten ausrichtet, ist daher davon auszugehen, dass die Vervielfältigungen nach der Erfüllung ihres Zwecks innerhalb des KI-Systems zur Löschung vorgesehen sind. Die Speicherung urheberrechtlich geschützter Inputdaten kann damit vorübergehend im Sinne von § 44b UrhG sein.<sup>754</sup>

#### b) Wesentlicher und integraler Bestandteil eines technischen Verfahrens

Ohne Inputdaten kann das KI-System keinen Output generieren, also seinen Anwendungszweck nicht erfüllen. Es besteht daher auch die notwendige Kausalbeziehung zwischen den Vervielfältigungshandlungen und dem

---

753 Dustmann, in: Fromm/Nordemann, Urheberrecht, UrhG § 44a Rn. 9; vgl. EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

754 Soll eine Sekundärverwertung der erfassten Werke stattfinden, richtet sich die urheberrechtliche Zulässigkeit der Werknutzung nach den Besonderheiten der jeweiligen Sekundärverwertungsart. Dies ist aber nicht Gegenstand der vorliegenden Untersuchung.

technischem Verfahren.<sup>755</sup> Die Vervielfältigungshandlungen werden zudem allein durch das KI-System beziehungsweise vorgelagerte und unmittelbar mit dem KI-System verbundene technische Verfahren ausgeführt.<sup>756</sup> Demzufolge ist die Speicherung urheberrechtlich geschützter Inputdaten auch integraler und wesentlicher Bestandteil eines technischen Verfahrens, der Anwendung eines KI-Systems.

c) Flüchtiger oder begleitender Charakter der Vervielfältigungen

Die Vervielfältigungen müssen darüber hinaus flüchtig oder begleitend sein. Flüchtig sind die Vervielfältigungen dann, wenn ihre Speicherdauer derart auf das für das technische Verfahren notwendige Maß beschränkt ist, dass die Vervielfältigungen nach Erfüllung ihres Zwecks innerhalb des technischen Verfahrens automatisch gelöscht werden und deshalb von einer besonders kurzweiligen Vervielfältigung gesprochen werden kann.<sup>757</sup>

Werden die Inputdaten nicht zu Sekundärzwecken verwendet, wovon im Rahmen der vorliegenden Untersuchung auszugehen ist, spricht nichts dagegen, dass die Vervielfältigungen nach ihrer Vorverarbeitung und Eingabe in das KI-System, also nach Erfüllung ihres Zwecks innerhalb des technischen Verfahrens, automatisch wieder gelöscht werden. Die Vervielfältigungen, die bei der Speicherung urheberrechtlich geschützter Inputdaten angefertigt werden, können demzufolge flüchtigen Charakters im Sinne des § 44a UrhG sein.

Solange mit den urheberrechtlich geschützten Inputdaten keine Sekundärzwecke verwirklicht werden sollen, dienen die Vervielfältigungen auch keinem anderen, eigenständigen Zweck als der Anwendung des KI-Systems. Sie sind unselbstständig in den Anwendungsvorgang des KI-Systems eingebunden. Die Speicherung urheberrechtlich geschützter Inputdaten ist daher auch begleitend zur Anwendung des KI-Systems im Sinne von § 44a UrhG.<sup>758</sup> Beide Tatbestandsalternativen sind daher erfüllt.

---

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

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

757 EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 64 – Infopaq; EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA; KG, Urt. v. 30.04.2004 - 5 U 98/02, GRUR-RR 2004, 228 (231) – Ausschnittsdienst.

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

d) Rechtmäßige Nutzung

Weil Zweck der Vervielfältigungen nicht die Übertragung von Inhalten zwischen Dritten ist, müssen die Vervielfältigungen für ein Eingreifen des Schrankentatbestands gemäß § 44a Nr. 2 UrhG ausschließlich dazu dienen, die rechtmäßige Nutzung der Werke zu ermöglichen. Rechtmäßig sind solche Nutzungen, die vom Rechtsinhaber zugelassen oder nicht vom urheberrechtlichen Ausschließlichkeitsrecht erfasst und Dritten damit gestattet sind.<sup>759</sup> Rechtmäßige Nutzung ist insbesondere der rezeptive Werkgenuss, der von vornherein nicht Teil der urheberrechtlichen Verwertungsrechte und für Nutzer damit frei möglich ist.<sup>760</sup> Der freie Werkgenuss kann auch in einer Informationsentnahme münden. Denn die im Werk enthaltenen semantischen Informationen genießen grundsätzlich ebenfalls keinen urheberrechtlichen Schutz.<sup>761</sup> Damit ist auch die Analyse eines Werkes, beispielsweise zu wissenschaftlichen Zwecken, als rechtmäßige Nutzung von § 44a Nr. 2 UrhG umfasst.<sup>762</sup>

Zweck der Vervielfältigungen, die bei der Speicherung urheberrechtlich geschützter Inputdaten stattfinden, ist die Ermöglichung der Anwendung von KI-Systemen. Urheberrechtlich geschützte Inputdaten werden dabei nach einer Umwandlung in numerische Darstellungsformen als Grundlage für Berechnungen innerhalb der KI genutzt. Das Berechnungsergebnis wird als Output ausgegeben. Werke als Berechnungsgrundlage für Algorithmen zu verwenden, ist eine besondere Form der Analyse der im Werk enthaltenen Informationen. Die Art der Verwendung ist demzufolge nicht dem urheberrechtlichen Ausschließlichkeitsrecht zugewiesen.<sup>763</sup> Das gilt für eine händische Berechnungen, aber auch für automatisierte Prozesse.<sup>764</sup> Schließlich kommt es für die Berechnung sogar mitunter nicht auf die schöpferischen Elemente des Werkes an.<sup>765</sup> Die algorithmische Verwendung von Werken als Inputdaten einer KI als solche ist, unabhängig von dabei angefertigten Vervielfältigungshandlungen, daher nicht vom urheberrechtlichen Ausschließlichkeitsrecht umfasst. Demzufolge dienen die

---

759 ErwGr. 33 InfoSoc-RL.

760 Dustmann, in: Fromm/Nordemann, Urheberrecht, UrhG § 44a Rn. 17; vgl. beispielhaft EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 171 – Football Association Premier League.

761 EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 44 – Infopaq II.

762 Vgl. ErwGr. 9 S. 2 DSM-RL

763 Wie hier zum KI-Training auch Hofmann, WRP 2024, II Rn. 14.

764 Vgl. ErwGr. 9 S. 2 DSM-RL, außerdem BT-Drs. 18/12329, S. 40.

765 Jager, Artificial Creativity?, S. 341.

Vervielfältigungen urheberrechtlich geschützter Werke zur Verwendung als Inputdaten eines KI-Systems ausschließlich einer rechtmäßigen Nutzung im Sinne von § 44a Nr. 2 UrhG.

e) Keine eigenständige wirtschaftliche Bedeutung

Darüber hinaus dürfen die Vervielfältigungen der Werke gemäß § 44a UrhG auch keine eigenständige wirtschaftliche Bedeutung haben. Das ist der Fall, wenn die Vervielfältigungshandlungen einen wirtschaftlichen Vorteil generieren, der über den aus der rechtmäßigen Nutzung resultierenden Vorteil hinausgeht.<sup>766</sup> Hiergegen spricht, wenn die Vervielfältigungshandlungen untrennbarere Bestandteile der rechtmäßigen Nutzung sind. Das ist beispielsweise dann der Fall, wenn der Nutzer auf die Vervielfältigungshandlungen schon deswegen keinen Einfluss nehmen kann, weil sie ihm nicht bewusst sind.<sup>767</sup>

aa) Eigene wirtschaftliche Bedeutung unmittelbar aus der Nutzung der Werke als Inputdaten

Der durchschnittliche Nutzer eines autonomen Fahrzeugs ist sich beispielsweise nicht darüber bewusst, dass das Fahrzeug mit Kamerasensoren am Wegesrand befindliche Werke der Baukunst vervielfältigt. Urheberrechtlich geschützte Inputdaten können also unter Umständen ohne Kenntnis des KI-Systemnutzers generiert werden. Sofern der Nutzer im Beispieldfall nicht den Einsatz des autonomen Fahrzeugs insgesamt unterlassen will, kann er auch keinen Einfluss darauf nehmen, ob urheberrechtlich geschützte Werke bei der Anwendung des KI-Systems vervielfältigt werden. In Fallkonstellationen wie dieser sprechen demzufolge Gründe dafür, dass die vorübergehenden Vervielfältigungshandlungen untrennbare Bestandteile der Anwendung des KI-Systems sind. Dann käme ihnen keine eigene wirtschaftliche Bedeutung zu.

Bei der Anwendung anderer Arten von KI-Systemen könnte der Umgang mit urheberrechtlich geschützten Inhalten für Nutzer allerdings sichtbarer

---

766 EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 175 – Football Association Premier League.

767 Vgl. EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 176 – Football Association Premier League.

### C. Speicherung urheberrechtlich geschützter Inputdaten

sein. Denkbar ist das insbesondere beim Einsatz von Webcrawlern zur Speicherung von Inputdaten. Diese laden aus frei zugänglichen Internetquellen mitunter große Mengen an Text- oder Bildinhalten herunter. Hierbei wird zumindest billigend in Kauf genommen, dass urheberrechtlich geschütztes Material mitvervielfältigt wird. Unbewusst sind die Vervielfältigungen in diesen Fallkonstellationen damit nicht. Nichtsdestotrotz ist die Speicherung urheberrechtlich geschützter Inputdaten unselbstständig in die Anwendung des KI-Systems integriert. Die Speicherung der Inputdaten ist für die Anwendung der KI notwendig. Die Daten können nach der Umwandlung in mathematische Tensoren auch wieder gelöscht werden. Die Vervielfältigungen sind demzufolge auch in dieser Fallkonstellation elementare Bestandteile der Anwendung der KI-Systeme, also der rechtmäßigen Nutzungen. Da es keinen etablierten Markt für Inputdaten gibt, spart sich der KI-Nutzer durch die Speicherung der Werke auch keine eigenen Aufwendungen. Grundsätzlich bringt die Speicherung der geschützten Inputdaten abgesehen von der Ermöglichung der KI-Anwendung weder für KI-Nutzer noch für Entwickler oder Anbieter unmittelbar einen über die rechtmäßige Nutzung hinausgehenden wirtschaftlichen Vorteil.

#### bb) Mittelbare eigene wirtschaftliche Bedeutung aus der Möglichkeit zur Sekundärverwertung von Inputdaten

Allerdings können die Inputdaten auch zu Sekundärzwecken verwendet werden. Hierzu lassen sich die Vervielfältigungsstücke ohne größeren technischen Aufwand dauerhaft abspeichern. Die Vervielfältigungsstücke können dann insbesondere zum Training von KI genutzt werden. Verbreitet ist beispielsweise der Einsatz der Inputdaten zur Optimierung des KI-Systems selbst.<sup>768</sup> Auch denkbar ist, dass die generierten Vervielfältigungsstücke zu Protokollierungs- und Beweissicherungszwecken archiviert werden. Solche Sekundärverwertungen finden häufig bei autonomen Fahrzeugen statt. Sie können auch wirtschaftlich relevant sein. Das gilt insbesondere für die Verwendung der Inputdaten als Trainingsmaterial.<sup>769</sup> Findet eine Sekundärnutzung der generierten Inputdaten statt, kommt den Vervielfältigungshandlungen infolgedessen eine aus der Sekundärnutzung der Werke

---

<sup>768</sup> Vgl. als Beispiele *Hoffmann/Vogt*, MMR-Aktuell 2020, 431466; *Shah*, AIwire v. 8.3.2023; *Stecklow/Cunningham/Jin*, Reuters v. 6.4.2023.

<sup>769</sup> Dazu siehe oben unter 2. Teil D. I. I. sowie zum Training I. Teil B. II. 3.

resultierende eigene wirtschaftliche Bedeutung zu. Sie geht über den wirtschaftlichen Vorteil aus der primären Anwendung des KI-Systems hinaus.

Fraglich ist, ob bereits die *Möglichkeit* der Sekundärverwertung der generierten Inputdaten und damit der Begründung einer eigenständigen wirtschaftlichen Bedeutung der Vervielfältigungen neben der rechtmäßigen Nutzung, also der Anwendung des KI-Systems, ausreicht, damit die Schrankenbestimmung aus § 44a UrhG nicht anwendbar ist.<sup>770</sup>

Dafür spricht zunächst, dass es für Urheber nur einen geringen Unterschied macht, ob der wirtschaftliche Vorteile aus der Vervielfältigung unmittelbar realisiert wird oder nur mittelbar realisierbar ist. Dass die in § 44a UrhG umgesetzte InfoSoc-RL (Art. 5 Abs. 1 InfoSoc-RL) ein hohes Schutzniveau für Urheber gewährleisten will,<sup>771</sup> spricht dafür, den Urhebern bereits Schutz vor einer mittelbaren, also drohenden Beeinträchtigung ihrer Verwertungsinteressen zu bieten. Das wird auch bei der funktionsbezogenen Auslegung des Vervielfältigungsrechts deutlich.<sup>772</sup>

Allerdings ist Art. 5 Abs. 1 InfoSoc-RL und damit auch § 44a UrhG richtlinienkonform so auszulegen, dass die praktische Wirksamkeit der Schrankenbestimmung erhalten bleibt.<sup>773</sup> Die Erfassung von potenziellen Weiterverwertungsmöglichkeiten unter dem Tatbestandsmerkmal der eigenen wirtschaftlichen Bedeutung im Sinne des § 44a UrhG würde dazu führen, dass eine Anwendung der Schrankenbestimmung in den allermeisten Fällen ausscheidet. Denn vorübergehende Vervielfältigungen können selbst aus dem Cache (flüchtiger Zwischenspeicher) eines Computers geborgen werden. Damit würde potenziell eine eigenständige Verwertung der Vervielfältigungen ermöglicht.<sup>774</sup> Wenn selbst diese Vervielfältigungen nicht mehr unter § 44a UrhG gefasst werden könnten, verbliebe kaum noch Anwendungsspielraum für den Erlaubnistatbestand. Die praktische Wirksamkeit der Schrankenbestimmung wäre damit erheblich beeinträchtigt. Darüber hinaus hatte der Richtliniengeber auch neue Medien als Anwendungsfall für die Schrankenbestimmung vor Augen.<sup>775</sup> Diese wären der weiten Lesart zur Folge aber zumeist ebenfalls nicht von § 44a UrhG umfasst.

---

770 Zu dieser Frage im Kontext des Streaming bereits *Stieper*, MMR 2012, 12.

771 ErwGr. ErwGr. 4, 9 InfoSoc-RL.

772 Siehe oben unter 3. Teil B. I. 1. a).

773 EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 163 – Football Association Premier League.

774 Dazu genauer *Stieper*, MMR 2012, 12 (16).

775 ErwGr. 31 InfoSoc-RL.

### *C. Speicherung urheberrechtlich geschützter Inputdaten*

Weiterhin sind potenzielle Verwertungsmöglichkeiten alternative Sachverhalte, deren Verwirklichung vom Inhaber des Vervielfältigungsstücks abhängt. Sie bei der Bewertung der eigenen wirtschaftlichen Bedeutung einer Vervielfältigung im Rahmen des § 44a UrhG einzubeziehen, bedeutete daher, die urheberrechtliche Zulässigkeit einer Werknutzung auf Basis eines hypothetischen Sachverhalts zu bewerten, dessen Verwirklichung nicht sicher ist. Dies kann keine Basis für eine rechtliche Beurteilung sein. Erst der Eingriff in den bestehenden technischen Ablauf, also die Absonderung der generierten, urheberrechtlich geschützten Inputdaten zum Zwecke einer tatsächlich stattfindenden Sekundärverwertung, hat eine neue, eigenständige Nutzungsmöglichkeit zur Folge.<sup>776</sup> Ohne diesen Eingriff ist die Verwertungsmöglichkeit zu abstrakt, als dass sie Beurteilungsgrundlage sein könnte.

Darüber hinaus entgehen den Urhebern auch erst durch eine tatsächliche Sekundärverwertung, beispielsweise zum Training von KI, Lizenzentnahmen.<sup>777</sup>

#### *cc) Schlussfolgerungen*

Alles in allem sprechen damit durchgreifende Erwägungen insbesondere aus der Auslegung des in § 44a UrhG umgesetzten Art. 5 Abs. 1 Info-Soc-RL für eine enge Auslegung des Tatbestandsmerkmals der eigenen wirtschaftlichen Bedeutung. Bloße entfernt-hypothetische, im technischen Prozess nicht angelegte Verwertungsmöglichkeiten genügen demzufolge nicht, damit Vervielfältigungen neben der rechtmäßigen Nutzung eine eigene wirtschaftliche Bedeutung zukommt.<sup>778</sup> Für die Speicherung von urheberrechtlich geschützten Inputdaten bedeutet das, dass die Vervielfältigungen erst dann eine eigene wirtschaftliche Bedeutung erlangen, wenn eine Sekundärverwertung der Inputdaten beispielsweise zum Training von KI durch eine Absonderung der Inhalte hierfür tatsächlich vorbereitet wird. Die eigentliche Speicherung urheberrechtlich geschützter Inputdaten

---

<sup>776</sup> Vgl. *Stieper*, MMR 2012, 12 (16).

<sup>777</sup> Vgl. zur Beeinträchtigung der Verwertungsinteressen durch die Verwendung als Trainingsdaten 2. Teil B. I. 1.

<sup>778</sup> *Eichelberger*, K&R 2012, 393 (397); *Stieper*, MMR 2012, 12 (16); anders wohl *Dreier*, in: *Dreier/Schulze*, Urheberrechtsgesetz, UrhG § 44a Rn. 9; *Loewenheim*, in: *Schrieker/Loewenheim*, Urheberrecht, UrhG § 44a Rn. 13; *Fangerow/Schulz*, GRUR 2010, 677 (680); tendenziell weiter wohl auch *Wiebe*, in: *Spindler/Schuster*, Recht der elektronischen Medien, UrhG § 44a Rn. 6.

hat neben der Anwendung des KI-Systems daher keine eigenständige wirtschaftliche Bedeutung im Sinne des § 44a UrhG.

f) Zwischenergebnis

Insgesamt sind die bei der Speicherung von Inputdaten angefertigten vorübergehenden Vervielfältigungen urheberrechtlich geschützter Werke grundsätzlich<sup>779</sup> von der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG gedeckt.<sup>780</sup>

g) Erfassung rechtswidrig verfügbarer Werke

Bei der typischerweise automatisiert ablaufenden Sammlung von Inputdaten insbesondere mit Hilfe automatisierter Webcrawler kann nicht zweifelsfrei zwischen rechtmäßig und unrechtmäßig zugänglich gemachten Werken unterschieden werden. Unklar ist vor diesem Hintergrund, ob die Freistellungswirkung des § 44a Nr. 2 UrhG auch dann eingreift, wenn – insbesondere mit Hilfe von Webcrawlers – rechtswidrig zugänglich gemachte Werke zur Verwendung als Inputdaten vorübergehend vervielfältigt werden. Das im zweiten Teil der Untersuchung identifizierte Freistellungsbedürfnis differenziert nicht zwischen diesen Werkgruppen. Vielmehr ist eine uneingeschränkte Freistellung der Speicherung von urheberrechtlich geschützten Inputdaten interessengerecht.

Auch wenn die vorübergehende Vervielfältigung von rechtswidrig öffentlich zugänglich gemachten Werken als Inputdaten für die Anwendung von KI-Systemen einer rechtmäßigen Nutzung im Sinne des § 44a Nr. 2 UrhG dient,<sup>781</sup> könnte die Freistellung dieser Vervielfältigungshandlungen nicht mit dem im Rahmen von § 44a UrhG zu berücksichtigenden Dreistufentest aus Art. 5 Abs. 5 InfoSoc-RL vereinbar sein. Das ist der Fall, wenn die vorübergehenden Vervielfältigungen eine Verringerung rechtmäßiger

---

<sup>779</sup> Zum möglichen Sonderfall, der Anwendung generativer KI-Systeme, siehe unten unter 3. Teil C. II. 4.

<sup>780</sup> So zum Web Scraping von Werken als Inputdaten für KI-Textgeneratoren im journalistischen Kontext auch *Gräfe/Kahl*, MMR 2021, 121 (124); die Möglichkeit sehen auch *Nägele/Apel*, in: *Kaulartz/Braegelmann*, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 7.1 (KI und Urheberrecht) Rn. 32.

<sup>781</sup> Dazu siehe unmittelbar oben.

### *C. Speicherung urheberrechtlich geschützter Inputdaten*

Transaktionen zur Folge haben, die die betroffenen Urheber unzumutbar beeinträchtigen.<sup>782</sup> Da es keinen etablierten Markt gibt, auf dem im erheblichen Umfang Transaktionen über Inputdaten getätigt werden können,<sup>783</sup> kommt eine Verringerung solcher rechtmäßiger Transaktionen aber bereits von vornherein kaum in Betracht. Darüber hinaus findet die Speicherung von Inputdaten in bestimmten Anwendungsszenarien automatisiert und in großem Umfang statt. In diesen Fällen ist es dem Datenverarbeiter nicht möglich, im Einzelnen die Rechtmäßigkeit der Zugänglichmachung der gespeicherten Werke zu überprüfen. Im Unterschied zu Fällen wie dem Linksetzen auf urheberrechtsverletzend öffentlich zugänglich gemachte Werke erfolgt auch keine weitere Verbreitung der rechtswidrigen Inhalte. Die Werknutzung beschränkt sich vielmehr auf eine technische Verarbeitung der Werke. Aus alldem folgt, dass der Dreistufentest gemäß Art. 5 Abs. 5 InfoSoc-RL im konkreten Fall nicht verletzt ist. Seine Anwendung führt demzufolge nicht zu einer einschränkenden Auslegung der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen, nach der nur rechtmäßig zugänglich gemachte Inhalte für die Speicherung von Werken als KI-Inputdaten verwendet werden dürfen. Auch urheberrechtswidrig öffentlich verfügbar gemachte Werke dürfen daher als Inputdaten für KI gemäß § 44a Nr. 2 UrhG vorübergehend vervielfältigt werden.<sup>784</sup>

## 2. Weitere Schrankenbestimmungen

Die Speicherung urheberrechtlich geschützter Inputdaten könnte daneben noch in den Anwendungsbereich weiterer Schrankenbestimmungen fallen. Hiervon sind aber keine umfangreicheren Freistellungswirkungen zu erwarten. Deswegen sollen die übrigen Schrankenbestimmungen nur übersichtsartig dargestellt werden.

### a) Schrankenbestimmung für Text und Data Mining (§ 44b UrhG)

Die Speicherung urheberrechtlich geschützter Inputdaten könnte insbesondere auch von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG erfasst sein.

---

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

783 Dazu bereits siehe oben unter 2. Teil D. I. I.

784 A.A. Käde, Kreative Maschinen und Urheberrecht, S. 84 f.

### aa) Anwendbarkeit der Schrankenbestimmung

Damit § 44b UrhG auch auf die Verarbeitung von urheberrechtlich geschützten Inputdaten Anwendung finden kann, muss auch die *Anwendung* einer KI unter den Text und Data Mining Begriff im Sinne des § 44b Abs. 1 UrhG subsumiert werden können. Text und Data Mining beschreibt gemäß § 44b Abs. 1 UrhG die automatisierte Analyse von einzelnen digitalen oder digitalisierten Werken, um „daraus“ Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen. Nach einer richtlinienkonformen Auslegung des Tatbestandsmerkmals genügt es hierbei, wenn die Informationen nicht unmittelbar *aus* dem analysierten Werk, sondern *mit dessen Hilfe* gewonnen werden. Notwendig, aber ausreichend ist demnach eine Kausalbeziehung zwischen der automatisierten Analyse eines Werkes und der Speicherung von Informationen.<sup>785</sup>

Bei der Verwendung von Werken als Inputdaten werden die Werke zunächst in numerische Tensoren umgewandelt. Dann können sie als Berechnungsgrundlage des Algorithmus verwendet werden. Dabei werden die einzelnen Bestandteile der Werke extrahiert und als Anknüpfungspunkte für die beim Training ermittelten Muster und Zusammenhänge verwendet. Auf ihrer Grundlage können statistische Vorhersagen getroffen werden. Sie sind im Output des KI-Systems repräsentiert. Die eingegebenen Werke werden also im Lichte der bereits ermittelten Muster und Zusammenhänge ausgewertet. In der Anwendungsphase eines KI-Systems findet damit eine automatisierte Analyse der Inputdaten im Sinne von § 44b UrhG statt.<sup>786</sup> Ziel der automatisierten Analyse der Inputdaten ist die Generierung von Output. Dieser kann beispielsweise in Form von Klassifikationen, Vorhersagen oder auch Kreativinhalten<sup>787</sup> ausgestaltet sein. Ziel des Analyseprozesses ist daher die Gewinnung einer Information durch den Output des KI-Systems.<sup>788</sup> Bezugspunkt der Informationsgewinnung ist demzufolge auch hier

---

785 Dazu siehe oben unter 3. Teil A. III. 2. i) ee) (3) bb).

786 Zum Begriff der Analyse siehe *Kleinkopf*, Text- und Data-Mining, S. 30. Zum Tatbestandsmerkmal der Automatisierung vgl. *Bomhard*, in: Ahlberg/Götting/Lauber-Rönsberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 44b Rn. 11.

787 Zur Anwendung der hier ermittelten Grundsätze auf generative KI-Systeme siehe unten 3. Teil C. II. 4. Zum Informationsgewinnungsbegriff und dessen Reichweite insbesondere im Kontext der Erzeugung fiktiver (Kreativ-)inhalte bereits oben unter 3. Teil A. III. 2. c).

788 So auch *Jager*, Artificial Creativity?, S. 351; a. A. *Schippan*, ZUM 2024, 670 (675), nach dem ein Output selbst dann keine gewonnene Information im Sinne der Bestimmung verkörpert, wenn er sich in einer Musteranalyse erschöpft.

### C. Speicherung urheberrechtlich geschützter Inputdaten

die Erzeugung des Outputs durch das KI-System. Die Output-Informationen werden dabei durch, also mit Hilfe der automatisierten Analyse der Inputdaten generiert. Infolgedessen besteht auch die notwendige Kausalbeziehung zwischen der Analyse der urheberrechtlich geschützten Inputdaten und der Informationsgewinnung. Mithin ist auch die Verarbeitung von Werken als Inputdaten eines KI-Systems Text und Data Mining im Sinne des § 44b Abs. 1 UrhG.

Die Speicherung, also Vervielfältigung urheberrechtlich geschützter Inputdaten ist infolgedessen grundsätzlich<sup>789</sup> auch von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 UrhG freigestellt.<sup>790</sup> Da die Inhalte nach Eingabe in das KI-System gelöscht werden können, ist hierfür die Löschungspflicht aus § 44b Abs. Abs. 2 S. 2 UrhG kein Hindernis. Auch bei der Vervielfältigung der Werke als Inputdaten ist jedoch ein gemäß § 44b Abs. 3 UrhG erklärter Nutzungsvertrag zu beachten.

#### bb) Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen an dritte Datenverarbeiter

Auch im Vorfeld der Speicherung urheberrechtlich geschützter Inputdaten können Zugangsinhaber Werkzugänge an Dritte weitergeben oder diesen Dritten Vorlagen zur weiteren Vervielfältigung von Werken zur Verfügung stellen, die sie auf Grundlage der Werkzugänge angefertigt haben. Die Dritten stellen im Anschluss hieran die eigenen Vervielfältigungsstücke der Werke her, die als Inputdaten verwendet werden sollen. Denkbar sind solche Fälle beispielsweise im Rahmen einer KI-gestützten Analyse von Werken im Auftrag eines Zugangsinhabers. Hierfür vermittelt dieser dem die Analyse ausführenden Datenverarbeiter Zugang zu den zu analysierenden Werken. Hiermit konkretisiert er den Analysegegenstand.

Auch in solchen Fällen stellt sich die Frage, inwieweit vom originären Zugangsinhaber begünstigte Dritte, die die Werke als Inputdaten vervielfältigen, rechtmäßigen Zugang zu den vom Zugangsinhaber vermittelten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben. Allgemein kann dabei

---

789 Zum möglichen Sonderfall, der Anwendung generativer KI-Systeme, siehe unten 3. Teil C. II. 4.

790 In diesem Sinne wohl auch *Jager, Artificial Creativity?*, S. 351; in einem engeren Kontext außerdem *Hofmann*, ZUM 2024, 166 (174), wobei in Einzelfällen eine Anwendung der Schrankenbestimmung abgelehnt wird; a. A. hingegen *Schippan*, ZUM 2024, 670 (675); wohl auch *Maamar*, ZUM 2023, 481 (487).

zwischen der Vermittlung des Zugangs selbst, beispielsweise durch die Weitergabe von Login-Daten zu einer geschützten Datenbank, und der Bereitstellung von seitens des Zugangsinhabers angefertigten Vervielfältigungsvorlagen unterschieden werden, auf deren Grundlage der Datenverarbeiter im Folgenden neue Werkstücke anfertigt. Die Vervielfältigung der urheberrechtlich geschützten Inputdaten kann dabei als Hilfstätigkeit für den Zugangsinhaber oder im Eigeninteressen des Datenverarbeiters ausgeführt werden.<sup>791</sup>

Da die sich hieraus ergebenden vier Fallkonstellationen (Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen der Werke sowie jeweils im Rahmen einer Hilfstätigkeit oder im Eigeninteresse des Datenverarbeiters) den Weitergabe- beziehungsweise Vermittlungsszenarien bei der Speicherung von Werken als Trainingsdaten entsprechen, können die hierfür gewonnenen Erkenntnisse auf die vorliegenden Fallkonstellationen übertragen werden. Demzufolge ist die Tatbestandsvoraussetzung der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG so auszulegen, dass die Zugangsvermittlung für eine Speicherung von urheberrechtlich geschützten Inputdaten im Interesse des Zugangsinhabers, also als Hilfstätigkeit für diesen, vom Begriff der rechtmäßigen Zugänglichkeit mitumfasst ist. Datenverarbeiter, denen zur Herstellung von Vervielfältigungsstücken im Interesse des Zugangsinhabers Zugang zu Werken vermittelt wurde, haben demzufolge rechtmäßigen Zugang zu diesen geschützten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG.<sup>792</sup> Sie müssen die aufgestellten Zugangsvoraussetzungen nicht in ihrer eigenen Person erfüllen. Insbesondere müssen die also keine eigenen Zugangsdaten erwerben und nicht selbst ein Zugangsabonnement abschließen.<sup>793</sup> Werden dem Datenverarbeiter statt Zugangsinformationen Vervielfältigungsvorlagen zur Verfügung gestellt, die durch den Zugangsinhaber auf Basis seiner Werkzugänge angefertigt wurden, hat der Datenverarbeiter bei der Herstellung seiner eigenen Vervielfältigungsstücke ebenfalls rechtmäßigen Zugang zu den Werken im Sinne des § 44b Abs. 2 S. 1 UrhG.<sup>794</sup> Auf das Verhältnis zwischen Zugangs- und Rechtsinhaber sowie den Ursprung der verwendeten Vervielfältigungsvorlagen kommt es nicht an.<sup>795</sup> Das Tatbestandsmerkmal der rechtmäßigen Zugänglichkeit ist für den Verarbeiter der Inputdaten bereits deswegen er-

---

791 Vgl. zu den Fallgruppen ausführlicher oben unter 3. Teil A. III. 2. f) bb).

792 Vgl. oben unter 3. Teil A. III. 2. f) bb) (1).

793 Vgl. oben unter 3. Teil A. III. 2. f) bb) (2).

794 Vgl. oben unter 3. Teil A. III. 2. f) bb) (2).

795 Vgl. oben unter 3. Teil A. III. 2. f) bb) (2).

### *C. Speicherung urheberrechtlich geschützter Inputdaten*

füllt, weil er selbst beim Zugriff auf die genutzten Vervielfältigungsvorlagen des Zugangsinhabers keine Zugangsschranken überwinden muss.

Anders sind jedoch die Fallkonstellationen zu beurteilen, in denen die Werke nicht im Interesse des Zugangsinhabers, sondern im Eigeninteresse des Datenverarbeiters vervielfältigt werden. Aufgrund einer Anwendung des Dreistufentests aus Art. 7 Abs. 2 DSM-RL i. V. m. Art. 5 Abs. 5 InfoSoc-RL können die Datenverarbeiter für ihre Vervielfältigungshandlungen bei der Speicherung der urheberrechtlich geschützten Inputdaten dann nicht von der Schrankenbestimmung aus § 44b UrhG profitieren, wenn für sie objektiv erkennbar ist, dass die Werke einer zugangsgeschützten Quelle entstammen.<sup>796</sup> Das ist stets der Fall, wenn einem Inputdatensammler der Zugang zu einer geschützten Quelle weitervermittelt wird, beispielsweise durch die Weitergabe von Login-Daten zu einer Datenbank.<sup>797</sup> Eine Freistellung der Vervielfältigungshandlungen des Datenverarbeiters über § 44b UrhG kann dann nicht stattfinden. Eine Ausnahme hiervon muss nur dann gemacht werden, wenn für den Urheber oder Rechtsinhaber bei der Einräumung des Zugangs zu den geschützten Inhalten erkennbar gewesen ist, dass der Zugangsinhaber beabsichtigt, den Werkzugang an dritte Datenverarbeiter zum Zwecke des Text und Data Minings weiterzugeben.<sup>798</sup> Werden dem im Eigeninteresse tätigen Datenverarbeiter durch den Zugangsinhaber Vervielfältigungsvorlagen für die Herstellung von Werkstücken zur Verfügung gestellt, kommt es für eine Freistellung seiner Vervielfältigungshandlungen nach den vorgenannten Grundsätzen auf die Umstände des konkreten Einzelfalls an.<sup>799</sup> Eine Subsumtion seiner Vervielfältigungshandlungen zur Speicherung von urheberrechtlich geschützten Inputdaten unter § 44b UrhG ist aber grundsätzlich möglich.

Alles in allem können von Zugangsinhabern begünstigte Inputdatenverarbeiter wie beispielsweise Betreiber KI-gestützter Analysesysteme rechtmäßigen Zugang zu den vermittelten Werken im Sinne des § 44b Abs. 2 S. 1 UrhG haben. Dabei kommt es darauf an, unter welche der vier herausgearbeiteten Fallgruppen die konkrete Sachverhaltskonstellation gefasst werden kann.

---

796 Vgl. oben unter 3. Teil A. III. 2. f) bb) (3).

797 Vgl. oben unter 3. Teil A. III. 2. f) bb) (iv).

798 Vgl. dazu ausführlich oben unter 3. Teil A. III. 2. f) bb) (3).

799 Vgl. oben unter 3. Teil A. III. 2. f) bb) (3).

- cc) Fernwirkung des Nutzungsvorbehalts bei Zugangsvermittlung oder Bereitstellung von Vervielfältigungsvorlagen an dritte Datenverarbeiter

Bei der durch Zugangsinhaber initiierten Vermittlung von Werkzugängen an dritte Datenverarbeiter, beispielsweise durch eine Weitergabe von Login-Daten zu einer Datenbank, oder der Bereitstellung von auf Grundlage der Zugänge angefertigten Werkstücken als Vorlagen für weitere Vervielfältigungshandlungen durch Dritte stellt sich im Übrigen die Frage, ob und inwieweit ein gegenüber dem Zugangsinhaber erklärter Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG auch die Freistellung der Vervielfältigungshandlungen des dritten Datenverarbeiters beendet. Technisch ist dabei zwischen der echten Zugangsvermittlung, also insbesondere durch die Weitergabe von Login-Daten zu einer Datenbank, und der Anfertigung von Vervielfältigungen durch den Zugangsinhaber zu unterscheiden, die dem Datenverarbeiter als Vorlage für eigene Vervielfältigungshandlungen zur Verfügung gestellt werden. In beiden Fällen kann, muss der Datenverarbeiter keine Kenntnis vom erklärten Nutzungsvorbehalt haben. Dies ist vom jeweiligen Einzelfall und der konkreten Umsetzung der Vorbehaltserklärung abhängig.

Maßgeblich für die Beurteilung der Fernwirkung des Nutzungsvorbehalts ist die wiedergabebezogene Wirkung der Vorbehaltserklärung.<sup>800</sup> Bei der Zurverfügungstellung von Vervielfältigungsvorlagen für den Datenverarbeiter führt sie dazu, dass gegenüber dem Zugangsinhaber erklärte Nutzungsvorhalte, beispielsweise in den Nutzungsbedingungen einer Webseite, keine Wirkung zulasten des Datenverarbeiters entfalten, wenn die Vorbehaltserklärung nicht mit den vom Zugangsinhaber angefertigten Vervielfältigungsvorlagen verbunden ist. Denn nur diese Wiedergaben der betreffenden Werke werden vom Datenverarbeiter verwendet. Deswegen kann der Datenverarbeiter trotz eines ursprünglich erklärten Nutzungsvorbehalts für seine Vervielfältigungshandlungen von der Freistellungswirkung des § 44b Abs. 2 UrhG profitieren.<sup>801</sup> Hiergegen spricht auch nicht die Anwendung des Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL.<sup>802</sup> Einzelfall-abhängig kann der Datenverarbeiter aber als Teilnehmer für die Urheberrechtsverletzung des Zugangsinhabers mitverantwortlich sein.

---

800 Dazu siehe oben unter 3. Teil A. III. 2. i) ff).

801 Dazu ausführlich siehe oben unter 3. Teil A. III. 2. i) ee) (3) hh) (10) (i).

802 Zur Anwendung des Dreistufentests oben ausführlich unter 3. Teil A. III. 2. i) ee) (3) hh) (10) (i).

### *C. Speicherung urheberrechtlich geschützter Inputdaten*

Wird dem Datenverarbeiter lediglich der Zugang zu den Werken vermittelt, die vervielfältigt werden, so greift er für seine Vervielfältigungs-handlungen auf die originären Werkstücke beziehungsweise öffentlichen Wiedergaben der Werke zurück. An diesen ist der Nutzungsvorbehalt erklärt worden. Dies führt dazu, dass auch der Datenverarbeiter bei der Vervielfältigung der vermittelten Inhalte die erklärten Nutzungsvorbehalte beachten muss. Ist eine wirksame Vorbehaltserklärung abgegeben, erlischt demzufolge auch für ihn die Schrankenwirkung des § 44b Abs. 2 UrhG.<sup>803</sup> Im Ergebnis muss sich der Datenverarbeiter bei der Vermittlung eines Zugangs zu Werken infolgedessen selbst vergewissern, dass beispielsweise in den Nutzungsbedingungen keine wirksamen Nutzungsvorbehalte erklärt worden sind.<sup>804</sup>

#### b) Anwendungsspezifische Schrankenbestimmungen

Abhängig von der konkreten Fallkonstellation könnte die Speicherung urheberrechtlich geschützter Inputdaten für die Anwendung von KI-Systemen zudem insbesondere noch von den Schrankenbestimmungen aus §§ 48 (öffentliche Reden), 53 (Privatkopie), 59 (Werke an öffentlich verfügbaren Plätzen), 61 (verwaiste Werke) oder 61d UrhG (nicht verfügbare Werke) erfasst sein. Da diese Schrankenbestimmungen aber keine umfassende, fallgruppenunabhängige Freistellung der Speicherung von Werken als Inputdaten zur Folge haben, werden die Regelungen hier nicht näher beleuchtet.

### 3. Verhältnis der Schrankenbestimmungen aus § 44a Nr. 2 UrhG und § 44b UrhG

Die Speicherung, also Vervielfältigung von Werken als Inputdaten für die Anwendung von KI-Systemen wird im Grundsatz sowohl von der Schrankenbestimmung für Text und Data Mining aus § 44b UrhG als auch von der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG urheberrechtlich freigestellt. Die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining kann jedoch mit Hilfe

---

803 Dazu ausführlich siehe oben unter 3. Teil A. III. 2. i) ee) (3) hh) (10) (ii).

804 Vgl. zur Abgabe von Vorbehaltserklärungen in Nutzungsbedingungen siehe 3. Teil A. III. 2. i) aa).

eines Nutzungsverbauchs im Sinne des § 44b Abs. 3 UrhG beendet werden. Fraglich ist vor diesem Hintergrund, ob die Erklärung des Nutzungsverbauchs auch Auswirkungen auf die Freistellung der Vervielfältigungshandlungen zur Verwendung von Werken als KI-Inputdaten über § 44a Nr. 2 UrhG hat.

Denkbar wäre, dass die Erklärung des Nutzungsverbauchs auf einer dinglichen Ebene auch die Schrankenwirkung des § 44a Nr. 2 UrhG beendet. Dafür spricht, dass der Nutzungsverbauch gegen die Vervielfältigung von Werken als Inputdaten praktisch leerlaufen würde, wenn eine Freistellung der im Rahmen des § 44b UrhG vorbehaltenen Nutzungshandlung auch über § 44a Nr. 2 UrhG erfolgen könnte. Allerdings beinhaltet ErwGr. 18 UAbs. 2 S. 3 DSM-RL eine Klarstellung, nach der sich die quas dingliche Beendigungswirkung des Nutzungsverbauchs nur auf die Schrankenbestimmung für Text und Data Mining erstreckt.<sup>805</sup> Ohnehin kann der Urheber die Nutzung seines Werkes dinglich nicht begrenzen, wenn die Nutzungshandlung zu keinem Zeitpunkt seiner Rechtsposition zugeordnet war.<sup>806</sup> Schrankenbestimmungen führen jedoch dazu, dass dem Urheber von vornherein nur ein beschränktes Recht gewährt wird.<sup>807</sup> Sie gestalten den Umfang des Urheberrechts von Anfang an aus.<sup>808</sup> Es findet keine nachträgliche Einschränkung einer zunächst umfassend gewährten Rechtsposition statt, wie es bei der Gewährung eines ausschließlichen Nutzungsrechts durch den Urheber der Fall ist. Weil die Schrankenbestimmungen deswegen nicht mit vertraglichen Nutzungsrechtseinräumungen des Urhebers vergleichbar sind, ist auch § 31 Abs. 1 S. 2 UrhG auf gesetzliche Erlaubnisse nicht anwendbar.<sup>809</sup> Die Regelung sieht eine inhaltliche Beschränkbarkeit vertraglicher Nutzungsrechte vor. Auch eine analoge Anwendung scheidet

---

805 Dazu bereits ausführlicher mit einer abweichenden Ansicht zum Verständnis des Erwägungsgrunds unter 3. Teil A. III. 2. i) dd).

806 *Gräbig*, Abdingbarkeit und vertragliche Beschränkungen urheberrechtlicher Schranken, S. 59; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 204.

807 *Zech*, in: *Leible/Ohly/Zech*, Wissen - Märkte - Geistiges Eigentum, S. 187 (192); *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 129 ff.

808 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 129 ff.; a. A. *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236.

809 *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 200 ff.; a. A. *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236 f.

aus, weil die Rechtsfragen mit Hilfe einer teleologischen Auslegung der urheberrechtlichen Schrankentatbestände beantworten werden können. Damit fehlt es an einer Regelungslücke.<sup>810</sup> Zudem geht das UrhG gemäß § 32 Abs. 1 S. 2 UrhG davon aus, dass eine rechtsgeschäftliche Lizenz im Regelfall mit der Gewährung einer angemessenen Vergütung einhergeht. Im Gegensatz dazu sind die meisten gesetzlichen Schrankenbestimmung vergütungsfrei ausgestaltet. Auch daran lässt sich erkennen, dass das Urheberrecht beiden Regelungskomplexen eine jeweils eigene Regelungskonzeption zugrunde legt. Das zeigt, dass zwischen den urheberrechtlichen Schrankentatbeständen und den Regelungen zu rechtsgeschäftlichen Nutzungsrechten grundsätzlich auch keine vergleichbare Interessenlage besteht. Im Einzelfall kann das mitunter anders beurteilt werden. Insbesondere gewahrt das Urheberrecht aber eine absolute, quasi-dingliche Rechtsposition.<sup>811</sup> Damit unterliegt es auch einem Typenzwang.<sup>812</sup> Eine dingliche Beendigung der Schrankenwirkung kraft Disposition des Urhebers ist mit diesem Typenzwang grundsätzlich nicht vereinbar. Alles in allem folgt hieraus, dass eine quasi-dinglich wirkende Ausgestaltung des urheberrechtlichen Ausschließlichkeitsrechts durch Vertrag oder einseitige Erklärung des Rechtshabers insbesondere im Bereich der Schrankenbestimmungen grundsätzlich nicht möglich ist.<sup>813</sup> Dieser Rechtsgedanke spiegelt sich auch in § 137 S. 1 BGB wider.<sup>814</sup>

Eine Ausnahme hiervon kann nur dann gemacht werden, wenn die quasi-dingliche Rechtsposition im UrhG ausdrücklich zugunsten einer fakultativen Dispositionsbefugnis geöffnet wird. Dies ist beispielsweise in

---

810 A. A. aber *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236 f.

811 *Loewenheim*, in: Schricker/Loewenheim, Urheberrecht, Einleitung zum UrhG Rn. 25; *Wolf/Wellenhofer*, Sachenrecht § 2 Rn. 14.

812 *Stieper*, in: Schricker/Loewenheim, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 56; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 204; *Hofmann*, UFITA 2014, 381 (388).

813 *Stieper*, in: Schricker/Loewenheim, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 56; *Lüft*, in: *Wandtke/Bullinger*, Urheberrecht, Vor UrhG §§ 44a ff. Rn. 3; *Zech*, in: *Leible/Ohly/Zech*, Wissen - Märkte - Geistiges Eigentum, S. 187 (192); *Gräbig*, Abdingbarkeit und vertragliche Beschränkungen urheberrechtlicher Schranken, S. 59; *Stieper*, Rechtfertigung, Rechtsnatur und Disponibilität der Schranken des Urheberrechts, S. 204 f.; sich anschließend hieran *Hofmann*, UFITA 2014, 381 (388); vgl. zum allgemeinen Sachenrecht und dem Zypenzwang *Wolf/Wellenhofer*, Sachenrecht § 3 Rn. 3; a. A. *Zurth*, Rechtsgeschäftliche und gesetzliche Nutzungsrechte im Urheberrecht, S. 236 f.

814 *Hofmann*, UFITA 2014, 381 (388 f.).

§ 44b Abs. 3 UrhG und § 49 Abs. 1 S. 1 UrhG der Fall. Eine entsprechende Öffnungsklausel sieht § 44a UrhG jedoch nicht vor. Demzufolge kann die Freistellungswirkung der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen nicht fakultativ durch den Rechtsinhaber eingeschränkt werden. Ein gemäß § 44b Abs. 3 UrhG erklärter Nutzungsvorbehalt beendet demzufolge nicht quasi-dinglich die Wirksamkeit der Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG. Vielmehr sind die Vervielfältigungshandlungen der Inputdatensammler auch dann gemäß § 44a Nr. 2 UrhG urheberrechtlich freigestellt, wenn ein wirksamer Nutzungsvertrag im Sinne des § 44b Abs. 3 UrhG erklärt worden ist.

#### 4. Anwendung der Schrankenbestimmungen auf generative KI-Systeme

Fraglich ist jedoch, ob die herausgearbeiteten Ergebnisse zur Anwendung insbesondere der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen (§ 44a Nr. 2 UrhG) sowie der Schrankenbestimmung für Text und Data Mining (§ 44b UrhG) auch auf die Speicherung von Werken als Inputdaten für generative KI-Systeme übertragen werden können. Hierbei handelt es sich um KI-Systeme, die Kreativinhalte wie Texte oder Bilder erzeugen.

Einer parallelen Behandlung der Speicherung von Werken als Inputdaten für generative KI-Systeme würde entgegenstehen, wenn technische Unterschiede in der Verarbeitung der Inputdaten, die spezifische Art des von generativen KI-Systemen erzeugten Outputs oder eine Anwendung des urheberrechtlichen Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL eine abweichende Bewertung der vorgenommenen Vervielfältigungshandlungen notwendig machen.

##### a) Technische Perspektive auf die Verarbeitung als Inputdaten

Aus technischer Perspektive werden auch Inputdaten generativer KI-Systeme als Berechnungsgrundlage eines Algorithmus verwendet. Er besteht auch bei generativen KI-Systemen aus mathematischen Funktionen, die aus den Inputdaten abgeleitete numerische Aktivierungswerte verarbeiten. Hierfür werden die Inputdaten ebenfalls vorverarbeitet und insbesondere in numerische Tensoren umgewandelt.

### *C. Speicherung urheberrechtlich geschützter Inputdaten*

Aus technischer Perspektive besteht demzufolge kein Unterschied zwischen der Verarbeitung von Werken als Inputdaten sonstiger sowie generativer KI-Systeme. Bei der Verwendung von Werken als Inputdaten generativer KI-Systeme werden die Werke daher insbesondere auch rechtmäßig im Sinne des § 44a Nr. 2 UrhG als Berechnungsgrundlage genutzt. Die technische Perspektive auf den Verarbeitungsprozess auf Eingabeseite rechtfertigt daher keine abweichende Beurteilung.

#### b) Erzeugung von Kreativinhalten im Output

Die Erzeugung gegebenenfalls fiktiver Kreativinhalte im Output generativer KI-Systeme könnte zum einen Auswirkungen darauf haben, ob eine Informationsgewinnung im Sinne von § 44b Abs. 1 UrhG stattfindet. Dies ist für die Subsumtion der KI-Anwendung unter den Text und Data Mining-Begriff der Schrankenbestimmung notwendig. Bereits an anderer Stelle konnte allerdings gezeigt werden, dass § 44b Abs. 1 UrhG nicht nach der Art der erzeugten Informationen unterscheidet. Insbesondere auch die Generierung von Kreativinhalten wie Bildern oder Texten ist als Informationsgewinnung erfasst. Auch eine Fiktion der gewonnenen Informationen ist unschädlich.<sup>815</sup> Die spezielle Art des Outputs generativer KI-Systeme hat damit keinen Einfluss darauf, dass die tatbestandlich notwendige Informationsgewinnung im Sinne von § 44b Abs. 1 UrhG stattfindet.

Zum anderen ist aber auch fraglich, ob die Erzeugung möglicherweise urheberrechtsverletzenden Outputs durch das generative KI-System schon die Freistellung der Vervielfältigung von Werken als Inputdaten über die Schrankenbestimmungen aus §§ 44a, 44b UrhG beenden muss.

Hiergegen spricht allerdings, dass das Urheberrecht als techniksensibles Recht traditionell zwischen verschiedenen technischen Verarbeitungsschritten differenziert und diese jeweils einer konkreten, individuellen Beobachtung unterzieht. Aus diesem Grunde beurteilt sich beispielsweise die Vervielfältigung im Arbeitsspeicher eines Computers eigenständig und bedarf der Anwendung einer Schrankenbestimmung (§ 44a Nr. 2 UrhG), auch wenn mit Hilfe dieser Vervielfältigung nur der private, urheberrechtlich freie Werkgenuss bezweckt wird. In gleicher Weise bedarf die Übertragung in einem Netz zwischen Dritten durch einen Vermittler einer gesonderten Schrankenbestimmung (§ 44a Nr. 1 UrhG), auch wenn der soziale Sinnge-

---

<sup>815</sup> Dazu bereits oben unter 3. Teil A. III. 2. c).

halt der Fallkonstellation durch die Nutzung des Übermittlungsdienstes seitens eines Endnutzers und dessen urheberrechtskonformen oder rechtsverletzenden Verhaltens bestimmt wird. Die Generierung von möglicherweise urheberrechtsverletzenden Outputs ist ebenfalls eine eigene Vervielfältigungshandlung.<sup>816</sup> Sie ist daher im Lichte der dargestellten Regelungskonzeption ebenfalls eigenständig zu beurteilen.

§ 44a Nr. 1 UrhG, der die Werknutzung des Vermittlers unabhängig davon privilegiert, ob die Nutzer seines Übertragungsdienstes mitunter eine Urheberrechtsverletzung begangen haben, bringt darüber hinaus zum Ausdruck, dass die Rechtmäßigkeit einer Vorbereitungshandlung nicht zwingend das Schicksal der nachgelagerten Nutzung teilt. Ein entsprechender Mechanismus besteht im Urheberrecht nicht. Das ist auch bei der Anwendung generativer KI-Systeme und der urheberrechtlichen Bewertung einer Speicherung von Werken als Inputdaten – einer Vorbereitungshandlung – zu berücksichtigen. Die Rechtmäßigkeit der Speicherung von Werken als Inputdaten ist damit von der Rechtmäßigkeit der späteren Output-Generierung zu trennen. Umgekehrt bedeutet dies aber auch, dass ein urheberrechtsverletzender und sanktionsfähiger Output auch dann erzeugt werden kann, wenn die Verarbeitung des Werkes auf Inputebene rechtmäßig gewesen ist. Die Erzeugung des Outputs ist als eigenständige Nutzungshandlung eigenständig zu beurteilen.

Schließlich ist die Erzeugung konkreter Urheberrechtsverletzungen durch generative KI-Systeme einzelfallabhängig, sodass die urheberrechtliche Bewertung ebenfalls auf die jeweiligen Einzelfälle bezogen sein sollte.<sup>817</sup> Beispielsweise ist die Erzeugung von Abstracts mittels generativer KI-Systeme in gewissen Grenzen urheberrechtlich frei möglich, wenn im Abstract keine schutzbegründenden Elemente des Werks identifizierbar sind. Auch die Tatsache, dass im Output generativer KI-Systeme Kreativinhalte und mitunter vereinzelt sogar Urheberrechtsverletzungen erzeugt werden, verlangt deswegen keine abweichende Beurteilung.

c) Dreistufentest

Klärungsbedürftig bleibt damit, ob die Anwendung des Dreistufentests aus Art. 5 Abs. 5 InfoSoc-RL – gegebenenfalls in Verbindung mit Art. 7 Abs. 2 DSM-RL – einer Anwendung der einschlägigen Schrankenbestimmungen

---

816 Dazu bereits ausführlicher oben unter 3. Teil B. IV. 3.

817 Vgl. dazu bereits oben unter 2. Teil B. I. 2. d) aa).

### C. Speicherung urheberrechtlich geschützter Inputdaten

auf die Speicherung von Werken als Inputdaten *generativer* KI-Systeme entgegensteht. Danach darf die Schrankenbestimmung nur in bestimmten Sonderfällen angewandt werden, in denen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt und die berechtigten Interessen des Rechteinhabers nicht ungebührlich verletzt werden. Letztendlich kommt es hierbei auf eine umfassende Abwägung von Urheber- und Nutzerinteressen an.<sup>818</sup>

Im Rahmen dieser ist im Kontext der Speicherung von Werken als Inputdaten generativer KI-Systeme insbesondere zu berücksichtigen, dass im Output der KI-Systeme möglicherweise Urheberrechtsverletzungen erzeugt werden. Zudem besteht auch unabhängig von konkreten Rechtsverletzungen die Gefahr, dass Konkurrenzprodukte zu den als Inputdaten verwendeten Werken erstellt werden.

Beide Aspekte sind jedoch schon umfassend bei einer Abwägung von Urheber- und Nutzerinteressen im Kontext der Freistellung des Web Scrapings von Werken als Trainingsdaten berücksichtigt worden.<sup>819</sup> Sie konnten entkräftet werden und haben im Ergebnis nicht zu einem Überwiegen der Urheberinteressen geführt. Aus diesem Grund steht der Dreistufentest auch keiner Freistellung des Web Scrapings von Werken als Trainingsdaten entgegen.<sup>820</sup> Die Interessenabwägung in der vorliegenden Fallkonstellation ist vergleichbar, sodass auch hier aus den oben genannten Gründen kein Überwiegen der Urheberinteressen festgestellt werden kann. Es kann deswegen angenommen werden, dass der Dreistufentest auch bei der Speicherung von Werken als Inputdaten für generative KI-Systeme nicht verletzt ist. Auch das rechtfertigt daher keine abweichende Beurteilung.

#### d) Schlussfolgerungen

Alles in allem sind keine Gründe dafür ersichtlich, die herausgearbeiteten Ergebnisse nicht auf die Speicherung von Werken als Inputdaten für den Einsatz generativer KI-Systeme anzuwenden. Auch diese Verwendungsform ist deswegen – sofern sie vorübergehend erfolgt und die weiteren Tatbestandsvoraussetzungen erfüllt sind – ebenfalls von den Schrankenbestimmungen aus § 44a Nr. 2 UrhG sowie § 44b UrhG freigestellt.

---

818 Vgl. zur methodischen Herleitung oben unter 3. Teil A. III. 2. d).

819 Siehe 2. Teil B. I. 2. d) aa)

820 Vgl. oben unter 3. Teil A. III. 2. d).

#### IV. Zwischenfazit zur Speicherung urheberrechtlich geschützter Inputdaten

Alles in allem werden bei der Speicherung urheberrechtlich geschützter Inputdaten Vervielfältigungsstücke von Werken hergestellt. Damit findet bei der Speicherung von Werken als Inputdaten für KI ein Eingriff in das urheberrechtliche Vervielfältigungsrecht aus § 16 UrhG statt. Das im zweiten Teil der Untersuchung identifizierte Regelungsbedürfnis nach einer Freistellung durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte wird de lege lata demzufolge nicht erfüllt. De lege ferenda ist dies durch eine Einschränkung des urheberrechtlichen Vervielfältigungsrechts in § 16 UrhG zu lösen.

Nichtsdestotrotz ist zumindest das Ob des Freistellunginteresses im geltenden Urheberrecht durch die Anwendbarkeit von Schrankenbestimmungen abgebildet. Denn die Speicherung urheberrechtlich geschützter Inputdaten beispielsweise durch Kamerasensoren oder Webcrawler ist als vorübergehende Vervielfältigungshandlung von § 44a UrhG erfasst, solange die Inhalte nicht zur Verwirklichung von Sekundärzwecken aus der technischen Umgebung des KI-Systems extrahiert werden. Da auch die Verwendung von Werken als Inputdaten in der Anwendungsphase eines KI-Systems unter den Text und Data Mining-Begriff aus § 44b Abs. 1 UrhG gefasst werden kann, fallen die Vervielfältigungshandlungen darüber hinaus in den Anwendungsbereich der Schrankenbestimmung aus § 44b UrhG. Das gilt allerdings nur, solange kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt worden ist. Die Freistellungswirkungen der Schrankenbestimmungen aus §§ 44a, 44b UrhG greifen dabei unabhängig davon ein, welche Art von KI-System durch die Verwendung von Werken als Inputdaten konkret eingesetzt werden soll. Auch die Anwendung generativer KI kann auf Inputseite grundsätzlich unter die Schrankentatbestände gefasst werden. Der konkret generierte Output ist hiervon allerdings zu trennen.

## D. Vorverarbeitung von urheberrechtlich geschützten Inputdaten durch Normalisierung und Feature Encoding

Auch Inputdaten müssen in der Regel vorverarbeitet, also insbesondere normalisiert und in ein numerisches Format umgewandelt werden, bevor

## *D. Vorverarbeitung der Inputdaten (Normalisierung und Feature Encoding)*

sie als Berechnungsgrundlage eines KI-Systems verwendet werden können.<sup>821</sup> Wie auch die Vorverarbeitung für das Training von KI sollten auch diese Vorverarbeitungsschritte aus regulatorischen Gesichtspunkten urheberrechtlich freigestellt sein. Dabei ist zweckmäßig, die Freistellung im geltenden Urheberrecht durch eine unmittelbar begrenzende Ausgestaltung der Verwertungsrechte abzubilden.<sup>822</sup>

### *I. Normalisierung der Inputdaten*

Damit eine Freistellung der Vorverarbeitung entsprechend den identifizierten Regelungsbedürfnissen erfolgen kann, darf die Normalisierung der Inhalte nicht das urheberrechtliche Ausschließlichkeitsrecht beeinträchtigen.

#### 1. Eingriff in das urheberrechtliche Verwertungsrecht

Da die Normalisierungsverfahren von Inputdaten einerseits und Trainingsdaten andererseits dem gleichen Zweck dienen sowie identische Techniken verwendet werden, besteht zwischen ihnen kein wesensmäßiger Unterschied. Mit der Normalisierung urheberrechtlich geschützter Trainingsdaten kann ein Datenverarbeiter in das Ausschließlichkeitsrecht der betroffenen Urheber in Form des Vervielfältigungsrechts aus § 16 UrhG eingreifen.<sup>823</sup> Im Detail hängt das von den angewendeten Normalisierungsschritten ab. Solange schutzbegründende Elemente der verwendeten Werke in den normalisierten Erzeugnissen wiedererkennbar bleiben, haben hierauf auch Veränderungen des Werke im Normalisierungsverfahren keinen Einfluss. Vervielfältigungen im Sinne des § 16 UrhG, die im Zuge der Normalisierung gegebenenfalls hergestellt werden, sind daher erlaubnispflichtig. Diese Ergebnisse können auch auf die Normalisierung von Inputdaten übertragen werden. Das hat zur Folge, dass das urheberrechtliche Regelsregime auch für die Normalisierung von Inputdaten den identifizierten Freistellungsinteressen nicht gerecht wird.

---

821 Zur Vorverarbeitung technisch ausführlicher siehe oben unter I. Teil B. II. 2.

822 Zur Herleitung siehe oben unter 2. Teil E.

823 Dazu ausführlich siehe oben unter 3. Teil B. I. 1. a).

## 2. Anwendung urheberrechtlicher Schrankenbestimmungen

Sofern bei der Normalisierung der als Inputdaten verwendeten Werke ein Eingriff in das urheberrechtliche Vervielfältigungsrecht stattfindet, könnte dieser jedoch von einer urheberrechtlichen Schrankenbestimmung freigestellt werden.

### a) Schrankenbestimmung für vorübergehende Vervielfältigungs-handlungen (§ 44a UrhG)

Denkbar wäre zum einen, dass die Vervielfältigung von Werken im Rahmen der Normalisierung von Inputdaten durch die Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen gemäß § 44a UrhG freigestellt wird. Im Rahmen der vorliegenden Untersuchung wird dabei davon ausgegangen, dass keine Sekundärverwertung der normalisierten Werke stattfindet, die mitunter auch eine dauerhafte Speicherung der Inhalte notwendig macht.

#### aa) Grundvoraussetzungen des Schrankentatbestands

Wie auch bei der Speicherung der urheberrechtlich geschützten Inputdaten ist auch die Funktion der normalisierten Vervielfältigungsstücke nach Beendigung des Berechnungsvorgangs innerhalb des KI-Systems erfüllt. Nach Ablauf der konkreten KI-Anwendung können sie deswegen – von einer möglichen Sekundärverwertung der Inhalte abgesehen – gelöscht werden. Auch die bei der Normalisierung der Inputdaten möglicherweise angefertigten Vervielfältigungsstücke können daher vorübergehender Natur im Sinne des § 44a UrhG sein.<sup>824</sup>

Zwar kann der Algorithmus aus einer technischen Perspektive darüber hinaus auch auf der Grundlage nicht-normalisierter Inputdaten in Gang gesetzt werden. Dies kann aber zu einer Verringerung der Ausgabegenaugkeit

---

<sup>824</sup> Vgl. zur Voraussetzung *Dustmann*, in: Fromm/Nordemann, Urheberrecht, UrhG § 44a Rn. 9; außerdem EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA.

des KI-Systems führen.<sup>825</sup> Eine Normalisierung der Daten ist daher für eine effiziente und einwandfreie Funktion des KI-Systems notwendig. Damit sind auch die bei der Normalisierung von Inputdaten angefertigten Vervielfältigungsstücke integraler und wesentlicher Bestandteil der Anwendung eines KI-Systems.<sup>826</sup> Da normalisierte Vervielfältigungsstücke grundsätzlich auch automatisiert nach Ablauf des Algorithmus gelöscht werden können und neben der Anwendung des KI-Systems keinen eigenständigen Zweck erfüllen, sind sie darüber hinaus auch sowohl flüchtiger als auch begleitender Natur.<sup>827</sup> Schließlich ermöglicht das Normalisierungsverfahren mit der informationsbezogenen Analyse der in der Anwendungsphase der KI als Input vorgesehenen Werke auch eine rechtmäßige Nutzung im Sinne des § 44a Nr. 2 UrhG.<sup>828</sup>

bb) Eigene wirtschaftliche Bedeutung der Vervielfältigungen

Grundsätzlich kommt den Vervielfältigungshandlungen bei der Normalisierung der Inputdaten auch keine eigenständige wirtschaftliche Bedeutung zu, die einer Anwendung des § 44a UrhG entgegensteht.<sup>829</sup> Dass die Möglichkeit besteht, die normalisierten Inputdaten potenziell auch zu Sekundärzwecken wie insbesondere dem Training von KI zu verwenden, führt nicht zu einer eigenständigen wirtschaftlichen Bedeutung der Vervielfältigungshandlungen, solange keine Absonderung aus dem technischen Verfahren zur Verwirklichung von Sekundärverwertung stattfindet.<sup>830</sup> Die bei der Normalisierung der Werke vorgenommenen Änderungen sind vielmehr außerdem rein technisch bedingt. Außerhalb der Anwendung des KI-Systems erlangen sie keine eigenständige Bedeutung. Auch die im Rah-

---

825 Werden die eingegebenen Zahlenwerte zu groß oder klein, können sie nicht mehr in unterschiedungsfähige Aktivierungswerte der einzelnen Neuronen übersetzt werden, siehe Paß/Hecker, Künstliche Intelligenz, S. 102.

826 Vgl. zur Voraussetzung EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 36 – PRCA/NLA.

827 Vgl. zur Flüchtigkeit EuGH, Urt. v. 16.07.2009 - C-5/08, GRUR 2009, 1041 Rn. 64 – Infopaq; EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 28 – PRCA/NLA; KG, Urt. v. 30.04.2004 - 5 U 98/02, GRUR-RR 2004, 228 (231) – Ausschnittsdienst. Zum begleitenden Charakter einer Vervielfältigung vgl. EuGH, Urt. v. 05.06.2014 - C-360/13, GRUR 2014, 654 Rn. 43 – PRCA/NLA.

828 Dazu ausführlich oben unter 3. Teil C. II. 1. d).

829 Dazu ausführlich oben unter 3. Teil C. II. 1. e).

830 Dazu bereits ebenfalls siehe oben unter 3. Teil C. II. 1. e)

men des Normalisierungsverfahrens stattfindenden Änderungen an den digitalen Vervielfältigungsstücken führen demzufolge nicht dazu, dass die geänderten Vervielfältigungsstücke als solche *neben* der rechtmäßigen Nutzung, also der Anwendung des KI-Systems, verwertet werden sollen.<sup>831</sup> Eine eigene wirtschaftliche Bedeutung der Vervielfältigungshandlungen folgt deswegen auch nicht aus der Veränderung der betreffenden Werke im Rahmen der Normalisierung. Alles in allem ist deswegen nicht davon auszugehen, dass den Vervielfältigungshandlungen bei der Normalisierung urheberrechtlich geschützter Inputdaten neben Anwendung der KI eine eigene wirtschaftliche Bedeutung zukommt.<sup>832</sup>

cc) Schlussfolgerungen zur grundsätzlichen Anwendbarkeit von § 44a Nr. 2 UrhG

Alles in allem erfüllen die Vervielfältigungshandlungen, die bei der Normalisierung von urheberrechtlich geschützten Inputdaten für die Anwendung von KI stattfinden können, alle Voraussetzungen des § 44a Nr. 2 UrhG. Sie sind daher grundsätzlich als vorübergehende Vervielfältigungshandlung von der Schrankenbestimmung aus § 44a Nr. 2 UrhG freigestellt. Eine Sekundärverwertung der als Inputdaten normalisierten Werke, die mitunter auch eine dauerhafte Speicherung notwendig macht, ist hier aber nicht berücksichtigt.

dd) Verändernde Vervielfältigungen und das Änderungsverbot (§ 62 UrhG)

Bei der Normalisierung der Inputdaten können weiterhin veränderte Vervielfältigungsstücke hergestellt werden. Wie auch bei der Vorverarbeitung der zum Training vervielfältigten Werke ist auch hier fraglich, ob diese Änderungen vor dem Hintergrund des in § 62 UrhG verankerten Änderungsverbots durchgeführt werden können. Entscheidend ist hierfür aufgrund

---

831 Davon, dass eine Veränderung an einem Werk automatisch zu einer eigenen wirtschaftlichen Bedeutung der Vervielfältigung führt, geht aber in unzutreffender Weise der EuGH aus, siehe EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 53 – Infopaq II. Im Anschluss hieran auch *Dregelies*, GRUR 2024, 1484 (1485).

832 Zu den Anforderungen vgl. EuGH, Urt. v. 04.10.2011 - C-403/08, 429/08, GRUR 2012, 156 Rn. 175 – Football Association Premier League. Kritisch allerdings *Maamar*, ZUM 2023, 481 (487).

des in § 62 Abs. 1 S. 2 UrhG angelegten Verweises auf § 39 UrhG, ob die Änderungen zur Erfüllung des jeweiligen Schrankenzwecks erforderlich und zumutbar sind.<sup>833</sup> Darüber hinaus muss in richtlinienkonformer Auslegung aber auch berücksichtigt werden, ob die der Schrankenregelung zugrundeliegende Richtlinienbestimmung verändernde Vervielfältigungshandlungen urheberrechtlich freistellen wollte. Im Anwendungsbereich des § 44a UrhG kommt es hierbei auf Art. 5 Abs. 1 InfoSoc-RL an.<sup>834</sup>

Dafür, dass Art. 5 Abs. 1 lit. b InfoSoc-RL für die Ermöglichung rechtmäßiger Nutzungen auch verändernde Vervielfältigungshandlungen freistellt, spricht, dass gemäß ErwGr. 33 S. 3 Hs. 2 InfoSoc-RL eine Veränderung von Informationen im Kontext des Art. 5 Abs. 1 lit. a InfoSoc-RL in bestimmten Fallkonstellationen unzulässig sein soll. In Bezug auf die Ermöglichung rechtmäßiger Nutzungen über Art. 5 Abs. 1 lit. b InfoSoc-RL kann der Richtlinie ein solcher Aussagegehalt hingegen nicht entnommen werden. Aus der ausdrücklichen Regelung zur Veränderung von Inhalten im Kontext des Art. 5 Abs. 1 lit. a InfoSoc-RL kann demzufolge geschlussfolgert werden, dass eine Veränderung der genutzten Werke im Rahmen des Art. 5 Abs. 1 lit. b InfoSoc-RL nicht zu einem Ausbleiben der Freistellungswirkung führen soll.<sup>835</sup> Dies muss im Regelungsbereich des § 44a Nr. 2 UrhG im Wege richtlinienkonformer Auslegung auch bei der Anwendung des § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG berücksichtigt werden.

Darüber hinaus ist der Schrankenzweck des § 44a Nr. 2 UrhG respektive Art. 5 Abs. 1 lit. b InfoSoc-RL, rechtmäßige Nutzungen zu ermöglichen. Eine solche ist auch die Verwendung von Werken als Berechnungsgrundlage einer KI.<sup>836</sup> Hierfür müssen die Werke im Rahmen des vorgeschalteten Normalisierungsverfahrens verändert werden.<sup>837</sup> Die Herstellung der veränderten Vervielfältigungsstücke im Normalisierungsverfahren ist dem-

---

833 OLG Hamburg, Urt. v. 05.06.1969 - 3 U 21/69, GRUR 1970, 38 (39) – Heintje; Engels, in: Götting/Lauber-Rönsberg/Ahlberg, Beck'scher Online-Kommentar Urheberrecht, UrhG § 62 Rn. 10 f.; Dreier, in: Dreier/Schulze, Urheberrechtsgesetz, UrhG § 62 Rn. 6, 12; Bullinger, in: Wandtke/Bullinger, Urheberrecht, UrhG § 62 Rn. 10. Zur Herleitung dieser Voraussetzung siehe oben unter 3. Teil B. I. 1. b) bb) (2).

834 Zur Herleitung dieser Voraussetzung siehe oben unter 3. Teil B. I. 1. b) bb) (2).

835 Zu einer anderen Sachverhaltskonstellation und methodisch eingebettet in die eigene wirtschaftliche Bedeutung der verändernden Vervielfältigungshandlungen abweichende Ansicht EuGH, Beschl. v. 17.01.2012 - C-302/10, ZUM 2012, 398 Rn. 53 – Infopaq II, wobei das Gericht hier einem Fehlschluss unterliegt. Siehe dazu oben unter 3. Teil D. I. 2. a) bb).

836 Dazu siehe 3. Teil C. II. 1. d).

837 Dazu oben unter 3. Teil D. I. 2. a) aa).

zufolge auch zur Erfüllung des mit § 44a Nr. 2 UrhG respektive Art. 5 Abs. 1 lit. b InfoSoc-RL verfolgten Schrankenzwecks erforderlich. Da die Werknutzung nicht über die rein technische Verarbeitung der veränderten Werkstücke hinausgeht und insbesondere keine menschliche Rezeption der veränderten Inhalte stattfindet, sind die verändernden Werknutzungen für betroffene Rechtsinhaber darüber hinaus auch zumutbar.

Beide Gründe sprechen im Ergebnis dafür, dass die Herstellung der veränderten Vervielfältigungsstücke im Rahmen der Normalisierung von Inputdaten zur Anwendung von KI, die § 44a Nr. 2 UrhG dem Grunde nach freistellt, gemäß § 62 Abs. 1 S. 2 UrhG i. V. m. § 39 Abs. 2 UrhG vom urheberrechtlichen Änderungsverbot ausgenommen ist. Auch unter Berücksichtigung der stattfindenden Veränderungen an den Werken findet demzufolge eine Freistellung der Nutzungshandlung über § 44a Nr. 2 UrhG statt.

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

Vervielfältigungen von Werken bei der Normalisierung von Inputdaten sind, wie auch die wesensmäßig verwandten Vervielfältigungen bei der Normalisierung von urheberrechtlich geschützten Trainingsdaten, darüber hinaus von der Schrankenbestimmung für Text und Data Mining aus § 44b Abs. 2 UrhG gedeckt. Hierfür ist ebenfalls unschädlich, dass bei der Normalisierung der Inputdaten unter Umständen veränderte Vervielfältigungsstücke der Werke hergestellt werden.<sup>838</sup>

### 3. Zwischenergebnis zur Normalisierung der Inputdaten

Weil das Vervielfältigungsrecht bei der Normalisierung von urheberrechtlich geschützten Inputdaten beeinträchtigt werden kann, erfüllt das gelende Urheberrecht auch bei der Normalisierung von als Inputdaten verwendeten Werken nicht die im zweiten Teil der vorliegenden Untersuchung identifizierten Regelungsbedürfnisse. Dennoch greifen die Schrankenbestimmungen sowohl für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG als auch für das Text und Data Mining aus § 44b UrhG ein. Dies hat zur Folge, dass es für die Normalisierung von als Inputdaten

---

838 Dazu ausführlich vgl. oben unter 3. Teil B. I. 1. b) bb).

## *D. Vorverarbeitung der Inputdaten (Normalisierung und Feature Encoding)*

verwendeten Werken keiner rechtsgeschäftlichen Erlaubnis des Urhebers bedarf. Demzufolge wird der geltende Rechtsrahmen in Bezug auf die Normalisierung urheberrechtlich geschützter Inputdaten zumindest dem Ob des identifizierten Freistellungsbedürfnisses gerecht.

### *II. Feature Encoding*

Nach der Normalisierung werden die erstellten Eingabedaten in Tensoren umgewandelt. Dieser Vorgang wird als Feature Encoding bezeichnet. Tensoren sind numerische Darstellungen der Eingabedaten. Sie sind erforderlich, weil nur diese zahlenbasierten Repräsentationen im KNN als statistischen Modell verarbeitet werden können.

Wie die Normalisierung ist auch das Feature Encoding urheberrechtlich geschützter Inputdaten mit dem Feature Encoding zum Training verwendet. Der Werke wesensverwandt. Deswegen können die für den Umgang mit Trainingsdaten entwickelten Grundsätze auch im vorliegenden Fall angewendet werden.<sup>839</sup> Aus diesem Grund ist davon auszugehen, dass die beim Feature Encoding urheberrechtlich geschützter Inputdaten verwendeten Werke aufgrund ihrer numerischen Repräsentation in den Tensoren mangels tatsächlicher Identifizierbarkeit nicht mehr im urheberrechtlichen Sinne wiedererkennbar sind. Damit werden sie in den Tensoren nicht im Sinne des § 16 UrhG vervielfältigt. Dies führt dazu, dass das Feature Encoding der geschützten Inputdaten nicht das urheberrechtliche Ausschließlichkeitsrecht berührt. Das gilt auch für Vervielfältigungen der Tensoren im Rahmen der Anwendung des KI-Systems. Hinsichtlich des Feature Encodings wird der geltende Rechtsrahmen den identifizierten Freistellungsbedürfnissen daher gerecht. Eine urheberrechtlich relevante Vervielfältigung der Werke findet erst bei der Rückumwandlung der Tensoren in menschlich wahrnehmbare Werkrepräsentationen statt.

### *III. Zwischenfazit zur Vorverarbeitung urheberrechtlich geschützter Inputdaten*

Es erwies sich als zweckmäßig, dass die Vorverarbeitung urheberrechtlich geschützter Inputdaten insgesamt im Wege einer unmittelbar begrenzenden

---

<sup>839</sup> Dazu insgesamt siehe oben unter 3. Teil B. I. 2.

Ausgestaltung der urheberrechtlichen Verwertungsrechte erlaubnisfrei gestellt sein sollte. Während der Normalisierung als Inputdaten verwendeter Werke können – je nach angewendetem Verfahren – jedoch Vervielfältigungen im Sinne des § 16 UrhG hergestellt werden. Infolgedessen besteht ein regulatorisches Defizit hinsichtlich der Normalisierung von Inputdaten.

Allerdings ermöglicht eine Anwendung sowohl der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen aus § 44a UrhG als auch der Schrankenregelung für Text und Data Mining nach § 44b UrhG eine Freistellung der urheberrechtlich relevanten Normalisierungsschritte. Damit wird das geltende Urheberrecht jedenfalls dem Ob des regulatorischen Interesses gerecht. Letztere ist jedoch nur anwendbar, wenn kein Nutzungsvorbehalt im Sinne des § 44b Abs. 3 UrhG erklärt wurde. De lege ferenda könnte eine Anpassung der Reichweite des Vervielfältigungsrechts in § 16 UrhG Abhilfe schaffen.

Im Gegensatz hierzu werden die für das anschließende Feature Encoding der urheberrechtlich geschützten Inhalte ermittelten Regelungsbedürfnisse vom geltenden Rechtsrahmen vollständig erfüllt. Die Erstellung der numerischen Werkrepräsentationen greift nicht in das Vervielfältigungsrecht aus § 16 UrhG ein.

## E. Zusammenfassung des dritten Teils

Für das Web Scraping urheberrechtlich geschützter Trainingsdaten hält das geltende Urheberrecht mit § 44b UrhG eine Schrankenbestimmung bereit, deren Freistellungswirkung die Nutzungshandlung dem Grunde nach umfasst. Der Einsatz von Webcrawlern zur Vervielfältigung von Werken für das Training von KI bedarf daher grundsätzlich keiner rechtsgeschäftlichen Erlaubnis. Dies deckt sich insoweit auch mit den regulatorischen Freistellungsbedürfnissen, die im zweiten Teil der Untersuchung identifiziert worden sind. Das gilt auch für den persönlichen Anwendungsbereich der Schrankenbestimmung. Er umfasst auch Web Scraping-Aktivitäten kommerzieller Trainingsdatensammler. Sowohl bereits angefertigte Werkstücke als auch Zugänge zu Werkdatenbanken können bei Einhaltung bestimmter Voraussetzungen auch an Dritte Datenverarbeiter weitergegeben werden, ohne dass die Freistellungswirkung der Schrankenbestimmung für Text und Data Mining-Aktivitäten für diese Dritten erlischt. Hierfür kommt es

## *E. Zusammenfassung des dritten Teils*

auf die Auslegung des Tatbestandsmerkmals der rechtmäßigen Zugänglichkeit im Sinne des § 44b Abs. 2 S. 1 UrhG an.

Dennoch führen sowohl die in § 44b Abs. 2 S. 2 UrhG geregelte Lösungspflicht, die Vergütungsfreiheit der Nutzung sowie die in § 44b Abs. 3 UrhG angelegte Möglichkeit, einen die Schrankenwirkung beenden den Nutzungsvorbehalt zu erklären, dazu, dass das geltende Urheberrecht hinter den regulatorischen Freistellungsbedürfnissen zurückbleibt. Der ex nunc wirkende Nutzungsvorbehalt kann dabei auch in natürlicher Sprache abgegeben werden. Wirksamkeitsvoraussetzung ist aber, dass die Erklärung in einer Art und Weise digital hinterlegt ist, die eine unmittelbare Verarbeitung durch Computer ermöglicht. Sie muss darüber hinaus sowohl angemessen als auch ausdrücklich erfolgt sein. Für eine interessengerechte Freistellung des Web Scrapings von Werken zum Training von KI sind alles in allem Anpassungen im urheberrechtlichen Rechtsrahmen vorzunehmen.

Das eigentliche Training von KI-Systemen auf Basis von Werken kann in vier Teilschritte unterteilt werden: Zunächst müssen die als Trainingsdaten gesammelten Werke vorverarbeitet, also für die maschinelle Verarbeitung aufbereitet werden. Hierzu gehören die Normalisierung, also die Transponierung auf eine vergleichbare Skala, sowie das Feature Encoding, also die Umwandlung der Werke in numerische Tensoren. Nur diese numerischen Repräsentationen der Werke können durch das KNN verarbeitet werden. An die Vorverarbeitung der Trainingsdaten schließt sich die eigentliche Trainingsanalyse an. Bei ihr werden ggf. vorübergehende Vervielfältigungen der in numerische Tensoren umgewandelten Trainingsdaten angefertigt. Der Trainingsprozess mündet in der Anpassung der Gewichtungswerte und anderen Parameter innerhalb eines KNN.

Im Rahmen der beschriebenen Teilschritte des Trainingsprozesses können KI-Entwickler allein bei der Normalisierung der als Trainingsdaten verwendeten Werke möglicherweise in das urheberrechtliche Ausschließlichkeitsrecht eingreifen. Ob Vervielfältigungsstücke im Sinne des § 16 UrhG hergestellt werden, hängt von den im Einzelnen angewendeten Normalisierungsverfahren ab. Alle übrigen Teilschritte des Trainingsprozesses fallen nicht in den Anwendungsbereich eines Verwertungsrechts. Damit erfüllt das geltende Urheberrecht hier im Wesentlichen die regulatorischen Bedürfnisse. Nach ihnen sollte das Training im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Ausschließlichkeitsrechte freigestellt sein. Die während der Normalisierung urheberrechtlich geschützter Trainingsdaten möglicherweise stattfindenden Vervielfältigungs handlungen sind im Übrigen zwar durch die Schrankenbestimmung für

Text und Data Mining aus § 44b UrhG erlaubnisfrei gestellt. Hierdurch wird das Urheberrecht jedoch nur dem *Ob*, nicht aber dem *Wie* der im zweiten Teil der Untersuchung identifizierten Freistellungsbedürfnisse gerecht. Hinsichtlich der Normalisierung von Trainingsdaten macht dies de lege ferenda Anpassungen der Rechtslage notwendig.

Auch die Speicherung urheberrechtlich geschützter Inputdaten, beispielsweise durch Sensortechnik wie Kameras oder Webcrawler, bedarf aus einer regulatorischen Perspektive der Freistellung im Wege einer unmittelbar beschränkenden Ausgestaltung der urheberrechtlichen Verwertungsrechte. Zwar ist die Speicherung als Inputdaten verwendeter Werke de lege lata sowohl über die Schrankenbestimmung für Text und Data Mining, § 44b UrhG, als auch über die Schrankenbestimmung für vorübergehende Vervielfältigungen aus § 44a Nr. 2 UrhG erlaubnisfrei gestellt. Dennoch wird dies, wie auch bei der Normalisierung urheberrechtlich geschützter Trainingsdaten, nicht dem *Wie* des identifizierten Freistellungsbedürfnisses gerecht. Daher bedarf es auch hier de lege ferenda einer Anpassung des Urheberrechts. Durch sie muss die Speicherung urheberrechtlich geschützter Inputdaten für eine Verarbeitung durch KI im Wege einer unmittelbar begrenzenden Ausgestaltung der urheberrechtlichen Verwertungsrechte aus dem Schutzbereich des urheberrechtlichen Ausschließlichkeitsrechts ausgenommen werden.

Für die gleichermaßen notwendige Vorverarbeitung der Inputdaten gilt im Wesentlichen das für die Vorverarbeitung von Trainingsdaten Gesagte. Zwar ist die Normalisierung urheberrechtlich geschützter Inputdaten auch von der Schrankenbestimmung für vorübergehende Vervielfältigungshandlungen, § 44a Nr. 2 UrhG, erlaubnisfrei gestellt. Im Ergebnis gilt aber auch hier, dass die Vorverarbeitung inklusive der Normalisierung der als Inputdaten verwendeten Werke aus einer regulatorischen Perspektive durch eine unmittelbar begrenzende Ausgestaltung der urheberrechtlichen Verwertungsrechte freigestellt sein sollte. Die Normalisierung urheberrechtlich geschützter Inputdaten, Teil des Vorverarbeitungsprozesses, kann aber auch hier in das Vervielfältigungsrecht betroffener Urheber aus § 16 UrhG eingreifen. Auch in Bezug auf die Vorverarbeitung urheberrechtlich geschützter Inputdaten wird das Freistellungsbedürfnis daher de lege lata nicht vollständig erfüllt. Damit müssen de lege ferenda auch in Bezug auf die Vorverarbeitung, konkret die Normalisierung urheberrechtlich geschützter Inputdaten Anpassungen im UrhG erfolgen.

