

Hanjo Hamann*

Nutzungsvorbehalte für KI-Training in der Rechtsgeschäftslehre der Maschinenkommunikation

Dogmatische und praktische Schwächen von Art. 4 Abs. 3 DSM-RL
und § 44b Abs. 3 UrhG

Die Entwicklung Künstlicher Intelligenzen und großer Sprachmodelle (LLMs) setzt riesige Textbestände als Trainingsmaterial voraus. Müssen die Urheberinnen solcher Texte dulden, dass ihre Werke für das KI-Training vervielfältigt werden? Das richtet sich danach, ob ihr Vervielfältigungsrecht (§ 16UrhG) einer urheberrechtlichen Schranke unterfällt – wie derjenigen für kommerzielles „Text und Data Mining“ (§ 44b UrhG) nach europäischem Vorbild (Art. 4 DSM-RL). Auch wenn das neue europäische KI-Gesetz den Streit über die Anwendbarkeit der TDM-Schranke auf das KI-Training geklärt haben dürfte, beginnen damit gerade erst die praktischen Probleme. Aus Ausgleichs- und Anreizgründen reicht die TDM-Schranke nur soweit wie sich Urheber diese Nutzung nicht selbst wirksam vorbehalten haben. Zahlreich waren die Stimmen, die deshalb ein „Leerlaufen“ der TDM-Schranke befürchteten. Ebenso zahlreich allerdings die Stimmen, die ganz im Gegenteil ein „Leerlaufen“ des Nutzungsvorbehalts prognostizierten. Eine Entscheidung dieser Kontroverse setzt die Klärung zahlreicher dogmatischer Einzelprobleme voraus – etwa die Konkretisierung des Rechtsbegriffs „maschinenlesbare Form“, des unausgesprochenen Ausdrücklichkeitserfordernisses, des systematischen Zusammenhangs mit § 49 UrhG, der noch ungeklärten Erklärungsbefugnis und der zahlreichen Zweifel an der Richtlinienkonformität des § 44b UrhG, die sich ganz anders bewahrheiten als bislang angenommen. Ist die dogmatische Kontur des § 44b UrhG geklärt, bleibt die alles andere als triviale Frage, wie Maschinenlesbarkeit technisch konkret zu erreichen wäre. Nach Art. 53KI-Gesetz sollen Nutzungsvorbehalte mit „modernsten Technologien“ erkannt werden, also fragt sich umgekehrt, mittels welcher Technologien Nutzungsvorbehalte überhaupt wirksam erklärt werden können. Die Studie sichtet sieben technische Protokolle, von denen nur drei die rechtlichen Anforderungen überhaupt erfüllen könnten, aber statt zur Standardisierung zum Wildwuchs neigen. Daher dürfte es momentan kaum möglich sein, einen Nutzungsvorbehalt gerichtsfest wirksam zu erklären.

* Prof. Dr. Dr., JSM (Stanford), Qualifikationsprofessor für Bürgerliches Recht, Wirtschafts- und Immaterialgüterrecht an der EBS Universität in Wiesbaden. Der Beitrag beruht auf Vorträgen an der Humboldt-Universität zu Berlin am 19. Juni 2023 und auf der Tagung „Generative KI im Spiegel des Urheberrechts“ am 23. Februar 2024 in Berlin (dazu *Suilmann*, RuZ 2024, 61). Ich danke meinen Mitveranstalterinnen Katharina de la Durantaye, Franz Hofmann und Benjamin Raue für die angenehme und vertrauensvolle Zusammenarbeit und für weiterführende Hinweise zum vorliegenden Beitrag. Ebenso danke ich Anna Bernzen, Linda Kuschel, Alexander Peukert, Jonathan Pukas, Tristan Radtke und Maren Wöbbeking für hilfreiche Verbesserungsvorschläge.



A. Ein kleiner Bot speichert das Internet	114
B. Problemstellung und Überblick	117
C. Maßgaben des geltenden Rechts für KI-Training	118
I. Eingrenzung des Untersuchungsgegenstands	118
II. Die Suche nach dem rechten Rechtsmaßstab	120
III. Begrenzung der TDM-Schranke im Spiegel ihrer Nah- und Fernziele	122
IV. Die Furcht vor dem „Leerlaufen“ – aber wovon?	124
D. § 44b Abs. 3 S. 2 UrhG und die H2M-Rechtsgeschäftslehre	127
I. Rechtsfolge: „nur dann wirksam, wenn“	127
II. Tatbestand: „in maschinenlesbarer Form“	128
III. Normgenese: Ausdrücklichkeitserfordernis (Art. 4 Abs. 3 DSM-RL)	134
IV. Systematik: „Vorbehalt der Rechte“ in § 49 Abs. 1 S.1 UrhG	135
V. Das klärungsbedürftige Problem der Erklärungsbefugnis	137
VI. Richtlinienkonforme Auslegung: „Inhalte“ statt „Werke“	140
VII. Zwischenergebnis: Die Dogmatik des § 44b Abs. 3 UrhG	144
E. Wie lässt sich KI-Training maschinenlesbar untersagen?	144
I. Impressum oder AGB	146
II. Robots Exclusion Protocol (robots.txt)	149
III. Spawning-Protokoll (ai.txt)	154
IV. HTTP Response Header	155
V. HTML Metadaten-Tags	157
VI. JSON (tdmrep.json, Reichs ai.txt, C2PAs)	159
VII. Technische Zugangshürden (Paywall, CAPTCHA, etc.)	162
VIII. Zwischenergebnis: „Standards are great. Everyone should have one!“	164
F. Fazit und Ausblick	166

A. Ein kleiner Bot speichert das Internet

Die folgende Geschichte beruht auf einer wahren Begebenheit.

Es war einmal ein kleines Computerprogramm. Das sollte Textmaterial sammeln, um eine große Künstliche Intelligenz zu trainieren, ein sog. „Large Language Model“. Deshalb bekam das kleine Programm eine einfache Aufgabe: Gehe zu einer Website im Internet, kopiere ihren gesamten Inhalt in eine Datenbank, dann folge allen Links von dieser Website zu anderen und fange dort jeweils von vorn an. Dieses Herumkriechen im Internet (*crawlen*) war eine sehr repetitive Aufgabe, aber das Computerprogramm erledigte sie mit sturer Disziplin wie ein Roboter. Deswegen nannte man es „Bot“. Der kleine Crawler-Bot erledigte die ihm gestellte Aufgabe zuverlässig, obwohl sie technisch gar nicht so simpel war wie es zunächst klang:

Immer wenn der Bot eine Website ansteuerte und an der Tür Einlass begehrte, musste er dem digitalen Diener („Server“) seinen Namen mitteilen – unser Bot nannte sich vielleicht „CCbot“ oder „GPTBot“ oder „anthropic-ai“. Eines Neujahrsmorgens kam der Bot zu einem Server und fragte nach guter alter Bot-

Tradition¹ zuerst nach den Hausregeln. Der Server teilte ihm mit, dass er solche habe und herauszugeben bereit sei. Wörtlich sagte er in Maschinensprache:²

```
HTTP/2 200
server: myracloud
date: Mon, 01 Jan 2024 09:10:11 GMT
accept-ranges: bytes
tdm-policy: https://rsw.beck.de/beck-online-service/tdm-vorbehalt
tdm-reservation: 1
content-security-policy: [...] etag: [...] x-content-type-options: [...]
X-Firefox-Spdy: h2
```

Zusammen mit dieser Antwort überreichte der Server die gewünschten Hausregeln als Textdatei („robots.txt“). Unser Bot las sie sogleich, darin stand:³

```
User-agent: CCBot
User-agent: GPTBot
User-agent: ChatGPT-User
Disallow: /
```

Diesen Text kannte der Bot bereits, denn über zwei Fünftel der Nachrichtenportale weltweit (40,7 %) verwenden genau diesen Text.⁴ Danach fand unser Bot noch zwei weitere Textzeilen,⁵ erkannte aber an den vorangestellten Rautezeichen (#), dass dieser Text und für ihn ohnehin nicht verständlich, sondern nur für Menschen geschrieben war.

Nun bat er den Server als nächstes um die Startseite. Normalerweise hieß sie „index.html“ oder so ähnlich, aber in diesem Fall schlicht „/ Home“. Der Server wusste jedenfalls, was der Bot meinte und überreichte eine Datei, die der Bot ebenfalls begierig las. Etwa die ersten achtzig Zeilen waren in Maschinensprache verfasst, beginnend mit:⁶

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html lang="de" class="">
<head>[...]
  <title>Homepage - beck-online</title>
  <meta name="format-detection" content="telephone=no" /> [...]
  <meta name="tdm-reservation" content="1">
```

1 Vgl. *Peukert*, ZUM 2023, 233 (246): „... liest ein Webcrawler beim Auffinden einer Webseite immer zuerst die Datei robots.txt im Stammverzeichnis einer Domain.“

2 HTTP Response Header von <https://beck-online.beck.de/robots.txt> (letzter Abruf 4.3.2024). Näher unten E.IV.

3 Dateinhalt von <https://beck-online.beck.de/robots.txt> (letzter Abruf 4.3.2024); frühere Version zitiert bei *Maamar*, ZUM 2023, 481 (484). Näher unten E.II.

4 Berechnung und Nachweise unten bei Fn. 195, 196.

5 Wörtlich: “# Legal notice: Verlag C.H.BECK oHG expressly reserves the right to use its content for commercial text and data mining (§ 44b Urheberrechtsgesetz). – # The use of robots or other automated means to access our websites or collect or mine data without the express permission of Verlag C.H.BECK oHG is strictly prohibited.”

6 Dateinhalt von <https://beck-online.beck.de/Home> (letzter Abruf 4.3.2024). Näher unten E.V.

```
<meta name="tdm-policy" content="https://rsw.beck.de/beck-online-
service/tdm-vorbehalt">
<meta name="robots" content="noai, noimageai">
```

Der Bot speicherte den Inhalt dieser Datei in seine Datenbank und machte sich daran, alle darin enthaltenen Links aufzurufen. Einer der Links war mit „Impressum“ beschriftet und führte zur Datei „/Impressum“, die der Bot sich vom Server geben ließ. Auch diese Datei begann mit Maschinensprache, ging dann aber in ein Gemisch von Maschinen- und Menschensprache über. Zum Beispiel fand der Bot diese Zeichenkette:⁷

```
<p><b>Text und Data Mining gem&auml;&szlig; &sect; 44b UrhG | [...]<br>
</b>Der Verlag beh&auml;&lt; sich das Recht zu Vervielf&auml;&lt;igungen
f&uuml;r das Text und Data Mining gem&auml;&szlig; &sect; 44b UrhG
vor.<br>
```

Damit konnte der Bot nichts anfangen. Er verstand die Sprache der Menschen nicht. Allenfalls hätte er die Maschinensprachsnipsel nutzen können, um einen für Menschen lesbaren Text anzuzeigen.⁸ Aber da war niemand, der das hätte sehen wollen. Deshalb befolgte der Bot brav seinen Arbeitsauftrag, speicherte die gesamte Datei ab und ging weiter zum nächsten Link. Der war mit „AGB“ beschriftet und führte zu einer ganz anderen Website mit eigenen (großzügigen) Hausregeln; auch deren Inhalt holte sich der Bot und fand darin ein ähnliches Potpourri von Maschinen- und Menschensprache:⁹

```
<div >[...] <h4>9. Schutzrechte</h4><p>[...]<br /></font>9.2 Der Ver-
lag beh&auml;&lt; sich gem&auml;&szlig; &sect; 44b Abs. 3 UrhG&nbsp;&nbsp;&nbsp;das
Recht vor, Vervielf&auml;&lt;igungen von Inhalten der Datenbank und der
E-Mail-Dienste&nbsp;&nbsp;&nbsp;i.S.v. &sect; 44b Abs. 2 Satz 1 UrhG und Ver-
vielf&auml;&lt;igungen eines nach Art oder Umfang wesentlichen Teiles
der Datenbank i.S.v. &sect; 87c Abs. 1 Nr. 4 i.V.m. &sect; 44b Abs. 2
Satz 1 UrhG jeweils zum Zwecke des Text und Data Mining vorzuneh-
men.<br /><br />
```

Auch das speicherte der Bot ab, weil er es niemandem anzeigen konnte.¹⁰ Es gab noch etliche weitere Links, denen der Bot folgte. Die meisten davon führten zu

⁷ Zitat aus <https://beck-online.beck.de/Impressum> (letzter Abruf 4.3.2024). Näher unten E.I.

⁸ Nämlich: „Text und Data Mining gemäß § 44b UrhG | [...] Der Verlag behält sich das Recht zu Vervielfältigungen für das Text und Data Mining gemäß § 44b UrhG vor.“

⁹ Zitat aus <https://rsw.beck.de/beck-online-service/agb-beck-online> (letzter Abruf 4.3.2024). Näher unten E.I.; „Hausregeln“ in <https://rsw.beck.de/robots.txt>.

¹⁰ Sonst hätte derjenige gelesen: „9. Schutzrechte [...] 9.2 Der Verlag behält sich gemäß § 44b Abs. 3 UrhG das Recht vor, Vervielfältigungen von Inhalten der Datenbank und der E-Mail-Dienste i.S.v. § 44b Abs. 2 Satz 1 UrhG und Vervielfältigungen eines nach Art oder Umfang wesentlichen Teiles der Datenbank i.S.v. § 87c Abs. 1 Nr. 4 i.V.m. § 44b Abs. 2 Satz 1 UrhG jeweils zum Zwecke des Text und Data Mining vorzunehmen.“

Dateien, die jede ungefähr 10 Kilobyte groß waren und für einen Menschen so ausgesehen hätten.¹¹

„Sie können das gewünschte Dokument [...] nur aufrufen, wenn Sie eingeloggt sind. Bitte geben Sie hierzu Ihren Benutzernamen und das Passwort in die Login-Maske ein. [...] Alternativ können Sie sich das Dokument auch einzeln freischalten, indem Sie sich bei beck-treffer anmelden.“

Jede dieser kleinen Dateien speicherte der Bot und besuchte anschließend noch viele andere Websites, die er in gleicher Weise kopierte: Datei für Datei, Link für Link. Schon bald hatte er Milliarden von Texten in seiner Datenbank. Und weil ein Bot nicht sterben kann, speichert er das Internet noch heute.

Und die Moral von der Geschichte? Wie ist das „Verhalten“ unseres kleinen Bots urheberrechtlich zu bewerten?

B. Problemstellung und Überblick

Die geschilderte Geschichte illustriert drei für das moderne Urheberrecht zentrale Erkenntnisse, dass nämlich (1) „das Internet“ für Menschen und Maschinen sehr unterschiedlich aussieht,¹² (2) grundverschiedene Verhinderungsstrategien denkbar sind, um eigene Texte vom Training fremder KI auszuschließen,¹³ und (3) die KI-Materialsammlung technisch kleinteilig („Datei für Datei, Link für Link“) und zugleich hochautomatisiert Milliarden Texte ohne menschliche Intervention erfasst.

Wie geht das Urheberrecht mit diesen Erkenntnissen um? Darf die Herrin und Meisterin des kleinen Bots die von ihm erstellte Datenbank überhaupt nutzen? Diesen Fragen geht der folgende Beitrag nach. Er identifiziert zunächst die einschlägigen rechtlichen Maßgaben (C.) und zeigt ihre Auslegungsschwierigkeiten auf (D.). Sodann bewertet er die in der Eingangsschilderung illustrierten KI-Verhinderungsstrategien auf ihre Tauglichkeit (E.). Daraus ergibt sich, dass das geltende Recht Sensibilität für die technischen Abläufe beim KI-Training vermissen lässt, dass Gesetzgeberin und Schrifttum in Zukunft erheblich größeres Gespür für die Funktionsweise von Maschinenkommunikation an den Tag legen müssen, und dass die Materialgewinnung für KI-Training zur Zeit gar nicht rechtssicher verhinderbar ist (F.).

¹¹ Zitat aus https://beck-online.beck.de/?vpath=bibdata%2Fkomm%2FDieNotKostBer_9%2Fcont%2FDieNotKostBer%2EglKap1%2EglA%2EglI%2Ehtm (letzter Abruf 4.3.2024). Näher unten E.VII.

¹² Vgl. die Zitate im Haupttext nach Fn. 7, 9 mit denjenigen in Fn. 8, 10.

¹³ Etwa durch maschinensprachliche Erklärungen (Zitate nach Fn. 2, 3, 6), durch Vorbehalte in natürlicher Sprache (Zitate in Fn. 5, 8, 10) oder durch technische Zugangshürden (bei Fn. 11).

C. Maßgaben des geltenden Rechts für KI-Training

I. Eingrenzung des Untersuchungsgegenstands

Künstliche Intelligenz wirft unzählige Rechtsfragen auf, deren Beantwortung in einer Geschwindigkeit voranschreitet, die kaum noch einen Gesamtüberblick zulässt. Um die vorliegende Untersuchung überschaubar zu halten, bedarf es deshalb zunächst einiger methodischer und thematischer Eingrenzungen:

Die vorliegende Untersuchung befasst sich ausschließlich mit der *Input*-Seite von Künstlicher Intelligenz, die mit Blick auf die KI-Wertschöpfungskette auch als „*Upstream*“ bezeichnet wird. Während viel darüber geschrieben wird, ob und wie der *Output* Künstlicher Intelligenz (also im „*Downstream*“ der Wertschöpfungskette) rechtlich geschützt ist oder werden kann,¹⁴ erscheinen Fragen der *Upstream* / *Input*-Regulierung „aus praktischer Perspektive viel drängender“.¹⁵ Zugleich gelten sie aber als „etwas weniger ansprechend“¹⁶ (weil technisch anspruchsvoller) und erhalten deshalb seit Jahren „kaum wissenschaftliche Aufmerksamkeit“.¹⁷

Dafür wählt die vorliegende Untersuchung einen *technikrechtlichen* Zugang, also die Anwendung dogmatisch gewonnener Maßstäbe auf technisch definierte Abläufe. Ob KI-Training gesellschaftspolitisch wünschenswert, ob es ökonomisch effizient oder human-ethisch vertretbar ist, bleibt hier ungeachtet der Wichtigkeit solcher Fragen ausgeblendet.

In ihren dogmatischen Maßstäben beschränkt sich die Untersuchung schließlich auf das *Urheberrecht*. Im gewerblichen Rechtsschutz sowie im Wettbewerbsrecht mögen ähnliche oder weitere Verhinderungsstrategien denkbar sein; darauf lässt sich hier nicht eingehen, ohne den Umfang eines Zeitschriftenbeitrags zu sprengen.

Innerhalb des Urheberrechts werden nur Vorgaben an *kommerziell* betriebenes KI-Training beleuchtet. Nur hier stellt sich die Frage nach einem Nutzungsvorbehalt; zugleich ist kommerzielles KI-Training der politisch brisan-

14 Monographisch etwa *Fierdag*, Die Aleatorik in der Kunst und das Urheberrecht, 2005, insb. Kap. 2; *Käde*, Kreative Maschinen und Urheberrecht, 2021, insb. Teil 4; *Maamar*, Computer als Schöpfer, 2021, insb. Kap. 7; *Grätz*, Künstliche Intelligenz im Urheberrecht, 2021, insb. Kap. 4; *Linke*, „Künstliche Intelligenz“ und Urheberrecht – quo vadis?, 2021, insb. Kap. 5; *Muhr*, KI-Schöpfungen und Urheberrecht, 2022, insb. Teile 2 und 5; *Klawonn*, Künstliche Intelligenz, Musik und das Urheberrecht, 2023, insb. Teil 2; *Jäger*, Artificial creativity?, 2023, insb. Kap. 4.

15 So *Schönberger*, ZGE 10 (2018), 35 (47) (“from a practical point of view [...] far more pressing”); ähnl. *Gernhardt*, GRUR-Prax 2022, 69 („für die Auftraggeber [...] von hoher praktischer Bedeutung“).

16 So *Schönberger*, ZGE 10 (2018), 35 (47) (“slightly less aesthetic”).

17 Vgl. nur *Schönberger*, ZGE 10 (2018), 35 (47): “to date these questions have attracted little academic attention” (ebenso S. 57) und *Schack*, NJW 2024, 113 (Rn. 2): „Deutlich seltener behandelt wird die Input-Seite“.

tere, praxisrelevantere und derzeit umstrittenere Anwendungsfall. Im Übrigen bedürfte die historisch bedingte Zweigleisigkeit von Vorgaben für wissenschaftliche Forschung (Art. 3 DSM-RL mit ErwGr. 8–17, sowie § 60d UrhG)¹⁸ und für sonstige Anwendungen (Art. 4 DSM-RL mit ErwGr. 18, sowie § 44b UrhG) eines steten Hin- und Herwanderns des Blicks und eines Maßes an Binnendifferenzierung, das die Übersichtlichkeit des Beitrags gefährden würde.

Innerhalb der Literatur zum Upstream-Urheberrecht kommerziellen KI-Trainings werden dafür nicht nur Studien und Publikationen aus der deutsch- und englischsprachigen Rechtswissenschaft berücksichtigt, sondern auch die zahlreichen *Stellungnahmen* gesellschaftlicher Akteure (in Fußnoten abgekürzt als „StN“) aus dem mehrstufigen Gesetzgebungsverfahren zur Schaffung der relevanten Vorschriften. Die in diesen Stellungnahmen aufgeworfenen Detailfragen erhielten seinerzeit „nur wenig Aufmerksamkeit“;¹⁹ umso mehr empfiehlt sich nun ihre Sichtung. Es handelte sich um

- 124 Stellungnahmen zu einer öffentlichen Konsultation durch das BMJV,²⁰
- 91 Stellungnahmen zum Diskussionsentwurf des späteren Gesetzes („DiskE“),²¹
- 104 Stellungnahmen zum entsprechenden Referentenentwurf („RefE“),²² sowie
- 14 im Rechtsausschuss des Deutschen Bundestages erörterte Stellungnahmen zum Regierungsentwurf („RegE“).²³

Diese Stellungnahmen wurden hier systematisch gesichtet,²⁴ wobei sich herausstellte, dass erstaunlich viele der später erörterten dogmatischen Problemfelder bereits früh im Gesetzgebungsverfahren identifiziert, aber vergebens angesprochen worden waren.

18 Dazu bspw. *Spindler*, ZGE 10 (2018), 273; *Kleinkopf*, Text und Data Mining, 2022; *Raue*, RuZ 2022, 4.

19 Dreier/Schulze/Dreier, 7. Aufl. 2022, § 44b UrhG Rn. 2, und weiter: „weil das Hauptaugenmerk auf Art. 17 DSM-RL zur Haftung der Anbieter von Online-Diensten konzentriert war“; dazu noch unten bei Fn. 62.

20 Zeitraum 28.6.–6.9.2019, StN verfügbar unter [bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2019_Konsultation_Umsetzung_EU_Richtlinien_Urheberrecht.html](https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2019_Konsultation_Umsetzung_EU_Richtlinien_Urheberrecht.html).

21 Veröffentlicht am 15.1.2020, StN verfügbar unter [bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2020_Gesetz_I_Anpassung-Urheberrecht-dig-Binnenmarkt.html](https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2020_Gesetz_I_Anpassung-Urheberrecht-dig-Binnenmarkt.html).

22 Veröffentlicht am 13.10.2020, StN verfügbar unter [bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2020_Gesetz_Anpassung-Urheberrecht-dig-Binnenmarkt.html](https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2020_Gesetz_Anpassung-Urheberrecht-dig-Binnenmarkt.html).

23 Veröffentlicht am 9.3.2021, StN verfügbar unter [bundestag.de/dokumente/textarchiv/2021/kw15-pa-recht-830028](https://www.bundestag.de/dokumente/textarchiv/2021/kw15-pa-recht-830028).

24 Die 333 Stellungnahmen wurden nach dem Begriff „maschinenlesbar“ durchsucht, um diejenigen zu identifizieren, die sich konkret mit der Form des Art. 4 Abs. 3 DSM-RL, § 44b Abs. 3 UrhG befassen. Die resultierenden 40 Dokumente (12 %) wurden genauer gelesen, 29 davon enthielten relevante Ausführungen, die der Beitrag an den passenden Stellen zitiert.

II. Die Suche nach dem rechten Rechtsmaßstab

Lange bestand Streit darüber, welche urheberrechtlichen Maßstäbe für die Sammlung von KI-Trainingsdaten überhaupt gelten. Bei Inkrafttreten der EU-Richtlinie 2019/790 über das Urheberrecht im digitalen Binnenmarkt („DSM-RL“) vor fünf Jahren war die enorme Konjunktur der Künstlichen Intelligenz, die mit der Veröffentlichung von ChatGPT am 30. November 2022 einsetzte, noch gar nicht absehbar. Weder die Richtlinie noch ihre Erwägungsgründe erwähnen das Konzept „Künstliche Intelligenz“ überhaupt – es „wurde offensichtlich übersehen“²⁵ und lag „damals noch außerhalb der Vorstellung des Gesetzgebers“.²⁶

Deshalb harmonisierte die Richtlinie seinerzeit nur die Anforderungen an das aus dem Forschungskontext bereits bekannte (seit 2018 in § 60d UrhG geregelte) „Text und Data Mining“ (TDM) als einer

„Technik für die automatisierte Analyse von Texten und Daten in digitaler Form, mit deren Hilfe Informationen unter anderem – aber nicht ausschließlich – über Muster, Trends und Korrelationen gewonnen werden können.“ (Art. 2 Nr. 2 DSM-RL)

Die Richtlinie sollte das TDM aber auch außerhalb der Forschung ermöglichen (Art. 4 DSM-RL) und führte zum kommerziellen TDM als „vorherrschender Technik in der Digitalwirtschaft“ (ErwGr. 8 S. 3 DSM-RL) in den Erwägungsgründen aus,

„dass die Nutzer des Text und Data Mining mit Rechtsunsicherheit hinsichtlich der Frage konfrontiert sein könnten, ob Vervielfältigungen und Entnahmen [...] vorgenommen werden dürfen [...]. Um in diesen Fällen für mehr Rechtssicherheit zu sorgen und auch in der Privatwirtschaft zu Innovationen anzuregen, sollte diese Richtlinie unter bestimmten Bedingungen eine Ausnahme oder Beschränkung [...] für die Zwecke des Text und Data Mining vorsehen“ (ErwGr. 18 S. 3, 4 DSM-RL)

In der Folge stritten sich die Gelehrten, ob KI-Training, das heutzutage meist in der „Black Box“ neuronaler Netze stattfindet, überhaupt als „Informationsgewinnung“ i.S.v. Art. 2 Nr. 2 DSM-RL gelten kann, ob es eine Form der Mustererkennung sei oder eine andere Art von Information (die Art. 2 Nr. 2 DSM-RL ja „nicht ausschließen“ sollte), und ob die zitierte Regelungsabsicht, „in der Privatwirtschaft zu Innovation anzuregen“ (ErwGr. 18 S. 3 DSM-RL) nicht ge-

25 Geiger, When the Robots (Try to) Take Over: Of Artificial Intelligence, Authors, Creativity and Copyright Protection, in: FS Hilty, 2024, S. 67 (77) (“has obviously been overlooked”).

26 So von Welser, GRUR-Prax. 2023, 516 (518 Rn. 19), und weiter: „Die TDM-Schranken wurde [sic] nicht für generative KI geschrieben.“ Sowie dort Rn. 22f.: „Tatsächlich haben weder der deutsche, noch der EU-Gesetzgeber generative KI, die 2022 ihren Durchbruch hatte, im Jahr 2019 vorhergesehen.“; zuvor schon Nordemann/Pukas, JIPLP 2022, 973 (974): “there is no provision in the Directive that expressly deals with the training of AI and the copyright-related aspects.”; seither noch Schack, NJW 2024, 113 (114 bei Fn. 20): „Das KI-Problem hatte der europäische Gesetzgeber 2019 schlicht noch nicht auf dem Schirm.“; Mezei, Eur. IP Rev. 46 (2024), 461 (465): “even if the TDM exceptions were designed in light of the developments of AI, they were not drafted in light of GenAI.”

rade die Erfassung ganz neuer und deshalb unvorhergesehener Anwendungen privilegieren müsse.

Die meisten Autorinnen sowohl in der europäischen²⁷ als auch in der deutschen²⁸ Diskussion befürworteten infolge solcher Überlegungen die Anwendung der TDM-Schranke auf das KI-Training. Einzelne allerdings ganz undifferenziert,²⁹ andere nur zähneknirschend, weil sie TDM lieber enger gefasst hätten.³⁰ Wieder andere lehnten die Anwendung der TDM-Schranke rundheraus ab,³¹ oder äußerten Zweifel an der Anwendbarkeit, ohne sich dafür oder dagegen zu entscheiden.³²

-
- 27 *Chiou*, JIPITEC 10 (2019), 398 (409 Rn. 22): “This exception could be invoked, *a priori*, within the framework of any ML project”; *Dusollier*, CML Rev. 57 (2020), 979 (984): “Artificial intelligence, based on machine-learning, is also deeply reliant on data mining and analysing.”; ähnl. *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (25 nach Fn. 42); *Vesala*, IIC 54 (2023), 351 (355); *Senfleben*, IIC 54 (2023), 1535 (1542 bei Fn. 33).
- 28 *Spindler*, CR 2019, 277 (278 Rn. 2); *Hacker*, ZGE 12 (2020), 239 (257) („gewinnorientierte Verwendungszwecke des KI-Trainings fallen [...] lediglich unter Art. 4 DSM-RL“); ausf. *Linke*, „From sainted to tainted data“ – Rechte Dritter an KI-Trainingsdaten“, in: Kuschel/Asmusen/Golla (Hrsg.), *Intelligente Systeme – Intelligentes Recht* (GRUR Junge Wissenschaft), 2021, S. 179 (188–190); *Käde* (Fn. 14), S. 81–83, 90 f.; *Kleinkopf* (Fn. 18), S. 46–48 (II.), 238 (bei Fn. 917); seitdem noch *Siems/Repka*, DSRITB 12 (2021), 517 (525); *Müller ter Jung/Rexin*, CR 2022, 169 (174 Rn. 37); *Söbbing/Schwarz*, RD 2023, 415 (418 Rn. 15); *Bernzen*, Beil. zu K&R 10/2023, 6 (7 bei Fn. 19); *Bomhard*, DSRITB 14 (2023), 255 (259); *Durantaye*, ZUM 2023, 645 (651); *Pesch/Böhme*, GRUR 2023, 997 (1006 bei Fn. 144); *Sigmüller/Gassner*, RD 2023, 124 (127 Rn. 12, bei Fn. 35); zuletzt ausf. *Hofmann*, WRP 2024, 11 (13 Rn. 15); *Durantaye*, AfP 2024, 9 (14 Rn. 28–30).
- 29 Insb. *Maamar*, ZUM 2023, 481 (483) bezeichnet Art. 4 DSM-RL umstandslos als „Schrankenregelung für maschinelles Lernen“ und meint in ErwGr. 18 DSM-RL das „Ziel“ zu finden, „einen sicheren Rechtsrahmen für KI-Anwendungen zu schaffen“. Beides folgt jedenfalls nicht unmittelbar aus dem Normwortlaut.
- 30 *Margoni/Kretschmer*, GRUR Int. 2022, 685 (686): “by employing an overly broad definition of text and data mining, the provisions under analysis regulate by way of a narrow exception [...] a vast field that includes most forms of modern artificial intelligence (AI) applications”.
- 31 *von Welsler*, GRUR-Prax. 2023, 516 (518 Rn. 22 f.): „diese Bestimmung passt erkennbar bei generativer KI nicht. [...] Die Vorschriften den [sic] TDM tragen dieser Entwicklung nicht Rechnung.“; *Schack*, NJW 2024, 113 (114): „Beim KI-Webscraping geht es jedoch gerade um den geistigen Inhalt der zu Trainingszwecken genutzten Werke und letztlich um die Schaffung inhaltsgleicher oder ähnlicher Konkurrenzzeugnisse. Das ist vom Telos der TDM-Schranke [...] nicht mehr gedeckt.“
- 32 *Schönberger*, ZGE 10 (2018), 35 (56): “a relationship might be seen [...] although ML is much further down the line than TDM”; *Picht/Thouvenin*, IIC 54 (2023), 916 (928): “unclear whether the exceptions also cover the use of copyright content for the development, training, and testing of AI systems”; *Nordemann/Pukas*, JIPLP 2022, 973 (974): “not without a degree of uncertainty”; auch für *Pukas*, GRUR 2023, 614 (615) war „zumindest nicht eindeutig, ob Trainingsdatensammler ihre Aktivitäten tatsächlich auf §§ 44b, 60d UrhG stützen können“; zust. *von Welsler*, GRUR-Prax. 2023, 516 (517) („keineswegs eindeutig“); *Wilmer*, K&R 2023, 233 (234) („könnte [...] in Betracht kommen“); zuletzt *Guadamuz*, GRUR Int. 2024, 111 (120): DSM-Richtlinie “should work to allow some machine learning operations to take place legally, but there will be some room for interpretation depending on the particulars of each situation.”

Die seit 2019 kontrovers geführte Debatte wurde auch dadurch nicht beendet, dass die deutsche Umsetzungsgesetzgeberin Anfang 2021 eine „neue [...] gesetzliche Erlaubnis“ für kommerzielles TDM in § 44b UrhG schuf und gezielt deren „besondere Bedeutung“ für „das maschinelle Lernen als Basis-Technologie für Künstliche Intelligenz“ betonte.³³ Gegner der Subsumtion des KI-Trainings unter die TDM-Definition (in § 44b Abs. 1 UrhG fast unverändert aus der Richtlinie übernommen) beirrte auch diese Gesetzesbegründung kaum.³⁴

Inzwischen dürfte sich der Streit allerdings endgültig erledigt haben.³⁵ Denn das vom EU-Parlament angenommene Gesetz über künstliche Intelligenz (KIG)³⁶ stellt klar:

„KI-Modelle mit allgemeinem Verwendungszweck [...] bedeuten einzigartige Innovationsmöglichkeiten, aber auch Herausforderungen [...] In diesem Zusammenhang können Text- und Data-Mining-Techniken in großem Umfang für das Abrufen und die Analyse solcher Inhalte, die urheberrechtlich und durch verwandte Schutzrechte geschützt sein können, eingesetzt werden.“ (ErwGr. 105 S. 1, 3 KIG)

Damit steht fest, dass die europäische Gesetzgeberin (ebenso wie die deutsche) die TDM-Schranke auch auf das KI-Training angewandt wissen will.³⁷ Um die weitere Untersuchung schlank zu halten, konzentriert sie sich auf das deutsche Umsetzungsgesetz; das Gesagte gilt, wo nicht anders angemerkt, *mutatis mutandis* für die europäische Rechtslage.³⁸

III. Begrenzung der TDM-Schranke im Spiegel ihrer Nah- und Fernziele

Rechtfertigt die Schranke des § 44b UrhG damit grundsätzlich die Datensammlung für kommerzielles KI-Training, kann dies gleichwohl nicht ausnahmslos gelten. Denn „im Rahmen des § 44b UrhG gilt eine Regel, die einige an die Grundrechtsdogmatik erinnern wird: Auch Schranken gelten nicht schrankenlos.“³⁹ Tatsächlich *erinnert* dies nicht nur an die Grundrechtsdogmatik, sondern

33 RegE zum DSM-RL-UmsG, BT-Drs. 19/27426 v. 9.3.2021, S. 60 <dserver.bundestag.de/btd/19/274/1927426.pdf>.

34 Insb. Schack, NJW 2024, 113 (114 Fn. 20) zitiert die Begründung unbeirrt.

35 So auch Bomhard/Siglmüller, RDi 2024, 45 (50 bei Fn. 44): „sodass die letzten Diskussionen dazu nunmehr beendet sein dürften.“; Peukert, GRUR Int. 2024, 497 (503 nach Fn. 88): „final nail in the coffin“; a.A. Geiger (Fn. 25), S. 77: „the discussion is not over“; Guadamuz, GRUR Int. 2024, 111: „growing debate“.

36 Legislative Entschließung des Europäischen Parlaments vom 13. März 2024 zu dem Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz), Dok. P9_TA(2024)0138 verfügbar unter <europarl.europa.eu/doceo/document/TA-9-2024-0138_DE.pdf>.

37 Peukert, GRUR Int. 2024, 497 (503 bei Fn. 90): „EU legislator confirmed this prevailing view qua lex posterior“.

38 Dazu demnächst Hamann, JIPITEC 15 (2024), im Erscheinen; zum KI-Gesetz im hiesigen Kontext schon Durantaye, AfP 2024, 9 (15–17 Rn. 35–40).

39 Baumann, NJW 2023, 3673 (3675 Rn. 17).

ist einfachrechtlich kodifizierte Dogmatik zur Auflösung einer echten Grundrechtskollision in Fällen des Text und Data Mining:

Die Schranke für kommerzielles TDM ist Ausdruck der Forschungs- und Berufsfreiheit der KI-Entwicklerinnen (Art. 5 Abs. 3 S. 1, 12 Abs. 1 GG).⁴⁰ Beide Grundrechte gelten nicht schrankenlos, sondern müssen mit den Persönlichkeits-, Kommunikations- und Eigentumsgrundrechten der Urheberinnen (Art. 1 Abs. 1 i.V.m. Art. 2 Abs. 1, Art. 5 Abs. 1, 3 sowie Art. 14 Abs. 1 GG) in praktische Konkordanz gebracht werden.⁴¹ Denn Urheberrecht im Allgemeinen und die europäische Harmonisierung des digitalen Binnenmarkts im Besonderen sollen einen „angemessenen Rechte- und Interessenausgleich zwischen den Urhebern und [...] Nutzern“ herbeiführen (ErwGr. 6 S. 1 DSM-RL), wozu vor allem der sog. Dreistufentest dient (ErwGr. 6 S. 2 DSM-RL).⁴²

Dieser grundrechtsdogmatisch begründete Ausgleich als Nahziel des § 44b UrhG korrespondiert mit dem Fernziel seiner utilitaristisch begründeten Anreizwirkung. Denn Urheberrecht setzt gerade auch „Anreize für Innovation, Kreativität, Investitionen und die Produktion neuer Inhalte“ (ErwGr. 2 DSM-RL). Wo die TDM-Schranke bezweckt, Nutzerinnen „in der Privatwirtschaft zu Innovationen anzuregen“ (ErwGr. 18 UA 1 S. 4 DSM-RL), muss sie dennoch zugleich den Urheberinnen Produktionsanreize belassen, indem diese „auch künftig Lizenzen für die Nutzung ihrer Werke oder sonstigen Schutzgegenstände erteilen können“ (ErwGr. 18 UA 1 S. 2 DSM-RL).

Folglich sprechen sowohl das grundrechtliche Nahziel (Ausgleich) als auch das utilitaristische Fernziel (Anreiz) für eine Entlastung der Urheberinnen im Gegenzug für ihre Belastung mit der vergütungsfreien TDM-Schranke. Dazu dient eine auf Art. 4 Abs. 3 DSM-RL gestützte Rückausnahme von der Schranke im dritten Absatz der Schrankenvorschrift:

„Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining [...] sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat.“ (§ 44b Abs. 3 S. 1 UrhG)

Dieser sog. Nutzungsvorbehalt soll es Urheberinnen ermöglichen, durch einseitige Erklärung gegenüber Nutzerinnen die TDM-Schranke mit relativer Wir-

40 Zudem schützt die TDM-Schranke (zumindest mittelbar) weitere Grundrechtspositionen wie die Kommunikationsgrundrechte (Art. 5 Abs. 1 GG) und Kunstfreiheit (Art. 5 Abs. 3 S. 1 GG) derjenigen, die „generative“ KI kreativ oder produktiv nutzen; für diesen Hinweis danke ich *Anna Bernzen*.

41 Dazu ausf. *Schack*, Urheber- und Urhebervertragsrecht, 10. Aufl. 2021, § 5.II („Persönlichkeitschutz“), § 5.III („Eigentumsgarantie“) und § 5.IV/V („Kunst- und Schaffens-“ sowie „Informationsfreiheit“). Dass TDM in das Eigentumsgrundrecht eingreift, wird meist eher unterstellt als begründet (vgl. etwa *Durantaye/Raue*, RuZ 2020, 83 [90 bei Fn. 33]; *Kleinkopf* [Fn. 18], S. 85 f., 198 f.; *Brockmeyer*, Text und Data Mining: Eine rechtsökonomische Analyse der neuen Schranken im Urheberrecht, 2022, S. 39, 155), kann hier aber auch nicht vertieft werden.

42 Kodifiziert in Art. 7 Abs. 2 DSM-RL i.V.m. Art. 5 Abs. 5 InfoSoc-RL, sowie Art. 9 Abs. 2 RBÜ.

kung *ex nunc*⁴³ außer Kraft zu setzen („*opt out*“). Dass dies gerade auch im KI-Kontext gilt, ergibt sich aus der Verpflichtung der Anbieter von KI-Modellen,

„eine Strategie [...] insbesondere zur Ermittlung und Einhaltung eines gemäß Artikel 4 Absatz 3 der Richtlinie (EU) 2019/790 geltend gemachten Rechtsvorbehalts, auch durch modernste Technologien, auf den Weg bringen“ (Art. 53 Abs. 1 lit. c KIG)

Auch in den zugehörigen Erwägungsgründen 104 bis 106 des europäischen KI-Gesetzes heißt es, dass „Anbieter von KI-Modellen mit allgemeinem Verwendungszweck“ für Inhalte mit vorbehaltener TDM-Nutzung „eine Genehmigung von den Rechteinhabern einholen“ müssen (ErwGr. 105 S. 7 KIG).

Das Schrifttum resümiert deshalb zu Recht, dass der Nutzungsvorbehalt für KI-bezogenes TDM einen „schonenden Ausgleich zwischen den betroffenen Rechtspositionen“⁴⁴ bezweckt (Nahziel) und zugleich einen „Markt für die Zustimmung zur Nutzung von Werken als KI-Trainingsmaterial“⁴⁵ schaffen soll (Fernziel). Vereinzelt sah man (als drittes Ziel) auch eine „urheberpersönlichkeitsrechtlich“ begründete Abwehrfunktion.⁴⁶

IV. Die Furcht vor dem „Leerlaufen“ – aber wovon?

Umstritten ist, ob die skizzierten Ziele des Nutzungsvorbehalts auch in der Praxis erreicht werden können. Insoweit stehen sich zwei genau gegenläufige Befürchtungen gegenüber:

Ein Lager befürchtet eine „erhebliche Zahl von pro forma-Ausschlussklärungen der Rechtsinhaber“,⁴⁷ „womit die Schranke wohl weitgehend leerlaufen dürfte.“⁴⁸ Schon in der Diskussion zur europäischen Vorbildvorschrift (Art. 4

43 RegE zum DSM-RL-UmsG (Fn. 33), S. 89: „Ein Nutzungsvorbehalt entfaltet Wirkung *ex nunc*.“; zudem Heesen/Jüngels, RuZ 2021, 45 (55): „Der Diskussionsentwurf sah noch vor, dass Rechteinhaber die Nutzung nach § 44b II 1 UrhG-E ‚untersagen‘ können. Durch den jetzt gewählten Begriff ‚vorbehalten‘ wird deutlich, dass diese von Beginn an aktiv tätig werden müssen und nicht etwa rückwirkend eingreifen können.“

44 Hacker, Digitale Marktordnung durch Urheber- und Datenschutzrecht, in: FS Hopt, 2020, S. 351 (367).

45 Maamar, ZUM 2023, 481 (486) m.Verw. auf Margoni/Kretschmer, GRUR Int. 2022, 685 (686, vgl. Zitat in Fn. 30); auch Baumann, NJW 2023, 3673 (3675 Rn. 19) versteht den Nutzungsvorbehalt als „Vehikel, um in Lizenzverhandlungen über die Nutzung der Werke als Trainingsdaten einzutreten“.

46 Insb. Durantaye, AfP 2024, 9 (14 f. Rn. 32) m.Verw. in Fn. 56 auf Durantaye, ZUM 2023, 645 (651) bei Fn. 60: „Nicht wenige Urheberinnen sind prinzipiell dagegen, dass ihre Werke für das Training generativer KI genutzt werden. Sie streben also keine Lizenzlösung an, sondern wollen von ihrem Verbotsrecht Gebrauch machen.“; RegE zum DSM-RL-UmsG (Fn. 33), S. 89: „Sinn und Zweck der Regelung ist es, [...] Rechtsinhabern die Möglichkeit zu eröffnen, die Nutzung auf Basis der gesetzlichen Erlaubnis zu untersagen.“

47 Pukas, GRUR 2023, 614 (615); Hartmann/Jacobsen, MMR-Aktuell 2021, Nr. 441332 in I. a.E.: „dass alle relevanten Anbieter von Inhalten einen entsprechenden Vorbehalt anbringen“; Hofmann, WRP 2024, 11 (14 Rn. 18): „dass dies in Zukunft regelmäßig geschieht“.

48 Spindler, ZGE 10 (2018), 273 (299).

DSM-RL) galt der Nutzungsvorbehalt als „a provision that may very well frustrate its efficacy“,⁴⁹ im Konsultationsverfahren als „empfindliche Einschränkung des kommerziellen TDM“⁵⁰ und die deutsche Umsetzungsgeberin räumte gar verschämt ein, dass sie wegen des Nutzungsvorbehalts die „Auswirkungen der Reform nicht verlässlich prognostizieren“ könne.⁵¹ Noch heute erscheint vielen der Nutzungsvorbehalt als „echte Hürde für Vervielfältigungen im Rahmen von Trainingsvorgängen“,⁵² die die TDM-Schranke „faktisch eines großen Teils des ihr eigentlich zuzuordnenden Anwendungsbereiches berauben“ werde.⁵³ Einzelne fordern sogar rundheraus die Abschaffung des § 44b Abs. 3 UrhG.⁵⁴

Dem steht ein anderes Lager gegenüber, das genau umgekehrt das „Leerlaufen des Opt-Out-Mechanismus“ befürchtet.⁵⁵ Denn Nutzungsvorbehalte können nach der im Einzelnen noch näher zu betrachtenden Formvorschrift des § 44b Abs. 3 Satz 2 UrhG (dazu gleich D.) nur „in maschinenlesbarer Form“ erklärt werden und schon im Gesetzgebungsverfahren wurde darauf hingewiesen, dass es „in der Praxis [...] schwierig“ werde, „von dieser Option im Einzelfall Gebrauch zu machen“:

49 Margoni/Kretschmer, GRUR Int. 2022, 685 (695); Bayamhoğlu, Eur. Rev. Priv. Law 2023, 329 (346): „major shortcoming of the provision which is likely to render it inefficient“; Picht/Thouvenin, IIC 54 (2023), 916 (928): „The scope of these exceptions is therefore limited.“; zuvor schon Geiger/Izyumenko, Am. U. Int. L. Rev. 35 (2019), 1 (18 f.): „will most likely leave the practice of commercial text and data mining for non-research purposes uncertain“; Ducato/Strowel, IIC 50 (2019), 649 (666): „the possibility for rightholders to make some reservations risks subordinating the legislative exception to private will“; Dusollier, CML Rev. 57 (2020), 979 (987): „The exception [...] is thus rather precarious and is subservient to its prohibition by rightholders“; Geiger (Fn. 25), S. 76: „usefulness of this provision might be rather limited [...] can make the provision rather ineffective“; Mezei, Eur. IP Rev. 46 (2024), 461 (464): „We cannot but agree with the reviewers’ frustration with the substance and the practical functionality of these rules.“

50 ZVEI, StN v. Sep. 2019 zur BMJV-Konsultation (Fn. 20), S. 13.

51 RegE zum DSM-RL-UmsG (Fn. 33), S. 73.

52 Hofmann, WRP 2024, 11 (14 Rn. 18), und weiter: „Kurzum: Auf § 44b UrhG können sich Entwickler von KI-Systemen nicht verlassen.“

53 Brockmeyer (Fn. 41), S. 113 bei Fn. 660; ähnl. Hartmann/Jacobsen, MMR-Aktuell 2021, Nr. 441332 vor I.: Der Vorbehalt werde TDM-Nutzungen „praktisch unmöglich machen und damit die Ziele der Schranke in ihr Gegenteil verkehren“.

54 So v.a. Brockmeyer (Fn. 41), S. 166–170; vorsichtiger schon ZVEI, StN v. Sep. 2019 zur BMJV-Konsultation (Fn. 20), S. 13 („Eine Abschaffung des Nutzungsvorbehalts bei kommerziellem TDM wäre ein Punkt für eine künftige Überarbeitung der CDSM-RL“); Mezei, Eur. IP Rev. 46 (2024), 461 (468 f.): „whether the CDSM Directive shall be amended, is far from being certain. [...] In general, Article 4(3) CDSM Directive shall be revisited to provide for more certainty [...] With the end of the von der Leyen Commission’s tenure in 2024, this time is not ‘ideal’ for any such updates.“

55 von Welser, GRUR-Prax. 2023, 516 (519 Rn. 31 f).

„Die wenigsten Urheber:innen verfügen über die nötigen Fähigkeiten und Kenntnisse, einen solchen Vorbehalt in einer maschinenlesbaren Form zu verfassen und dessen Einhaltung zu kontrollieren“⁵⁶

Später beobachtete man „in der Praxis“, dass „auch zwei Jahre nach Inkrafttreten des § 44b UrhG die allermeisten Rechtsinhaber keinen Gebrauch von ihrem Opt-out-Recht machen“.⁵⁷ Es fehle an „Möglichkeiten“, „den Opt-out wirksam zu erklären“, daher seien Nutzungsvorbehalte „in vielen Fällen [...] unwirksam“ und die gegenwärtige Rechtslage werde statt der TDM-Schranke vielmehr den „Nutzungsvorbehalt leerlaufen lassen“, so dass auch „der vom Gesetzgeber gewollte Schutz der Rechtsinhaber faktisch leerläuft.“⁵⁸

Beide Lager treibt die Sorge vor dem „Leerlaufen“, beide äußern gewichtige Befürchtungen. Während ein Leerlaufen der Schranke die Grundrechte von KI-Entwicklerinnen und die Innovationsanreize durch die TDM-Schranke verkürzen könnte, könnte umgekehrt ein Leerlaufen des Nutzungsvorbehalts die Grundrechte der Urheberinnen und die finanziellen Anreize durch die Monetarisierbarkeit ihrer Inhalte für das KI-Training verkürzen. Sorgenträger beider Lager sehen also jeweils beide Ziele der TDM-Schranke (Ausgleich und Anreiz) in Gefahr.⁵⁹ Nur das seltener behandelte urheberpersönlichkeitsrechtliche Ziel (Abwehr) würde durch ein Leerlaufen der Schranke gefördert, durch das Leerlaufen des Nutzungsvorbehalts dagegen gefährdet.

Welche dieser Sorgen berechtigt ist (wenn überhaupt eine), lässt sich freilich kaum pauschal feststellen. Erst die nähere Auseinandersetzung mit den formalen Voraussetzungen eines Nutzungsvorbehalts (dazu gleich D.) und dem Stand der Technik („modernste Technologien“ im Sinne des europäischen KI-Gesetzes), für und durch die Vorbehalte kommuniziert werden (E.), erlaubt eine informierte Einschätzung (F.).

⁵⁶ Deutscher Journalisten-Verband, StN v. 6.11.2020 zum RefE (Fn. 22), S. 8; ähnl. *Mezei*, Eur. IP Rev. 46 (2024), 461 (465, 468): “how such a reservation shall operate in real life is far from clear. Indeed [...] it is a doctrinal and practical minefield. [...] rights reservation is extremely time-consuming and consequently expensive”.

⁵⁷ *Bomhard*, DSRITB 14 (2023), 255 (265).

⁵⁸ *Maamar*, ZUM 2023, 481 (484, 486) – allerdings auch mit Neigungen zum anderen Lager: „Sobald der notwendige Standard für den Opt-out entsteht, ist jedoch mit umfassenden Nutzungsvorbehalten zu rechnen.“

⁵⁹ Hin- und hergerissen auch *Mezei*, Eur. IP Rev. 46 (2024), 461 (462): “Article 4(3) CDSM Directive cannot serve the purpose it was designed for – neither for the benefit of authors (who were the targeted beneficiaries of this provision), nor for the AI industry (whose contribution to humankind’s development is unquestionable).”

D. § 44b Abs. 3 S. 2 UrhG und die H2M-Rechtsgeschäftslehre

Die bereits kurz erwähnte Formvorschrift für den Nutzungsvorbehalt lautet:

„Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt.“ (§ 44b Abs. 3 S. 2 UrhG)

Diese Vorgabe, die manchen als „Musterbeispiel für technologie- bzw. umweltsensibles Recht“ erscheint⁶⁰ und bereits im Gesetzgebungsverfahren für analogiefähig auf digitale, aber nicht online zugängliche Werke gehalten wurde,⁶¹ wirft zugleich zahlreiche dogmatische Detailfragen auf. Sie blieben bislang ungeklärt, weil sich im Gesetzgebungsverfahren „die politische Aufmerksamkeit und die Grundsatzdiskussionen“ auf andere Themen „konzentriert hatten“.⁶² Zu beantworten bleibt insbesondere:

Welche Rechtsfolge und Beweislast ordnet die Vorschrift an? (1.) Was ist unter „maschinenlesbarer Form“ zu verstehen? (2.) Welche zusätzlichen Voraussetzungen neben dem Wortlaut ergeben sich aus der historisch-genetischen Auslegung? (3.) Welche anderen Vorschriften sind bei der systematischen Auslegung zu beachten? (4.) Wer ist befugt, den Nutzungsvorbehalt zu erklären? (5.) Bedarf § 44b UrhG einer richtlinienkonformen Auslegung? (6.)

Erst wenn diese Fragen geklärt sind, ergibt sich der normative Maßstab, anhand dessen konkrete Nutzungsvorbehalte geprüft werden können (7.).

I. Rechtsfolge: „nur dann wirksam, wenn“

Der Wortlaut des § 44b Abs. 3 S. 2 UrhG klingt zunächst wie eine sprachlich minimal variierte Abwandlung des § 125 S. 1 BGB und hätte ohne Weiteres identisch formuliert sein können: *Ein Nutzungsvorbehalt, welcher der maschinenlesbaren Form ermangelt, ist nichtig.* Dass § 44b Abs. 3 S. 2 UrhG dennoch genau anders herum formuliert ist, bleibt nicht ohne Folgen für die Beweislast: Während der Formmangel in § 125 BGB eine Nichtigkeitsvoraussetzung darstellt, deren Voraussetzungen vom dadurch Begünstigten zu beweisen wären, statuiert § 44b UrhG die Einhaltung der Form gerade umgekehrt als positiv festzustellende Wirksamkeitsvoraussetzung.

Damit begünstigt die Formwahrung die Rechteinhaberin und ist im Zweifel von ihr ebenso zu beweisen wie umgekehrt die Nutzerin das Fehlen eines Nut-

⁶⁰ Raue, ZUM 2019, 684 (693); ebenso („gelungenes Beispiel“) später Raue, ZUM 2021, 793 (797).

⁶¹ iRights, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 5: „Da TDM per definitionem technisch erfolgt – durch Software, Algorithmen – sollte diese Regel jedoch auch für auf anderem Weg veröffentlichte digitale Inhalte gelten. Ansonsten kann die Einhaltung solcher Rechtevorbhalte nicht gewährleistet werden.“

⁶² Raue, ZUM 2021, 793 (794) („zum Glück für die datengestützte Forschung“); ähnl. oben bei Fn. 19.

zungsvorbehalts aufgrund § 44b Abs. 3 S. 1 UrhG zu beweisen hätte („wenn der Rechteinhaber sich diese nicht vorbehalten hat“).⁶³ Wo ein Fehlen des Nutzungsvorbehalts nicht erweislich (oder sein Vorhandensein unstreitig) ist, muss die Rechteinhaberin also Umstände darlegen, aus denen sich dessen Maschinenlesbarkeit ergibt.⁶⁴ Weil dies keine geringe Hürde darstellt (dazu nach E.), dürfte § 44b Abs. 3 S. 2 UrhG die aufgrund der „unüberwindbaren Beweishürden“ des Satzes 1 bislang befürchtete Beweisnot für Nutzerinnen⁶⁵ lindern.

Lässt sich die Maschinenlesbarkeit im Rechtssinne nicht feststellen, so ist der Nutzungsvorbehalt im Umkehrschluss aus § 44b Abs. 3 S. 2 UrhG unwirksam. Dass die Gesetzgeberin den Nutzungsvorbehalt überhaupt in Kategorien von Wirksamkeit und Unwirksamkeit definiert (statt etwa von Unbeachtlichkeit oder Unanwendbarkeit), belegt überdies, dass sie ihn als rechtsgeschäftliche Erklärung auffasst. Hierauf wird in Kürze (D.IV. bei Fn. 123) zurückzukommen sein.

II. Tatbestand: „in maschinenlesbarer Form“

Ist die Rechtsfolge der Formvorschrift in § 44b Abs. 3 S. 2 BGB damit konkretisiert, harrt noch ihre Voraussetzung einer Definition. Denn §§ 126–129 BGB kennen keine „maschinenlesbare Form“. Im gesamten Urheberrecht ist sie nirgends legaldefiniert, selbst technisch versierte Rechtsberater beklagen „erhebliche Unklarheit in der Praxis“.⁶⁶ Auch systematische Seitenblicke helfen kaum weiter: Die Datenschutz-Grundverordnung, die im Kontext der Datenportabilität (Art. 20 Abs. 1 DSGVO) und in einem anderen Zusammenhang (dazu gleich nach Fn. 74) Maschinenlesbarkeit erfordert, bleibt eine Konkretisierung dieses Begriffs ebenfalls schuldig.⁶⁷ Sogar neueres vertragstheoretisches Schrifttum zu

63 Vgl. RegE zum DSM-RL-UmsG (Fn. 33), S. 89 („Die Beweislast für das Fehlen eines Nutzungsvorbehalts trägt der Nutzer.“).

64 So auch *Löblich u.a.*, JIPITEC 2023, 499 (501 Rn. 12); *Raue*, ZUM 2021, 793 (797) nimmt nur (aber immerhin) eine „sekundäre Darlegungslast“ eines Rechteinhabers dafür an, „wann und wie er den Vorbehalt erklärt hat“; zust. *Stieper*, ZUM 2021, 776 (778) bei Fn. 25; *Dreier / Schulze / Dreier*, 7. Aufl. 2022, § 44b UrhG Rn. 9.

65 So BRat, StN zum DSM-RL-UmsG, BT-Drs. 19/28171 v. 1.4.2021, Nr. 6 = S. 5 <[dserver.bundestag.de/btd/19/281/1928171.pdf](https://www.bundestag.de/btd/19/281/1928171.pdf)>, und weiter: „Gerade bei sogenannten ‚Webcrawlern‘, die in einer sehr kurzen Zeit eine sehr große Anzahl an Seiten aufrufen, um diese zu analysieren, wäre ein Beweis durch den Nutzer, dass kein entsprechender Vorbehalt bestand ist, nicht zu erbringen.“; genau sinngleich *Kleinkopf* (Fn. 18), S. 250; zust. auch *Brockmeyer* (Fn. 41), S. 172 bei Fn. 984.

66 *Bomhard / Siglmüller*, RD 2024, 45 (50 Rn. 25 dritter Bulletpunkt); zuvor schon *Bomhard*, DSRITB 14 (2023), 255 (266): „höchst unklar, welche genauen Anforderungen [...] zu stellen sind“.

67 Dazu nur *Gola / Heckmann / Piltz*, DSGVO / BDSG, 3. Aufl. 2022, Art. 20 Rn. 25 f.: „Was konkret mit ‚strukturiert‘, ‚gängig‘ und ‚maschinenlesbar‘ gemeint ist, wird in der DS-GVO nicht erläutert. [...] So ist in technischer Hinsicht die Anforderung des Art. 20 nach einem standardisierten Datenformat kaum zu erfüllen.“

sog. *smart contracts* meidet den Begriff der „Maschinenlesbarkeit“ trotz seiner in § 44b UrhG erkannten Relevanz als Rechtsbegriff.⁶⁸

Sprachlich kann „maschinenlesbar“ vieles bedeuten. Gängige Wörterbücher erschöpfen sich zwar in Tautologien,⁶⁹ aber zumindest zwei Extreme im Bedeutungsspektrum lassen sich benennen: Nach dem semantisch aufgeladenen Alltagsverständnis setzt „Lesbarkeit“ Intelligibilität voraus – Leser eines Textes ist nicht schon, wer Bedeutungszeichen und Trägermedium optisch unterscheiden kann, sondern nur, wer das Mitgeteilte sinngemäß erfasst. Für Maschinen „lesbar“ wäre danach nur, was Maschinen „verstehen“ können. Dagegen wäre nach einer weitgefassten technischen Betrachtungsweise alles durch Maschinen „lesbar“, was in ihren Arbeitsspeicher geladen werden kann, weil es dann zumindest für Menschen am Peripheriegerät (Monitor) intelligibel wird. Kurzum: „maschinenlesbar“ kann ebenso gut die Lesbarkeit „für Maschinen“ wie „durch Maschinen“ bezeichnen. Dementsprechend breit wurde das Bedeutungsspektrum des Adjektivs gezogen: Von binärkodierten und unmittelbar ausführbaren Prozessorbefehlen (sog. *nativer Maschinencode*) bis hin zu „jeder digital hinterlegten Information“, die „in einem Internetstandard für Text codiert wurde“.⁷⁰

Manche Autorinnen deuteten das letztgenannte Zitat einseitig verkürzend als „richtige Auslegung“ des § 44b UrhG,⁷¹ so dass ein maschinenlesbarer Nutzungsvorbehalt nur „in Form von durchsuchbarem Text erklärt werden“ müsse.⁷² Unausgesprochene Konsequenz dieser Ansicht wäre ein Zusammenfallen der „maschinenlesbaren“ Form in § 44b Abs. 3 UrhG mit einer digital gewährten „Textform“ i.S.v. § 126b BGB. Noch weiter ging die Ansicht mancher Branchenverbände im Gesetzgebungsverfahren, dass bereits „aus technischen Gründen“ in „Online-Angeboten ein Nutzungsvorbehalt stets maschinenlesbar“, also die Form des § 44b Abs. 3 UrhG „stets als erfüllt“ anzusehen sei:

„Jede Erklärung in Form eines Textes oder eines Piktogramms, die zum Ausdruck bringt, dass der Rechteinhaber von seinem Recht zum Nutzungsvorbehalt Gebrauch macht, erfüllt die Anforderungen“.⁷³

68 Vgl. *Wöbbeking* (Fn. 90), S. 10 Fn. 46.

69 Bspw. Duden Online, Stichwort „Maschinenlesbarkeit“: „das Maschinenlesbarsein“ <duden.de/rechtschreibung/Maschinenlesbarkeit>; Stichwort „maschinenlesbar“: „für einen Computer o. Ä. lesbar“ <duden.de/rechtschreibung/maschinenlesbar>.

70 Anschaulich *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, unter II.2.a), II.2.c).

71 So *Wilmer*, K&R 2023, 233 (234 bei Fn. 19) unter Berufung auf *Hartmann/Jacobsen* (vorige Fn.), die allerdings das von *Wilmer* als „richtige Auslegung“ Zitierte gerade ablehnen (in II.2.c); *Maamar*, ZUM 2023, 481 (484) wiederum beruft sich auf *Wilmer* und übernimmt sogar (in Fn. 29) sein unvollständiges Zitat „Hartmann, MMR-Aktuell 2021, 441332“ (ohne Koautor).

72 So *Akinci/Heidrich*, IPRB 2023, 270 (272).

73 BDZV/VDZ, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 18 = BDZV/VDZ/VDL, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 11 = BDZV/VDZ, StN v. 6.11.2020 zum RefE (Fn. 22), S. 3 = VG Media, StN v. 2.9.2020 zum RefE (Fn. 22), S. 2.

Nach dieser Ansicht wäre schon die strukturelle Kodierung als Text verzichtbar und Nutzungsvorbehalte in einer grafischen Form zulässig, auf die § 126b BGB allenfalls analog angewendet werden könnte.⁷⁴ Für diese Lesart spricht immerhin eine (kaum weniger problematische) europäische Parallelvorschrift, die datenschutzrechtliche Transparenzpflichten durch „standardisierte Bildsymbole [...] in elektronischer Form“ zu erfüllen erlaubt, wenn sie „maschinenlesbar“ sind (Art. 12 Abs. 7 S. 2 DSGVO). Schon diese Vorschrift warf allerdings praktisch kaum zu überwindende Probleme auf und gestattet mitnichten, *jedes* vom Browser darstellbare Piktogramm als maschinenlesbar anzusehen.⁷⁵ Dennoch übernahmen einzelne Kommentatoren des § 44b UrhG diese Extremdeutung wörtlich von der Zeitungsverlegerlobby,⁷⁶ ungeachtet ihrer offenkundigen Interessesgeleitetheit. Ähnliche Interessen prägen wohl das weiteste bislang dokumentierte Verständnis aus Kreisen der amerikanischen Filmindustrie, wonach schlicht „jede Nutzungsuntersagung, die von einem Menschen lesbar ist, auch maschinenlesbar“ sei.⁷⁷

Eine solche strukturelle Beliebigkeit der Form ließe sich, wenn sie § 44b Abs. 3 UrhG nicht geradewegs zur Leerformel degradieren soll, allenfalls durch syntaktische Einschränkungen abmildern, wie einige Autorinnen sie vorschlugen. Statt etwa „auch laienhaft formulierte Vorbehalte“ als maschinenlesbar anzusehen,⁷⁸ sei zumindest eine „einfache und standardisierte Syntax“ zu wählen (etwa die Formulierung „Text und Data Mining vorbehalten“),⁷⁹ oder eine „möglichst universell und leicht verständliche Formulierung [...] immer auch in englischer Sprache (lingua franca)“.⁸⁰

74 Vgl. BeckOGK BGB/*Primaczenko/Frohn*, Stand 1.5.2020, § 126b BGB Rn. 10 („Bildsymbole erfüllen ebenso wenig das Merkmal ‚lesbar‘.“) in Fn. 21 m.Verw. auf Staudinger BGB/*Hertel*, Neub. 2017, § 126b BGB Rn. 25; für punktuelle Analogie etwa *Koller*, TransportR, 11. Aufl. 2023, § 410 Rn. 3 zu „eindeutigen Piktogrammen, die aus dem Kontext ihren Urheber erkennen lassen“.

75 Vgl. *Löblich u.a.*, JIPITEC 2023, 499 (502 Rn. 17); in ähnl. Kontext *Gola/Heckmann/Franck*, DSGVO/BDSG, 3. Aufl. 2022, Art. 12 DSGVO Rn. 52: „Wie sich aus der Formulierung Abs. 7 S. 2 ergibt, sind elektronische Darstellbarkeit und Maschinenlesbarkeit keineswegs deckungsgleich. Es genügt also nicht, wenn die Symbole von einem Browser lediglich dargestellt werden können. Es muss vielmehr möglich sein, die Symbole automatisiert zu erkennen und interpretieren zu lassen.“

76 *Barudi/Steinrötter/Schauer*, Das neue Urheberrecht, 2021, § 44b Rn. 14 (S. 152).

77 Motion Picture Association, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 2; davor warnend zeitgleich LIBER, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 2 (Nr. 8): “anything on a computer screen is theoretically ‘machine readable’ – a PDF, terms and conditions etc. It is vitally important that it is clear this relates to widely used machine readable ‘standards’ [... else] requiring all terms and conditions on a website having to be read and interpreted by a human one by one.”

78 Krit. *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, in II.2.c) mit dem Beispiel eines „Unicode-Textfiles“ mit dem Inhalt „ich will nicht, dass Computer meine Inhalte ausforschen“.

79 Dafür aber, immerhin „neben einem eindeutigen Marker in den Metatags“ (dazu gleich E.V.), *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, in II.3.

80 Dafür *Bomhard*, DSRITB 14 (2023), 255 (266).

Letztlich beruhen alle vorgenannten Ansichten auf dem Versuch, eine „Erklärung in natürlicher Sprache“ als „maschinenlesbar“ einzuordnen, obwohl

„es keine einheitliche Formulierung gibt, die ein Computer prüfen könnte, und schon nicht klar ist, welche Sprache dafür maßgeblich sein soll. Maschinenlesbar ist insofern nur eine technisch codierte Information, wie sie automatische Crawler verarbeiten können.“⁸¹

Damit jedoch sind „Willenserklärungen in natürlicher Sprache zunächst unbeachtlich“.⁸² Denn auszugehen ist vom eben erwähnten Ziel der Maschinenlesbarkeit, eine automatische Verarbeitung durch Crawler (Bots) zu ermöglichen. Dies bestätigen auch frühere Legaldefinitionen des Begriffs im deutschen Recht. Danach liegt ein „maschinenlesbares Format“ vor, wenn

„Softwareanwendungen bestimmte Daten [...] leicht identifizieren, erkennen und extrahieren können“ (§ 2 Nr. 5 IWG 2006–2021)⁸³

„die Daten durch Software automatisiert ausgelesen und verarbeitet werden können“ (§ 3 Nr. 5 DNG)⁸⁴

„die enthaltenen Daten durch Software automatisiert ausgelesen und verarbeitet werden können. Die Daten sollen mit Metadaten versehen werden.“ (§ 12 I 2, 3 EGovG)⁸⁵

Diese Definitionen gelten zwar unmittelbar nur für die öffentliche Verwaltung.⁸⁶ Ihre europäischen Vorbildvorschriften wurden aber verschiedentlich zur Auslegung der urheberrechtlichen „Maschinenlesbarkeit“ herangezogen,⁸⁷ und erscheinen auch technisch zielführend. Auch für § 44b Abs. 3 S. 2 UrhG ist deshalb zu fragen, ob der Nutzungsvorbehalt automatisiert verarbeitet werden kann, ob also eine unmittelbare Weiterverarbeitung durch Software gewährleistet er-

81 *Maamar*, ZUM 2023, 481 (484); zust. *Bomhard*, DSRITB 14 (2023), 255 (266): „Daher spricht viel dafür, dass Erklärung[en] in natürlicher Sprache nicht ‚maschinenlesbar‘ [...] sind, weil es keine einheitliche Formulierung gibt, die ein Computer prüfen könnte.“

82 *Maamar*, ZUM 2023, 481 (484) (das hier weggelassene Attribut „widersprüchliche Willenserklärung“ ist missverständlich, weil der Autor wohl keinen in sich paradoxen Vorbehalt meint, sondern einen der TDM-Nutzung „widersprechenden“).

83 Informationsweiterverwendungsgesetz v. 13.12.2006, BGBl. I S. 2913; zur Vorbildvorschrift Art. 2 Nr. 6 PSI-RL (2003/98) i.d.F. Art. 1 Nr. 2, ErwGr. 21 PSI-Änd-RL (2013/37) = Art. 2 Nr. 13, ErwGr. 35 PSI-II-RL (2019/1024) bereits *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (29): „The term ‚machine readable‘ is also found in other pieces of EU legislation, such as in Art. 2 of the PSI-II Directive“.

84 Datennutzungsgesetz, Art. 2 des G. v. 16.7.2021, BGBl. I S. 2941, 2942.

85 E-Government-Gesetz v. 25.7.2013, BGBl. I S. 2749; dazu schon *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, in II.2.b).

86 § 1 Abs. 1 IWG (bei öffentlichen Stellen vorhandene Informationen), § 2 Abs. 1, 2 DNG (öffentliche Datenbereinsteller), § 1 Abs. 1 EGovG (öffentlich-rechtliche Verwaltungstätigkeit der Bundesbehörden).

87 *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (29): „Article 4 DSMD should be interpreted in combination with the PSI-II Directive“; *Bomhard/Siglmüller*, RD 2024, 45 (50) bei Fn. 45: „Mit Blick auf die Einheit der europäischen Rechtsordnung spricht jedenfalls bei online zugänglichen Werken einiges dafür, Maschinenlesbarkeit iSd ErwG 35 PSI-Richtlinie zu verstehen.“

scheint, ohne dass ein Mensch die abgegebene Erklärung noch von Hand kodieren (d.h. in ein strukturiertes Datenformat überführen) müsste. In der bewährten Typologie elektronischer Willenserklärungen mit ihrer viergliedrigen Unterscheidung zwischen computergestützt erstellten oder übertragenen und ein- oder zweiseitig automatisierten Willenserklärungen⁸⁸ erscheint der Nutzungsvorbehalt folglich als einseitig automatisierte „Mensch-Maschine-Kommunikation“ (Human-to-Machine, H2M).

Das bleibt nicht ohne Einfluss auf die einschlägige Auslegungsmethodik. Während manche Autoren schon davon ausgehen, dass sich die Auslegung einer Erklärung, die „lediglich in einer Programmiersprache abgebildet wird“ und „gegebenenfalls erst übersetzt werden muss“, ohne Weiteres „am objektiven Sinngehalt des Erklärten ausrichten“ müsse,⁸⁹ will eine neuere Ansicht den Auslegungsmaßstab teleologisch kontextualisieren, um auch in formalen Sprachen „nicht an dem buchstäblichen Sinne des Ausdrucks zu haften“ i.S.v. § 133 BGB.⁹⁰ Daher richten sich die zulässigen Auslegungselemente danach, „welche Zwecke mit der Formalisierung im Einzelfall angestrebt wurden“.⁹¹ Das ergibt sich für § 44b Abs. 3 UrhG unmittelbar aus der Gesetzesbegründung: Die Maschinenlesbarkeit sollte „sicherstellen, dass automatisierte Abläufe, die typisches Kriterium des Text und Data Mining sind, tatsächlich auch automatisiert durchgeführt werden können.“⁹² Weil sich Nutzungsvorbehalte also erkennbar an „maschinelle Rezipienten“ richten, kann für ihre Auslegung nur „an den durchschnittlichen Empfänger angeknüpft werden“.⁹³ Denn Zweck jeder derartigen „maschinenlesbaren Mensch-zu-Maschine-Kommunikation“ ist es „gerade“,

„eine möglichst weitgehende Komplexitätsreduktion zu gewährleisten. Daher müssen Umstände, die die konkreten Verhältnisse einzelner Kommunikationsteilnehmer betreffen oder außerhalb der technischen Kommunikation liegen, bei der Auslegung [...] unberücksichtigt bleiben. [Solche] Erklärungen haben ausschließlich die Bedeutung, die ihnen aufgrund der formalen Regeln des Sprachspiels zugewiesen sind. Einer Auslegung unter Berücksichtigung eines konkreten Empfängerhorizonts oder individueller Umstände oder Erklärungen sind sie nicht zugänglich.“⁹⁴

88 *Wiebe*, Die elektronische Willenserklärung, 2002, S. 202 f.; vereinfacht differenziert MüKoBGB/Säcker, 9. Aufl. 2021, Einl. BGB Rn. 190 f. zwischen „automatisierten“ und „elektronisch übermittelten“ Willenserklärungen.

89 *Busche*, Digitale Kommunikation in der Rechtsgeschäftslehre, in: FS Säcker, 2021, S. 15 (25 f.), und weiter: „Insoweit besteht eine gewisse Parallele zu Allgemeinen Geschäftsbedingungen.“

90 Ausf. *Wöbbeking*, Vertragssprache pro machina: Die Auslegung von Smart Contracts und anderen formalsprachlichen Verträgen im System des allgemeinen Vertragsrechts, 2024, S. 211–289.

91 *Wöbbeking* (vorige Fn.), S. 269.

92 RegE zum DSM-RL-UmsG (Fn. 33), S. 89.

93 So zu Recht *Kleinkopf* (Fn. 18), S. 250; a.A. wohl MüKoBGB/Säcker, 9. Aufl. 2021, Einl. BGB Rn. 190, der „keine eigenständige rechtliche Behandlung im Rahmen der Rechtsgeschäftslehre“ zugestehen will – bezeichnenderweise allerdings unter Berufung auf Quellen von 1985–1992 (ebd. in Fn. 486).

94 *Conrad/Schubert*, GRUR 2018, 350 (354 vor Fn. 50) (zur REP-Kommunikation, dazu unten E.II.); ebenso *Wiebe* (Fn. 88), S. 415: „Andere Kontext- und situationsbedingte Faktoren wer-

Kurzgesagt: Auslegung nach § 157 BGB, nicht § 133 BGB.

Um aber den nach der Verkehrssitte i.S.v. § 157 BGB „zugewiesenen“ Sinngehalt formalsprachlicher Äußerungen zu ermitteln, kommt es auf die Spezifikationen der jeweiligen Formalsprache an, die den Äußerungen in dieser Sprache erst ihre Bedeutung zuweisen.⁹⁵ Denn wo Formalsprache bestimmt, was normativ überhaupt gesagt werden kann („Code is Law“⁹⁶), wird die Sprachdefinition maßgeblich für die normative Auslegung:

„Nach den Grundsätzen der normativen Auslegung wird die Bedeutungsermittlung geprägt durch den verwendeten Code [...] so reduziert sich ein Interpretationsspielraum bei der Auslegung auf denjenigen, den der normativ zu verwendende Code läßt. [...] Technische Konventionen ersetzen sprachliche Konventionen.“⁹⁷

Um diese „konventionalisierten“ formalen „Regeln des Sprachspiels“ überhaupt verstehen zu können, muss wiederum klar sein, welche Sprache überhaupt gespielt wird – also welches Protokoll, welchen Standard oder welche Programmierschnittstelle eine Erklärung nutzen soll. Darin nun liegt das dogmatische Hauptproblem des § 44b Abs. 3 UrhG, denn „an einer solchen Spezifikation der Schnittstelle fehlt es in § 44b UrhG.“⁹⁸ Dies war schon im Gesetzgebungsverfahren wiederholt angemahnt worden:

„Hilfreich wäre, wenn der deutsche Gesetzgeber zumindest in der Begründung die (technischen) Anforderungen spezifizieren würde, denen Rechteinhaber gerecht werden müssen, die von einem maschinenlesbaren Nutzungsvorbehalt Gebrauch machen wollen.“⁹⁹

Dennoch finden sich weder im Wortlaut noch in den Gesetzesmaterialien nähere Bestimmungen zum Standard, nach dem der Nutzungsvorbehalt für Maschinen lesbar werden soll. Damit bleibt nichts anderes übrig, als die bisher vorgeschlagenen Standards zu sichten (unten E.) und auf ihre Vereinbarkeit mit den zweckbezogenen Wirksamkeitsvoraussetzungen des § 44b UrhG zu überprüfen.

den dann vollständig ausgeblendet oder in die Nachricht selbst integriert. Das gilt auch für den Zweck als zu berücksichtigenden Faktor“.

95 Beispiele hierfür unten in E.

96 Lessig, Code and Other Laws of Cyberspace, 1999, S. 6 <lessig.org/images/resources/1999-Code.pdf>, in Endnote 7 (S. 241) m. Verw. auf Mitchell, City of Bits: Space, Place, and the Infobahn, 1995, S. 111 u.a. Nachw.; neuere Lit. zu diesem „übermäßig häufig zitierten Ausspruch“ bei Wöbbeking (Fn. 90, S. 99 in Fn. 13).

97 Wiebe (Fn. 88), 414 f., 537 (sowie S. 419: „Das [...] gilt auch für den Fall der Automatisierung nur auf Empfängerseite.“); allg. zur „Auslegung einer Maschinenerklärung“ auch Pieper, GRUR-Prax. 2019, 298 (299) m.Verw. auf BGH, Urt. v. 16.10.2012 – X ZR 37/12.

98 Hartmann/Jacobsen, MMR-Aktuell 2021, Nr. 441332, in II.2.b); Akinci/Heidrich, IPRB 2023, 270 (272): „keine feste Formvorgabe durch das Gesetz“.

99 Börsenverein, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 3 = StN v. 31.1.2020 zum DiskE (Fn. 21), S. 11 = StN v. 5.11.2020 zum RefE (Fn. 22), S. 10; ähnl. Reda, StN v. 23.1.2020 zum DiskE (Fn. 21), S. 8 = Buczek, StN v. 30.1.2020 zum DiskE (Fn. 21), S. 15: „Leider wird hier versäumt, die Anforderung an ein maschinenlesbares Opt-Out für Online-Inhalte näher zu definieren.“

III. Normgenese: Ausdrücklichkeitserfordernis (Art. 4 Abs. 3 DSM-RL)

Eine weitere Wirksamkeitsvoraussetzung hat ihren Weg nicht von der europäischen Richtlinie bis in den Wortlaut von § 44b Abs. 3 S. 2 UrhG gefunden,¹⁰⁰ womit die deutsche Umsetzungsgesetzgeberin in Europa allerdings nicht allein-zustehen scheint.¹⁰¹ Dafür ergibt sie sich immerhin deutlich aus den Gesetzgebungsmaterialien und lässt sich daher im Rahmen der historisch-genetischen Auslegung berücksichtigen:¹⁰²

„Ein Nutzungsvorbehalt muss ausdrücklich erklärt werden“¹⁰³

Damit bezog sich die Umsetzungsgesetzgeberin unausgesprochen auf den Wortlaut der europäischen Vorbildregelung (Art. 4 Abs. 3 DSM-RL), die ebenfalls einen „ausdrücklich“ erklärten Nutzungsvorbehalt erfordert.¹⁰⁴ Diesen Begriff definiert die Richtlinie zwar nicht und verwendet ihn auch nur an einer einzigen anderen Stelle.¹⁰⁵ Deswegen lässt sich aber auch nicht behaupten, seine Bedeutung sei „völlig unklar“.¹⁰⁶

Denn zunächst kann kein Zweifel daran bestehen, dass das Ausdrücklichkeits-erfordernis „nur aus den Umständen ersichtliche, konkludente Nutzungsvor-behalte“ ausschließt.¹⁰⁷ Darüber jedoch geht der europäische Ausdrücklichkeits-begriff hinaus: Nach dem zugehörigen Erwägungsgrund sollen auch „andere Nutzungen [...] von dem Rechtsvorbehalt für die Zwecke des Text und Data Mining nicht betroffen sein.“ (ErwGr. 18 UA 2 S. 3 DSM-RL). Aus dieser Vorgabe schloss die Umsetzungsgesetzgeberin zu Recht:

„Ein Nutzungsvorbehalt nach § 44b Absatz 3 UrhG-E für eine Webseite darf nicht dazu führen, dass diese im Rahmen anderer Nutzungen ohne sachliche Rechtfertigung ungleich behandelt wird, beispielsweise bei der Anzeige als Suchmaschinentreffer.“¹⁰⁸

¹⁰⁰ Daher meinte *Raue*, ZUM 2020, 172 (173), sie „sollte aus Gründen der Rechtsklarheit und Rechtssicherheit ergänzt werden.“

¹⁰¹ Krit. schon *Margoni / Kretschmer*, GRUR Int. 2022, 685 (695) zu “linguistic divergences in its [Art. 4’s] transposition”, insb. der “omission of the ‘express’ element [of] the ‘express reservation’ mechanism” (Fn. 75); auch *Bomhard / Siglmüller*, RD 2024, 45 (50 Rn. 25 dritter Bullet-punkt) zum „Nutzungsvorbehalt, der in den Mitgliedstaaten unterschiedlich umgesetzt ist“.

¹⁰² A.A. *Raue*, ZUM 2020, 172 (173): „Klarstellung in der Gesetzesbegründung reicht dafür nicht aus.“

¹⁰³ RegE zum DSM-RL-UmsG (Fn. 33), S. 89.

¹⁰⁴ Zitiert unten in D.VI.; auf Englisch heißt es “expressly reserved”, auf Französisch «*expressément réservée*».

¹⁰⁵ ErwGr. 69 S. 2 DSM-RL: „Wenn Rechteinhaber Nutzern ausdrücklich die Erlaubnis erteilt haben, ...“; auf Englisch heißt es hier “explicitly authorised”, auf Französisch «*autorisé exp-pressément*».

¹⁰⁶ So aber *Bomhard*, DSRITB 14 (2023), 255 (265) (zumindest für nicht online zugängliche Werke).

¹⁰⁷ So zu Recht *Raue*, ZUM 2021, 793, (796) a.E.

¹⁰⁸ RegE zum DSM-RL-UmsG (Fn. 33), S. 89; zust. *Dreier / Schulze / Dreier*, 7. Aufl. 2022, § 44b UrhG Rn. 9.

Diese Formulierung wurde kritisiert, weil sie als übergreifige Überdehnung des Urheberrechts erscheinen könnte.¹⁰⁹ Bei wohlwollender Interpretation wollte die Gesetzgeberin hier aber lediglich das Ausdrücklichkeitserfordernis dahin konkretisieren, dass es konkret auf eine Nutzung zu TDM-Zwecken zugeschnitten sein muss. Mithin muss der Nutzungsvorbehalt sowohl *expressis verbis* (nicht-konkludent) als auch so zielgenau (konkret-individuell) erklärt werden, dass er zweifelsfrei einen bestimmten Inhalt und eine bestimmte Nutzung (TDM) erfasst. Damit sind insbesondere

„komplexe, verschachtelte oder nicht auf der konkreten Seite der Inhalte abrufbare Vorbehalte oder solche, die nicht ausdrücklich auf das Text und Data Mining abstellen, unangemessen und damit nach § 44b Abs. 3 S. 2 UrhG unwirksam.“¹¹⁰

IV. Systematik: „Vorbehalt der Rechte“ in § 49 Abs. 1 S. 1 UrhG

Man hat gesagt, der Nutzungsvorbehalt kehre „im Ergebnis den Regel-Ausnahme-Fall einer zustimmungsbedürftigen Nutzung“ um, weil er Rechteinhaberrinnen ein Tätigwerden zur Wahrung ihrer Rechte abverlange.¹¹¹ Dabei ist die Vorschrift keineswegs die erste im Urheberrechtsgesetz, die einen solchen „opt out“-Mechanismus nutzt. Das scheint nur in Genese und Rezeption des § 44b Abs. 3 UrhG bislang weitgehend übersehen worden zu sein.

Der Entwurf des Urheberrechtsgesetzes von 1965 enthielt gleich zwei solcher Opt-Out-Vorschriften: Sowohl die Schranke des § 5 Abs. 2 UrhG-E für „andere amtliche Werke, die im amtlichen Interesse zur allgemeinen Kenntnisnahme veröffentlicht worden sind“, als auch die Schranke des § 49 Abs. 1 S. 1 UrhG-E für Zeitungsartikel, die „politische, wirtschaftliche oder religiöse Tagesfragen betreffen“, sollten nur gelten, wenn die betreffenden Werke „nicht mit einem Vorbehalt der Rechte versehen sind“.

Dies begründete die Bundesregierung im Fall von § 5 UrhG damit, dass Behörden, die amtliche Werke veröffentlichen, deren „Nachdruck [...] überwachen“ können müssten, „etwa weil ein ungenauer oder entstellter Nachdruck beson-

109 *Raue*, ZUM 2021, 793 (797): „Allerdings lässt sich eine solche Vorgabe nicht aus dem UrhG, sondern bei entsprechender Marktmacht nur aus dem Kartellrecht, aus (mittelbaren) grundgesetzlichen Gleichbehandlungspflichten oder anderen Diskriminierungsverboten ableiten.“

110 *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, in II.3 a.E.; ebenso für Nutzungsvorbehalte an „unüblicher Stelle“ *Raue*, ZUM 2021, 793 (797) und *Kleinkopf* (Fn. 18), S. 249 bei Fn. 950.

111 *Siglmüller/Gassner*, RD 2023, 124 (127 Rn. 12); *Maamar*, ZUM 2023, 481 (483): „dreht damit das aus dem Urheberrecht bekannte System um“; zuvor schon *Hacker*, in: FS Hopt 2020, S. 351 (365): „Die normative Richtung der dispositiven Regel wird daher gegenüber dem sonst für Lizenzverträge anwendbaren Gefüge umgekehrt.“; *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, unter I. a.E.: „Umdrehung von Regel und Ausnahme“.

ders nachteilige Folgen haben würde“.¹¹² Die Möglichkeit zum Rechtevorbekhalt in § 49 UrhG hingegen wurde schlicht mit Vorgängervorschriften des älteren Urheberrechtsgesetzes (LUG)¹¹³ und der Berner Übereinkunft (RBÜ)¹¹⁴ begründet.¹¹⁵

Der in § 5 Abs. 2 UrhG-E vorgesehene Rechtevorbekhalt wurde noch während des Gesetzgebungsverfahrens vom Rechtsausschuss durch ein Änderungsverbot i.S.v. § 62 UrhG ersetzt,¹¹⁶ weil ein pauschaler Rechtevorbekhalt „auch aus rein fiskalischem Interesse“ benutzt zu werden drohe, so dass die „ungehinderte Verbreitung“ der im allgemeinen Interesse veröffentlichten amtlichen Werke gefährdet sei.¹¹⁷ Die zweite unter Rechtevorbekhalt gestellte Schranke hingegen (§ 49 UrhG-E) wurde sogar noch auf Rundfunkkommentare ausgedehnt¹¹⁸ und lautet heute nur minimal geändert:

„Zulässig ist die Vervielfältigung und Verbreitung [...] sowie die öffentliche Wiedergabe solcher Kommentare, Artikel und Abbildungen, wenn sie politische, wirtschaftliche oder religiöse Tagesfragen betreffen und nicht mit einem Vorbehalt der Rechte versehen sind.“ (§ 49 Abs. 1 S. 1 UrhG)

Inzwischen dient § 49 Abs. 1 S. 1 UrhG auch zur Umsetzung von Art. 5 Abs. 3 lit. c InfoSoc-RL, dessen Formulierung noch stärker an § 44b Abs. 3 UrhG erinnert, weil er die gesetzliche Lizenz gewährt, „sofern eine solche Nutzung nicht ausdrücklich vorbehalten ist“.

Dennoch spielte § 49 UrhG für die systematische Auslegung des § 44b UrhG bislang kaum eine Rolle. Die Gesetzgeberin erwähnte nur ganz beiläufig die Möglichkeit von „Nutzungsvorbehalten auf Basis sonstiger Rechtsgrundlagen, wie [...] § 49 Absatz 1 Satz 1 UrhG“,¹¹⁹ zog aber nicht in Erwägung, diese Rechtsgrundlage systematisch vergleichend heranzuziehen. Nur eine Stellungnahme im

112 RegE zum UrhG, BT-Drs. IV/270 v. 23.3.1962, S. 39 <dserver.bundestag.de/btd/04/002/040270.pdf>.

113 § 18 S. 1 Hs. 1 LUG: „Zulässig ist der Abdruck einzelner Artikel aus Zeitungen, soweit die Artikel nicht mit einem Vorbehalte der Rechte versehen sind“.

114 Die Gesetzgeberin von 1965 berief sich noch auf Art. 9 Abs. 2 S. 1 der Brüsseler Fassung von 1948, die inhaltlich Art. 10^{bis} RBÜ entspricht: „Der Gesetzgebung der Verbandsländer bleibt vorbehalten, die Vervielfältigung [...] von Artikeln über Tagesfragen wirtschaftlicher, politischer oder religiöser Natur; [...] zu erlauben, falls die Vervielfältigung [...] nicht ausdrücklich vorbehalten ist.“; näher hierzu demnächst *Hamann*, JIPITEC 15 (2024), im Erscheinen.

115 RegE zum UrhG (Fn. 112), S. 66.

116 RsA-Antrag zum UrhG-E, BT-Drs. IV/3401 v. 10.5.1965, S. 3 <dserver.bundestag.de/btd/04/034/0403401.pdf>.

117 RsA-Bericht (*Reischl*) zum UrhG-E, BT-Drs. zu IV/3401 v. 14.5.1965, S. 2 <dserver.bundestag.de/btd/04/034/0403401zu.pdf>.

118 RsA-Antrag zum UrhG-E (Fn. 116), S. 16 und RsA-Bericht (*Reischl*) zum UrhG-E (Fn. 117), S. 7: („Im Interesse einer Gleichbehandlung von Presse und Rundfunk hält es der Ausschuss für geboten, eine entsprechende Einschränkung des Urheberrechts auch für Rundfunkkommentare [...] vorzusehen“.

119 RegE zum DSM-RL-UmsG (Fn. 33), S. 89, und weiter: Solche „kann der Rechtsinhaber weiterhin erklären, gegebenenfalls zusätzlich zu einem Vorbehalt nach § 44b Absatz 3 UrhG-E.“

Gesetzgebungsverfahren regte für § 44b UrhG eine Formulierung an, „die dem Rechtevorbekalt in § 49 UrhG nachgebildet ist.“¹²⁰ Auch das spätere Schrifttum zog § 49 UrhG nur ganz vereinzelt und nur für einzelne Fragen (etwa nach Rechtsnatur oder Wortlaut des Vorbehalts) heran.¹²¹

Dabei ergibt die Diskussion zu § 49 UrhG durchaus relevante Einsichten, die in der weiteren Untersuchung an passender Stelle¹²² fruchtbar gemacht werden können. Hier sei vorerst nur darauf hingewiesen, dass dies auch die bereits angesprochene rechtsgeschäftliche Rechtsnatur des Nutzungsvorbehalts (D.I a.E.) betrifft. Zu § 49 UrhG ist nämlich umstritten, ob der Rechtevorbekalt als geschäftsähnliche Handlung anzusehen ist, auf die „die allgemeinen Vorschriften über Willenserklärungen entsprechende Anwendung finden.“¹²³ Die Gegenauffassung wurde dort allerdings nur mit der durch die Vorschrift bezweckten Rechtssicherheit begründet, die eine „Verteidigung des Entlehners im Prozess mit dem Argument“ ausschließen müsse, „der Vorbehalt sei mangels Geschäftsfähigkeit des Urhebers unwirksam.“¹²⁴ Dass jedoch eine ganz grundsätzliche Umqualifizierung der Rechtsnatur erforderlich sein soll, nur um im Bestreitensfall keinen Rechtsfähigkeitsbeweis führen zu müssen, leuchtet kaum ein.

V. Das klärungsbedürftige Problem der Erklärungsbezugnis

Aus dem systematischen Seitenblick auf § 49 UrhG ergibt sich zudem eine Rechtsfrage, die im Kontext des § 44b Abs. 3 UrhG bislang kaum thematisiert wurde. Gestritten wird nämlich darüber, ob der Rechtevorbekalt in § 49 Abs. 1 S. 1 UrhG wirksam „nur vom Urheber, dessen Rechtsnachfolger oder deren Bevollmächtigten erklärt werden“ kann,¹²⁵ oder „sogar der von einem Nichtberechtigten erklärte“ Vorbehalt wirksam ist, weil die Vorschrift „nur auf das objektive Vorhandensein des Vorbehalts abstellt.“¹²⁶ Dieses Problem stellt sich für den Nut-

120 BDZV/VDZ/VDL, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 10.

121 Bei Stieper, ZUM 2021, 776 (777) für die Rechtsnatur („wie bei § 49 Abs. 1 S. 1 UrhG a. E. im Sinne einer Erklärung“), bei Maamar, ZUM 2023, 481 (485) für den Wortlaut („Zu dem vergleichbaren Nutzungsvorbekalt bei § 49 UrhG ist anerkannt, dass der pauschale Vermerk ‚Alle Rechte vorbehalten‘ ausreichend ist.“); etwas ausführlicher zu der in Fn. 114 zitierten Vorbildvorschrift Art. 10^{bis} RBÜ immerhin noch *Senfleben*, IIC 54 (2023), 1535 (1545); zur Vorsicht insoweit mahnt Ginsburg, Berne-Forbidden Formalities and Mass Digitization, Boston L. Rev. 96 (2016), 745 (759): „The drafting history of the Berne Convention indicates that art. 10bis(1) is a ‘lex specialis,’ [... that] does not create a basis for generalization“.

122 Dazu unten bei Fn. 126, 179, 211.

123 Dafür Schrickler/Loewenheim/Melichar/Stieper, 6. Aufl. 2020, § 49 UrhG Rn. 24.

124 So Heidelb. Komm. UrhR / Dreyer, 4. Aufl. 2018, § 49 UrhG Rn. 19.

125 So Schrickler/Loewenheim/Melichar/Stieper, 6. Aufl. 2020, § 49 UrhG Rn. 24.

126 So BeckOK UrhR (vorm. Möhring/Nicolini)/Engels, 41. Ed. 2024, § 49 UrhG Rn. 14 mit dem etwas konstruierten Beispiel eines „vom Setzer eigenmächtig hinzugefügten Vorbehalts“; ebenso Heidelb. Komm. UrhR / Dreyer, 4. Aufl. 2018, § 49 UrhG Rn. 19: „Der Vorbehalt muss nicht vom [sic] dem Urheber, sondern kann auch vom Verleger oder Herausgeber stammen,

zungsvorbehalt nicht in gleicher Weise, weil § 44b Abs. 3 S. 1 UrhG voraussetzt, dass „der Rechtsinhaber sich [...] vorbehalten“ hat, wo § 49 Abs. 1 S. 1 UrhG nur danach fragt, ob Inhalte „mit einem Vorbehalt [...] versehen“ sind.

Der durch Nichtberechtigte eigenmächtig hinzugefügte Nutzungsvorbehalt – etwa von der redaktionell nicht verantwortlichen Webmasterin einer Internetseite – genügt demnach für § 44b UrhG nicht. Denn obwohl der Nutzungsvorbehalt keine Willenserklärung darstellt (deren Rechtsfolgen einträten, weil sie gewollt sind), sondern „nur“ eine rechtsgeschäftsähnliche Handlung (die unabhängig vom Willen des Erklärenden Rechtsfolgen zeitigt), muss der Erklärungstatbestand dennoch vom Befugten erfüllt werden – ganz ähnlich wie auch die Mahnung eines völlig Fremden keinen Verzug begründen kann.

Damit verschiebt sich das Problem allerdings nur auf die Frage nach der rechtlichen Befugnis, also „wer den Vorbehalt erklären darf“:¹²⁷ Sind auch diejenigen, die (ausschließliche) Nutzungsrechte von der Urheberin ableiten, zum Nutzungsvorbehalt befugt? Soweit die Frage im Kontext von § 44b UrhG überhaupt erörtert wird, heißt es mit Blick auf die vermögensrechtliche Komponente des Vervielfältigungsrechts:

„Den Nutzungsvorbehalt dürfen des Weiteren nach Art. 4 Abs. 3 DSM-RL die ‚jeweiligen Rechteinhaber‘ erklären, weshalb nicht nur der Urheber als solches hierzu berechtigt ist, sondern auch all jene, denen Nutzungsrechte im Sinne der Art. 31 ff. UrhG [sic] eingeräumt wurden.“¹²⁸

Dabei ist die Übertragungszwecklehre (§ 31 Abs. 5 S. 1 UrhG) zu berücksichtigen, so dass allein die in Verlagsverträgen gebräuchliche pauschale Übertragung der „räumlich und zeitlich unbeschränkten Rechte“ für alle Nutzungsformen einschließlich der „heute noch nicht bekannten“ keine rechtssichere Vorbehaltsbefugnis begründen dürfte. Denn richtigerweise stellt „die Vervielfältigung urheberrechtlich geschützter Werke, um sie als KI-Trainingsdaten zu verwenden, eine eigenständige Nutzungsart“ dar,¹²⁹ für die die Beschränkungen bei Einräumung unbekannter Nutzungsarten (§ 31a UrhG) gelten. Zwar liegt im bloßen Nutzungsvorbehalt noch keine eigene TDM-Nutzung, doch dient der Vorbehalt gerade als Vorbereitungshandlung für eine solche Nutzung und muss daher zumindest von § 31a UrhG analog erfasst werden. Mithin kann eine durch pau-

selbst wenn dieser nicht dazu ermächtigt ist. Er muss nicht einmal mit Billigung des Urhebers angebracht worden sein.“

127 So *Durantaye*, AfP 2024, 9 (14 Rn. 32), und weiter: „Verwerter:innen und Urheber:innen werden sie [diese Frage] nicht unbedingt gleich beantworten.“; mit Bezug darauf *Peukert*, GRUR Int. 2024, 497 (503 bei Fn. 91).

128 *Brockmeyer* (Fn. 41), S. 79; *Vesala*, IIC 54 (2023), 351 (357): “it can also be unclear whether reservations have been made by rightholders themselves or at their behest, or only by a service provider (in which case they would not prevent mining).”

129 *Pukas*, GRUR 2023, 614 (617 bei Fn. 32); ebenso *Schack*, NJW 2024, 113 (115 Rn. 16): „Bei der Vervielfältigung zu KI-Trainingszwecken handelt es sich um eine neue eigenständige Nutzungsart.“; zögernd noch *de la Durantaye*, ZUM 2023, 645 (651 bei Fn. 58), die das für „unklar“ hält.

schale Vertragsklauseln gebundene Urheberin die Befugnis zum Nutzungsvorbehalt entsprechend § 31a Abs. 1 S. 3 UrhG wieder an sich ziehen, solange keine gemeinsame Vergütungsregel nach §§ 31a Abs. 2 S. 2, 36 UrhG, keine schriftliche Mitteilung nach § 31a Abs. 1 S. 4 UrhG und keine gesonderte Nachtragsvergütung nach §§ 31a Abs. 2 S. 1, 32c Abs. 1 S. 1 UrhG erfolgt sind. Folglich dürften große Repertoireinhaberinnen (wie die Inhaberin der Verlagswebsite im Eingangsbeispiel) kaum ohne Einzelfallprüfung Nutzungsvorbehalte für Inhalte erklären können, die sie in der Vergangenheit unter Vertrag genommen haben. Erst seit wenigen Monaten scheinen Verlagsverträge auch ausdrücklich das Recht zum Nutzungsvorbehalt zu umfassen.¹³⁰

Ähnliches gilt für die Erklärungsbefugnis der Verwertungsgesellschaften. Der aus Dezember 2021 stammende Muster-Wahrnehmungsvertrag für Sprachwerke kann das kommerzielle TDM noch gar nicht erfassen – und sichert der VG Wort deshalb nur „den Vergütungsanspruch für Text und Data Mining (§§ 60d, 60h Abs. 1 UrhG)“ (§ 1 Abs. 1 Nr. 28 VGW-WahrnV),¹³¹ den es in der realen Welt gar nicht gibt (§ 60h Abs. 2 Nr. 3 UrhG). Der Berechtigungsvertrag für die Musikindustrie hingegen wurde zuletzt im Mai 2023 beschlossen und umfasst nun auch den TDM-Nutzungsvorbehalt:

„Die Rechteübertragung umfasst auch die Befugnis, Nutzungsvorbehalte gemäß § 44b Abs. 3 UrhG gegen Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining zu erklären.“ (§ 1 S. 1 lit. h S. 2 GEMA-BerechV)

Basierend auf dieser Bestimmung wird vertreten, dass die GEMA „mit Flächenwirkung für bestimmte Werkarten“ Nutzungsvorbehalte erklären könne.¹³² Dagegen wird unter ökonomischen Gesichtspunkten eingewandt, dass bei „Trainingsdatensätzen mit Milliarden von Werken“ auch Lizenzverhandlungen mit großen „Urhebervereinigungen“ – ganz zu schweigen von „einzelnen Urhebern“ – „schon wegen der hohen Transaktionskosten ineffizient“ wären:

„Das Gesetz droht so einen finanziellen Anreiz dafür zu setzen, dass die KI-Systeme der Zukunft mit älteren gemeinfreien Werken, freien Inhalten ohne Nutzungsvorbehalt oder KI-generiertem Material trainiert werden, was die Qualität von in der EU entwickelten KI-Anwendungen beeinträchtigen kann.“¹³³

130 An einer belastbaren empirischen Tatsachengrundlage fehlt es hier ebenso wie im gesamten Urhebervertragsrecht (dazu jüngst *Kankanhalli/Kwan*, J. Emp. Legal Stud. 21 [2024], 109); die Beobachtung stützt sich daher auf die letzten vom Verf. selbst abgeschlossenen Verlagsverträge.

131 Verfügbar unter vgwort.de/fileadmin/vg-wort/pdf/Wahrnehmungsvertrag/Wahrnehmungsvertrag_Muster_Dez_2021.pdf.

132 *Kögel*, DSRITB 14 (2023), 285 (289 Fn. 15); auch *Pukas*, GRUR 2023, 614 (615) zur „Möglichkeit, den Nutzungsvorbehalt auch durch kollektive Rechtewahrnehmungsorganisationen auszuüben. Ein entsprechendes Recht hat sich die GEMA bereits einräumen lassen“.

133 *Maamar*, ZUM 2023, 481 (486 bei Fn. 52).

Abgesehen von ökonomischen Einwänden lassen sich auch Bedenken nicht von der Hand weisen, dass Repertoireinhaber und Nutzungsberechtigte schon wegen einer persönlichkeitsrechtlich begründbaren – und damit indisponiblen – Abwehrfunktion des Nutzungsvorbehalts (dazu oben bei Fn. 46) keine ausschließliche Vorbehaltsbefugnis erwerben können. Jedenfalls bleibt die noch ungeklärte Frage nach der Erklärungsbefugnis eine echte dogmatische Hürde für wirksame Nutzungsvorbehalte.

VI. Richtlinienkonforme Auslegung: „Inhalte“ statt „Werke“

Aus verschiedenen Gründen wird immer wieder die Richtlinienwidrigkeit des § 44b Abs. 3 S. 2 UrhG behauptet.¹³⁴ Das europäische Vorbild Art. 4 Abs. 3 DSM-RL lautet:

„Die Ausnahmen und Beschränkungen [für kommerzielles TDM] finden Anwendung, sofern die jeweiligen Rechteinhaber die [...] genannten Werke und sonstigen Schutzgegenstände nicht ausdrücklich in angemessener Weise, etwa mit maschinenlesbaren Mitteln im Fall von online veröffentlichten Inhalten, mit einem Nutzungsvorbehalt versehen haben.“

Tatsächlich weicht die Umsetzung durch § 44b Abs. 3 S. 2 UrhG in mehrfacher Hinsicht von dieser Vorgabe ab und bedarf einer richtlinienkonformen Auslegung – allerdings ganz anders als bislang angenommen.

Zunächst ergibt sich die Richtlinienwidrigkeit nicht daraus, dass § 44b Abs. 3 UrhG im Gegensatz zu Art. 4 Abs. 3 DSM-RL keinen „ausdrücklichen“ Nutzungsvorbehalt verlangt. Insoweit wurde bereits festgestellt, dass das nicht ausgesprochene Ausdrücklichkeitserfordernis schon aufgrund genetischer Auslegung in die Norm hineinzulesen ist (oben D.III.). Viel häufiger stützte sich der Vorwurf der Richtlinienwidrigkeit deshalb auf die vermeintlich zu strenge Umsetzung der Formvorschrift in § 44b Abs. 3 S. 2 UrhG. Zwar lässt sich das nicht schon darauf stützen, dass § 44b Abs. 3 S. 2 UrhG eine maschinenlesbare „Form“ vorsieht, wo Art. 4 Abs. 3 DSM-RL die Maschinenlesbarkeit lediglich auf die „Mittel“ eines Nutzungsvorbehalts bezieht (ebenso ErwGr. 18 UA 2 S. 2 Hs. 1 DSM-RL), denn dabei handelt es sich um einen rechtlich folgenlosen terminologischen Unterschied. Ebensowenig liegt ein Umsetzungsdefizit darin, dass diese

¹³⁴ Außer Betracht bleibt hier die ebenfalls behauptete Richtlinienwidrigkeit von Satz 1 wegen der dort ausgesprochenen Beweislast der Nutzerin für das Fehlen eines Nutzungsvorbehalts (oben bei Fn. 63), die sich „angesichts der Formulierung des Artikel 4 Absatz 3 der DSM-Richtlinie nicht aufrechterhalten“ lasse (so BRat, StN zum DSM-RL-UmsG [Fn. 65], Nr. 6 = S. 5), weil „die DSM-RL in Art. 4 Abs. 3 DSM-RL keine Beweislastverteilung einführen sollte“ (Kleinkopf [Fn. 18], S. 250 f.; zuvor noch „unklar“ laut Kleinkopf/Pflüger, ZUM 2021, 643 [646 bei Fn. 28]). Da es sich um eine prozessuale Frage handelt, führt eine richtlinienkonforme Auslegung des materiellen Rechts ohnehin nicht weiter. Ebenfalls außer Betracht bleibt die Frage, ob die *ex-nunc*-Wirkung von § 44b UrhG (oben bei Fn. 43) mit der Richtlinie vereinbar ist; zu den „conflicting arguments that make the discussion of this issue extremely hard“ nun ausf. Mezei, Eur. IP Rev. 46 (2024), 461 (466).

Mittel ihrerseits in Art. 4 Abs. 3 DSM-RL nur als Regelbeispiel („etwa mit“) für die „angemessene Weise“ eines Nutzungsvorbehalts genannt sind und dass die vom Bundesrat vorgeschlagene Aufnahme dieser Generalklausel in § 44b Abs. 3 UrhG¹³⁵ von der Bundesregierung mit unklarer Begründung abgelehnt wurde.¹³⁶ Denn dass die Erklärungsweise „angemessen“ sein muss, folgt im nationalen Recht schon aus dem Rechtsgedanken von Treu und Glauben in §§ 157, 162, 226, 242 BGB, die auf den Nutzungsvorbehalt wegen seiner Rechtsgeschäftsähnlichkeit (dazu oben D.I.) zumindest entsprechend anwendbar sind.

Wohl aber könnte eine Richtlinienwidrigkeit daraus folgen, dass § 44b Abs. 3 S. 2 UrhG bei Onlineinhalten die Maschinenlesbarkeit als Wirksamkeitsvoraussetzung behandelt (s.o. D.I.), während Art. 4 Abs. 3 DSM-RL nur von Anwendbarkeit spricht („finden Anwendung“) und nach Ansicht von Branchenverbänden auch „im Falle von online veröffentlichten Werken“ die „maschinenlesbare Form nur als Beispiel eines angemessenen Mittels“ nenne.¹³⁷ Deshalb insistierten etliche Interessengruppen während des Gesetzgebungsverfahrens, dass die zwingende Form des § 44b Abs. 3 S. 2 UrhG „zu eng“ sei und der Richtlinie widerspreche, die „diese strikte Anordnung“ nicht treffe.¹³⁸

Dieses Missverständnis löst sich indessen schon bei genauer Normlektüre auf. In der Formulierung „in angemessener Weise, etwa mit maschinenlesbaren Mitteln im Fall von online veröffentlichten Inhalten“ bezieht sich die Partikel „etwa“ keineswegs auf „maschinenlesbare Mittel“¹³⁹ (und ließe dementsprechend alternative Erklärungsweisen zu),¹⁴⁰ sondern auf „online veröffentlichte Inhalte“ (und bestätigt so seine Anwendbarkeit auf andere Inhalte). Sprachlich weniger missverständlich wäre also folgende Umstellung gewesen: „in angemessener Weise, etwa im Fall von online veröffentlichten Inhalten mit maschinenlesbaren Mitteln“.

135 BRat, StN zum DSM-RL-UmsG (Fn. 65), Nr. 6 = S. 5: „§ 44b Absatz 3 Satz 1 ist wie folgt zu fassen: ‚Nutzungen nach Absatz 2 Satz 1 sind nicht zulässig, wenn der Rechteinhaber sich diese in angemessener Weise vorbehalten hat.‘“; krit. auch *Raue*, ZUM 2021, 793 (796): „Es hätte die Rechtssicherheit erhöht, wenn der deutsche Gesetzgeber die entsprechende Vorgabe aus Erwägungsgrund 18 S. 3, 4 DSM-RL klarer im Gesetzeswortlaut zum Ausdruck gebracht hätte.“

136 BReg, Gegenäußerung zu Nr. 6 BRat-StN (Fn. 65), S. 16.

137 So DFL, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 1f. = DFL/DFB, StN v. 6.11.2020 zum RefE (Fn. 22), S. 26.

138 BDZV/VDZ/VDL, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 10 = BDZV/VDZ, StN v. 6.11.2020 zum RefE (Fn. 22), S. 2 = VG Media, StN v. 2.9.2020 zum RefE (Fn. 22), S. 2; ebenso VAUNET Verband Privater Medien, StN v. 4.2.2020 zum DiskE (Fn. 21), S. 4 („Diese Einschränkung des Nutzungsvorbehalts des Rechteinhabers sieht die DSM-Richtlinie jedoch selbst nicht vor.“) und Motion Picture Association, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 2.

139 So wohl aber *Hacker*, ZGE 12 (2020), 239 (257): „Bei online veröffentlichten Inhalten kann dies nach dem Wortlaut etwa durch maschinenlesbare Formate geschehen.“

140 So neben den in Fn. 138 Genannten wohl auch *Gotthardt*, Die urheberrechtliche Beurteilung des Text und Data Mining in Europa und den USA aus rechtsvergleichender Sicht, 2023, S. 119.

Damit sind maschinenlesbare Mittel im Fall von online veröffentlichten Inhalten auch nach der Richtlinie die einzig „angemessene Weise“. Das bestätigen der zugehörige Erwägungsgrund,¹⁴¹ auf den sich die deutsche Gesetzgeberin berief,¹⁴² sowie ein Vergleich mit der englischen und französischen Sprachfassung der Richtlinie.¹⁴³ Beide lösen die eben skizzierte sprachliche Ambiguität zwar auch nicht im Normtext auf,¹⁴⁴ stellen aber im zugehörigen Erwägungsgrund noch „deutlicher als in der deutschen“ Sprachfassung klar,¹⁴⁵ dass für Onlineinhalte keine andere Erklärungsweise als die maschinenlesbare angemessen sei.¹⁴⁶ Auch darauf war im Gesetzgebungsverfahren von mehreren Seiten hingewiesen und die deutsche Richtlinienfassung für ihren „bedauerlichen [...] redaktionellen Fehler“ kritisiert worden.¹⁴⁷

Nichtsdestotrotz sind die vielfach geäußerten Zweifel an der Richtlinienkonformität des § 44b Abs. 3 S. 2 UrhG berechtigt – allerdings ganz anders als bisher gedacht. Denn entgegen den bisherigen Stimmen ist § 44b Abs. 3 S. 2 UrhG nicht zu eng geraten, weil er die Maschinenlesbarkeit zwingend verlangt, sondern weil er sie nur für „Werke“ verlangt. Denn das TDM ist nach Art. 4 DSM-RL an „Werken und sonstigen Schutzgegenständen“ zulässig, während § 44b Abs. 1 UrhG nur für „Werke“ gilt. Damit korrespondierend beziehen andere Sprachfassungen des Art. 4 Abs. 3 DSM-RL auch die Formbedürftigkeit des Nutzungsvorbehalts auf „online veröffentlichte Inhalte“ („content made publicly available online“,

141 ErwGr. 18 UA 2 S. 2 Hs. 1, S. 4 DSM-RL: „Wurden Inhalte im Internet öffentlich zugänglich gemacht, so sollte es als angemessen erachtet werden, einen Rechtsvorbehalt mit maschinenlesbaren Mitteln auszusprechen [...] In anderen Fällen [lies: als Internetveröffentlichungen, HH] kann es angemessen sein, einen Rechtsvorbehalt mit anderen Mitteln, etwa in vertraglichen Vereinbarungen oder durch eine einseitige Erklärung, auszusprechen.“

142 RegE zum DSM-RL-UmsG (Fn. 33), S. 89: „nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt (vergleiche ErwG 18 Unterabsatz 2 Satz 2 DSM-RL)“.

143 So *Kleinkopf* (Fn. 18), S. 235 bei Fn. 910, wenn auch ohne konkrete Zitate.

144 Auf Englisch heißt es „in an appropriate manner, such as machine-readable means in the case of content made publicly available online“, auf Französisch «de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne».

145 So *Raue*, ZUM 2020, 172 (173 Fn. 13); zust. *Brockmeyer* (Fn. 41), S. 79 bei Fn. 487.

146 ErwGr. 18 UA 2 S. 2 Hs. 1 DSM-RL lautet auf Englisch: „In the case of content that has been made publicly available online, it should *only* be considered appropriate to reserve those rights by the use of machine-readable means“, auf Französisch « En ce qui concerne les contenus mis à la disposition du public en ligne, la réservation de ces droits *ne devrait être jugée appropriée que si elle est effectuée au moyen de procédés lisibles par machine* » (Hervorhebung jeweils nur hier).

147 Google / Youtube, StN v. 11.9.2019 zur BMJV-Konsultation (Fn. 20), S. 5f.: „Bedauerlicherweise fehlt offenbar wegen eines redaktionellen Fehlers [...] das entscheidende Wort ‚nur‘, welches in den anderen Sprachversionen enthalten ist.“ m.Verw. auf *Raue*, ZUM 2019, 684 (693 Fn. 91): „In der deutschen Sprachfassung fehlt sinnetstellend ein ‚nur‘;“; ähnl. Bitkom, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 6 (zum Teil textgleich *Steinbrecher*, MMR 2019, 639 [640]); IBM Intellectual Property Law, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 3.

«contenus mis à la disposition du public en ligne») statt lediglich auf „Werke“ („works“, «œuvres») wie § 44b Abs. 3 S. 2 UrhG.¹⁴⁸ Diese Divergenz scheint zunächst rechtstechnisch bedingt: Dass die Schranke des § 44 Abs. 1 UrhG auch für sonstige Schutzgegenstände gilt, ergibt sich nämlich aus flächendeckenden Einzelverweisen in Teil 2 UrhG (Verwandte Schutzrechte). Dort wird in neun von zehn Fällen die „entsprechende Anwendung“ des § 44b UrhG angeordnet,¹⁴⁹ womit auch der Nutzungsvorbehalt und seine Formvorschrift bei online zugänglichen „Werken“ sinngemäß auf Nichtwerke anzuwenden wären. Probleme bereitet erst – dafür aber besonders praxisrelevant – der zehnte Fall:

Für Datenbanken i.S.v. § 87a Abs. 1 S. 1 UrhG, die nicht als Datenbankwerke i.S.v. § 4 Abs. 2 UrhG schutzfähig sind, ordnet § 87c Abs. 1 Nr. 4 UrhG nicht etwa die entsprechende Geltung des § 44b UrhG an, sondern erlaubt vielmehr „die Vervielfältigung eines [...] wesentlichen Teils [...] zu Zwecken des Text und Data Mining gemäß § 44b“. Nach diesem Wortlaut ist § 44b UrhG auf Datenbanken also nicht sinngemäß („entsprechend“) sondern direkt („gemäß“) anzuwenden – und damit womöglich ohne das auf „Werke“ beschränkte Vorbehaltsrecht. Man mag dies als sprachlichen Lapsus abtun und unter Hinweis auf den intendierten Normzweck durch Gesamtanalogie zu den neun anderen Verweisketten (Fn. 149) beheben wollen. Bessere Rechts(schutz)sicherheit vermittelt jedoch eine richtlinienkonforme Auslegung, die im Streitfall dem EuGH das letzte Wort beließe. Denn beim problematischen zehnten Fall (Datenbank-TDM wie im Eingangsbeispiel oben I.) handelt es sich zugleich um den wohl praxisrelevantesten. Dass sich

„eine neue Schrankenbestimmung [...] auch auf Computerprogramme (und Datenbanken) bezieht [...] ist umso bedeutender als vertreten wird, dass es sich bei Webseiten um Computerprogramme im urheberrechtlichen Sinn handelt. [...] Einschränkungen von TDM bezüglich Computerprogramme [sic] könnten daher dazu führen, dass Webseiten nicht mehr analysiert werden dürften.“¹⁵⁰

Um also rechtssicher klarzustellen, dass auch Fälle des § 87c Abs. 1 Nr. 4 i.V.m. § 44b UrhG eines maschinenlesbaren Nutzungsvorbehalts bedürfen, müsste zumindest eine der beiden Vorschriften dieser Verweiskette richtlinienkonform ausgelegt werden. Erst dann entspricht der zu eng gefasste § 44b Abs. 3 S. 2 UrhG seinem europäischen Vorbild Art. 4 Abs. 3 DSM-RL.

148 Soweit ersichtlich wies bislang nur *Frey*, StN v. 12.4.2021 zum RegE (Fn. 23), S. 42 Rn. 137 darauf hin, dass „ein Nutzungsvorbehalt nur hinsichtlich Inhalten mit Werkcharakter möglich“ sei, ohne allerdings die hier festgestellte Konsequenz in Erwägung zu ziehen.

149 § 70 Abs. 1, § 71 Abs. 1 S. 3 („sinngemäß“), § 72 Abs. 1, § 83, § 85 Abs. 4, § 87 Abs. 4, § 87i, § 94 Abs. 4, § 95 UrhG.

150 iRights, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 4 m.Verw. auf das damals anlautende Verfahren LG Hamburg, Urt. v. 14.1.2022, Az. 308 O 130/19 und *Kreutzer*, MMR 2018, 639 (639 ff.).

VII. Zwischenergebnis: Die Dogmatik des § 44b Abs. 3 UrhG

Die dogmatische Konkretisierung des eingangs zitierten Wortlauts von § 44b Abs. 3 UrhG ergibt also, dass die Gesetzgeberin den Nutzungsvorbehalt als rechtsgeschäftsähnliche Handlung ausgestaltet hat, deren Fehlen die Nutzerin beweisen müsste, deren Formvoraussetzungen aber von der Urheberin zu beweisen sind. Diese Formvoraussetzungen, die der Rechtsbegriff „maschinenlesbar“ zusammenfasst, decken sich nicht mit der Textform des § 126b BGB, sondern sollen als Mensch-Maschine-Kommunikation (H2M) die einseitige Teilautomatisierung des Erklärungsempfangs ermöglichen, wobei bislang ungeklärt ist, welche Schnittstelle dafür geeignet oder erforderlich sein könnte. Aus der Normgenese ergibt sich jedenfalls ein Ausdrücklichkeitserfordernis, das konkludente ebenso wie abstrakt-generelle Nutzungsvorbehalte verbietet, also stets einen auf einen konkreten Inhalt und die spezifische TDM-Nutzung bezogenen Vorbehalt erfordert. In systematischer Hinsicht lassen sich Parallelen zum „Vorbehalt der Rechte“ in § 49 Abs. 1 S. 1 UrhG fruchtbar machen, die neben der Rechtsnatur des Vorbehalts auch die Frage nach dessen Erklärungsbefugnis aufwerfen; richtigerweise dürften Repertoireinhaberinnen und Verwertungsgesellschaften nur dann Nutzungsvorbehalte erklären können, wenn sie sich diese Befugnis ausdrücklich haben einräumen lassen. Aus europäischer Perspektive schließlich erweisen sich Rufe nach einer richtlinienkonformen Auslegung des § 44b Abs. 3 UrhG als berechtigt – anders als vielerorts behauptet allerdings nicht zur Einschränkung der Vorschrift (etwa um nicht-maschinenlesbare Vorbehalte bei online zugänglichen Inhalten zu ermöglichen), sondern im Gegenteil zu ihrer Erweiterung (auf Inhalte, die nicht als Werke schutzfähig sind).

E. Wie lässt sich KI-Training maschinenlesbar untersagen?

Der gemeinsame Nenner des bisherigen Schrifttums zur maschinenlesbaren Form in § 44b Abs. 3 UrhG ist schiere Resignation. Soweit Autorinnen die Form nicht einfach zu überspielen oder für entbehrlich zu erklären versuchen,¹⁵¹ stellen sie einen „Mangel an belastbaren Standards“ fest.¹⁵² Daher sei es geradezu ein „Mysterium“,¹⁵³ „wie sich der Opt-out praktisch umsetzen lässt“.¹⁵⁴ Man kann nicht einmal sagen, dass die „Anforderungen an die Maschinenlesbarkeit“

151 Vgl. oben bei Fn. 73, 77.

152 *Bomhard*, DSRI/TB 14 (2023), 255 (265); *Durantaye*, ZUM 2023, 645 (651): „Momentan fehlt dafür ein spezieller technischer Standard.“

153 *Mezei*, Eur. IP Rev. 46 (2024), 461 (465): “the substance and the functioning of rights reservation is a mystery”.

154 *Maamar*, ZUM 2023, 481 (484); *Ducato/Strowel*, IIC 50 (2019), 649 (666): “questions remain as to what the reservations in a machine-readable format are, and how they could be implemented”.

ernsthaft „umstritten“ seien,¹⁵⁵ weil sie auf technischer Ebene nie konkret diskutiert werden. Denn derzeit weiß niemand, welche „modernsten Technologien“ das KI-Gesetz meint¹⁵⁶ und wie sich deren Leistungsfähigkeit in Zukunft entwickelt. Schon im deutschen Gesetzgebungsverfahren wurde es „im Lichte der technischen Entwicklung“ kaum für möglich gehalten, die Maschinenlesbarkeit „abschließend zu bestimmen“,¹⁵⁷ und noch vor kurzem hieß es:

“There are currently no generally recognized standards or protocols for the machine-readable expression of the reservation of rights provided for in Article 4 of the Directive.”¹⁵⁸

Trotz (oder wegen?) dieser schon für fehlgeschlagen erklärten Suche nach dem Erklärungsstandard findet sich kein einziger Versuch, die immerhin denkbaren (oder von Kennern der Technik sogar konkret vorgeschlagenen) Technologien systematisch zu sichten und zu vergleichen.¹⁵⁹ Manche Autorinnen benennen zwar konkrete Verfahren (worauf in den folgenden Abschnitten jeweils einzugehen ist); dies aber oft ohne Reflexion der jeweiligen Funktionsweise. Umgekehrt scheitern die von technisch Versierten vorgeschlagenen Möglichkeiten vielfach an den dogmatischen Anforderungen des § 44b Abs. 3 UrhG.

Um also die Wahrung der maschinenlesbaren Form *im Rechtssinne* festzustellen, bedarf es einer technikkrechtlichen Perspektive, die normative Anforderungen und technisches Prozessverständnis integriert. Zur Veranschaulichung konkreter Mechanismen soll das Eingangsbeispiel dienen (oben I.), das die meisten der gleich zu besprechenden Standards illustriert. Damit kommen wir nun, im letzten Abschnitt der Bearbeitung, auf unseren kleinen Bot zurück und bewerten seine Reise endlich juristisch:

Hat die Betreiberin des KI-Bots das Urheberrecht verletzt i.S.v. §§ 97 Abs. 1 S. 1, 99 UrhG? Welche der vielen Nutzungsvorbehalte, die der Bot ignoriert hat, waren nach dem Maßstab des § 44b Abs. 3 UrhG überhaupt wirksam?

155 So aber *Hofmann*, WRP 2024, 11 (14 Rn. 18); *Söbbing/Schwarz*, RD 2023, 415 (418 Rn. 15): „nicht ganz unumstritten“.

156 Vgl. schon *Mezei*, Eur. IP Rev. 46 (2024), 461 (469): “It is [...] far from being clear how the EU has imagined the respect of opt-out privileges via a ‘policy’”.

157 *Frey*, StN v. 12.4.2021 zum RegE (Fn. 23), S. 42 Rn. 136.

158 *Keller/Warso*, Defining Best Practices for Opting Out of ML Training (OpenFuture Policy Brief #5, 29.9.2023) <openfuture.eu/wp-content/uploads/2023/09/Best_practices_for_optout_ML_training.pdf>.

159 In Ansätzen immerhin *Keller/Warso* (vorige Fn.), S. 7–10 (“State of the Art”); *Löbbling u.a.*, JIPITEC 2023, 499 (505–509); jüngst vor allem *Mezei*, Eur. IP Rev. 46 (2024), 461 (465–468): “Practical experiences of rights reservation”, der seine Sichtung allerdings nach Sachfragen ordnet und deshalb nicht alle hier erwähnten Technologien als solche näher erläutert oder rechtlich bewertet.

I. Impressum oder AGB

Soweit die im Eingangsbeispiel ausgelesene Website einen Nutzungsvorbehalt im Impressum (bei Fn. 7) sowie in ihren Allgemeinen Geschäftsbedingungen (bei Fn. 9) erklärte, ist deren Maschinenlesbarkeit ganz offenkundig problematisch. Zwar weist die maschinenlesbare H2M-Kommunikation „eine gewisse Parallele zu Allgemeinen Geschäftsbedingungen“ auf.¹⁶⁰ Das kann allerdings keineswegs bedeuten, dass jede formularvertragliche Erklärung auch maschinenlesbar wäre. Davon ging auch die Gesetzgeberin nicht aus, als sie erklärte:

„[Der Nutzungsvorbehalt] kann auch im Impressum oder in den Allgemeinen Geschäftsbedingungen (AGB) enthalten sein, sofern er auch dort maschinenlesbar ist.“¹⁶¹

Diese Begründung wurde im späteren Schrifttum vielfach dahin missverstanden, dass „alle textlichen Ausdrucksformen, wie etwa ein Nutzungsvorbehalt im Impressum, erfasst sein sollen“,¹⁶² so dass sich die Nutzung eigener Werke zum KI-Training „ohne großen Aufwand ausschließen“ lasse, weil die Anforderungen an einen Nutzungsvorbehalt „nicht allzu hoch“ seien,¹⁶³ wenn er „auch niederschwellig im Impressum oder den AGB einer Website erklärt werden kann“.¹⁶⁴ In der Folge gingen viele umstandslos davon aus, dass Nutzungsvorbehalte in AGB oder Impressum stets den Vorgaben des § 44b Abs. 3 UrhG genügen.¹⁶⁵

Durchweg übersehen wurde dabei, dass die Gesetzesbegründung „die Maschinenlesbarkeit nicht spezifiziert, sondern voraussetzt“.¹⁶⁶ Denn durch ihre konditionale Formulierung („..., sofern...“) bringt sie gerade zum Ausdruck, dass die Maschinenlesbarkeit eine *zusätzliche* Anforderung *neben* dem möglichen Standort in AGB oder Impressum sein soll.¹⁶⁷ Nur dieses Verständnis entspricht auch der Richtlinie, die an der entsprechenden Stelle formuliert:

¹⁶⁰ Zitat und Nachw. oben in Fn. 89.

¹⁶¹ RegE zum DSM-RL-UmsG (Fn. 33), S. 89.

¹⁶² Baumann, NJW 2023, 3673 (3675 Rn. 19).

¹⁶³ Bernzen, Beil. zu K&R 10/2023, 6 (7 bei Fn. 22–24) m.Verw. auf den „Willen des Gesetzgebers“.

¹⁶⁴ Pukas, GRUR 2023, 614 (615); ähnl. Söbbing/Schwarz, RD 2023, 415 (418 Rn. 16): „einfaches Instrument [...] kann einen Zugriff eines Webcrawlers leicht verhindern.“

¹⁶⁵ Akinci/Heidrich, IPRB 2023, 270 (272); Kleinkopf (Fn. 18), S. 249 f. („Unter den Vorbehalt fallen jedenfalls Angaben [...] in den Allgemeinen Geschäftsbedingungen“); Gerecke, GRUR-Prax. 2023, 381(382 Rn. 11): „Solange der Rechteinhaber also [...] nicht] einen Vorbehalt [...] durch Hinweis im Impressum oder in seinen AGB erklärt, ...“; Senfleben, IIC 54 (2023), 1535 (1544): „AI trainers must take into account [...] terms and conditions of websites and online services.“; implizit auch Siglmüller/Gassner, RD 2023, 124 (127 Rn. 12, nach Fn. 35).

¹⁶⁶ Bomhard, DSRITB 14 (2023), 255 (266).

¹⁶⁷ Vgl. Löbbling u.a., JIPITEC 2023, 499 (502 Rn. 16). Präzise daher Raue, ZUM 2019, 684 (693) = Raue, ZUM 2021, 793 (797): „..., solange er maschinenlesbar ist.“; ebenso Brockmeyer (Fn. 41), S. 79; ähnl. Maamar, ZUM 2023, 481 (484): „... allerdings nur, soweit er dort maschinenlesbar ist.“; Söbbing/Schwarz, RD 2023, 415 (418 Rn. 16): „muss die Information [...] nur in einer maschinenlesbaren Form erfolgt sein.“

„Wurden Inhalte im Internet öffentlich zugänglich gemacht, so sollte es als angemessen erachtet werden, einen Rechtsvorbehalt mit maschinenlesbaren Mitteln auszusprechen; Das [sic] gilt auch für Metadaten und Geschäftsbedingungen einer Website oder eines Dienstes.“ (ErwGr. 18 UA 2 S. 2 DSM-RL)

Mit dem einleitenden „Das“ bezieht sich der zweite Halbsatz ersichtlich auf „mit maschinenlesbaren Mitteln auszusprechen“. Anders als weithin verstanden wollte die Richtlinie also gerade ausschließen, dass *allein* die Platzierung eines Nutzungsvorbehalts in Metadaten oder Geschäftsbedingungen für hinreichend gehalten würde,¹⁶⁸ wenn dies nicht *zugleich* auch maschinenlesbar geschieht.¹⁶⁹

Im Übrigen zeigt ein Vergleich des zitierten Erwägungsgrundes („Metadaten und Geschäftsbedingungen“) mit der deutschen Gesetzesbegründung („Impressum oder [...] AGB“), dass der Umsetzungsgesetzgeber das Impressum für eine (die einzig denkbare?) Art von „Metadaten“ hielt. Darin liegt eine reichlich unkonventionelle Deutung des Begriffs,¹⁷⁰ die als deutscher Sonderweg die bezweckte Harmonisierung im digitalen Binnenmarkt untergräbt.

Jedenfalls scheiden, wie oben (D.II.) dargestellt, alle Erklärungen in natürlicher Sprache aus – seien sie nun in AGB, Metadaten oder Impressum eingestellt. Das ist schon wegen der Vielzahl denkbarer Formulierungsvarianten¹⁷¹ sachgerecht. So war die im Eingangsbeispiel (in Fn. 10) zitierte AGB-Bestimmung noch wenige Monate zuvor ganz anders formuliert und stand an anderer Stelle derselben AGB.¹⁷² Auch sind Standort und Dateiformat von Geschäftsbedingungen in keiner Weise standardisiert, wie das Eingangsbeispiel ebenfalls

168 So aber *Vesala*, IIC 54 (2023), 351 (357): “online services use language in their terms of use that may [...] constitute an effective reservation (e.g. banning reverse engineering or similar methods, or the storing of available content)”; ähnl. *Mezei*, Eur. IP Rev. 46 (2024), 461 (465 bei Fn. 48) zum “risk that expressed terms of end-user licence agreements can exclude the lawfulness of TDM”; noch im Gesetzgebungsverfahren sprachen Verbände von „maschinenlesbaren Mitteln, einschließlich Metadaten und Geschäftsbedingungen“ (Motion Picture Association, StN v. 6.9.2019 zur BMJV-Konsultation [Fn. 20], S. 9) und meinten, „dass Metadaten sowie Geschäftsbedingungen [...] taugliche maschinenlesbare Mittel sind“ (DFL, StN v. 6.9.2019 zur BMJV-Konsultation [Fn. 20], S. 7 = StN v. 31.1.2020 zum DiskE [Fn. 21], S. 2 = DFL/DFB, StN v. 6.11.2020 zum RefE [Fn. 22], S. 26).

169 Zutr. daher Google/Youtube, StN v. 11.9.2019 zur BMJV-Konsultation (Fn. 20), S. 6, „dass eine allgemeine Formulierung auf einer Website oder in allgemeinen Geschäftsbedingungen nicht als Nutzungsvorbehalt in angemessener Weise gelten kann, wenn der Vorbehalt nicht durch einen maschinenlesbaren Code, der in automatisierter Weise verarbeitet werden kann, erklärt wird“; gegen “reservations not expressed in code (for instance when TDM restrictions are found in website terms and conditions in PDFs, images or as website text)” auch *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (30).

170 Zum Begriff „Metadaten“ gleich unten nach Fn. 221.

171 Beispiele von verschiedenen Websites bei *Müller ter Jung/Rexin*, CR 2022, 169 (174 Rn. 32, in Fn. 40–46); Untersuchung von 21 Plattformen dreier Branchen bei *Ducato/Strowel*, IIC 50 (2019), 649 (669–673).

172 Ziff. 10.9 S. 2 AGB v. 9.3.2022 (Stand 6.2.2023): „Der Verlag behält sich das Recht zu Vervielfältigungen für das Text und Data Mining gemäß § 44b UrhG vor.“ <web.archive.org/20230206004327/rsw.beck.de/beck-online-service/agb-beck-online>.

belegt: Die bei Fn. 9 zitierte AGB-Datei soll für die ausgelesene Website [beck-online.beck.de](https://online.beck.de) gelten, ist aber unter der ganz anderen (Sub-)Domain rsw.beck.de abgelegt – und soll überdies noch ein ganz anderes Medium („E-Mail-Dienste“) erfassen. Zudem werden AGB für ihre Druckbarkeit oft auf mehrere Dateien aufgeteilt oder im pdf-Format angeboten,¹⁷³ das entgegen verbreiteten Missverständnissen nur rudimentär standardisiert ist und technisch „fast unmöglich“ auszulesen ist:

“machine readability (as well as cognition) of online terms and conditions is difficult to near impossible for most systems except for the most sophisticated technology”¹⁷⁴

Wollte man also ernsthaft Nutzungsvorbehalte via AGB oder Impressum erklären, bliebe kaum etwas anderes übrig, als „wiederum der Einsatz von KI-Modellen“ zum Auffinden solcher Vorbehalte.¹⁷⁵ Damit aber erforderte die Suche nach dem Nutzungsvorbehalt eine Musteranalyse, die ihrerseits bereits den TDM-Tatbestand des § 44b Abs. 1 UrhG erfüllen würde. Die Katze bisse sich in den eigenen Schwanz: Ob eine Website automatisiert ausgelesen werden darf, wäre erst klar, nachdem man sie ausgelesen hat.

Selbst wenn es technisch gelänge, einen standardisierten Standort für AGB zu definieren und an diesem Standort einen Nutzungsvorbehalt in maschinenlesbarer Form zu erklären, bliebe noch das oben (D.III.) entwickelte Ausdrücklichkeitserfordernis im Sinne einer konkret-individuellen Erklärung. Der in einer solchen AGB erklärte Nutzungsvorbehalt könnte allenfalls für das AGB-Dokument selbst gelten, aber nicht pauschal für die (wie auch immer definierte) Menge aller Dokumente der Website.¹⁷⁶

Dies bestätigt auch ein systematischer Seitenblick auf den Rechtevorbekhalt nach der Parallelvorschrift § 49 UrhG (zu ihr oben D.IV.). Dieser muss nach allgemeiner Ansicht „in unmittelbarem räumlichen Zusammenhang mit dem Kommentar, Artikel usw.“ erklärt werden,¹⁷⁷ „und zwar“ gesondert für „jeden einzelnen von ihnen“.¹⁷⁸

173 Bsp. für beides etwa www.bundesanzeiger.de/pub/de/rechtliches.

174 IBM Intellectual Property Law, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 2.

175 So tatsächlich *Akinci/Heidrich*, IPRB 2023, 270 (272); auch *Löblich u.a.*, JIPITEC 2023, 499 (504) folgern aus einem Selbstversuch mit 100 Websites: “Effective opt-out management would require advanced NLP methods, which might still carry high error rates.” Einer privaten Korrespondenz mit *Jonathan Pukas* verdanke ich die Anregung, mittels einfacherer Texterkennung eine vordefinierte Menge bekannter Vorbehaltserklärungen abzuprüfen (wofür § 44a UrhG genügen dürfte), verbunden womöglich mit den bei Uploadfiltern erprobten Hashverfahren zur Erfassung von Formulierungsvarianten. Alle Einwände des Haupttextes könnte auch diese Lösung kaum beheben.

176 A.A. BDZV/VDZ/VDL, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 11 für den „Nutzungsvorbehalt an zentraler Stelle“ im Fall von „Sammlungen (zum Beispiel einer Internetseite), die mehrere Werke und Schutzgegenstände enthalten“.

177 Fromm/Nordemann/Nordemann, 12. Aufl. 2018, § 49 UrhG Rn. 7.

178 Dreier/Schulze/Dreier, 7. Aufl. 2022, § 49 UrhG Rn. 10.

„Ein genereller Hinweis, etwa [...] im Impressum einer Zeitung [...] genügt also nicht. Nur so ist sichergestellt, dass jeder, der für einen Artikel oder Kommentar von der Weiterverwendungsmöglichkeit nach § 49 Gebrauch machen will, klar und einfach erkennen kann, ob bezüglich dieses einzelnen Werks der Vorbehalt besteht.“¹⁷⁹

Das lässt sich bei § 49 UrhG bereits mit dem Wortlaut rechtfertigen,¹⁸⁰ folgt aber bei § 44b Abs. 3 UrhG nicht minder deutlich aus dem Ausdrücklichkeitserfordernis, das abstrakt-generelle Pauschalvorbehalte ausschließt.

II. Robots Exclusion Protocol (robots.txt)

Die im Eingangsbeispiel ausgelesene Website erklärte einen Nutzungsvorbehalt immerhin nicht nur in natürlicher Sprache, sondern auch in einer speziellen Syntax in der Datei robots.txt (oben bei Fn. 3). Dabei handelt es sich um eine jener „Textdateien“

„die von Web-Crawlern oder -Scrapern primär angefragt werden und die Angaben darüber enthalten, welche Inhalte nach Auffassung der Webseitenbetreiber abgegriffen werden dürfen und welche nicht“.¹⁸¹

Dafür kommt eine Syntax zum Einsatz, die schon 1994 durch den niederländischen Suchmaschinenpionier *Martijn Koster* definiert wurde und fast dreißig Jahre lang als „De-facto-Kommunikationsstandard [...] von Webseitenbetreibern“ galt.¹⁸² Im Jahr 2022 verabschiedete ihn die Internet Engineering Task Force (IETF) 2022 auf Betreiben von Google offiziell als „Robots Exclusion Protocol“ (REP).¹⁸³ Dieser Standard wird „von allen großen Suchmaschinen akzeptiert“,¹⁸⁴ und nach Schätzungen von Google aus dem Jahr 2019 auf einer halben Milliarde Websites eingesetzt.¹⁸⁵

179 Schrickler / Loewenheim / Melichar / Stieper, 6. Aufl. 2020, § 49 UrhG Rn. 23; Heidelb. Komm. UrhR / Dreyer, 4. Aufl. 2018, § 49 UrhG Rn. 19: „Ein allg Vorbehalt zB im Impressum ist wirkungslos“; Fromm / Nordemann / Nordemann, 12. Aufl. 2018, § 49 UrhG Rn. 7: „ein Vermerk im Impressum oder im Inhaltsverzeichnis genügt nicht“; Dreier / Schulze / Dreier, 7. Aufl. 2022, § 49 UrhG Rn. 10: „ein allgemeiner Vorbehalt etwa im Impressum oÄ reicht nicht aus“.

180 Wandtke / Bullinger / Lüft, 6. Aufl. 2022, § 49 UrhG Rn. 11: „Schon aus dem Wortlaut folgt, dass ein Einzelvorbehalt gefordert ist.“ m. Verw. auf BeckOK UrhR (vorm. Möhring / Nicolini) / Engels, 41. Ed. 2024, § 49 UrhG Rn. 14: „Diese Vorgabe ergibt sich bereits unmittelbar aus dem UrhG, das ausdrücklich fordert, dass die (zu übernehmenden) einzelnen Rundfunkkommentare bzw. einzelnen Artikel (‚sie‘) ‚mit einem Vorbehalt der Rechte versehen sind.‘“

181 Kleinkopf (Fn. 18), S. 51.

182 So Conrad / Schubert, GRUR 2018, 350 (351).

183 Koster / Illyes / Zeller / Sassman, Standard RFC 9309: Robots Exclusion Protocol (9/2022), dokumentiert unter <rfc-editor.org/rfc/rfc9309.html>.

184 Peukert, ZUM 2023, 233 (246).

185 Pierce, The text file that runs the internet (The Verge, 14.2.2024) <theverge.com/24067997/robots-txt-ai-text-file-web-crawlers-spiders>.

Schon in der internationalen Literatur wurde das REP – verstanden als “Private Ordering Through Opt-Outs”¹⁸⁶ – deshalb als Kommunikationsstandard für Nutzungsvorbehalte vorgeschlagen.¹⁸⁷ In der deutschen Umsetzungsdiskussion beriefen sich ebenfalls etliche Experten und Fachverbände auf das REP als „weltweit anerkannten Quasi-Standard“ und wollten Maschinenlesbarkeit ohne Weiteres auf das „Auslesen der Robots.txt Dateien“ beziehen.¹⁸⁸ Auch spätere Autorinnen hielten es für „technisch sinnvoller“ als die Platzierung in AGB oder Impressum, „den Vorbehalt in die sog. robots.txt aufzunehmen“.¹⁸⁹ Anders gewendet gehen viele Autorinnen davon aus, dass ein Nutzungsvorbehalt immer dann wirksam i.S.v. § 44b Abs. 3 S. 2 UrhG sei, wenn er „mittels Hinterlegung einer ‚robots.txt‘-Datei“ erklärt werde.¹⁹⁰ Dafür sprechen sich auch diejenigen aus, die in der vermeintlichen „robot.txt“ statt einer Textdatei eine „Norm“, einen „Zusatz“, eine „Information“ oder „eine Art von Metadaten“ sehen.¹⁹¹

Dabei sind allerdings zwei Missverständnisse zu vermeiden: Zum einen kann die robots.txt als einfache Textdatei beliebige Texte enthalten – auch solche in

¹⁸⁶ *Sag*, Nw. U. L. Rev. 103 (2009), 1607 (1666–1668).

¹⁸⁷ *Ducato/Strowel*, IIC 50 (2019), 649 (674): “an exclusion protocol that content providers can insert into the root directory to prevent crawling or indexing activities”; *Dusollier*, CML Rev. 57 (2020), 979 (987): “machine-readable means as robots.txt files”; *Tan/Lee*, Sing. Acad. Law J. 33 (2021), 1032 (1039): “owners may even adopt a Robots Exclusion Protocol”; *Löbbling u.a.*, JIPITEC 2023, 499 (505 Rn. 36): “setting up a robots.txt file can express an opt-out” (deutlich auch ebd., 506 Rn. 41); *Senfileben*, IIC 54 (2023), 1535 (1544): “AI trainers must take into account metadata, such as robots.txt files”; *Mezei*, Eur. IP Rev. 46 (2024), 461 (467): “inclusion of relevant computer-readable language in the robots.txt file”.

¹⁸⁸ *Bitkom*, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 3 f. = StN 4 v. 6.11.2020 zum RefE (Fn. 22), S. 2; *Reda*, StN v. 23.1.2020 zum DiskE (Fn. 21), S. 8 = *Buczek*, StN v. 30.1.2020 zum DiskE (Fn. 21), S. 15: „In der Praxis ist der robots.txt-Standard für solche Anwendungen etabliert und sollte zwecks Sicherung von Interoperabilität gesetzlich vorgeschrieben werden.“; *EARE*, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 2: „Vorbehalt in einem geeigneten maschinenlesbaren Format (Metadaten, robots.txt, etc.)“; *IBM Intellectual Property Law*, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 2: “robots.txt file or equivalent, which is a protocol/format that is used widely by web crawlers and web robots today”; *GFF*, StN v. 3.11.2020 zum RefE (Fn. 22), S. 5; *LIBER*, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 2 (Nr. 8).

¹⁸⁹ *Akinci/Heidrich*, IPRB 2023, 270 (272).

¹⁹⁰ So *Gerecke*, GRUR-Prax. 2023, 381 (382 Rn. 11); *Kleinkopf* (Fn. 18), S. 249 f.: „Unter den Vorbehalt fallen jedenfalls [...] auch robot.txts“; *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, in II.3; *Baumann*, NJW 2023, 3673 (3675 Rn. 19): „Jedenfalls zu beachten sind Anweisungen an Crawler über robots.txt-Dateien“; *Maamar*, ZUM 2023, 481 (484) bei Fn. 34; ausf. *Bomhard*, DSRITB 14 (2023), 255 (267); *Schack*, NJW 2024, 113 (116 Rn. 20): „mittels robots.txt oder ähnlichen Dateien“; *Konertz/Schönhof*, WRP 2024, 289 (296 Rn. 33, bei Fn. 110): Maschinenlesbarkeit „ist zumindest dann der Fall, wenn dies über entsprechende Einträge bei Webseiten geschieht, beispielsweise über den sog. ‚Robots Exclusion Standard‘ (robots.txt)“.

¹⁹¹ *EARE*, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 2: „maschinenlesbare ‚Normen‘ [...] wie zum Beispiel robot.txt“; *Hacker*, in: FS Hopt 2020, S. 351 (365 Fn. 70): „Als Vorbild kann etwa der Zusatz robots.txt dienen“; *Söbbing/Schwarz*, RD 2023, 415 (418 Rn. 15): „Eine eindeutige und einfache Möglichkeit ist jedenfalls der Hinweis mittels der sog. ‚robot.txt‘-Information.“; *Griffiths/Synodinou/Xalabarder*, GRUR Int. 2023, 22 (25 nach Fn. 42): “machine-readable means, including [...] robot.txt type metadata”.

natürlicher Sprache, wie das Eingangsbeispiel (oben in Fn. 5) belegt. Aufgrund der bereits dargelegten Untauglichkeit natürlicher Sprache (oben bei Fn. 82) kann die Hinterlegung in der robots.txt jedenfalls allenfalls insoweit wirksame Nutzungsvorbehalte begründen, wie „Webcrawler beim automatisierten Auslesen von robots.txt Dateien erklärte Nutzungsvorbehalte tatsächlich auffinden und erkennen können“.¹⁹² Maschinenlesbarkeit ist also nicht *Folge*, sondern *Voraussetzung* einer Deklaration in der robots.txt.

Zum anderen ist festzustellen, dass das Robots Exclusion Protocol keine technische Zugangshürde errichtet (Ziff. 1 UA 4 REP: “not a form of access authorization”), sondern freiwillig zu beachtende Regeln definiert (Ziff. 1 UA 3 REP: “rules [...] that crawlers are requested to honor”). Selbst große Bot-Betreiberinnen ignorieren diese Regeln mitunter demonstrativ.¹⁹³ Deshalb wurde zu Recht festgestellt:

„Wer das REP nutzt, *handelt* nicht durch Technik, sondern *kommuniziert* durch Code.“¹⁹⁴

Soweit nach dem REP „durch Code“ kommuniziert wird, handelt es sich zweifelsfrei um einen Standard für Maschinensprache. Fraglich ist allerdings, ob dieser sich auch für die Zwecke des § 44b Abs. 3 UrhG eignet. Zwar zeigt eine fortlaufende empirische Beobachtung von 886 US-amerikanischen und 273 weiteren Nachrichtenportalen aus 31 Ländern,¹⁹⁵ dass gut zwei Fünftel davon (40,7 %) in der robots.txt denselben Bots Hausverbot erteilen wie die Website in unserem Eingangsbeispiel (bei Fn. 3), und dass mehr als jede zweite (54,3 %) zumindest einen dieser drei Bots oder den von Google maschinenlesbar fernzubleiben heißt.¹⁹⁶ Dennoch sprechen sowohl technische als auch dogmatische Gründe *gegen* die Eignung des REP für Nutzungsvorbehalte.

In technischer Hinsicht stellen sich gleich mehrere Probleme: Erstens ist der Name eines Bots (sog. *product token*) frei wählbar (Ziff. 2.2.1 REP: “Crawlers set

192 Müller *ter Jung/Rexin*, CR 2022, 169 (174 Rn. 29, 31); ähnl. zuvor schon Bitkom (Fn. 188): „dass Webcrawler beim Auslesen der Robots.txt Dateien bereits erkennen, wo Untersagungen erklärt wurden.“

193 Pierce (Fn. 185): “The Internet Archive, for example, simply announced in 2017 that it was no longer abiding by the rules of robots.txt. [...] And that was that.”

194 Conrad/Schubert, GRUR 2018, 350 (352) (Hervorhebungen im Original); Ducato/Strowel, IIC 50 (2019), 649 (674): “The nature of robots.txt as a TPM [technological protection measure] is controversial. [...] Some authors have argued that the protocol contains instructions that do not qualify as a technical barrier because any software agent can simply ignore the ‘Disallow’ command without actively forcing any digital fence.” m.w.N.

195 Welsh, Who blocks OpenAI, Google AI and CC? (palewire.re) <palewi.re/docs/news-home/pages/openai-gptbot-robotstxt.html> abgerufen am 2.4.2024; aus Deutschland werden dort beobachtet: Bild, Spiegel, Stern, Zeit, SZ, chess24, ChessBase, DW und MDR Sachsen-Anhalt.

196 Eigene Auswertung der Daten von Welsh (vorige Fn.): 629 von 1.159 Nachrichtenportalen blockieren entweder Google AI („Google Extended“), OpenAI („GPTBot“, „ChatGPT-User“) oder Common Crawl („CCBot“). 472 blockieren beide letztgenannten, 421 alle drei genannten. Unter den deutschsprachigen blockieren vier von neun (44,4 %) sowohl OpenAI als auch Common Crawl.

their own name”), weshalb es im Eingangsbeispiel (vor Fn. 1) hieß: „unser Bot nannte sich vielleicht ...“. Dabei geben sich nicht alle Bots auf ihrer Reise durch das Internet zu erkennen; manche verschleiern ihre Identität sogar gezielt.¹⁹⁷ Zweitens wurde die Liste der *product token* seit 2011 nicht mehr aktualisiert, so dass allein die Identifikation der heute gebräuchlichen KI-Datensammler eine aufwändige Verkehrsbeobachtung erfordert.¹⁹⁸ Darauf immerhin haben die großen Anbieter rasch reagiert¹⁹⁹ – wohl nicht zuletzt in der Hoffnung, durch Einhaltung des alten REP-Standards einer effektiveren Regulierung zu entgehen.

Denn die Eignung des REP zur Kommunikation von Nutzungsvorbehalten ist auch dogmatisch fragwürdig. Nach Ziff. 2.2.2 REP sieht der Standard nämlich lediglich zwei Modalitäten vor: „Allow“ für Inhalte, die automatisiert ausgelesen werden dürfen, sowie „Disallow“ für Inhalte, bei denen dies verboten sein soll.²⁰⁰ Andere Modalitäten können zwar vorgesehen werden,²⁰¹ sind aber im Protokoll gerade nicht standardisiert. Damit fragt sich nun, wie ein „Disallow“-Vorbehalt nach dem insoweit primär relevanten²⁰² objektiven Empfängerhorizont des § 157 BGB auszulegen wäre.

In Ziff. 1 UA 3 REP heißt es zum Zweck solcher Vorbehalte lediglich, dass das Einlesen von Inhalten (*Crawling*) den Diensteanbietern „Unannehmlichkeiten“ verursachen könne.²⁰³ Dies erklärt sich aus dem ursprünglichen Zweck des REP, in der Frühzeit des Internet die Überlastung von Bandbreiten und die für ungeplantes Datenvolumen anfallenden Kosten zu vermeiden.²⁰⁴ Dieser Zweck hat sich weitgehend überlebt, so dass das REP heute vor allem dazu genutzt wird, ganz persönliche Präferenzen zu kommunizieren, welche Bot-Betreiberin die Inhalte eines Internetangebots nutzen soll und welche nicht.²⁰⁵ Da das REP

197 So *Wiese*, Robots.txt is not the answer (Search Engine Land, 18.7.2023) <searchengineland.com/robots-txt-new-meta-tag-llm-ai-429510>; ebenso *Pierce* (Fn. 185): “many others attempt to operate in relative secrecy [...] finding a sneaky crawler is needle-in-haystack stuff”; *Tan/Lee*, Sing. Acad. Law J. 33 (2021), 1032 (1070): “a key scenario is when web robots maliciously bypass REPs”.

198 *Waldvogel*, How to block AI crawlers with robots.txt (netfuture.ch, 9.7.–31.12.2023) <netfuture.ch/2023/07/blocking-ai-crawlers-robots-txt-chatgpt> m.Verw. auf die frühere Liste unter www.robotstxt.org/db.html.

199 OpenAI taufte am 8. August 2023 seinen „GPTBot“ <platform.openai.com/docs/gptbot>, Google definierte zum 28. September 2023 das *product token* “Google-Extended” <developers.google.com/search/docs/crawling-indexing/overview-google-crawlers>.

200 Ausführlichere Erläuterung bei *Conrad/Schubert*, GRUR 2018, 350 (351 unter 2.).

201 Ziff. 2.2.4 REP: “Crawlers MAY interpret other records that are not part of the robots.txt protocol – for example, ‘Sitemaps’”.

202 Vgl. ausf. Zitat oben bei Fn. 89.

203 “It may be inconvenient for service owners if crawlers visit the entirety of their URI space.”

204 *Conrad/Schubert*, GRUR 2018, 350 (351); neuerdings anschaulich *Pierce* (Fn. 185): “all it took was a few robots overzealously downloading your pages for things to break and the phone bill to spike.”

205 So weiter *Pierce* (Fn. 185): “It’s been a while since ‘overloaded servers’ were a real concern for

für diesen Zweck allerdings nie geschaffen wurde, erweist sich die grobkörnige Allow/Disallow-Syntax als viel zu holzschnittartig und unflexibel.²⁰⁶ Schon der Zweck, für den Inhalte ausgelesen werden sollen, spielt im REP keine Rolle.²⁰⁷ Soweit also Bots mehrere Zwecke erfüllen (bspw. Suchmaschinenindizierung und KI-Datensammlung)²⁰⁸ erlaubt das REP keine zwischen diesen Zwecken differenzierende Kommunikation:

„Bisher ist das Problem, dass die robots.txt Datei keine eigenständige Kategorie für Erklärungen zu Text und Data Mining kennt. Ein umfassender Vorbehalt würde also dazu führen, dass die Webseite auch nicht mehr bei Suchmaschinen auffindbar ist. Genau das wollte der Gesetzgeber eigentlich verhindern.“²⁰⁹

Damit eignet sich das REP nicht zur Wahrung des Ausdrücklichkeitserfordernisses, das eine „ausdrücklich auf das Text und Data Mining“ bezogene Erklärung erfordert.²¹⁰ Zudem beinhaltet Ausdrücklichkeit – wie schon mit Blick auf Nutzungsvorbehalte in AGB festgestellt (oben E.I.) – die konkret-individuelle Bezeichnung eines bestimmten Inhalts. Auch daran scheitert der Standard, denn nach Ziff. 2.5 REP dürfen Bots sich darauf beschränken (“parsing limit to protect their systems”), nur 512.000 Zeichen einer robots.txt-Datei zu berücksichtigen (“parsing limit must be at least 500 kibibytes”). Wenn aber für jeden Inhalt einer Website eine Entscheidung über Zulassung oder Abweisung jedes in Betracht kommenden Bots kommuniziert werden soll, ist diese Grenze schon bei mittelgroßen Internetangeboten von wenigen Hundert Dateien schnell erreicht.

most people. ‘Nowadays, it’s usually less about the resources that are used on the website and more about personal preferences,’ says John Mueller, a search advocate at Google.”

206 Deshalb verweigern manche Gedächtnisinstitutionen schon die Befolgung des REP, weil ihre Bedürfnisse sich von denen der Suchmaschinenbetreiberinnen unterscheiden, vgl. *Graham* zit. in *Pierce* (Fn. 185): “we have observed that the robots.txt files that are geared toward search engine crawlers do not necessarily serve our archival purposes”.

207 Ganz zu schweigen von der fehlenden Kommunizierbarkeit etwaiger Lizenzgebühren o.ä., dazu noch unten bei Fn. 240.

208 *Bomhard*, DSRITB 14 (2023), 255 (267) nennt als Beispiel den „Bingbot“ von Microsoft.

209 *Maamar*, ZUM 2023, 481 (484); im Gesetzgebungsverfahren schon Bundesverband professioneller Bildanbieter, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 7: „aktuell kann dies per Robots.txt auf Webseiten erfolgen, was jedoch den unerwünschten Nebeneffekt haben könnte, dass die gesamte Webseite von Suchmaschinen nicht indexiert und damit nicht gefunden werden kann“; ebenso *Durantaye*, ZUM 2023, 645 (651 bei Fn. 64); *Schack*, NJW 2024, 113 (116 Rn. 20): „zweischneidiges Schwert, wenn es verhindert, dass das Werk von Suchmaschinen gefunden wird.“; *Bomhard*, DSRITB 14 (2023), 255 (267): „Denn wenn man bei robots.txt über das Ziel hinausschießt (was schnell passiert), das kann zu einem dramatischen Absinken in den Suchmaschinenrankings führen“; *Durantaye*, AfP 2024, 9 (15 Rn. 33): „Wer auf diese Weise maschinenlesbar kommuniziert, dass er nicht für Trainingszwecke gescraped werden möchte, dessen Werk erscheint auch nicht mehr in Suchmaschinen. Damit ist man im Internet gar nicht mehr auffindbar.“

210 Zitat oben bei Fn. 110; ähnl. Kritik bei *Löbling u.a.*, JIPITEC 2023, 499 (505 Rn. 36) mit einem eigenen Reformvorschlag (507–509), der aber die weiteren hier aufgezeigten Einwände nicht erwähnt.

Würde dagegen wie im Eingangsbeispiel (bei Fn. 3) der pauschale Ausschluss von der gesamten Website erklärt (“Disallow: /”) oder speichersparende Dateinamenmuster verwendet, scheiterte ein solcher Nutzungsvorbehalt wiederum an der fehlenden Individualisierung der Inhalte. Dies bestätigt wiederum ein systematischer Seitenblick auf die Parallelnorm § 49 UrhG, wo es mit Blick auf die Folgen eines „allgemeinen“ Vorbehalts heißt, er werde „leicht ohne Einzelprüfung und damit voraussichtlich zu oft erklärt“, dadurch aber „entwertet und die Regelung ausgehebelt.“²¹¹

Besonders deutlich wird das in dem Fall, dass eine Rechteinhaberin sich (nur) mit Wirkung für die Zukunft entscheiden möchte, die TDM-Nutzung neuer Werke nicht länger vorzubehalten. Sie müsste dann entweder jeden neuen Inhalt zum TDM zulassen (“Allow”), was genau die Umkehrung des in § 44b Abs. 3 UrhG angelegten Regel-Ausnahme-Verhältnisses bedeuten würde. Oder sie müsste den allgemeinen Vorbehalt aufgeben (“Allow: /”), aber für jeden bis dahin vorbehaltenen Altinhalt nachträgliche Individualvorbehalte definieren, was schon aus arbeitsökonomischen Gründen kaum realistisch erscheint.²¹² Soll also der Nutzungsvorbehalt als echte Entscheidung der Rechteinhaberin zwischen realistischen Alternativen ernstgenommen werden, bleibt nichts anderes übrig, als ihn individualisiert für jeden vorbehaltenen Inhalt auszusprechen.

III. Spawning-Protokoll (ai.txt)

Neben der Nutzung der REP wurde im juristischen Schrifttum vereinzelt die Möglichkeit erwogen, einen „vergleichbaren Standard“ mit einem „granularen Zustimmungssystem [...] in Bezug auf jedes Werk“ zu etablieren.²¹³ Tatsächlich existiert ein solcher Standard bereits. Ein in Minneapolis ansässiges Startup von Medienschaffenden entwickelt unter dem Namen „Spawning“²¹⁴ derzeit „Data Governance for Generative AI“ mit dem selbstgesteckten Ziel, „die“ Infrastruktur für KI-Einwilligungen zu entwickeln (“Building the consent layer for AI”). Dazu arbeitet das Unternehmen auf der einen Seite mit großen KI-Anbieterinnen

²¹¹ BeckOK UrhR (vorm. Möhring / Nicolini) / Engels, 41. Ed. 2024, § 49 UrhG Rn. 14.

²¹² Ähnl. schon Mezei, Eur. IP Rev. 46 (2024), 461 (468): “Rightholders shall be allowed to decide whether they want to include [...] new contents in their earlier reservations or [...] change their mind and want to allow certain TDM activities for third parties. In such a case, a general / full reservation of rights could conflict with the rightholders’ altruism to offer selected contents for ML.”

²¹³ So Maamar, ZUM 2023, 481 (484); mit Verweis darauf erhofft auch Hofmann, WRP 2024, 11 (14 Rn. 18, bei Fn. 36) einen „Standard (vergleichbar dem von ‚robots.txt-Dateien‘ zur automatisierten Kommunikation mit Suchmaschinen)“.

²¹⁴ Die Gründungsumstände sind nur vage bekannt, vgl. Dredge, Holly Herndon reveals plans for her AI-focused startup Spawning (music:)ally, 16.11.2023) <musically.com/2023/11/16/h>.

(namentlich Hugging Face und Stability AI), auf der anderen mit großen Repertoire-Inhaberinnen (namentlich Shutterstock und ArtStation) zusammen.²¹⁵

Eines der ersten Spawning-Produkte ist eine am 30. Mai 2023 vorgestellte Technik unter der Bezeichnung ai.txt,²¹⁶ die genau wie robots.txt funktionieren soll (also als im Wurzelverzeichnis abzulegende Textdatei, die Bots freiwillig beachten sollen), die aber die vorbeschriebene Überinklusivität des REP zu beheben verspricht (“Keep yourself searchable, while restricting AI training”).²¹⁷ Dabei scheint sich weniger die Syntax von derjenigen des Robots Exclusion Protocol zu unterscheiden – was mangels öffentlich dokumentierten Standards letztlich unklar bleibt –, sondern vor allem die Modalitäten des Dateiaufrufs: Während die robots.txt nur beim erstmaligen Betreten einer Website geprüft wird (also nur wenn Inhalte nicht per Direktlink von anderen Websites heruntergeladen werden), soll die ai.txt mithilfe einer eigenen Schnittstelle für KI-Entwicklerinnen („Spawning API“) bei jedem Aufruf eines Inhalts geprüft werden.²¹⁸

Das allein behebt wohl kaum die oben skizzierten Einwände gegen das REP; immerhin stellt die Absonderung der ai.txt von der robots.txt aber sicher, dass Vorbehaltserklärungen sich spezifisch auf TDM-Nutzungen zum KI-Training beziehen. Die im Eingangsbeispiel zitierte Website nutzt diesen Standard allerdings noch nicht und seine öffentliche Erläuterung ist alles andere als transparent.

IV. HTTP Response Header

In Rechtsprechung und Schrifttum nicht völlig unbekannt,²¹⁹ aber im juristischen Schrifttum zu § 44b Abs. 3 UrhG anscheinend noch nie diskutiert,²²⁰ ist die Möglichkeit der Kommunikation durch sog. Hypertext Transfer Protocol (HTTP) Response Header. Dabei handelt es sich schlicht um die maschinenles-

215 About Spawning <spawning.ai/about> abgerufen am 7.3.2024.

216 Dazu bereits *Keller/ Warso* (Fn. 158), S. 8 f.; *Mezei*, Eur. IP Rev. 46 (2024), 461 (467 bei Fn. 61).

217 Spawning ai.txt <spawning.ai/ai-txt> sowie <site.spawning.ai/spawning-ai-txt> abgerufen am 7.3.2024.

218 So *Miller*, ai.txt: A new way for websites to set permissions for AI (Spawning Blog, 30.5.2023) <spawning.substack.com/p/aitxt-a-new-way-for-websites-to-set>.

219 Aus der Rspr. bspw. LG Berlin, Urt. v. 25.7.2017 – 15 O 251/16 zur „X-Frame-Option des HTTP Response Headers“; LG Köln, Urt. v. 1.12.2022 – 14 O 163/22 zur „Verwendung des View http Response Header“; BKartA, Beschl. v. 6.2.2019 – B6-22/16, Fn. 113 („Der HTTP-Client (Browser) sendet seine Anfrage (HTTP-Request) an den HTTP-Server (Webserver/ Web-Server). Dieser bearbeitet die Anfrage und schickt seine Antwort (HTTP-Response) zurück.“); zum Schrifttum vgl. nächste Fn.

220 Nur für Art. 4 DSM-RL vgl. *Mezei*, Eur. IP Rev. 46 (2024), 461 (467) zu “declaring a choice in an HTTP response”.

bar strukturierte Antwort eines Servers auf eine Datenanfrage²²¹ – wie im Eingangsbeispiel bei Fn. 2 illustriert.

Mit dieser Antwort²²² kommuniziert der Server zuerst einen Statuscode (im obigen Beispiel „200“ für „OK“) und sodann weitere sog. „Meta“-Daten (von griech. μετά für „mit, bei, zu“ im Sinne von „beigegebenen“ Daten, die die eigentlich angefragten begleiten, beschreiben oder kategorisieren). Damit erlaubt das HTTP-Protokoll eine auf einzelne Dateien bezogene Kommunikation zwischen Server und Browser, die dem Ausdrücklichkeitserfordernis besser gerecht werden könnte als die bislang erörterten Möglichkeiten der pauschalen Erklärung in Textdateien (oben E.II. und E.III.). Nicht umsonst erwähnte auch die Richtlinie „Metadaten“ als denkbaren Standort eines Nutzungsvorbehalts (ErwGr. 18 UA 2 S. 2 DSM-RL) und ein großes IT-Unternehmen bezeichnete Metadaten als „praktikabelste Methode zur Überprüfung von Nutzungsvorbehalten.“²²³

Folgerichtig war das HTTP-Protokoll einer von drei Kommunikationsstandards, die eine Arbeitsgruppe des World Wide Web Consortium (W3C) im Februar 2024 für die Erklärung von Nutzungsvorbehalten vorschlug – das sog. *TDM Reservation Protocol* (TDM ReP).²²⁴ Danach ließe sich ein Nutzungsvorbehalt dadurch erklären, dass das Metadatum „tdm-reservation“ mit dem Wert 1 in die Antwort des Servers eingefügt und genauere Bestimmungen in einer Datei hinterlegt werden, deren Adresse im Metadatum „tdm-policy“ mitgeteilt wird (Ziff. 6.2 TDM ReP). Im Eingangsbeispiel (oben nach Fn. 2) lautete die Adresse der TDM Policy rsw.beck.de/beck-online-service/tdm-vorbehalt, wo derzeit eine HTML-Datei wortgleichen Inhalts wie der oben bei Fn. 7 zitierte Text zu finden ist:

„Text und Data Mining gemäß § 44b UrhG: Der Verlag behält sich das Recht zu Vervielfältigungen für das Text und Data Mining gemäß § 44b UrhG vor.“

221 Vgl. schon *Stiemerling/Lachenmann*, ZD 2014, 133 (136): „Wird eine Webseite aufgerufen, sendet der eigene Browser eine Anfrage (HTTP-Request) an den Server der Webseite. Dieser kann dann in Verbindung mit seiner Antwort (HTTP-Response) – also der Darstellung der Webseite – ein kleines Datenpaket an den Browser senden.“

222 Zu ihrer Rechtsnatur *Peukert*, ZUM 2023, 233 (246 bei Fn. 108): „Bei dieser Antwort handelt es sich um eine automatisierte, aber gleichwohl ‚echte‘ Willenserklärung des für die Webseite Verantwortlichen, die nach ganz hM den allgemeinen Grundsätzen der Rechtsgeschäftslehre unterliegt.“

223 IBM Intellectual Property Law, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 2: “the most feasible method for checking reservation of rights for online content is by using common metadata. Using metadata would overcome the issue of readability as tools to parse metadata can be implemented fairly trivially and economically.”

224 W3C TDMReP Final Community Group Report (2.2.2024) <w3c.github.io/cg-reports/tdmrep/CG-FINAL-tdmrep-20240202> – im Folgenden zitiert als “TDM Reservation Protocol”; dazu schon *Maamar*, ZUM 2023, 481 (484); *Bomhard*, DSRITB 14 (2023), 255 (267 f.); *Durantaye*, ZUM 2023, 645 (651 Fn. 64); *Keller/Warso* (Fn. 158), S. 7–8; *Löbbling u.a.*, JIPITEC 2023, 499 (507 Rn. 51 f.); *Mezei*, Eur. IP Rev. 46 (2024), 461 (467 bei Fn. 60).

In diesem konkreten Fall enthält die Datei also keine genaueren Bestimmungen (zu dieser Möglichkeit erst unten 6.) sondern einen „menschenslesbaren“ Vorbehalt,²²⁵ den der Bot ebenso ignorieren muss wie den Text von Impressum oder AGB.

Im Unterschied zum fast dreißig Jahre älteren REP ist das TDM ReP keineswegs alternativlos. Auch andere Formen des Vorbehalts per HTTP Response Header kursieren in Fachkreisen – etwa die Verwendung des ebenfalls für Suchmaschinen entwickelten Metadatum „X-Robots“ mit einem Wert wie „user-agent: CC-BY“.²²⁶

V. HTML Metadaten-Tags

Eine andere Art von Metadaten illustriert das Eingangsbeispiel bei Fn. 6: Die auch im juristischen Schrifttum als möglicher Standard für die Kommunikation von Nutzungsvorbehalten erörterten „Meta-Elemente einer HTML-konformen Website“.²²⁷

HTML (Hypertext Markup Language) ist als sog. „Auszeichnungssprache“ (*markup language*)²²⁸ ein auch für Menschen lesbares Textformat, das neben *semantischem* Inhalt auch umfangreiche *syntaktische* Informationen zu verschriftlichen erlaubt. So wie natürliche Sprache durch bestimmte Syntaxelemente einen Text strukturiert (diese Klammern zum Beispiel nebensächliche Kommentare oder erläuternde Beispiele aus dem eigentlichen Textfluss heraushalten), lassen sich auch in der HTML-Sprache inhaltlich irrelevante, aber den Text strukturierende Auszeichnungselemente (*tags*) „ausklammern“. Dazu dienen die zu <spitzen Klammern> umfunktionierten kleiner/größer-als-Zeichen. Als Beispiel mögen die im Eingangsbeispiel vor Fn. 6 erwähnten „achtzig Zeilen [...] in Maschinensprache“ dienen.²²⁹

225 Ziff. 5.2 TDM ReP: “A TDM Policy is considered human readable if its content-type is text/html. It is considered machine-readable if its content-type is either application/json or application/ld+json”; ausdr. auch *Löbbling u.a.*, JIPITEC 2023, 499 (507 Rn. 51): “if the information at this URL is solely available in HTML or text formats, it is not considered machine-readable. To achieve machine-readability, policies must be articulated using JSON or JSON-LD”.

226 So *Wiese* (Fn. 197), wobei dieser Vorschlag keine Option zur TDM-Versagung vorsieht.

227 *Hartmann/Jacobsen*, MMR-Aktuell 2021, Nr. 441332, in II.3; *Baumann*, NJW 2023, 3673 (3675 Rn. 19): „Jedenfalls zu beachten sind Anweisungen an Crawler [...] in den Metadaten einer Website“; *Kleinkopf* (Fn. 18), S. 249 f.: „Unter den Vorbehalt fallen jedenfalls Angaben im Quelltext [...] in HTML-kompatibler Textform“; ebenso *Konertz/Schönhof*, WRP 2024, 289 (296 Rn. 33, nach Fn. 110) zu „eingebetteten Metainformationen“; *Löbbling u.a.*, JIPITEC 2023, 499 (506 Rn. 40): “meta tags could serve as suitable machine-readable methods to accurately convey opt-outs for TDM”.

228 Zu diesen ausf. *Wöbbeking* (Fn. 90), S. 80–84.

229 *Allg. Peukert*, ZUM 2023, 233 (246 bei Fn. 107): „Seitenbetreiber können [...] mithilfe von sogenannten <meta>-Tags im HTML-Quelltext festlegen, ob und inwieweit auf die automatische Abfrage eines Suchbots geantwortet wird.“; auch *IBM Intellectual Property Law*, StN

Anders als die HTTP-Metadaten werden HTML-Metadaten nicht einer übermittelten Datei „beigegeben“, sondern innerhalb dieser Datei ihrem „eigentlichen“ Inhalt. Mit einer Analogie aus der physischen Welt könnte man sagen, dass HTTP-Metadaten dem Lieferschein eines Buches entsprechen, HTML-Metadaten dagegen seinem Impressum: Letzteres befindet sich zwar innerhalb des Buches, ist aber dennoch „beigegeben“ im Sinne einer „nicht inhaltstragenden“ Seite. Die Analogie verdeutlicht, dass zwischen Lieferschein und Impressum genau unterschieden werden muss, auch wenn sich manche Informationen (Titel, Erscheinungsjahr, etc.) an beiden Stellen finden können, während andere nur an einer der beiden Stellen sinnvoll Platz finden (etwa das Lieferdatum nur im Lieferschein, die Illustratorennamen nur im Impressum).

HTML-Metadaten sind der zweite Standard im TDM Reservation Protocol für die Erklärung von Nutzungsvorbehalten. Dazu sieht Ziff. 6.3 TDM ReP vor, dass sie in einem Auszeichnungselement der Klasse <meta ...> mit denselben Eigenschaften angegeben werden wie schon im HTTP Response Header, also den Namen „tdm-reservation“ bzw. „tdm-policy“ und den Werten 1 bzw. dem Dateistandort. Da innerhalb einer Datei auch mehrere <meta>-Tags verwendet werden können, könnten auf diese Art sogar verschiedene Absätze desselben Textes unterschiedlich ausgezeichnet werden. Damit erlaubt es dieser Kommunikationsstandard nun endgültig, einen wirklich auf einzelne Inhalte bezogenen (und damit „ausdrücklichen“) Nutzungsvorbehalt zu erklären. Dies allerdings – wie oben bei Fn. 227 bereits angedeutet – nur in „HTML-konformen“ Dateien. Als einziges anderes Dateiformat behandelt das TDM Reservation Protocol elektronische Bücher im .epub-Format (Ziff. 6.4 TDM ReP), während etwa für Bilddateien doch wieder andere Standards gefunden werden müssten.²³⁰

Zudem ist der Wildwuchs an denkbaren Metadaten-Tags noch größer als in anderen Kommunikationsstandards. So finden sich einschließlich des TDM ReP-Vorschlags mindestens fünf verschiedene seit 2012 vorgeschlagene Metadaten-Attribute zur Kommunikation eines Nutzungsvorbehalts:

```
<meta name="CCBot" content="nofollow">231
<meta name="robots" content="noai, noimageai">232
<meta name="usage-rights" content="CC-BY-SA" />233
```

v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 2 zum “structuring the metadata according to a predefined format /syntax (e.g., using tags)”.

230 Bspw. Bundesverband professioneller Bildanbieter, StN v. 31.1.2020 zum DiskE (Fn. 21), S. 7: „Die Untersagung auf Dateiebene bei Fotodateien könnte in einem künftig erweiterten Standard der IPTC-Foto-Metadaten erfolgen“; ebenso *Maamar*, ZUM 2023, 481 (484).

231 Common Crawl FAQ seit 6.12.2012 <commoncrawl.org/faq>.

232 DeviantArt, UPDATE All Deviations Are Opted Out of AI Datasets (11.11.2022) <deviantart.com/team/journal/UPDATE-All-Deviations-Are-Opted-Out-of-AI-Datasets-934500371>; das “robots“-Attribut diene ursprünglich ebenso wie die robots.txt (oben E.II.) zur Steuerung von Suchmaschinenbots.

233 *Wiese* (Fn. 197); dieser Vorschlag sieht keine Option zur TDM-Versagung vor.

```
<meta name="generative-ai" content="notraining">234
<meta name="tdm-reservation" content="1"> <meta name="tdm-policy"
content="...">235
```

Selbst große, kommerzielle und bekanntermaßen rechtesensible Verlagshäuser nutzen nicht alle fünf standardisierten Vorbehaltsvarianten auf ihrer Website (vgl. Eingangsbeispiel nach Fn. 6).

VI. JSON (tdmrep.json, Reichs ai.txt, C2PAs)

Der dritte Kommunikationsstandard, den die W3C-Arbeitsgruppe vorschlug, war die JavaScript Object Notation (JSON) – „ein verbreitetes Format zur Strukturierung von Daten zur Datenbereitstellung bzw. Datenaustausch, das von vielen Programmiersprachen verarbeitet werden kann“.²³⁶ Es ist seit 1997 in zwei technischen Standards (RFC 8259 und ECMA-404) dokumentiert. Die im Eingangsbeispiel genutzte Website scheint es bislang nicht zu nutzen.

Nach Ziff. 6.1 TDM ReP wäre im Wurzelverzeichnis der Website eine Textdatei mit dem Dateinamen *tdmrep.json* abzulegen, in der jede Information durch ein Attribut (z.B. „vcard:hasEmail“) und einen Wert (z.B. „mailto:contact@provider.com“) repräsentiert wird. Diese Attribut / Wert-Zuordnungen können beliebig verschachtelt und so auch rechtliche Befugnisse kodiert werden, wofür das TDM Reservation Protocol den Standard „Open Digital Rights Language“ (ODRL) in der Version 2.2 nutzt.²³⁷ Damit wäre folgende Beispielerklärung denkbar:²³⁸

```
"permission": [ {
  "action": "tdm:mine",
  "duty": [ {
    "action": "compensate"
  } ]
}]
```

Dieser Code-Ausschnitt definiert also eine Gestattung („permission“) für die Handlung („action“) des Text und Data Mining („tdm:mine“), mit einer Gegenverpflichtung („duty“) zur Entschädigung der Rechteinhaberin („compensate“) – also nach §§ 133, 157 BGB einen Antrag auf Abschluss einer vergütungspflichtigen

²³⁴ Bustos, Generative AI in web development. A new AI meta tag? (LinkedIn 29.7.2023) <linked in.com/pulse/generative-ai-web-development-new-meta-tag-eduardo-bustos>; dieser Standard hatte weniger § 44b UrhG im Blick als die Kennzeichnung von KI-Output, der zur Vermeidung von Rückkopplungen (*feedback loops*) von weiterem Training ausgeschlossen werden sollte.

²³⁵ Ziff. 6.3 TDM ReP; weiterer Vorschlag ohne Nutzung des <meta>-Auszeichnungselements (sondern eines neuen „data-notdm“-Attributs) bei *Löbbling u.a.*, JIPITEC 2023, 499 (509 Rn. 68).

²³⁶ *Käde* (Fn. 14), S. 139 Fn. 343.

²³⁷ Ausf. ODRL Information Model 2.2 (W3C Recommendation, 15.2.2018) <w3.org/TR/odrl-model>.

²³⁸ Aus Bsp. 14 in Ziff. 7.1.5.3 TDM ReP.

Lizenz,²³⁹ der dank seiner Maschinenlesbarkeit automatisiert angenommen werden könnte. Insofern beinhaltet dieses Erklärungsschema also eine Spezialsyntax für sog. *smart contracts* und beantwortet die bei Schaffung des § 44b Abs. 3 UrhG viel zu selten erörterte Frage,

„ob nur binäre Botschaften (Nutzung erlaubt / untersagt) oder auch komplexere maschinenlesbare Anweisungen (Nutzung erlaubt unter Bedingungen, die z.B. an die Person des Nutzers, einen von ihm gezahlten Preis oder ein von ihm verfolgtes Nutzungsziel anknüpfen) möglich sind.“²⁴⁰

Trotz dieser im Gesetzgebungsverfahren aufgeworfenen Frage und des Marktschaffungszwecks des Nutzungsvorbehalts (oben bei Fn. 45) ist selten die Rede davon, dass weitere Vertragsbedingungen maschinenlesbar kommuniziert werden müssten als die binäre Entscheidung „Erlaubnis oder Verbot“.²⁴¹ Insofern erweist sich der TDM ReP-Standard als geradezu visionär.²⁴²

Selbst dieser Standard jedoch ist in der JSON-Syntax nicht alternativlos. In einem am 18. Mai 2023 ins GitHub-Repositoryum eingestellten Entwurf für einen „Guide for Preparing Website Content for Large Language Models“²⁴³ schlug KI-Unternehmer *Robert Reich* einen anderen – wohl ebenfalls JSON-basierten – Standard vor, der „eher wie RSS als wie robots.txt“ funktionieren²⁴⁴ und ebenfalls durch eine ai.txt benannte Datei kommuniziert werden soll.²⁴⁵ Über eine Zweckbeschreibung und eine Beispieldatei scheint dieser Standard anders als das TDM ReP bislang nicht hinausgekommen zu sein.

Zudem kommen JSON und ähnliche (etwa davon abgeleitete) Formate auch in viel komplexeren Software-Architekturen zum Einsatz. Ein solcher Fall, der im rechtlichen Kontext vereinzelt erwähnt wurde,²⁴⁶ ist das sog. C2PA-Protokoll der industrienahen „Coalition for Content Provenance and Authenticity“.²⁴⁷ Es dient neben anderen Zwecken auch der Kommunikation von TDM-Nutzungs-

239 Im Beispiel fehlen noch die Lizenzkosten als *essentiale negotii*, die mittels des „payment“-Elements und seiner Attribute zu definieren wären, vgl. Bsp. 21 im ODRL Information Model (Fn. 237).

240 Börsenverein, StN v. 6.9.2019 zur BMJV-Konsultation (Fn. 20), S. 3 = StN v. 31.1.2020 zum DiskE (Fn. 21), S. 11 = StN v. 5.11.2020 zum RefE (Fn. 22), S. 10.

241 Nur *Maamar*, ZUM 2023, 481 (484) wollte unterscheiden, „ob die Nutzung verboten, erlaubt oder nur gegen Vergütung erlaubt sein soll.“

242 Lediglich die in Ziff. 6.1.2 TDM ReP vorgesehene pauschale Ausschließbarkeit jeglicher Inhalte mit dem Wert „/“ im „location“-Attribut, auf die mich *Tristan Radtke* hinwies, begegnet denselben Einwänden wie die „Disallow: /“-Erklärung im REP (dazu oben E.II. am Ende).

243 Benutzer *menro*, ai.txt <github.com/menro/ai.txt> abgerufen am 7.3.2024.

244 Benutzer *menro* (Fn. 243): „In essence, AI.txt is more akin to RSS feeds than robots.txt.“

245 Hier drohen der Maschinenlesbarkeit neue Hürden: Wenn sowohl Spawning AI als auch Robert Reich mittels einer „ai.txt“ kommunizieren, bräuchte es wiederum eines Metastandards, um die in der Datei verwendete Syntax vorweg mitzuteilen, damit ein Bot sie überhaupt einlesen (*parsen*) kann.

246 *Keller / Warso* (Fn. 158), S. 9; *Löblich u.a.*, JIPITEC 2023, 499 (507 Rn. 54).

247 Technische Spezifikationen (April 2023) unter c2pa.org/specifications/specifications/1.3.

vorbehalten (Ziff. 19.21 C2PA-Spezifikation), wie zum Beispiel in einer von JSON abgeleiteten Syntax (CBOR Diagnostic):²⁴⁸

```
{
  "entries":
    "c2pa.ai_training" : {
      "use" : "allowed"
    },
    "c2pa.ai_generative_training" : {
      "use" : "notAllowed"
    },
    "c2pa.data_mining" : {
      "use" : "constrained",
      "constraint_info" : "may only be mined on days whose names end
in 'y'"
    }
  }
}
```

Das Protokoll definiert für Nutzungsvorbehalte vier mögliche Modalitäten (Ziff. 19.21 C2PA-Spezifikation), von denen das Beispiel drei illustriert. Jede davon kann das „use“-Attribut mit einem von drei möglichen Werten belegen, wobei der Wert „constrained“ wie im Beispiel beliebig komplexe Nutzungsbedingungen gestattet.

Was diese JSON-Erklärung von den zuvor genannten unterscheidet, ist die Art ihrer Übermittlung an den Empfänger: Sie soll nicht in einer Plaintext-Datei abgelegt werden, die jeder geübte Nutzer mit einem simplen Texteditor lesen und bearbeiten könnte. Stattdessen definiert das C2PA-Protokoll einen technisch anspruchsvollen Prozess, um den zitierten Text in einen kryptographisch signierten „Claim“ zu verpacken und (ggf. zusammen mit anderen) als „Manifest“ in eine binärkodierte Datei einzubetten. Ungeachtet technischer Feinheiten ergeben sich daraus rechtsrelevante Konsequenzen:

Zum einen bedeutet das kryptographische Element (Ziff. 14 C2PA-Spezifikation), dass C2PA-Nutzungsvorbehalte nicht allein durch Menschen erzeugt oder gelesen werden können. Der „Claim“ ist vielmehr zwingend durch einen „nicht-menschlichen (Hardware- oder Software-) Akteur“ zu erzeugen (Ziff. 2.1.3 C2PA-Spezifikation). Was aus Ingenieursperspektive vertrauensgenerierend erscheinen mag, könnte aus juristischer Perspektive das Gegenteil bewirken und wirft deshalb die noch nicht erörterte Frage auf, ob Nutzungsvorbehalte nicht zumindest *auch* menschenlesbar sein müssen. Während Maschinenlesbarkeit die gleichzeitige Menschenlesbarkeit ganz sicher nicht ausschließt,²⁴⁹ spricht das Ausdrücklichkeitserfordernis wohl dafür (vgl. Zitat oben bei Fn. 110), dass sie sie sogar *voraussetzt*.

²⁴⁸ Beispiel aus Ziff. 19.12.1 C2PA-Spezifikation (vorige Fn.).

²⁴⁹ So zu Recht *Mezei*, Eur. IP Rev. 46 (2024), 461 (466 nach Fn. 49): “machine-readable means do not exclude human-readability of the reservation”.

Eine zweite Konsequenz der skizzierten Software-Architektur liegt darin, dass das resultierende „Manifest“ zwar in binärkodierte Dateien eingebettet werden kann (Ziff. 3.4 C2PA-Spezifikation nennt Bildformate und .pdf), für die sich andere Protokolle weniger eignen (dazu oben bei und in Fn. 230). Umgekehrt bedeutet das aber, dass derselbe Nutzungsvorbehalt nicht auch in einfacheren Formaten (wie Textdateien auf dem Server) oder generell Plaintext-basierten Inhalten (wie Websites) eingebettet werden kann. Für Textinhalte, wie sie unser Eingangsbeispiel illustrierte, ist das C2PA-Protokoll daher von vornherein ungeeignet. Das lässt vermuten, dass ein allgemeiner und dateiformatübergreifender Standard für Nutzungsvorbehalte kaum zu erwarten ist. Schon deshalb erscheint eine bleibende Fragmentierung von Standards unvermeidlich.

VII. Technische Zugangshürden (Paywall, CAPTCHA, etc.)

Eine letzte Möglichkeit, das KI-Training zu verhindern, liegt schlicht darin, die Inhalte der ausgelesenen Website hinter ein Login-Formular oder eine Paywall zu verbannen, so dass sie für Bots unsichtbar werden. Dies illustriert das Eingangsbeispiel bei Fn. 11. Solche technischen Zugangshürden werden für § 44b UrhG in zweierlei Hinsicht relevant:

Zum einen dürfen Rechtsinhaberinnen technische „Maßnahmen treffen, mit denen sie sicherstellen, dass etwaige Nutzungsuntersagungen Beachtung finden“.²⁵⁰ Hier wird also ein mit anderen Mitteln (d.h. rechtsgeschäftlich) erklärter Nutzungsvorbehalt mithilfe „wirksamer technischer Maßnahmen“ i.S.v. § 95a II 1, 2 UrhG durchgesetzt.²⁵¹ Demgegenüber ist die TDM-Schranke allerdings insofern technikfest gestaltet, als den Nutzerinnen „die notwendigen Mittel zur Verfügung“ gestellt werden müssen, um von der Schranke trotz technischer Schutzmaßnahmen „in dem erforderlichen Maße Gebrauch machen [zu] können“ (§§ 69f Abs. 3, 95b Abs. 1 Nr. 1 UrhG).²⁵²

Zum anderen werden technische Zugangshürden bisweilen selbst als eine Art von Nutzungsvorbehalt gedeutet.²⁵³ Dafür berufen sich einzelne Autorinnen²⁵⁴ auf die EuGH-Entscheidung zum Website-Framing, in der es heißt:

„Um die Rechtssicherheit und das ordnungsgemäße Funktionieren des Internets zu gewährleisten, ist es dem Urheberrechtsinhaber nicht gestattet, seine Erlaubnis auf andere Weise als

²⁵⁰ RegE zum DSM-RL-UmsG (Fn. 33), S. 89 m. Verw. auf ErwGr. 18 UA 2 S. 5 DSM-RL.

²⁵¹ Krit. *Mezei*, Eur. IP Rev. 46 (2024), 461 (465): “Another open issue is the way how rights reservation might correlate with the existing rules on technical protection measures (TPM) and rights management information (RMI) under the InfoSoc Directive.”

²⁵² Zu dieser „Schrankendurchsetzung gegenüber technischen Schutzmaßnahmen“ ausf. *Kleinkopf* (Fn. 18), S. 251.

²⁵³ *Meys*, GRUR-Int. 2020, 457 (466 Fn. 157): “it can be argued that machine-readable means refer to the systems used to prevent the algorithm from mining the contents of a source. In the absence of such systems, it should be assumed that the source can be freely mined.”

²⁵⁴ Insb. *Rosati*, Copyright in the Digital Single Market, 2021, S. 90 als “reasoning by analogy”.

durch wirksame technische Maßnahmen im Sinne von Art. 6 Abs. 1 und 3 der Richtlinie 2001/29 zu beschränken.“²⁵⁵

Würden diese Erwägungen, die ja ganz allgemein auf das „Funktionieren des Internets“ gemünzt waren, auch für den Kontext des TDM-Nutzungsvorbehalts gelten, wären alle bislang besprochenen Kommunikationsprotokolle untauglich, da sie als „Kommunikation durch Code“ (oben bei Fn. 194) vom Nutzer ignoriert werden können und damit nicht „durch eine Zugangskontrolle oder einen Schutzmechanismus [...] unter Kontrolle gehalten“ werden im Sinne der vom EuGH herangezogenen Legaldefinition des Art. 6 Abs. 3 S. 2 InfoSocRL (§ 95a Abs. 2 S. 2 UrhG). Denkbar wären Nutzungsvorbehalte dann wohl nur noch als „Login- oder Paywall-Schranken“²⁵⁶ oder etwa CAPTCHA-Abfragen,²⁵⁷ bei denen der Zugriff auf dynamische Webseiteninhalte vom Bestehen eines Tests abhängig gemacht wird, der etwa das Erkennen von Symbolen auf einem Bild verlangt.“²⁵⁸

Solche Vorschläge indes lassen sich angesichts der oben (DVII.) entwickelten Dogmatik des § 44b UrhG kaum aufrechterhalten. Inwiefern nämlich in der rein faktischen Zugangerschwerung eine rechtsgeschäftsähnliche Vorbehaltserklärung liegen könnte, erklären die zitierten Autoren nicht.²⁵⁹ Die Einrichtung technischer Zugangshürden mag viele Gründe haben, eine Erklärung i.S.v. § 44b Abs. 3 UrhG ist damit keineswegs notwendig verbunden; jedenfalls wäre sie konkludent und entbehrte damit wiederum der Ausdrücklichkeit. Auch die Vorgabe des europäischen KI-Gesetzes, dass KI-Anbieter eine Strategie zur Einhaltung von Nutzungsvorbehalten benötigen (Art. 53 Abs. 1 lit. c KIG), ergibt nur dann Sinn, wenn Nutzungsvorbehalte nicht selbstvollziehend („wirksam“ i.S.v. § 95a Abs. 3 S. 2 UrhG) sein müssen. Die Grundsätze des EuGH zum Website-Framing können also für TDM-Nutzungsvorbehalte nicht gelten.

Ebenso wenig liegt ein Nutzungsvorbehalt in der rein technischen Selbsthilfe durch das Auslegen vergifteter KI-Köder (sog. *poisoning*), die für Bilder schon intensiv erprobt wird. So wurde an der Universität Chicago die Software *Glaze* vorgestellt, die „promptspezifische Giftangriffe auf Bildgeneratoren“ ermöglicht,²⁶⁰

255 EuGH, Urt. v. 9.3.2021 – C-392/19 (*VG Bild-Kunst v. Stiftung Preussischer Kulturbesitz*), Rn. 46.

256 Dafür wohl *Gerecke*, GRUR-Prax. 2023, 381 (382 Rn. 11): „Solange der Rechteinhaber also seine Werke nicht hinter eine Login- oder Paywall-Schranke stellt, [...] darf auf die Werke zugegriffen werden.“

257 Dafür wohl *Kleinkopf* (Fn. 18), S. 249 f.: „Unter den Vorbehalt fallen jedenfalls [...] CAPTCHA-Programme.“

258 *Kleinkopf* (Fn. 18), S. 51 bei Fn. 129.

259 Genauer deshalb *Dreier/Schulze/Dreier*, 7. Aufl. 2022, § 44b UrhG Rn. 10: „Auch ohne einen Vorbehalt erklärt zu haben, können Rechteinhaber fremdes Text und Data Mining rein faktisch dadurch verhindern, dass sie ihre Schutzgegenstände durch technische Schutzmaßnahmen gegen Zugang und Vervielfältigung schützen.“ (Hervorhebungen nur hier).

260 So der Titel von *Shan u.a.*, Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models (arXiv, 20.10.2023 [Version 1], 16.2.2024 [Version 2]) <arxiv.org/abs/2310.13828v2>.

indem sie Bilder für Menschen unsichtbar verändert, dadurch aber den KI-Algorithmen statt eines Hundes plötzlich eine Katze, statt einer Handtasche einen Toaster und statt eines STOP-Zeichens einen Vogel vorgaukelt.²⁶¹ Die sicherheitsrelevanten Konsequenzen gerade des letztgenannten Beispiels könnten immerhin Druck auf KI-Entwicklerinnen ausüben, kein unlizenziertes Material zu verwenden. Für Texte scheint diese Strategie wenig erfolgversprechend, doch das bereits erwähnte Startup Spawning bietet neben dem ai.txt-Protokoll (dazu oben E.III.) auch ein Programm namens *Kudurru* an, das Serveranfragen von KI-Datensammlern identifizieren und mit nutzlosen Pseudodaten quittieren soll.²⁶² Dies wäre auch für Texte denkbar, aber wiederum nur als faktische Hürde und nicht als Vorbehaltserklärung.²⁶³

VIII. Zwischenergebnis: “Standards are great. Everyone should have one!”

Auch wenn unsere Gesamtschau der bisher vorgeschlagenen (in der Rechtsliteratur nur zum Teil thematisierten und fast nie detailliert erörterten) Kommunikationsstandards sicher kein „exponentielles Wachstum“ denkbarer Opt-Out-Standards belegt,²⁶⁴ bewahrheitet sie doch einen alten Ingenieurswitz:

“The nice thing about standards is that you have so many to choose from”²⁶⁵

Diesem Befund lässt sich je nach Erkenntnisinteresse unterschiedlich begegnen:

Aus technologischer Perspektive ist zu fragen, welches Vorrangverhältnis zwischen verschiedenen Standards gelten soll und wie denkbare Widersprüche aufzulösen sind, wenn Vorbehalte auf mehreren unterschiedlichen Kanälen kommuniziert werden. Zu dieser Frage der “Processing Priority” sieht beispielsweise das TDM Reservation Protocol eine eigene Regelung vor (Ziff. 6.5 TDM ReP).²⁶⁶

Aus rechtspraktischer Perspektive muss nach dem Grundsatz des sichersten Weges wohl einstweilen zur Nutzung *jedes* denkbaren Standards geraten werden: „Nutzer sind derzeit gut beraten, alle Formen von Nutzungsvorbehalten zu beachten; Rechteinhaber, alle Formen von Nutzungsvorbehalten zu erklären.“²⁶⁷

261 Vgl. Abbildung 7 auf S. 8 der pdf-Version von *Shan u.a.* (Fn. 260).

262 *Knibbs*, A New Tool Helps Artists Thwart AI – With a Middle Finger (WIRED, 12.10.2023) <[wired.com/story/kudurru-ai-scraping-block-poisoning-spawning](https://www.wired.com/story/kudurru-ai-scraping-block-poisoning-spawning/)>.

263 So auch *Mezei*, Eur. IP Rev. 46 (2024), 461 (467): “far from being right[s] reservations per the CDSM Directive”.

264 So aber *Mezei*, Eur. IP Rev. 46 (2024), 461 (465): “the number of different opt-out models has grown exponentially.”

265 *Tanenbaum*, Computer Networks, 1981, S. 168.

266 Ausf. jüngst *Wöbbeking* (Fn. 90) zur „relativen Auslegungsrelevanz“ (S. 270) und „Hierarchie formalsprachlicher Versionen zum selben Regelungsgegenstand“ (S. 281 f.), wobei aber eher „verschiedene übereinander liegende Versionen“ (S. 260) desselben Protokolls gemeint waren, also die Konkurrenz verschiedener Erklärungsschichten (*layer*), als die Konkurrenz verschiedener Erklärungsprotokolle.

267 *Baumann*, NJW 2023, 3673 (3675 Rn. 19).

Aus rechtstheoretischer Perspektive sind zwei verschiedene Herangehensweisen denkbar: Eine optimistische (unter der Prämisse, dass „sich über kurz oder lang ein akzeptierter Standard entwickeln“ werde²⁶⁸), aber auch eine pessimistische (unter der Prämisse, dass „der Gesetzgeber sehenden Auges die Regelung leerlaufen lassen wollte“²⁶⁹) – je nachdem, ob man lieber die technische oder die politökonomische Desillusionierung riskiert.

Aus technikatrichterlicher Perspektive lässt sich immerhin feststellen, dass einige Standards eklatant ungeeignet sind, den dogmatischen Vorgaben des § 44b Abs. 3 UrhG zu entsprechen (oben I.–III., VII.). Unter den danach noch denkbaren Erklärungsprotokollen (IV.–VI.) herrscht allerdings reger Wildwuchs, dessen restlose Bereinigung ebenso wenig zu erhoffen ist wie die Standardisierung aller anderen Formate im Internet. Denn wo ein Markt für Vorbehalte geschaffen wird (dazu oben bei Fn. 45), werden schon die kreativen Kräfte des Wettbewerbs neue Standards gebären – wie den unter III. erörterten Vorschlag eines Startups, der letztlich nur als Vehikel zum Verkauf der eigenen Schnittstelle dient.

Was bleibt also, außer diese und weitere Standards fortlaufend zu sichten und zu sammeln?²⁷⁰ Letztlich wohl nur die Einsicht, dass die Gesetzgeberin in § 44b Abs. 3 UrhG mit dem Wort „maschinenlesbar“ eine Büchse der Pandora eingebaut hat, die nur „marktstarke Inhaber von Nutzungsrechten“ – und damit gerade nicht die große Mehrheit der kleinen Kreativen, sondern allenfalls finanzkräftige Repertoireinhaberinnen – wieder zu schließen vermögen.²⁷¹ Ein halbwegs gerichtsfester Vorbehalt von TDM-Nutzungen setzt nämlich voraus (was derzeit nicht einmal dem im Eingangsbeispiel angedeuteten Großverlag gelingt), dass der Berechtigte die Entwicklung neuer Standards laufend im Blick behalten, technisch verstehen (was in manchen Fällen „trivial“ sein mag,²⁷² aber etwa bei JSON-Standards geradezu anspruchsvoll erscheint) und zeitnah einsetzen kann.

268 So Hofmann, WRP 2024, 11 (14 Rn. 18); ähnl. Durantaye, AfP 2024, 9 (15 Rn. 33): „möglicherweise wird sich in absehbarer Zeit ein Standard entwickeln“.

269 So Hartmann/Jacobsen, MMR-Aktuell 2021, Nr. 441332, in II.2.c) mit der Alternative „... oder auf die verständige Auslegung der Norm durch die Gerichte vertraute“.

270 Dies tut bspw. Benutzer *healsdata*, AI Training Opt Out <github.com/healsdata/ai-training-opt-out> abgerufen am 7.3.2024.

271 Schack, NJW 2024, 113 (115 Rn. 16): „Die Existenz eines Verbotungsrechts schafft einen Markt für Lizenzen, der aber nur funktionieren kann, wenn die Lizenzierung eine rechtssichere Nutzung verbürgt. [...] Deren Lizenzierung wäre marktstarken Inhabern von Nutzungsrechten, insbesondere an großen Datenbanken, durchaus möglich; für die Masse der im Netz gescriperten Werke scheidet eine individuelle Lizenzierung jedoch praktisch aus.“; Geiger (Fn. 25), S. 78: “the author will not necessarily benefit directly from this situation as it will likely be the big rightholders that will license the uses”; Durantaye, AfP 2024, 9 (15 Rn. 33): „Selbst ein Vorbehalt, der den Vorgaben entsprechend kommuniziert wurde, ist für Kreative aber nicht mit großen Vorteilen verbunden. [...] Selbst wenn sich herausstellen sollte, dass ein Vorbehalt nicht beachtet wurde, werden die Konsequenzen für die einzelne Urheberin kaum spürbar sein.“

272 So Sag, Nw. U. L. Rev. 103 (2009), 1607 (1667): “The monetary cost of using the Robots Exclusion Protocol is zero and the information costs are not significantly higher. Adding a robots.txt file to a website is trivial”.

F. Fazit und Ausblick

Mit der „maschinenlesbaren Form“ in § 44b Abs. 3 S. 2 UrhG ist das Urheberrecht um einen schillernden Rechtsbegriff reicher, der sich nur in genauer Auseinandersetzung mit den in Betracht kommenden technischen Protokollen sachgerecht konturieren lässt. Einstweilen konkurrieren zahlreiche Standards in jeweils mehreren Varianten, die sich zur Erfüllung der dogmatischen Anforderungen von § 44b UrhG unterschiedlich gut eignen. Nur wirtschaftsstarke Repertoireinhaberinnen werden die technische Expertise entwickeln können, um in diesem Dschungel den Überblick zu wahren; zugleich ist ihre Erklärungsbefugnis alles andere als gesichert.

Denkbar wäre immerhin, dass das aufgrund von Art. 64 KIG im Januar 2024 eingerichtete europäische Amt für künstliche Intelligenz („AI Office“)²⁷³ einige dieser Probleme „quasi-gesetzlich“²⁷⁴ löst. Denn nach Art. 56 Abs. 1, 2 und ErwGr. 116 KIG soll das Amt „die Ausarbeitung von Verhaltenskodizes“ fördern, die „mindestens die in den Artikeln 53 und 55 vorgesehenen Pflichten abdecken“ – also auch die Pflicht zur Einhaltung von Nutzungsvorbehalten gem. Art. 53 Abs. 1 lit. c KIG. Folglich fiele eine amtliche Definition der Maschinenlesbarkeit und eines für Nutzungsvorbehalte geeigneten Standards zweifelsohne in die Regelungsbefugnis des KI-Amtes. Zumal das Amt nach Art. 56 Abs. 8 S. 2 KIG die „verfügbaren Normen“ bewerten helfen soll, womit ausweislich der englischen Fassung („available standards“) ein mittelbarer Standardisierungsauftrag verbunden ist. Werden solche Standards/Normen in einen Verhaltenskodex aufgenommen, können Modellanbieter nach Art. 54 Abs. 4 KIG die Einhaltung ihrer Pflichten „bis zur Veröffentlichung einer harmonisierten Norm“ durch die Befolgung dieses Kodex’ nachweisen.

Selbst wenn aber zuguterletzt EU-weit harmonisierte Standards/Normen entstünden und die Technikoptimistinnen Recht behielten, fehlt es im nächsten Schritt an Infrastrukturen für die Verwaltung und Verteilung etwaiger Einnahmen. Denn dafür bräuchte es „ein zumindest EU-weites Clearinghaus, das die Ansprüche gesammelt durchsetzt“:²⁷⁵

273 Vgl. Commission Decision C (2024) 390 of 24.1.2024 establishing the European Artificial Intelligence Office, verfügbar unter <ec.europa.eu/newsroom/dae/redirection/document/101625>.

274 Vgl. Peukert, GRUR Int. 2024, 497 (504): „in an abstract, quasi legislative way“, und weiter: „meta-regulation of the AI Act could help to resolve these open issues much faster than the conventional copyright system“.

275 Maamar, ZUM 2023, 481 (486) m.Verw. auf Kop, The Right to Process Data for Machine Learning Purposes in the EU, Harv. J. of Law & Tech. 34 (2021), 1 (7): „creation of an online one-stop-shop clearinghouse with mandatory or statutory licensing for machine learning training datasets alike a pan-European, multi-territorial collective rights agency“ m.w.N. in Fn. 39.

“Even if standardized rights reservation protocols – capable of expressing remuneration wishes and modalities – become available, it is unclear whether copyright holders and collecting societies will ever manage to create efficient, pan-European rights clearance solutions that offer reliable and well-functioning payment interfaces with the technical safeguards”²⁷⁶

Vielleicht aber ist die Ohnmacht des Rechts gegenüber diesem technischen Fortschritt nicht das Schlechteste. Denn nach der ökonomischen Theorie des Rechtsschutzes²⁷⁷ gibt es eben nicht nur Verbotungsrechte (*property rules*) und Vergütungsrechte (*liability rules*), deren relative Effizienz sich nach den zu erwartenden Transaktionskosten richtet,²⁷⁸ sondern auch unveräußerliche Rechte (*inalienable entitlements*), zu denen nach internationalem Vorbild²⁷⁹ auch das TDM für das Training künstlicher Intelligenzen zählen könnte. Denn wem sollte wirtschaftlich der Wert einer Nutzung zugewiesen werden, die mehr ist als die Summe ihrer Teile?²⁸⁰ Entstehen durch Nutzungsvorbehalte wirklich die behaupteten Produktionsanreize für Kreative, oder nicht nur innovationshinderliche Hold-up-Situationen zugunsten großer Rechteinhaber? Dies sind weiterführende Fragen, die ökonomischer und empirischer Analyse vorbehalten bleiben müssen und in technikrechtlichen Überlegungen keinen Platz mehr finden.

Summary

Developing artificial intelligence and large language models (LLMs) requires huge amounts of text as training material. Do authors of such texts have to tolerate their works being reproduced for AI training? This depends on whether their reproduction right (§ 16 German Copyright Act) is subject to a copyright exception – such as that for commercial “text and data mining” (§ 44b German Copyright Act) transposed from a European model (article 4 CDSM Directive). Although the new European AI Act may have ended a long dispute over whether the TDM exception can be applied to AI training, the practical hurdles of the exception have only begun to be appreciated. For rights- and incentive-based reasons, the TDM exception only applies insofar as rightholders have not effectively reserved this use for themselves.

²⁷⁶ *Senftleben*, IIC 54 (2023), 1535 (1546); dazu etwa *Durantaye*, AfP 2024, 9 (15 Rn. 33, in Fn. 58).

²⁷⁷ Wegweisend *Calabresi/Melamed*, Property Rules, Liability Rules, and Inalienability: One View of the Cathedral, Harvard Law Review 85 (1972), 1089–1128.

²⁷⁸ *Krier/Schwab*, Property Rules and Liability Rules: The Cathedral in Another Light, NYU Law Review 70 (1995), 440, 451: “When transaction costs are low, use property rules; when transaction costs are high, use liability rules.”

²⁷⁹ Zur rechtsökonomisch begründeten „Vorbildfunktion der japanischen TDM-Schranke für Europa“ ausf. *Brockmeyer* (Fn. 41), S. 129 f. m.Verw. auf *Ueno*, GRUR Int. 2021, 145 (149) u.a.; rechtsvergleichend auch *Spindler*, GRUR 2016, 1112 (1117 f.); *Fil-Flynn u.a.*, Science 378 (2022), 951.

²⁸⁰ Vgl. *Dreier/Schulze/Dreier*, 7. Aufl. 2022, § 44b UrhG Rn. 1: zum „informationellen Gehalt“, der „den Inhabern der einzelnen Werke [...] als solcher nicht zugewiesen ist“; auch *Guadamuz*, GRUR Int. 2024, 111 (117) zur “fundamental challenge in copyright law concerning AI: discerning the individual value of works in a vast dataset versus the collective value extracted from the aggregation of these works.”

Numerous were the scholars who feared that the TDM exception would “run dry”. Equally as numerous, however, were scholars suggesting on the contrary that the reservation of rights would “run dry” instead. Any stance in this controversy requires clarifying various doctrinal problems – such as interpreting the term of art “machine-readable means”, introducing the unwritten requirement of expressivity, exploiting systematic parallels with § 49 German Copyright Act, clarifying who may even declare reservations, and resolving various claims of inconformity with European law – which turn out to be founded in a completely different manner than previously assumed. Once the doctrinal contours of § 44b(3) German Copyright Act are clarified, we are left with the (all but trivial) question of how machine-readability could be technically achieved. According to article 53 AI Act, AI developers need to identify reservations “through state of the art technologies”, conversely raising the question which technologies are even suitable to communicate reservations effectively. The study explores seven such communication standards, of which but three could even meet the legal requirements; instead of standardization they tend towards proliferation. Therefore, it becomes almost impossible to reserve rights in a manner that courts will safely uphold.