

E. Urheberrechtliche Anforderungen an die Verwendung von Aktenauszügen zu Trainingszwecken

Die Verwendung von Aktenauszügen zum Training eines Sprachmodells wird im Folgenden auf der Grundlage und am Maßstab des UrhG zu prüfen sein.

I. Urheberrechtlicher Schutz und Zuordnung

Zunächst ist die Frage aufzuwerfen, ob und inwieweit Aktenauszüge urheberrechtlich geschützt sind. Als Schöpfer (§ 7 UrhG) kommt stets nur eine natürliche Person in Betracht, wobei Nutzungsrechte auch juristischen Personen (wie etwa dem Arbeitgeber) eingeräumt werden können.

Schwerpunktmäßig sind Schriftsätze in den Blick zu nehmen (unter 1.), wobei etwaige Besonderheiten für Aktenauszüge im Übrigen anschließend adressiert werden (unter 2.).

1. Schriftsätze

Mit Blick auf anwaltliche Schriftsätze sind neben dem Schutz als persönliche geistige Schöpfung weitere (Schutz-)Rechte und Einschränkungen aus dem UrhG zu erörtern.

a. Schutz als Werk nach § 2 UrhG

Als Werke³⁶⁰ geschützt sind nach der nicht-abschließenden³⁶¹ Aufzählung des § 2 Abs. 1 UrhG insbesondere „Sprachwerke, wie Schriftwerke“

360 Noch nicht vollständig unionsrechtlich harmonisiert nach *Schmid*, KIR 2025, 69 (76); weitgehende Harmonisierung durch die Infopaq-Entscheidung annehmend *Wandtke*, MMR 2017, 367 (369); s. auch *EuGH*, GRUR 2019, 73 (Rn. 35 ff.) – *Levola Hengelo BV/Smilde Foods BV*.

361 Statt aller *Raue*, in: *Dreier/Schulze*, § 2 UrhG Rn. 101 f.

(Nr. 1),³⁶² sofern es sich um eine persönliche geistige Schöpfung handelt (§ 2 Abs. 2 UrhG). Über das Merkmal der persönlichen geistigen Schöpfung schützt das Urheberrecht das Vorkommen von einzelnen Zeichen (Syntax) als eine konkrete Form, sofern diese Syntax nach ihrer Bedeutung (Semantik) den (kreativen) geistigen Gehalt einer natürlichen Person (i.e. des Urhebers) zum Ausdruck bringt (so insbesondere „Form und Art der Sammlung, Einteilung und Anordnung des dargebotenen Stoffes“³⁶³ sowie „schöpferische Gedankenformung und -führung des dargebotenen Inhalts“³⁶⁴).³⁶⁵ Ein solcher geistiger Gehalt ist einem (Sprach-)Werk unter Berücksichtigung eines Freihaltebedürfnisses³⁶⁶ nicht zu entnehmen, wenn und soweit die Form des Werks „aus der Natur der Sache oder nach den Gesetzen der Zweckmäßigkeit vorgegeben ist“.³⁶⁷ Zudem folgt auch aus einem „innovativen Charakter“ des Inhalts eines Schriftsatzes noch keine persönliche geistige Schöpfung.³⁶⁸

Grundsätzlich können anwaltliche Schriftsätze als Sprachwerk urheberrechtlichen Schutz genießen.³⁶⁹ Der BGH stellte zwar zunächst aufgrund des begrenzten Spielraums für schöpferische Kreativität an wissenschaftliche Sprachwerke und insbesondere auch Anwaltsschriftsätze erhöhte Schutzanforderungen³⁷⁰ über die sonst angenommene sog. kleine Münze³⁷¹ hinaus (so z.B. über die chronologische Schilderung eines Sachverhalts, auf der Hand liegende rechtliche Anforderungen und Anträge hinaus).

An den Maßgaben aus dieser Rechtsprechung dürfte unter Berücksichtigung der unionsrechtlichen Vorgaben³⁷² aber nicht mehr festgehalten wer-

362 Ein Datenbankwerk i.S.d. § 4 Abs. 2 UrhG wird mangels systematischer und methodischer Anordnung von Literaturquellen in vereinzelt Fußnoten in den Anwaltsschriftsätzen nicht in Betracht kommen.

363 OLG Düsseldorf, MDR 2023, 1401 (1401); BGH, MMR 2011, 182 (Rn. 36).

364 BGH, GRUR 1986, 739 (740).

365 Zech, Information als Schutzgegenstand, S. 37 f., 246 ff., 355 ff.; Hofmann, WRP 2024, 11 (14); Hofmann, ZUM 2024, 166 (167).

366 Bullinger, in: Wandtke/Bullinger, § 2 UrhG Rn. 25.

367 BVerwG, GRUR 2020, 189 (Rn. 19); BGH, GRUR 2002, 958 (960).

368 BVerwG, GRUR 2020, 189 (Rn. 19).

369 Schon BGH, GRUR 1986, 739 (740); Rauer/Bibi, in: Götting/Lauber-Rönsberg/Rauer, § 2 UrhG Rn. 195; OVG Hamburg, GRUR-RS 2021, 34061; OLG Düsseldorf, MDR 2023, 1401.

370 BGH, GRUR 1986, 739.

371 Statt aller BGH, GRUR 2014, 175 (Rn. 18).

372 S. etwa EuGH, GRUR 2019, 73 – Levola Hengelo BV/Smilde Foods BV.

den können.³⁷³ Es sind daher für Sprachwerke, wie es der BGH schon für Werke der angewandten Kunst entschieden hat,³⁷⁴ grundsätzlich einheitliche Schutzanforderungen zu stellen, wobei gegebenenfalls in Abhängigkeit von dem jeweiligen Werktyp und fehlendem Raum für Kreativität ein geringerer Schutzzumfang anzunehmen sein kann.³⁷⁵

Unter Anwendung dieses Maßstabs verneinte jüngst das OLG Düsseldorf den Schutz für eine Klageschrift, die lediglich „eine chronologische Darstellung des prozessrelevanten Sachverhalts mit einer Bezugnahme auf beigefügte Anlagen [enthält] und [...] im Folgenden kurz die angekündigten Klageanträge auf Zahlung eines angemessenen Schmerzensgeldes, auf Begleichung außergerichtlicher Rechtsanwaltskosten gegenüber der Rechtsschutzversicherung sowie auf Feststellung einer weitergehenden Schadensersatzpflicht [erläutert]“.³⁷⁶ Für die weiteren verfahrensgegenständlichen Schriftsätze wurde ein Schutz ebenfalls abgelehnt, da diese Schriftsätze „soweit sie inhaltlich zum Prozessstoff Stellung nehmen, auf das Vorbringen der Gegenseite“ erwidern.³⁷⁷ Hervorzuheben ist, dass im konkreten Fall der verfassende Rechtsanwalt allerdings nicht substantiiert zu den Schriftsätzen als Ausdruck einer schöpferischen Tätigkeit vorgetragen hatte.

Das OVG Hamburg hingegen nahm im Zuge einer recht großzügigen Auslegung³⁷⁸ des § 2 Abs. 2 UrhG einen Schutz für einen Schriftsatz an und stützte sich hierzu unter anderem auf den Umfang (acht Seiten), eine zwar nicht besonders kreative, aber zumindest auch nicht zwingende Gliederung (z.B. „b. Werbung/und weiter hilfsweise auch für die Packung i) ‚Organic‘ ii) ‚aus ökologischem Anbau‘“) und die nicht vorgegebene Wortwahl.³⁷⁹

Aus alledem ergibt sich, dass anwaltliche Schriftsätze zwar urheberrechtlich geschützt sein können; ein solcher Schutz ist aber jedenfalls nicht zwingend und gegebenenfalls in seinem Umfang begrenzt. Die Frage des Vorliegens eines solchen Schutzes und von dessen Reichweite lässt sich

373 Vgl. etwa *BVerwG*, GRUR 2020, 189 (Rn. 22); s. auch *OVG Hamburg*, GRUR-RS 2021, 34061; *OLG Düsseldorf*, MDR 2023, 1401 (1401).

374 BGH, GRUR 2014, 175; GRUR-RS 2025, 2426 (Rn. 26): für alle Werkarten wird „eine nicht zu geringe Schöpfungshöhe“ vorausgesetzt.

375 Ausf. *Rauer/Bibi*, in: Götting/Lauber-Rönsberg/Rauer, § 2 UrhG Rn. 80 m.w.N.; *Loewenheim/Leistner*, in: Schricker/Loewenheim, § 2 UrhG Rn. 60; *Nordemann-Schiffel/Nordemann*, in: Peifer/Kubis/Stieper/Raue, Alles Käse oder vielleicht doch mehr?, S. 241; s. zur Schutzzfähigkeit einer Datenschutz-Erklärung *OLG München*, MMR 2023, 974 (Rn. 19).

376 *OLG Düsseldorf*, MDR 2023, 1401 (1402).

377 *OLG Düsseldorf*, MDR 2023, 1401 (1402).

378 *Raue*, in: Dreier/Schulze, § 2 UrhG Rn. 146: „sehr großzügig“.

379 *OVG Hamburg*, GRUR-RS 2021, 34061 (Rn. 55 ff.).

nicht abstrakt anhand äußerer Merkmale (z.B. über den Umfang eines Schriftsatzes) beantworten, sondern bedarf vielmehr jeweils einer Einzelfallbetrachtung. Für die automatisierte Verwertung von Schriftsätzen sind daher urheberrechtliche Anforderungen im Ausgangspunkt grundsätzlich zu beachten.

b. Kein Ausschluss nach § 5 Abs. 1 UrhG

Nach § 5 Abs. 1 UrhG genießen „Gesetze, Verordnungen, amtliche Erlasse und Bekanntmachungen sowie Entscheidungen und amtlich verfaßte Leitsätze zu Entscheidungen“ keinen urheberrechtlichen Schutz. Die Regelung trägt dem Allgemeininteresse an dem Zugang zu amtlichen Dokumenten Rechnung.³⁸⁰ Hierdurch werden sämtliche gerichtliche Entscheidungen unabhängig von ihrer Art in Bezug genommen, das umfasst insbesondere Urteile i.S.d. §§ 300 ff. ZPO, Beschlüsse und Verfügungen im Gesamten einschließlich der Entscheidungsgründe³⁸¹ (z.B. auch Hinweisbeschlüsse und wohl auch Terminladungen). Soweit Anwaltsschriftsätze zum Bestandteil der Gerichtsentscheidung werden (z.B. durch eine wörtliche Übernahme von Textpassagen oder die ausdrückliche Inbezugnahme des Anwaltsschriftsatzes), unterfallen auch sie § 5 Abs. 1 UrhG.³⁸²

Im Übrigen stellt der Akteninhalt einschließlich der anwaltlichen Schriftsätze allerdings keine gerichtliche Entscheidung dar und ist daher nicht nach § 5 Abs. 1 UrhG vom Urheberrechtsschutz ausgenommen.³⁸³ Zur Begründung verweist die Rechtsprechung insoweit auf die Ausnahme für öffentliche Reden in § 48 Abs. 1 Nr. 2 UrhG, die grundsätzlich den Schutz von vor Gericht vorgetragenen Plädoyers und damit auch erst recht von anwaltlichen Schriftsätzen voraussetzt.³⁸⁴

Im Regelfall sind anwaltliche Schriftsätze daher schutzfähig und an den Vorgaben des UrhG zu messen.

380 Dreier, in: Dreier/Schulze, § 5 UrhG Rn. 3.

381 Arnold, Amtliche Werke im Urheberrecht, S. 94; Marquardt, in: Wandtke/Bullinger, § 5 UrhG Rn. 12; zust. Ahlberg/Lauber-Rönsberg, in: Götting/Lauber-Rönsberg/Rauer, § 5 UrhG Rn. 16; Katzenberger/Metzger, in: Schricker/Loewenheim, § 5 UrhG Rn. 46, wenngleich mit weitergehender Diff. in Rn. 37.

382 LG Köln, GRUR-RR 2011, 5 (6).

383 BGH, GRUR 1986, 739 (740); s. auch die Gesetzeshistorie Katzenberger, GRUR 1972, 686 (690 f.).

384 BGH, GRUR 1986, 739 (740).

c. Verwandte Schutzrechte

Ein etwaiges Leistungsschutzrecht aufgrund der *Zusammenstellung* der Akte als Datenbank i.S.d. §§ 87a ff. UrhG wäre der Justiz als Datenbankherstellerin³⁸⁵ und nicht dem jeweiligen Rechtsanwalt zuzuordnen. Aus den §§ 87a ff. UrhG ergeben sich also keine im untersuchungsgegenständlichen Fall erhöhten Anforderungen an die Verwertung von anwaltlichen Schriftsätzen.

d. Zwischenergebnis

Anwaltliche Schriftsätze können urheberrechtlichen Werkschutz genießen, wenngleich der Schutzzumfang regelmäßig auf einzelne Abschnitte beschränkt sein und verschiedene Schriftsätze ausschließen dürfte.

2. Aktenauszüge im Übrigen

Von den oben unter D.VI.1 dargestellten Akteninhalten sind zahlreiche Inhaltstypen bereits den gerichtlichen Entscheidungen i.S.d. § 5 Abs. 1 UrhG zuzuordnen³⁸⁶ oder verwertungsrechtlich dem jeweiligen (Justiz-)Arbeitgeber zugeordnet (vgl. § 43 UrhG).³⁸⁷ Insbesondere Anlagen zu den Schriftsätzen sowie eingeholte Gutachten stechen insoweit allerdings heraus und bedürfen daher einer weitergehenden Erörterung.

385 S. § 87a Abs. 2 UrhG und ErwGr. 41 Datenbank-RL. Hersteller ist in der Regel der Arbeitgeber, s. *Paul*, in: Hilber/Borges, § 87a UrhG Rn. 23; im Überblick zum Begriff des Datenbankherstellers *Vogel*, in: Schricker/Loewenheim, § 87a UrhG Rn. 71 f. m.w.N.

386 S. zuvor unter E.I.1.b.

387 Der Arbeitgeber bzw. Dienstherr erwirbt hiernach i.V.m. dem jeweiligen Arbeitsverhältnis und unter Berücksichtigung der Zweckübertragungslehre Nutzungsrechte an den in Ausübung der Tätigkeit hergestellten Werken. s. hierzu etwa *Wandtke*, in: Wandtke/Bullinger, § 43 UrhG Rn. 49; das betrifft auch Richter, *Nordemann/Obergfell*, in: Loewenheim, § 69 Sonderfragen bei Arbeits- und Dienstverhältnissen, Rn. 10; *Klass*, GRUR 2019, 1103 (1106), wobei die Vorschrift für Hochschullehrer wegen deren eigenverantwortlicher Tätigkeit für nicht anwendbar gehalten wird. Angesichts der auf die gerichtliche Entscheidung i.S.d. § 5 UrhG gerichteten Tätigkeit und insoweit vorbereitenden Tätigkeiten dürften aber auch Richter dem § 43 UrhG unterfallen; s. jüngst zur Unzulässigkeit eines über § 43 UrhG hinausgehenden gesetzlichen Übergangs der Rechte *EuGH*, GRUR-RS 2025, 3103 – *Orchestre national de Belgique*.

Die Anlagen zu den Schriftsätzen können grundsätzlich urheberrechtlich geschützt sein als Sprachwerke bzw. als Lichtbildwerke (§ 2 Abs. 1 Nr. 1, 5 UrhG) oder leistungsschutzrechtlich geschützt sein mit (noch) geringeren Anforderungen als Lichtbilder (§ 72 UrhG).³⁸⁸ Eine pauschale Bewertung ist wegen der Vielgestaltigkeit der Form dieser Inhalte nicht möglich. Bei einer automatisierten Verarbeitung dieser Inhalte wird daher zugrunde zu legen sein, dass jedenfalls Teile der betreffenden Anlagen urheber- bzw. leistungsschutzrechtlich geschützt sein dürften.

Für Gutachten als Form einer wissenschaftlichen Ausarbeitung gelten die oben dargestellten Grundsätze zum urheberrechtlichen Schutz von Anwaltsschriftsätzen entsprechend. Ein Gutachten basierend auf (zwingenden) mathematischen Berechnungen zu der Geschwindigkeit(sspanne) eines verunfallten Kraftfahrzeugs ohne nennenswerten Textanteil dürfte regelmäßig keinen urheberrechtlichen Schutz genießen.³⁸⁹ Demgegenüber liegt für ein umfangreiches psychologisches oder sonst textbasiertes Gutachten die Annahme eines Urheberrechtsschutzes grundsätzlich näher. Vor diesem Hintergrund wird also regelmäßig von einem urheberrechtlichen Schutz auszugehen sein. Im Übrigen können potenziell nicht schutzfähige Gutachten mit höherem Anteil an mathematischen Berechnungen und geringerem Textanteil (z.B. Gutachten zur Wertberechnung) gegebenenfalls anhand der Fachdisziplin identifiziert und geclustert werden.

II. Nutzung innerhalb der Justiz

Zunächst ist eine Verwendung der Aktenauszüge für das Training eines Sprachmodells zu untersuchen, das ausschließlich justizintern eingesetzt wird. Mit Blick auf die geschützten Werke sind zunächst die betroffenen Verwertungsrechte des jeweiligen (Urheber-)Rechts am Werk zu thematisieren (unter 1.), bevor in einem zweiten Schritt untersucht wird, ob sich diese Verwertungs- bzw. Nutzungshandlungen belastbar auf Nutzungsrechtseinräumungen oder urheberrechtliche Schrankenbestimmungen stützen lassen (unter 2.), wobei die Einbindung weiterer Stellen zu berücksichtigen ist (unter 3.). Hieraus werden sodann Bedingungen für das Training und den Einsatz des Sprachmodells abgeleitet (unter 4.).

388 S. etwa *LG Hamburg*, ZUM-RD 2010, 80 (82).

389 Schutz für ein Wertermittlungsgutachten bejahend *LG Hamburg*, ZUM-RD 2010, 80 (82).

1. Betroffene Verwertungsrechte

Für die untersuchungsgegenständlichen Szenarien und Use-Cases ist mit Blick auf urheberrechtliche Verwertungsrechte³⁹⁰ zwischen der Zusammenstellung der Trainingsdaten, dem eigentlichen Training i.w.S. und der Generierung der Ausgabe zu differenzieren. Das Training selbst als „Umwandlung“ von Sprachwerken in eine mathematische, wahrscheinlichkeitsbasierte Repräsentation ist urheberrechtlich nicht relevant,³⁹¹ ebenso wenig der reine Werkgenuss;³⁹² urheberrechtlich von Bedeutung sind vielmehr etwaige Vervielfältigungen i.S.d. § 16 UrhG vor, während und nach dem (KI-)Trainingsprozess.

a. Zusammenstellung des Trainingskorpus und Vervielfältigungen im Trainingsprozess

Die urheberrechtliche Vervielfältigung betrifft sämtliche „körperlichen Festlegungen, die geeignet sind,“ wiedererkennbare, prägende Bestandteile des Werks den „menschlichen Sinnen“ wahrnehmbar zu machen.³⁹³ Entscheidend sind somit die Wahrnehmbarkeit und – vor allem auch im Hinblick auf veränderte Werkbestandteile – die Wiedererkennbarkeit³⁹⁴ des Werks oder von prägenden Werkteilen, die ihrerseits die Voraussetzungen des § 2 UrhG erfüllen.³⁹⁵ Nach dem deutschsprachlichen Begriffsverständnis soll auch die Eignung zur mittelbaren Wahrnehmung genügen,³⁹⁶ worauf nachfolgend zurückzukommen sein wird (unter bb.). Demgegenüber

390 Zu dem Handlungsbezug der Verwertungsrechte s. *Sesing-Wagenpfeil*, ZGE 16 (2024), 212 (221) m.w.N.

391 *Schack*, NJW 2024, 113 (Rn. 8); *Bomhard*, InTeR 2023, 174 (175); vgl. BT-Drs. 18/12329, 40.

392 Statt aller *Sucker*, Der digitale Werkgenuss im Urheberrecht, S. 50 ff., 59, 87 ff.; *Beurskens*, RD 2025, 1 (Rn. 7); *Schack*, NJW 2024, 113 (Rn. 9).

393 BT-Drs. IV/270, 47; *BGH*, GRUR 2017, 793 (Rn. 41); GRUR 2009, 942 (Rn. 25).

394 *EuGH*, GRUR 2019, 929 (Rn. 36 ff.) – Pelham ua [Metall auf Metall III]; *Grisse*, RuZ 2020, 143 (147).

395 *Raue*, in: Dreier/Schulze, § 16 UrhG Rn. 14; zur str. Abgrenzung zu § 23 UrhG *Raue*, in: Dreier/Schulze, § 23 UrhG Rn. 16.

396 BT-Drs. IV/270, 47; *BGH*, GRUR 2017, 793 (Rn. 41); GRUR 2009, 942 (Rn. 25).

sind der Zweck, das Format³⁹⁷ oder die beabsichtigte Speicherdauer³⁹⁸ ohne Belang für die Frage nach dem Vorliegen einer Vervielfältigung.

aa. Vervielfältigungen im Einzelnen

Falls Aktenauszüge in Vorbereitung des Trainingsprozesses zunächst digitalisiert werden, ist hierin bereits eine urheberrechtliche Vervielfältigung zu sehen.³⁹⁹ Das Zusammenführen der digitalen Aktenauszüge an einem Ort⁴⁰⁰ in Vorbereitung des Trainings des Sprachmodells geht ebenfalls mit Vervielfältigungen einher.⁴⁰¹

Im GSJ-Projekt lassen sich Verarbeitungen gegebenenfalls durch die Weiterverarbeitung an der Dokumentquelle und abgeleitete Textformate⁴⁰² vermeiden.⁴⁰³ Die Weiterverarbeitung an der Dokumentquelle meint, dass möglichst viele Verarbeitungs- und Trainingsschritte lokal ohne erneute Zwischenspeicherung in einer Datei vorgenommen werden. Nach Abschluss des Verarbeitungsprozesses werden die benötigten Daten (i.e. nicht das ursprüngliche Dokument) sodann als abgeleitetes Textformat an die KI-Trainingsserver übermittelt. Dieses abgeleitete Textformat könnte mangels Wahrnehmbarkeit und Wiedererkennbarkeit nicht mehr als Abbildung eines urheberrechtlich geschützten Werkes anzusehen sein (hierzu sogleich unter bb.),⁴⁰⁴ sodass die Daten grundsätzlich ohne urheberrechtli-

397 Etwa BGH, GRUR 2010, 628 (Rn. 17) m.w.N.

398 Das zeigt insb. die in § 44a UrhG niedergelegte Schrankenbestimmung, die temporäre Vervielfältigungen adressiert und mithin die Anwendbarkeit des UrhG auf diese temporären Vervielfältigungen unterstellt, s. Sucker, Der digitale Werkgenuss im Urheberrecht, S. 75 ff.; Käde, Kreative Maschinen und Urheberrecht, S. 69.

399 Vgl. etwa OLG Hamburg, GRUR-RR 2008, 378 (380 f.); BGH, GRUR 1999, 325 (327).

400 Z.B. das Anlegen eines digitalen Trainingskorpus oder das Hochladen einer Datei auf einen KI-Trainingsserver. S. aber zur Rechtfertigung nach § 44a UrhG nachfolgend unter E.II.2.d.

401 Vgl. auch La Durantaye, ZUM 2023, 645 (647).

402 Hierzu etwa Raue/Schöf, RuZ 2020, 118; Grisse, RuZ 2020, 143 m.w.N.; Organisiak/Downie, in: Golub/Y.-H. Liu, Research access to in-copyright texts in the humanities; Schöch et al., Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen; Schöch et al., RuZ 2020, 160; Brockmeyer, Text und Data Mining, S. 200 ff.

403 Im Überblick zu der Frage, ob sich Vervielfältigungen vermeiden lassen, s. Jager, Artificial Creativity?, S. 329 m.w.N.

404 Raue/Schöf, RuZ 2020, 118 (119): „durch eine gezielte Informationsreduktion so zu transformieren, dass die urheberrechtlich geschützte Textstruktur irreversibel verloren geht“; s. zur gegebenenfalls möglichen Rekonstruierbarkeit von abgeleiteten

che Implikationen verwertet werden könnten.⁴⁰⁵ Da Informationsverluste denkbar sind,⁴⁰⁶ bedarf die Frage der Realisierbarkeit im vorliegend zu untersuchenden Sachverhalt einer technischen Prüfung, die allerdings nicht Gegenstand dieses Rechtsgutachtens ist.

bb. Vervielfältigung bei möglicher Rekonstruierbarkeit aus abgeleiteten Textformaten oder anhand eines Sprachmodells

Die Auswirkungen einer Rekonstruierbarkeit auf die Annahme einer Vervielfältigung (insbesondere über eine Eignung zur mittelbaren Wahrnehmbarkeit) betreffen nicht nur abgeleitete Textformate, sondern auch und gerade Sprachmodelle. Analog zur datenschutzrechtlichen Einordnung⁴⁰⁷ könnten auch Sprachmodelle in Abhängigkeit von den möglichen Ausgaben als Vervielfältigung anzusehen sein. Sprachmodelle ähneln den abgeleiteten Textformaten, indem ihre Gewichte eine – allerdings wahrscheinlichkeitsbasierte – mathematische Repräsentation von Wort-Vektoren darstellen.

Es ist nicht abschließend geklärt, ob und unter welchen Voraussetzungen eine mögliche Rekonstruierbarkeit des Werks anhand der abgeleiteten Textformate⁴⁰⁸ oder eines Sprachmodells⁴⁰⁹ über eine mittelbare Wahrnehm-

Textformaten *Kugler et al.*, in: *InvBERT: Reconstructing Text from Contextualized Word Embeddings by inverting the BERT pipeline*.

405 Das steht unter dem Vorbehalt, dass diese Daten nicht wieder zusammengeführt werden (können) und das ursprüngliche Werke erkennbar abbilden.

406 Zu möglichen Formaten *Grise*, RuZ 2020, 143 (152 ff.); Term-Dokument-Matrizen, Lemmatisierung, gestörte Sequenzinformationen, N-Gramm-Tabellen und Wort-Vektoren; zu den Wort-Vektoren auf Basis von Word Embedding im Natural Language Processing *Schöch et al.*, RuZ 2020, 160 (171 ff.).

407 S. oben unter D.I.1.b.bb.

408 Offen mit Tendenz zur Berücksichtigung der Rekonstruierbarkeit *Grise*, RuZ 2020, 143 (150 f.); *Raue/Schöf*, RuZ 2020, 118 (124): nicht „mit verhältnismäßigem Aufwand rekonstruierbar“.

409 Ein KI-Modell stellt kein Vervielfältigungsstück dar, nach *La Durantaye*, AfP 2024, 9 (Rn. 19); *Leistner*, GRUR 2024, 1665 (1668 f.): wegen wahrscheinlichkeitsbasierendem Aufbau nicht vergleichbar mit einer Kopie; ähnl. *Konertz/Schönhof*, WRP 2024, 534 (Rn. 8); *Antoine*, GRUR 2025, 118 (120); implizit *Maamar*, ZUM 2023, 481 (486), wonach ein einmal im Ausland trainiertes Sprachmodell ohne weitere urheberrechtliche Implikationen in der EU angeboten werden kann; ähnl. *Peukert*, GRUR Int. 2024, 497 (506); Tendenz gegen eine Vervielfältigung bei *Baumann*, NJW 2023, 3673 (Rn. 10 ff.); „im Regelfall“ abzulehnen nach *Schmid*, KIR 2025, 69 (72); abhängig von der Konzeption und intendierten Einsatzes des Modells *Käde*,

barkeit zur Annahme einer Vervielfältigung führt. Diese Unsicherheit ist auch auf die fehlende (EuGH-)Rechtsprechung zu der konkreten Fragestellung zurückzuführen; es ist unklar, ob und inwieweit die Eignung zur mittelbaren menschlichen Wahrnehmung für die Annahme einer Vervielfältigung genügt. Im Ausgangspunkt adressiert Art. 2 InfoSoc-RL jede „unmittelbare oder mittelbare, vorübergehende oder dauerhafte Vervielfältigung auf jede Art und Weise und in jeder Form“ und wurde durch den EuGH bereits weit ausgelegt mit Blick auf Werkteile.⁴¹⁰

Aus der bisher insoweit ergangenen Rechtsprechung des EuGH lässt sich allerdings jedenfalls eine Tendenz ableiten, auch subjektive Aspekte in die Prüfung einer Vervielfältigung einzustellen⁴¹¹ (z.B. das fehlende Interesse einer Person, ein Werk tatsächlich zu rekonstruieren) und die objektiv ermittelte Rekonstruierbarkeit nicht ausreichen zu lassen. Insoweit wird zutreffend insbesondere auch darauf hingewiesen,⁴¹² dass der EuGH – allerdings auf Ebene des Schutzzumfangs statt mit Blick auf die Verwertung – über den Werkbegriff eine „Ausdrucksform des urheberrechtlichen Schutzobjekts [erfordert], die es mit hinreichender Genauigkeit und Objektivität identifizierbar werden lässt“.⁴¹³ Eine solche Ausdrucksform ist bei stets wahrscheinkeitsbasierten Sprachmodellen nicht gegeben; dieser Befund gilt auch bei den hier untersuchungsgegenständlichen Use-Cases und dem Einsatz von Ausgabefiltern.

Ergänzend legt ein teleologischer Vergleich des jeweils europarechtlich determinierten sowie präformierten Urheberrechts und Datenschutzrechts

ZUM 2024, 174 (182); an einer Vervielfältigung zweifelnd *Hofmann*, WRP 2024, 11 (Rn. 12); a.A. *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 54 f.; *Jäger*, Artificial Creativity?, S. 333; *Sesing-Wagenpfeil*, ZGE 16 (2024), 212 (236 ff.): soweit sich Ausgaben nachweisbar auf Trainingsdaten zurückführen lassen; *Beurskens*, RD 2025, 1 (Rn. 9); *Pesch/Böhme*, GRUR 2023, 997 (1004 f.), nach denen auch die „Schwierigkeit der Rekonstruktion“ nicht schaden soll; womöglich auch *Welser*, GRUR-Prax 2023, 516 (Rn. 8), der darauf abstellt, dass Modelle Trainingsdaten „speichern“ können.

410 *EuGH*, GRUR 2009, 1041 (Rn. 38 ff.) – Infopaq; *Spindler*, in: *Peifer/Kubis/Stieper/Raue*, Künstliche Intelligenz und Urheberrecht aus europäischer Perspektive, S. 349.

411 Zum subjektiven Vervielfältigungsrecht *Raue*, ZGE 9 (2017), 514; *Grisse*, RuZ 2020, 143 (151) mit einem Überblick über die EuGH-Rechtsprechung in Fn. 51; objektiv hingegen nach *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 51.

412 *Leistner*, GRUR 2024, 1665 (1669); *EuGH*, GRUR 2019, 73 (Rn. 40) – *Levola Hengelo BV/Smilde Foods BV*.

413 *EuGH*, GRUR 2019, 73 (Rn. 40) – *Levola Hengelo BV/Smilde Foods BV*.

die zurückhaltende Annahme einer Vervielfältigung durch das Sprachmodell (oder abgeleitete Textformate) nahe.

Für das Datenschutzrecht ist die Verknüpfung von Datum und natürlicher Person grundlegend (i.e. Art. 4 Nr. 1 DSGVO, Art. 8 GRCh, Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG). Gerade diese Verknüpfung ermöglicht Rückschlüsse auf die natürliche Person und Auswirkungen auf deren Handeln sowie ihr Privat- und Familienleben (Art. 7 GRCh, Art. 8 EMRK), die ein entscheidender Grund für die grundrechtliche Gewährleistung in Art. 8 GRCh sind (siehe auch Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG).⁴¹⁴ Der EuGH legt die DSGVO vor diesem Hintergrund weit aus.⁴¹⁵ Bereits eine nur mittelbare Verknüpfung im Wege der Identifizierbarkeit mit verfügbaren Mitteln kann hiernach die betroffene Person tangieren.

Dem Urheberrecht ist eine derart enge Verknüpfung zwischen Werk und Urheber nicht fremd. So stellt sich das Urheberpersönlichkeitsrecht nach §§ 12 ff. UrhG dar als Ausdruck der ideellen Beziehung des Urhebers zu seinem Werk aus dem allgemeinen Persönlichkeitsrecht.⁴¹⁶ Das Werk ist eine individuelle schöpferische Leistung des Urhebers und damit Ausdruck seiner Persönlichkeit. Zugleich und von vorrangiger Bedeutung für das untersuchungsgegenständliche Projekt sieht das Urheberrecht Verwertungsrechte in §§ 15 ff. UrhG vor, die nicht primär persönlichkeitsrechtlich unterlegt sind, sondern vielmehr Vermögensinteressen ausgehend von dem geschaffenen Werk schützen (Art. 14 GG). Ein Werkformat, aus dem sich das ursprüngliche Werk rekonstruieren lässt, berührt vor diesem Hintergrund noch nicht die Verwertungsinteressen des Urhebers. Die bloße Wahrscheinlichkeit einer möglichen Beeinträchtigung der Verwertungsinteressen genügt insoweit gerade nicht; im Fall einer *tatsächlichen* Rekonstruktion liegt eine Vervielfältigung nach § 16 UrhG vor und ermöglicht dem Urheber die Kontrolle der Verwertung. Die (mittelbare) Wahrnehmbarkeit als Voraussetzung einer Vervielfältigung ist hierbei vor dem Hintergrund zu sehen, dass die menschliche Wahrnehmung eines Werks (der „Werkgenuss“) nicht

414 Beispielhaft *EuGH*, MMR 2023, 105 (Rn. 57 ff.) – Recht auf Vergessenwerden; *Ehmann/Selmayr*, in: *Ehmann/Selmayr*, Einl. Rn. 30 f.; *Buchner*, in: *Kühling/Buchner*, Art. 1 DSGVO Rn. 10: „Recht auf Privatleben ist Grundvoraussetzung für die Entwicklung einer freien und selbstbestimmten Persönlichkeit“.

415 Z.B. mit Blick auf personenbezogene Daten, vgl. *EuGH*, NJW 2018, 767 (Rn. 33) – Nowak m.w.N.; mit Blick auf den Begriff des Verantwortlichen ZD 2024, 328 (Rn. 54 f.) – IAB Europe.

416 *Specht-Riemenschneider*, in: *Dreier/Schulze*, Vor § 12 UrhG Rn. 1, 5 ff.; *Peukert*, in: *Schricker/Loewenheim*, Vor § 12 UrhG Rn. 29 ff.

geschützt ist⁴¹⁷ und daher bereits der unmittelbar einer Wahrnehmung vorgelagerte Akt erfasst sein soll. Die Konzentration des Urheberrechts auf Verwertungsformen (§ 15 UrhG) sperrt eine Erstreckung des Urheberrechts auf das Vorfeld einer Verwertung, ebenso wie auf den Nachgang in Form des Werkgenusses.

Das Datenschutzrecht beschränkt sich nicht auf einzelne Verwertungsformen, sondern umfasst nach Art. 4 Nr. 2 DSGVO sämtliche (teil-)automatisierten Vorgänge im Zusammenhang mit personenbezogenen Daten. Anders als das Urheberrecht verlangt die DSGVO umfangreiche technische und organisatorische Maßnahmen (Art. 24 ff. DSGVO) zur sicheren Aufbewahrung und Verarbeitung personenbezogener Daten. Diese datenschutzrechtlichen Anforderungen drohten konterkariert zu werden, wenn Daten mit Identifizierungsmöglichkeit der betroffenen Person ohne jede Sicherheitsmaßnahme gespeichert werden könnten und die DSGVO-Vorgaben erst ab der tatsächlichen Identifizierung eingreifen würden.

Hinzu kommt, dass der EuGH in seiner einschlägigen Rechtsprechung im Urheberrecht verschiedene Wertungen in die Auslegung der Verwertungsrechte einfließen lässt.⁴¹⁸ Die Berücksichtigung der vorgenannten Wertungen bei der Auslegung der urheberrechtlichen Vervielfältigung steht also nicht im Widerspruch zur bisherigen EuGH-Judikatur. Es liegt daher nahe, dass sowohl abgeleitete Textformate als auch trainierte Sprachmodelle grundsätzlich keine Vervielfältigung von urheberrechtlichen Werken darstellen werden. Unberührt hiervon bleibt eine tatsächlich erfolgende Vervielfältigung aufgrund des Sprachmodells in der Ausgabe des KI-Systems.⁴¹⁹

Da diese Frage durch den EuGH zum jetzigen Zeitpunkt (noch) ungeklärt ist, besteht ein (Rest-)Risiko, dass die Rechtsprechung geringere Anforderungen an die Annahme einer Vervielfältigung stellt. Um diese Unsicherheit zu adressieren, sollten die bereits mit Blick auf die DSGVO unter DVIII.4 dargestellten Mechanismen auch zur Verhinderung einer Rekonstruierbarkeit urheberrechtlich geschützter Werke zum Einsatz kommen. Diese Empfehlung betrifft insbesondere die Implementierung eines geeigneten System Prompts und der Ausgabefilter.

417 *Beurskens*, RD i 2025, 1 (Rn. 7); zum Hintergrund auch *Schack*, NJW 2024, 113 (Rn. 9).

418 *Ausf. Grisse*, RuZ 2020, 143 (151) m.w.N. in Fn. 51.

419 S. nachfolgend unter E.II.1.b.

b. Einsatz und Ausgabe des KI-Systems

Bei dem Einsatz des KI-Systems sind Vervielfältigungen unter Anwendung der vorstehenden Maßgaben systemintern regelmäßig abzulehnen.⁴²⁰ Denn die mathematischen Operationen basieren nicht auf wahrnehmbaren oder rekonstruierbaren Werkbestandteilen.

Allerdings kann die Ausgabe eines KI-Systems durchaus Bestandteile aus Trainingsdaten enthalten,⁴²¹ die durch die Ausgabe vervielfältigt werden (§ 16 Abs. 1 UrhG). Sofern die Ausgabe im Einzelfall über eine Internetseite öffentlich zugänglich gemacht wird, greift die speziellere Vorschrift des § 19a UrhG ein.⁴²² Eine solche Öffentlichkeit⁴²³ dürfte im untersuchungsgegenständlichen Szenario der Sachverhaltsaufbereitung in den Use-Cases 1-3 jedoch kaum in Betracht kommen, sodass es regelmäßig bei einer möglichen Vervielfältigung bleiben wird.

Die Wahrscheinlichkeit einer solchen Vervielfältigung kann durch verschiedene Mechanismen reduziert werden, wie etwa durch den System Prompt und durch die Implementierung von Ausgabefiltern. Ferner können entsprechende (Dienst-)Anweisungen zum Umgang mit dem KI-System Einfluss auf die Eingabeaufforderungen nehmen und hierüber die Wahrscheinlichkeit einer Vervielfältigung erheblich reduzieren. Sofern es im Einzelfall dennoch zu Vervielfältigungen kommt (in der Regel im Zusammenhang mit missbräuchlichen Eingabeaufforderungen), bestehen Zweifel an der Passivlegitimation der Projektverantwortlichen.⁴²⁴ Im Übrigen kann für vereinzelte Vervielfältigungen auch die Schranke für unwesentliches Beiwerk nach § 57 UrhG in den Blick genommen werden.⁴²⁵

420 Anders etwa *Schack*, NJW 2024, 113 (Rn. 4).

421 Etwa *Karamolegkou et al.*, Copyright Violations and Large Language Models. S. auch schon oben unter D.IV.1.b.cc.

422 *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 55; vgl. auch *Raue*, CR 2017, 656 (660).

423 *Dreier*, in: *Dreier/Schulze*, § 60d UrhG Rn. 10.

424 Hierzu auch unter E.III.

425 Dies diskutierend im Hinblick auf das Textkorpus *Raue/Schöf*, RuZ 2020, 118 (123 f.); krit. *Brockmeyer*, Text und Data Mining, S. 203 f.

c. Bearbeitungen des Werkes im Rahmen des Trainingsprozesses und des Einsatzes des KI-Systems

Wenn und soweit das Werk im Laufe des Trainingsprozesses und bei dem Einsatz des KI-Systems verändert wird, ist darüber hinaus § 23 UrhG⁴²⁶ mit gesonderten Anforderungen an Bearbeitungen und Umgestaltungen eines Werks zu beachten.⁴²⁷ Technisch bedingte Änderungen bei der automatisierten Analyse von Werken (dem sog. Text und Data Mining)⁴²⁸ sind allerdings nach § 23 Abs. 3 UrhG ausgenommen. Das betrifft nach vorzugswürdiger Auffassung beispielsweise die modellinterne Verarbeitung der Werke und zufällige Bearbeitungen von rekonstruierbaren Werkbestandteilen, repräsentiert als Zahlen im Modell.

d. Zwischenergebnis

Aus urheberrechtlicher Sicht sind insbesondere Vervielfältigungen der in den Aktenauszügen enthaltenen Werke relevant. Sofern diese Vervielfältigungen bereits in digitaler Form vorliegen, lässt sich eine Anwendbarkeit des Urheberrechts über abgeleitete Textformate verhindern. Eine solche Verhinderung steht ihrerseits unter dem Vorbehalt der technischen Machbarkeit. Nach vorzugswürdiger, wenngleich teilweise bestrittener Ansicht liegt auch dann regelmäßig keine Vervielfältigung vor, wenn sich anhand des gespeicherten Sprachmodells allfällige Werkteile mit nicht nur unerheblichem Aufwand rekonstruieren lassen. Insoweit ist mit Blick auf die Vervielfältigungen ein zugunsten der urheberrechtlichen Zulässigkeit großzügigerer Maßstab als unter der DSGVO mit Blick auf die Identifizierung natürlicher Personen anzusetzen.

In jedem Fall empfiehlt es sich, durch (Dienst-)Anweisungen zum Umgang mit dem KI-System, durch Ausgabefilter und den System Prompt Vervielfältigungen im Rahmen der Ausgabe des KI-Systems möglichst weitgehend zu verhindern. Sofern es dennoch zu entsprechenden Vervielfältigungen kommt, sind etwaige Nutzungsrechtseinräumungen und die urheberrechtlichen Schrankenbestimmungen zu prüfen.

426 Unabhängig davon, ob dem Bearbeitungsrecht noch eine eigenständige Bedeutung zukommt, s. *Schack*, ZGE 15 (2023), 263 (271 f.); vgl. auch *BGH*, GRUR 2010, 628 (Rn. 17); gegen eine eigenständige Bedeutung etwa *Raue*, in: *Dreier/Schulze*, § 23 UrhG Rn. 16; a.A. *Götting*, in: *Götting/Lauber-Rönsberg/Rauer*, § 16 UrhG Rn. 10.

427 *Schack*, NJW 2024, 113 (Rn. 4); *Jäger*, Artificial Creativity?, S. 333 ff.

428 S. nachfolgend unter E.II.2.c.

Die Veränderung der Werke durch das Sprachmodell ist wegen § 23 Abs. 3 UrhG und der vorrangigen Vervielfältigung nach § 16 UrhG von untergeordneter Bedeutung.

2. Zulässigkeit aufgrund von Nutzungsrechtseinräumungen oder Schrankenbestimmungen

Allfällige Verwertungshandlungen (wie etwa die Vervielfältigung) sind gegebenenfalls aufgrund einer Nutzungsrechtseinräumung oder einschlägigen Schrankenbestimmungen unter Beachtung des Dreistufentests i.S.d. Art. 5 Abs. 5 InfoSoc-RL⁴²⁹ zulässig.

a. Nutzungsrechtseinräumungen

Während für innerhalb der Justiz erarbeitete Inhalte regelmäßig bereits ein Nutzungsrecht eingeräumt worden sein wird (vgl. § 43 UrhG),⁴³⁰ ist eine (konkludente)⁴³¹ Rechtseinräumung nach § 29 Abs. 2, § 31 UrhG für das KI-Training nicht durch die Einreichung des Schriftsatzes oder eines angehängten Gutachtens bei Gericht anzunehmen. Selbst wenn bereits die Einreichung eines Schriftsatzes als eine solche Rechtseinräumung anzusehen wäre,⁴³² wird eine solche Einräumung entsprechend dem Zweck der Einreichung auf die Verwertung mit Blick auf das konkrete gerichtliche Verfahren beschränkt sein (vgl. § 31 Abs. 5 UrhG). Bei gerichtlich beauftragten Sachverständigengutachten (z.B. §§ 411, 411a ZPO) kann eine solche Nut-

429 S. u.a. Art. 5 Abs. 5 InfoSoc-Richtlinie: „Die [...] Beschränkungen dürfen nur [1.] in bestimmten Sonderfällen angewandt werden, in denen [2.] die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt wird und [3.] die berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden.“

430 S. oben unter E.I.2.

431 Nur *Paul*, in: *Holznagel/Hoeren/Sieber*, Teil 7.4, Rn. 93.

432 Gegebenenfalls ist eher eine Rechtseinräumung zugunsten des Mandanten denkbar, vgl. *OVG Hamburg*, BeckRS 2020, 30248.

zungsrechtseinräumung zwar tendenziell eher angedacht werden,⁴³³ scheint aber im Schrifttum kaum Anklang zu finden.⁴³⁴

Grundsätzlich denkbar ist die Beschaffung entsprechender Nutzungsrechte mit Blick auf Werke von Anwälten und Gutachtern, die gemäß dem geänderten Wahrnehmungsvertrag der VG Wort die sog. „KI-Lizenz“ eingeräumt haben.⁴³⁵ In der Praxis dürfte gegenwärtig nur eine geringe Anzahl der Anwälte und Gutachter bei der VG Wort gemeldet sein sowie eine entsprechende Lizenz eingeräumt haben.

b. Rechtspflege (§ 45 UrhG)

Nach § 45 Abs. 1 UrhG dürfen „einzelne Vervielfältigungsstücke von Werken zur Verwendung in Verfahren vor einem Gericht“ hergestellt werden und nach Abs. 3 ist „unter den gleichen Voraussetzungen [...] auch die Verbreitung, öffentliche Ausstellung und öffentliche Wiedergabe der Werke zulässig“. ⁴³⁶ Die Schranke stützt sich auf Art. 5 Abs. 3 lit. e, Abs. 4 Info-Soc-RL, die mitgliedstaatliche Schrankenbestimmungen zur „Sicherstellung des ordnungsgemäßen Ablaufs von [...] Gerichtsverfahren“ zulässt. Die Regelung soll die Aufgabenerfüllung der Gerichte sicherstellen,⁴³⁷ indem Vervielfältigungen zu Beweis Zwecken und als sonstige Hilfsmittel für eine Entscheidung ermöglicht werden.⁴³⁸

Der historische Gesetzgeber ging von einer bloß gerichtswerten und auf das jeweilige Verfahren beschränkten Verwertung aus und schloss daher unveröffentlichte Werke ein.⁴³⁹ Dementsprechend verlangt auch die (wohl) herrschende Auffassung zutreffend eine Verwendung in konkreten

433 Für in der Vergangenheit erstattete Gutachten wäre nicht ohne Weiteres auch eine Übertragung der Rechte mit Blick auf Text und Data Mining anzunehmen, sondern § 31a UrhG anzuwenden, vgl. *Hamann*, ZGE 16 (2024), 113 (138 f.).

434 Etwa *Ulrich*, DSB 2011, 308 (315), der keine Nutzungsrechteinräumung annimmt, sondern § 45 UrhG anwendet; s. aber zu Privatgutachten *T. Heinrich*, NZV 2015, 68 (68 f.).

435 S. <https://www.vgwort.de/veroeffentlichungen/aenderung-der-wahrnehmungsbedingungen/fragen-/antworten-zur-aenderung.html>.

436 Zum Änderungsverbot und der Quellenangabe nach §§ 62, 63 UrhG *Lüft*, in: *Wandtke/Bullinger*, § 45 UrhG Rn. 1.

437 *Lüft*, in: *Wandtke/Bullinger*, § 45 UrhG Rn. 1.

438 BT-Drs. IV/270, 63.

439 BT-Drs. IV/270, 63; *BVerwG*, GRUR 2020, 189 (Rn. 43); *Melichar/Stieper*, in: *Schricker/Loewenheim*, § 45 UrhG Rn. 13.

Gerichtsverfahren,⁴⁴⁰ wobei die Anzahl der Vervielfältigungen auf den Bedarf für das konkrete Verfahren zu beschränken ist.⁴⁴¹ Da der Begriff der Vervielfältigung nicht zwischen digitalen und analogen Niederlegungen differenziert,⁴⁴² sind im Ausgangspunkt auch digitale Vervielfältigungen einschließlich der Digitalisierung bei Führung einer digitalen Akte umfasst. Das gilt ebenso für Vervielfältigungen aus übergebenen (neuen) Akten in den Use-Cases 1-3.

Eine Digitalisierung von bestehenden Aktenauszügen nur zur Durchführung des KI-Trainings weist allerdings keinen konkreten Verfahrensbezug auf, weshalb dieser Vorgang nicht auf § 45 UrhG gestützt werden kann.

Unter Berücksichtigung des Schutzzwecks und des angelegten konkreten Verfahrensbezugs kann die Vorschrift zudem auch nicht im Wege einer Analogie auf Vervielfältigungen für das Training eines Sprachmodells als Schritt zur Verbesserung der Arbeitsabläufe in allen Gerichtsverfahren erstreckt werden. Denn die maßgebliche Interessenlage unterscheidet sich maßgeblich durch den weitergehenden Verwertungsumfang im Fall des Trainings eines Sprachmodells, selbst wenn das Sprachmodell zwar nur justizintern, aber eben gerichts- und verfahrensübergreifend zum Einsatz kommt. Zwar könnte der deutsche Gesetzgeber womöglich im Einklang mit dem Dreistufentest und der InfoSoc-RL die Schranke des § 45 UrhG, gegebenenfalls i.V.m. einer Vergütungspflicht, erweitern. Die derzeitige Regelung durch den deutschen Gesetzgeber gibt aber (noch) keine solche Auslegung her.

c. Text und Data Mining (§§ 44b UrhG und 60d UrhG)

Die § 44b, § 60d UrhG lassen gestützt auf⁴⁴³ Art. 3, 4 DSM-Richtlinie Vervielfältigungen für das sog. Text und Data Mining ohne Einschränkungen

440 Lüft, in: Wandtke/Bullinger, § 45 UrhG Rn. 3; BGH, GRUR 2021, 711 (713); LG Düsseldorf, GRUR-RR 2007, 193 (194); Dreier, in: Dreier/Schulze, § 45 UrhG Rn. 6; Melichar/Stieper, in: Schricker/Loewenheim, § 45 UrhG Rn. 6; Paul, in: Hilber/Borges, § 45 UrhG Rn. 1; Schulz, in: Götting/Lauber-Rönsberg/Rauer, § 45 UrhG Rn. 8; vgl. auch BGH, GRUR 2010, 623; Wiebe, in: Spindler/Schuster, § 45 UrhG Rn. 2; wohl auch Konertz, ZUM 2024, 355 (363).

441 Schulz, in: Götting/Lauber-Rönsberg/Rauer, § 45 UrhG Rn. 5.

442 Statt aller Heerma, in: Wandtke/Bullinger, § 16 UrhG Rn. 4.

443 S. zum Begriff der Ausnahme und Beschränkung in der DSM-Richtlinie und den Konsequenzen für die Mitgliedstaaten Brockmeyer, Text und Data Mining, S. 70 f.; schon zuvor Schack, GRUR 2021, 904.

des Zwecks mit Widerspruchslösung (§ 44b Abs. 2, 3 UrhG) bzw. für wissenschaftliche Forschungszwecke ohne Widerspruchsmöglichkeit zu (§ 60d UrhG). Text und Data Mining ist nach § 44b Abs. 1 UrhG „die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“.⁴⁴⁴

aa. KI-Training als Text und Data Mining

Nicht nur die Digitalisierung von analogen Dokumenten,⁴⁴⁵ sondern auch das weit verstandene⁴⁴⁶ KI-Training, unter Einschluss eines Sprachmodells, ist mit der (wohl) herrschenden Auffassung⁴⁴⁷ als Text und Data Mining i.S.d. § 44b UrhG zu qualifizieren.

Die Diskussion um die Anwendung und Reichweite des Text und Data Mining wurde unlängst vor allem durch ein Werk von *Dornis und Stober* neu entfacht.⁴⁴⁸ Unbestritten konnte der europäische Gesetzgeber bei Verabschiedung der DSM-RL vom 17. April 2019 noch nicht die Dimension des Scrapings für das Training von KI-Modellen vorhersehen,⁴⁴⁹ die in

444 S. auch die materiell übereinstimmende Definition aus Art. 2 Nr. 2 DSM-RL: „Text und Data Mining bezeichnet eine Technik für die automatisierte Analyse von Texten und Daten in digitaler Form, mit deren Hilfe Informationen unter anderem — aber nicht ausschließlich — über Muster, Trends und Korrelationen gewonnen werden können.“ sowie *T. Radtke*, ZGE 17 (2025), 1 (30 f.).

445 *Raue*, CR 2017, 656 (658); BT-Drs. 18/12329, 41.

446 D.h. die Zusammenstellung des Trainingskorpus, Pre-Training, Finetuning und weitere Schritte auf dem Weg zu einem KI-Modell, das in ein KI-System implementiert wird, sind umfasst.

447 S. hierzu insb. die Nachweise bei *T. Radtke*, ZGE 17 (2025), 1 (36) in Fn. 176; *Hamann*, ZGE 16 (2024), 113 (121) m.w.N.; *Bomhard*, in: Götting/Lauber-Rönsberg/Rauer, § 44b UrhG Rn. 11b ff. m.w.N.; s. auch BT-Drs. 19/27426, 60: „für das maschinelle Lernen als Basis-Technologie für Künstliche Intelligenz von besonderer Bedeutung“; *Schwartmann/Köhler*, in: Schwartmann/Keber/Zenner, 3. Kapitel, Rn. 113; einen Abgleich i.R.d. der Kuratierung von Trainingsdaten als TDM einordnend, die Frage aber im Übrigen offenlassend *LG Hamburg*, GRUR 2024, 1710 (Rn. 40 ff.); a.A. *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 94 ff.; außerdem sei die Diskussion noch nicht beendet nach *Geiger*, in: Thouvenin/Peukert/Jaeger/Geiger, When the Robots (Try to) Take Over: Of Artificial Intelligence, Authors, Creativity and Copyright Protection, S. 77.

448 *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle.

449 Zurecht aber auf die absehbare (Weiter-)Entwicklung von generativer KI hinweisend *Leistner*, GRUR 2024, 1665 (1669); zur Diskussion im Vorfeld der Regelung s. auch *Beurskens*, RD 2025, 1 (Rn. 17).

Chatbots wie ChatGPT zum Einsatz kommen. Maßgeblich aus diesem Umstand leitet *Dornis* ab,⁴⁵⁰ dass Text und Data Mining womöglich zwar Teilaspekte eines KI-Trainingsprozesses umfassen können, nicht aber sämtliche Vervielfältigungen im Zusammenhang mit dem KI-Training.⁴⁵¹ Schon zuvor wurden in der Literatur verschiedentlich Zweifel mit Blick auf einen Informationsgewinn unmittelbar durch den Trainingsprozess und⁴⁵² eine gebotene enge Auslegung als Ausnahmebestimmung⁴⁵³ geäußert.

Dieser Auffassung ist allerdings entgegenzuhalten, dass es dem Gesetzgeber grundsätzlich freisteht, ein weites Begriffsverständnis zu unterstellen. Die Definition des Text und Data Mining ist dementsprechend weitgefasst⁴⁵⁴ und schon vor einigen Jahren wurde ein enormes Wertschöpfungspotenzial in Text und Data Mining erblickt.⁴⁵⁵ Ferner bestätigen Wortlaut, Systematik unter Abgrenzung des Text und Data Mining zu wissenschaftlichen Forschungszwecken mit einem Erkenntnisgewinn und Genese der entsprechenden Vorschriften ein weites Verständnis des Text und Data Mining. Diesem Verständnis steht zudem auch nicht der Dreistufentest i.S.d. Art. 5 Abs. 5 InfoSoc-RL entgegen, da es sich (1.) bei dem Text und Data Mining weiterhin um einen Sonderfall handelt und insbesondere mangels der Entwicklung eines Konkurrenzprodukts auch (2. und 3.) die normale Verwertung des Werks und die sonstigen berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden.⁴⁵⁶ Zutreffend wird insoweit darauf hingewiesen, dass aus ökonomischer Perspektive bloße Kopien zu verhindern sind, nicht aber – wie regelmäßig im Rahmen von KI-Ausgaben von Bedeutung – eine Substitution, die auf (abstrakten) Ideen und Gestaltungen bestehender Werke aufbaut.⁴⁵⁷

Vor allem aber stützen die jüngeren Art. 53 Abs. 1 lit. c, ErwGr. 105 S. 2 ff. KI-VO eine solche Auslegung.⁴⁵⁸ Nach Art. 53 Abs. 1 lit. c KI-VO sind Anbieter von KI-Modellen mit allgemeinem Verwendungszweck ver-

450 Aus diesem Grund auch zweifelnd *Baumann*, NJW 2023, 3673 (Rn. 14).

451 *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 77, 87 ff.; *Dornis*, KIR 2024, 156 (158 ff.).

452 *Schack*, NJW 2024, 113 (Rn. 8).

453 *Welser*, GRUR 2024, 1406 (1413).

454 Etwa *Kleinkopf*, Text- und Data-Mining, S. 34.

455 *Geiger/Frosio/Bulayenko*, CEIPI Research Paper 2019 08 (19) m.w.N.; vgl. auch *Steinrötter/Schauer*, in: Barudi, § 4 Text und Data Mining, Forschung und Lehre, S. 146.

456 Ausf. *T. Radtke*, ZGE 17 (2025), 1 (34 ff.).

457 *Stieper/Denga*, GRUR 2024, 1473 (1475) m.w.N.

458 Überblick zum Meinungsstand bei *T. Radtke*, ZGE 17 (2025), 1 (36) in Fn. 177; etwa auch *Leistner*, GRUR 2024, 1665 (1669); *Hamann*, ZGE 16 (2024), 113 (122);

pflichtet, „eine Strategie zur Einhaltung des Urheberrechts der Union und damit zusammenhängender Rechte und 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 [zu bringen]“ (Hervorhebung d. d. Verf.). Hierdurch wird unterstellt, dass einem Vorbehalt nach Art. 4 Abs. 3 DSM-RL bzw. § 44b Abs. 3 UrhG entscheidende Bedeutung zukommt, also der Anwendungsbereich der Schrankenbestimmung auch und gerade die KI-Trainingsprozesse umfasst. Wäre hingegen das KI-Training als Ganzes nicht von der Schrankenbestimmung umfasst oder lediglich ein Teilaspekt des KI-Trainings als Text und Data Mining einzuordnen, hätte es gerade nicht der Herausstellung des betreffenden Vorbehalts bedurft.

Soweit *Dornis* der vorgenannten Vorschrift keinen Willen des Gesetzgebers entnehmen will, mit der Verabschiedung der KI-VO konkludent den Anwendungsbereich der DSM-RL zu erweitern,⁴⁵⁹ folgt dies wiederum (insoweit konsequent) der unzutreffenden Prämisse eines zuvor enger gefassten Begriffs des Text und Data Mining in der DSM-RL. Nach zutreffender Lesart hat der europäische Gesetzgeber durch die KI-VO in Kenntnis der jüngsten KI-Entwicklungen allerdings bloß die Reichweite des Text und Data Mining aus der DSM-RL *bestätigt*.⁴⁶⁰

Dementsprechend lehnt auch das überwiegende Schrifttum die durch *Dornis* vorgebrachten Einwände ab und hält zurecht an einer weiten Auslegung des Begriffs des Text und Data Mining fest.⁴⁶¹ Das KI-Training ist daher grundsätzlich als Text und Data Mining i.S.d. §§ 44b, 60d UrhG einzustufen.

Peukert, GRUR Int. 2024, 497 (504); Beurskens, RD 2025, 1 (Rn. 17); wohl auch Buchalik/Gehrmann, CR 2024, 145 (Rn. 59); anders vor Verabschiedung der KI-VO Jäger, Artificial Creativity?, S. 336, 349 ff.

459 *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 95 ff.

460 T. Radtke, ZGE 17 (2025), 1 (36 f.).

461 Stieper/Denga, GRUR 2024, 1473 (1474 ff.); Leistner, GRUR 2024, 1665 (1669 f.); Käde, KIR 2024, 162 (168); Dreier, in: Dreier/Schulze, § 44b Rn. 5; wohl auch Antoine, GRUR 2025, 118 (122) bei Fn. 59; LG Hamburg, GRUR 2024, 1710 (Rn. 49); Pesch, LTZ 2025, 72 (76): „flawed argument“; nicht Position beziehend, aber zumindest eine große Bedeutung des TDM für das Training generativer KI annehmend Schmid, KIR 2025, 69; offengelassen in der KI-VO nach Welser, GRUR 2024, 1406 (1415); Geiger, in: Thouvenin/Peukert/Jaeger/Geiger, When the Robots (Try to) Take Over: Of Artificial Intelligence, Authors, Creativity and Copyright Protection, S. 77.

bb. KI-Einsatz und Ausgabe als Text und Data Mining

Sofern für den konkreten Einsatz eines KI-Systems durch eine Eingabeaufforderung innerhalb des Rechenprozesses des Sprachmodells allfällige Vervielfältigungen angenommen werden,⁴⁶² sind diese als Teil des Analyseprozesses anzusehen. Denn derartige Vervielfältigungen stellen die Grundlage dar für den späteren Informationsgewinn aus trainiertem Sprachmodell und Eingabe.

Vervielfältigungen im Rahmen der durch das KI-System generierten Ausgabe erfolgen allerdings regelmäßig nicht (mehr), „um daraus [mittels automatisierter Analyse] Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“ (§ 44b Abs.1 UrhG), sprich Text und Mining vorzunehmen. Vielmehr wurden die in Rede stehenden Informationen bereits analysiert und dürften nunmehr dem (manuellen) Werkgenuss dienen. Dieses Ergebnis bestätigt auch der Dreistufentest auf den Stufen zwei und drei, wonach bei der wörtlichen Übernahme von Schriftsätzen und Gutachten die normale Verwertung des Werks gefährdet und die berechtigten Interessen des Rechtsinhabers verletzt sein könnten.

An wörtlichen, eng an den Trainingsdaten orientierten Ausgaben dürfte im untersuchungsgegenständlichen GSJ-Projekt regelmäßig allerdings ohnehin kein Interesse bestehen. Denn es wird eine spezifische Befassung mit neuen Anliegen und Aktenauszügen gewollt sein (z.B. im Rahmen der Sachverhaltsaufbereitung und -zusammenfassung in den Use-Cases 1-3), nicht dagegen eine Ausgabe eng am Stil und Wortschatz eines Anwalts oder Gutachters. Entsprechende Ausgaben sind nach den Ausführungen zum Datenschutzrecht⁴⁶³ hiernach durch angemessene technische und organisatorische Maßnahmen zu verhindern bzw. zu minimieren. Sollten ausnahmsweise durch einzelne Anwender gezielt rechtsverletzende Ausgaben durch Eingabeaufforderungen generiert werden (z.B. durch sog. Prompt Injections), liegt zudem mit Blick auf diese Vervielfältigung und eine hieraus resultierende Urheberrechtsverletzung eine Passivlegitimation primär des Anwenders anstelle des GSJ-Projektbetreibers nahe.⁴⁶⁴

462 S. aber oben unter E.II.1.a.bb

463 S. oben unter C.VI.4.

464 T. Radtke, ZGE 17 (2025), 1 (12); s. auch Determann/Paal, KI-Recht international, S. 139.

cc. Anwendbarkeit des § 44b Abs. 2 S. 1 UrhG

Die Anwendbarkeit des § 44b Abs. 2 S. 1 UrhG ist sowohl aus intertemporaler Perspektive als auch mit Blick auf einen möglicherweise erklärten Vorbehalt i.S.d. § 44b Abs. 3 UrhG zu untersuchen.

(1) Intertemporale Anwendbarkeit

Nach Art. 26 Abs. 1, 2 DSM-RL findet das dort niedergelegte unionsrechtliche Vorbild für § 44b UrhG auch Anwendung auf Werke, die (auch) ab dem 7. Juni 2021 geschützt sind, nicht aber auf „Handlungen und Rechte, die vor dem 7. Juni 2021 abgeschlossen bzw. erworben wurden“. Das deutsche Umsetzungsgesetz ist am gleichen Tag in Kraft getreten,⁴⁶⁵ greift die Regelung des Art. 26 Abs. 2 DSM-RL zwar nicht auf, ist aber unter Berücksichtigung der unionsrechtlichen Vorschrift auszulegen. Danach sind grundsätzlich alle geschützten Werke und Schutzgegenstände umfasst.⁴⁶⁶ Lediglich KI-Modelle, die auf Vervielfältigungen vor dem 7. Juni 2021 basieren, können unter Beachtung von Art. 26 Abs. 2 DSM-RL bei Auslegung des § 44b UrhG nicht auf die Schrankenbestimmung gestützt werden.⁴⁶⁷

Indem § 44b Abs. 2 S. 1 UrhG die rechtmäßige Zugänglichkeit von Werken voraussetzt, könnte die Anwendbarkeit auf schon vor dem 7. Juni 2021 rechtmäßig zugängliche Werke als erworbenes Recht infragestehen. Allerdings kann ein derart weitreichender Anwendungsbereich dem Art. 26 Abs. 2 DSM-RL mit Blick auf den Zusammenhang zu Art. 26 Abs. 1 DSM-RL nicht entnommen werden. Vielmehr bleiben gerade zuvor erworbene Rechte „unberührt“ und werden nicht durch Art. 26 Abs. 2 DSM-RL eingeschränkt.

(2) Vorbehalt der Vervielfältigungen zu Zwecken des Text und Data Mining

Rechteinhaber können sich nach § 44b Abs. 3 S. 1 UrhG ihre Rechte an Vervielfältigungen zu Zwecken des Text und Data Mining vorbehalten.

465 Vgl. Art. 5 des Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des Digitalen Binnenmarkts vom 31. Mai 2021, Bundesgesetzblatt 2021 Teil I Nr. 27, 1204 (4. Juni 2021).

466 Vgl. *Beurskens*, RDi 2025, 1 (Rn. 16).

467 *Schack*, NJW 2024, 113 (Rn. 7); *Welser*, GRUR-Prax 2023, 516 (Rn. 41); *Dregelies*, GRUR 2024, 1484 (1487).

(a) Form des Vorbehalts

Die Anforderung der Maschinenlesbarkeit⁴⁶⁸ des Vorbehalts sieht § 44b Abs. 3 S. 2 UrhG nur für über das Internet zugängliche Werke vor.⁴⁶⁹ Für die hier untersuchungsgegenständlichen, gerade nicht über das Internet frei zugänglichen Aktenauszüge schließen daher auch andere Formen eines ausdrücklichen⁴⁷⁰ Vorbehalts die Nutzung ex nunc⁴⁷¹ aus. Ein solcher Vorbehalt dürfte allerdings in Ansehung der für das Text und Data Mining in der Justiz bisher typischerweise kaum sensibilisierten Adressaten allenfalls (sehr) selten⁴⁷² vorkommen. In einem Schriftsatz oder Gutachten wären daher typischerweise Vorbehalte in natürlicher Sprache zu berücksichtigen, die dem Dokument aufgrund einer entsprechenden Formulierung, Positionierung und gegebenenfalls Hervorhebung „ohne allzu große Anstrengungen zu entnehmen sind“.⁴⁷³ Die Voraussetzung der „Ausdrücklichkeit“ schließt konkludente Vorbehalte aus sowie verlangt aber zugleich einen konkreten Bezug zu Text und Data Mining.⁴⁷⁴ Die Darlegungs- und Beweislast für einen fehlenden Vorbehalt soll nach der gesetzgeberischen Intention der Nutzende tragen.⁴⁷⁵

Konkrete Vorgaben zur Formulierung des Vorbehalts macht der Gesetzgeber nicht, sodass vielgestaltige Formulierungen denkbar sind, wie etwa: „Kein Text oder Data Mining“,⁴⁷⁶ „Kein TDM“ oder „Kein KI-Training“. Grenzfälle dürften allgemeine Vorbehalte darstellen, wie etwa „Alle Rechte vorbehalten“,⁴⁷⁷ „Vervielfältigung vorbehalten“ oder die in Gutachten übliche⁴⁷⁸ Untersagung der Weitergabe des Gutachtens. Ein bloßer Urheberrechtshinweis (z.B. „© Autor, 2025“) ist weder im Hinblick auf das Text und Data Mining noch einen Vorbehalt hinreichend ausdrücklich,

468 Hierzu etwa *LG Hamburg*, GRUR 2024, 1710 (1713 f.); *Hamann*, ZGE 16 (2024), 113 (128 ff.).

469 S. auch ErwGr. 18 UAbs. 2 DSM-RL.

470 S. Art. 4 Abs. 3 DSM-RL: „ausdrücklich in angemessener Weise“.

471 *Paul*, in: Hilber/Borges, § 44b UrhG Rn. 5; *Hamann*, ZGE 16 (2024), 113 (124) m.w.N.

472 S. allgemein *Bomhard*, DSRITB 2023, 255 (265).

473 *Bomhard*, in: Götting/Lauber-Rönsberg/Rauer, § 44b UrhG Rn. 25; *Bomhard*, InTeR 2023, 174 (178).

474 *Hamann*, ZGE 16 (2024), 113 (134 f.); vgl. BT-Drs. 19/27426, 89.

475 BT-Drs. 19/27426, 89.

476 *Paul*, in: Hilber/Borges, § 44b UrhG Rn. 5.

477 *Paul*, in: Hilber/Borges, § 44b UrhG Rn. 5; wohl aber nicht nach *Hamann*, ZGE 16 (2024), 113 (134 f.), der einen konkret-individuellen Bezug verlangt.

478 S. *T. Heinrich*, NZV 2015, 68.

sondern kann vielmehr auch als bloße Nennung des Rechtsinhabers bzw. gegebenenfalls auch des Urhebers i.S.d. § 13 UrhG verstanden werden.⁴⁷⁹ Die Vervielfältigung für eine automatisierte Analyse von Dokumenten,⁴⁸⁰ um einen Vorbehalt zu identifizieren, ist zwar ebenfalls Text und Data Mining, aber in systematischer Hinsicht mit Blick auf § 44b Abs. 3 UrhG und gegebenenfalls auch als vorübergehende Vervielfältigung nach § 44a UrhG⁴⁸¹ zulässig.

(b) Rechtsinhaber

Die Rechte vorbehalten kann sich nur der „jeweilige“ Rechtsinhaber i.S.d. Art. 3 Abs. 3 DSM-RL, der nicht mit dem Urheber identisch sein muss. Für die Frage der Rechtsinhaberschaft kann es daher darauf ankommen, wem Nutzungsrechte i.S.d. §§ 31 ff. UrhG eingeräumt wurden.⁴⁸² Unter Berücksichtigung der Zweckübertragungslehre sind die dem Vorbehalt zugrunde liegenden Rechte mit Blick auf Text und Data Mining für Rechteeinräumungen in der Vergangenheit nach einer Stimme in der Literatur nicht ohne Weiteres erfasst, sondern sollen als neue Nutzungsart zu behandeln sein (insbesondere § 31a UrhG).⁴⁸³

Sofern mit Blick auf Schriftsätze ein Rechteübergang auf den Mandanten angenommen wird,⁴⁸⁴ umfasste dieser Rechteübergang jedenfalls bis vor wenigen Jahren nicht die dem Vorbehalt zugrunde liegenden Rechte. Ein etwaiger Vorbehalt (konnte) also regelmäßig durch den jeweiligen Rechtsanwalt oder Gutachter als Rechtsinhaber erklärt werden. Ein ähnliches Ergebnis ergibt sich, wenn man die Übertragung von Nutzungsrechten verlangt und gegebenenfalls die Wertungen aus §§ 177 ff. BGB für eine etwaige (rückwirkende) Genehmigung eines nicht durch den jeweiligen Rechtsinhaber erklärten Vorbehalts heranzieht.⁴⁸⁵

Für Aktenauszüge aus jüngerer Vergangenheit, insbesondere solche nach dem 7. Juni 2021, kann auch ein Übergang der zum Vorbehalt berechtigten

479 Hierzu etwa Mantz, in: Dreier/Schulze, § 10 UrhG Rn. 13.

480 S. etwa LG Hamburg, GRUR 2024, 1710 (Rn. 67 f.).

481 S. nachfolgend unter E.II.2.d

482 Leistner, GRUR 2024, 1665 (1671); Vesala, IIC 2023, 351 (357); Brockmeyer, Text und Data Mining, S. 79.

483 Hamann, ZGE 16 (2024), 113 (138 f.); a.A. Leistner, GRUR 2024, 1665 (1672): allgemeine Rechtsgeschäftslehre.

484 S. unter E.II.2.a.

485 Vgl. Leistner, GRUR 2024, 1665 (1671 f.).

Nutzungsrechte vorliegen. In der Folge könnte nur der neue Rechtsinhaber (z.B. der Mandant) den Vorbehalt erklären.

Der hieraus resultierenden erhöhten Komplexität kann praktisch etwa dadurch begegnet werden, dass die Wirksamkeit sämtlicher in den Aktenauszügen erklärten Vorbehalte unterstellt wird und entsprechende Dokumente gegebenenfalls von dem Trainingsprozess ausgenommen werden.

(c) Rechtmäßig zugängliche Werke

Die Schrankenbestimmung ermöglicht die Vervielfältigungen von „rechtmäßig zugänglichen Werken“. Unter Berücksichtigung von ErwGr. 14 S. 1, 2 DSM-RL⁴⁸⁶ ist auf die Rechtmäßigkeit des Zugangs, nicht jedoch darauf abzustellen, „ob das Werk mit oder ohne Zustimmung des Rechteinhabers zugänglich gemacht wurde“.⁴⁸⁷ Es schadet der Verwirklichung des Tatbestandsmerkmals also nur, wenn Schutzmaßnahmen umgangen wurden.⁴⁸⁸ Für die Justiz wird im GSJ-Projekt regelmäßig ein solcher rechtmäßiger Zugang gegeben sein.⁴⁸⁹

Entsprechende Überlegungen für weitere Akteure, denen die Aktenauszüge durch die Justiz übermittelt werden, wobei gegebenenfalls weitere Anforderungen zu beachten sind.⁴⁹⁰

(d) Weitere Anforderungen

Die weiteren Anforderungen an einen wirksamen Vorbehalt betreffen einerseits einen möglichen Ausschluss abstrakt-genereller Vorbehalte aufgrund des Ausdrücklichkeitserfordernisses,⁴⁹¹ andererseits (zulässige) nach-

486 In ErwGr. 14 S. 1, 2 DSM-RL wird ein „rechtmäßig[er] Zugang“ verlangt.

487 So zutreffend *Raue*, CR 2017, 656 (658); *Specht-Riemenschneider*, OdW 2018, 285 (286); *Döhl*, RuZ 2020, 195 (217); anders hingegen *Schwartmann/Köhler*, in: *Schwartmann/Keber/Zenner*, 3. Kapitel, Rn. 118.

488 *Dornis/Stober*, Urheberrecht und Training generativer KI-Modelle, S. 98 f.; anschaulich *Talke*, Urheberrecht in Bildung, Wissenschaft und Kulturerbe, S. 30 f.; *K. Wagner*, MMR 2024, 298 (299).

489 Insb. auch mit Blick darauf, dass keine Zugangsbeschränkungen umgangen werden, s. *Paul*, in: *Hilber/Borges*, § 44b UrhG Rn. 3.

490 S. hierzu unter D.II.3.

491 So etwa *Hamann*, ZGE 16 (2024), 113 (148 f.)

träglich⁴⁹² erklärte Vorbehalte. Nur in Ausnahmefällen, so insbesondere wenn ein Rechtsinhaber nachträglich gegenüber den jeweiligen Gerichten⁴⁹³ einen Vorbehalt erklärt, gegebenenfalls pauschal bezogen auf seine sämtlichen Schriftsätze, werden sich diese Fragen stellen.⁴⁹⁴ Die Frage nach einer *ex tunc*- oder *ex nunc*-Wirkung⁴⁹⁵ ist ebenfalls zu vernachlässigen, da selbst eine *ex nunc*-Wirkung faktisch zurückwirken kann, wenn das einmal vervielfältigte Werk nach Erklärung des Vorbehalts für ein weiteres Training erneut vervielfältigt wird.

Die Möglichkeit eines vor Inkrafttreten des § 44b UrhG erklärten Vorbehalts⁴⁹⁶ dürfte für das untersuchungsgegenständliche Projekt in Bezug auf in den Aktenauszügen enthaltene Schriftsätze und Gutachten gleichfalls kaum von Relevanz sein.

(3) Zwischenergebnis

Die Regelung des § 44b Abs. 2 S. 1 UrhG kann belastbar das Training des Sprachmodells ermöglichen, setzt allerdings die Filterung von Werken voraus, für die ein Nutzungsvorbehalt erklärt wurde. Da tendenziell solche Erklärungen nur äußerst selten vorkommen dürften, besteht ein sehr geringes Restrisiko für nicht identifizierte und sodann übergegangene Vorbehalte. Selbst in diesen Fällen kann zudem gegebenenfalls eine andere Schrankenbestimmung eingreifen, was nachfolgend näher zu prüfen ist.

Die Regelung des § 44b UrhG dürfte in der Praxis zur Folge haben, dass ab Bekanntwerden entsprechender KI-Projekte in der Anwaltschaft pauschal entsprechende Vorbehalte, etwa in Kanzlei-Briefköpfen erklärt werden.⁴⁹⁷ Die hieraus folgende Begünstigung des GSJ-Projekts durch „Ge-

492 Nachträglich in diesem Sinne meint, einen Vorbehalt separat nach Einreichung eines Schriftsatzes oder Gutachtens zu erklären, der *ex nunc* zukünftiges Text und Data Mining verhindern könnte.

493 Eine allgemeine Bekanntmachung über die Website des Rechtsinhabers genügt nicht den Anforderungen an die einfache Auffindbarkeit des ausdrücklichen Vorbehalts.

494 Eine etwaige Genehmigung eines durch eine nichtberechtigte Person erklärten Vorbehalts kann nach § 184 BGB zurückwirken *Leistner*, GRUR 2024, 1665 (1671f.) unter Verweis auf die LAION-Entscheidung des LG Hamburg.

495 *Brockmeyer*, Text und Data Mining, S. 79 f.

496 S. die Konstellation in *LG Hamburg*, GRUR 2024, 1710.

497 Gegebenenfalls besteht insoweit aber auch kein bzw. ein nur eingeschränkter Markt und kein Interesse an der Erklärung eines Vorbehalts, vgl. *Raue*, ZUM 2020, 172 (173).

heimhaltung“ hat keine unmittelbaren urheberrechtlichen Implikationen (siehe aber zum Datenschutzrecht unter D.V.6.a), denn der Vorbehalt nach § 44b Abs. 3 UrhG hängt nicht von einem konkreten Projekt ab, sondern konnte bereits vorher bezogen auf beliebige Werke durch die Anwaltschaft und Gutachter erklärt werden. Die Frage des Umgangs mit dem jeweiligen Projekt dürfte daher vor allem eine Frage der Öffentlichkeitsarbeit und der öffentlich-rechtlichen Anforderungen an die Auftraggeber sein.

Vorliegend nicht im Detail zu erörtern sind ferner Fragen nach einem all-fälligen Umgestaltungsbedarf des besonderen elektronischen Anwaltspost-fachs (beA) zur Möglichkeit der einfachen Erklärung eines Vorbehalts.⁴⁹⁸

dd. Anwendbarkeit des § 60d UrhG

§ 60d UrhG erlaubt Text und Data Mining unabhängig von einem erklärten Vorbehalt für wissenschaftliche Forschungszwecke durch Forschungsorganisationen. Unter den Begriff der „Forschungsorganisationen“ fallen insbesondere Hochschulen nach § 60d Abs. 2 S. 2 UrhG.

Allerdings genügt es entgegen einem insoweit kritisch zu betrachtenden Urteil des LG Hamburg nicht, dass der durch Text und Data Mining generierte Datensatz später einmal für wissenschaftliche Forschungszwecke verwendet werden *kann*.⁴⁹⁹ Mit Blick auf die Abgrenzung von Text und Data Mining einerseits und wissenschaftlichen Forschungszwecken andererseits bedarf es gerade des Bezugs zu einem konkreten Forschungsprojekt.⁵⁰⁰ Für anwendungsbezogene Forschung, wie im GSJ-Projekt, ist darüber hinaus die Frage aufzuwerfen, ob es genügt, wenn die wissenschaftlichen Forschungszwecke von vornherein ein bloßes Durchgangsstadium für die spätere Umsetzung in der Praxis sind. Ein derartiges Verständnis würde dazu führen, dass letztlich jeder kommerzielle Einsatz gerechtfertigt wäre, wenn und soweit sich Elemente aus dem Entwicklungsstadium auch als Forschung einer Forschungsorganisation einordnen ließen. Für die Anwendung der Schrankenbestimmung aus § 44b UrhG bliebe somit kaum noch Raum.

498 In ErwGr. 18 UAbs. 2 S. 5 DSM-RL heißt es nämlich: „Die Rechteinhaber sollten in der Lage sein, Maßnahmen zu treffen, mit denen sie sicherstellen, dass ihre diesbezüglichen Vorbehalte Beachtung finden.“

499 *LG Hamburg*, GRUR 2024, 1710 (Rn. 76 f.); begrüßenswert hingegen nach *Leistner*, GRUR 2024, 1665 (1667).

500 Hierzu ausf. *T. Radtke*, ZGE 17 (2025), 1 (44 ff.).

ErwGr. II S. 2 DSM-RL lässt sich ein Hinweis darauf entnehmen, dass die Zusammenarbeit mit der Privatwirtschaft der persönlichen Anwendbarkeit der Vorschrift nicht schaden muss.⁵⁰¹ Forschungsergebnisse aus öffentlich-privaten Partnerschaften werden typischerweise in der Praxis umgesetzt. Außerdem verlangen § 60d UrhG, Art. 3 DSM-RL nach ihrem Wortlaut im Gegensatz zu anderen Schrankenbestimmungen (wie etwa § 44a UrhG, Art. 5 Abs. 1 InfoSoc-RL („alleiniger Zweck“)) keine Exklusivität der Verwendung zu wissenschaftlichen Forschungszwecken. Zugleich verfolgte Sekundärzwecke stehen daher der Anwendung des § 60d UrhG grundsätzlich nicht entgegen.

Gleichwohl bestehen mit Blick auf das Konkurrenzverhältnis zum allgemeinen § 44b UrhG gegebenenfalls Bedenken, wenn das Forschungsprojekt von *vornherein* nicht primär auf den wissenschaftlichen Erkenntnisgewinn ausgerichtet ist, sondern auf den Einsatz des trainierten Sprachmodells in der (Justiz-)Praxis zur Sachverhaltsaufbereitungen in konkreten Verfahren (siehe Use-Cases 1-3) und durch Forschung lediglich begleitet werden soll.⁵⁰² Denn in dieser Konstellation sind die Vervielfältigungen unmittelbare Ausgangsbasis für das in der Praxis eingesetzte Produkt und nicht ein bloßes Forschungsobjekt, auf dessen Grundlage sodann ein separates Produkt für den operativen Einsatz entwickelt wird. Es ist somit festzuhalten: Nur weil eine Tätigkeit im Rahmen der Produktentwicklung auch wissenschaftliche Erkenntnisse zu liefern vermag, kann die produktbezogene Tätigkeit nicht ohne Weiteres pauschal und einschließlich der hierzu notwendigen Verwertungshandlungen den Schranken für wissenschaftliche Forschungszwecke zugerechnet werden.

Diese Ablehnung wissenschaftlicher Forschungszwecke für das GSJ-Projekt wird zudem bestätigt durch die Inbezugnahme des Dreistufentests (vgl. insoweit unter anderem Art. 5 Abs. 5 InfoSoc-RL): Eine weite Auslegung der wissenschaftlichen Forschungszwecke, wonach begleitende Forschung durch Text und Data Mining genügt bei einem von vornherein gerade auf

501 „Im Einklang mit der derzeitigen Forschungspolitik der Union, die Hochschulen und Forschungsinstitute zur Zusammenarbeit mit der Privatwirtschaft anhält, sollten auch Forschungsorganisationen eine solche Ausnahme nutzen dürfen, sofern ihre Forschungstätigkeit im Rahmen öffentlich-privater Partnerschaften durchgeführt wird“.

502 Vgl. auch T. Radtke, ZGE 17 (2025), 1 (21, 44 ff.); sowie schon den deutschen Gesetzgeber in BT-Drs. 18/12329, 39: „Forschung, die ein Unternehmen betreibt, um Waren oder Dienstleistungen zu entwickeln [sic!] und diese dann zu vermarkten, dient allerdings kommerziellen Zwecken“.

den operativen Einsatz ausgelegt und entsprechend beauftragten Endprodukt (i.e. das KI-Modell), würde sich gerade nicht mehr auf bestimmte Sonderfälle⁵⁰³ im Sinne der ersten Stufe des Dreistufentests beschränken. Eine solche Auslegung würde vielmehr die umfassende Anfertigung von Vervielfältigungen als Ausgangspunkt für die weitere Verwertung ermöglichen. Die Vervielfältigung i.S.d. § 60d UrhG würde hiernach im Ergebnis zu einem Regelfall. Diese weite Auslegung ist daher auch mit Blick auf die dritte Stufe des Dreistufentests bedenklich, sprich die ungebührliche Verletzung der berechtigten Interessen des Rechtsinhabers unter Berücksichtigung der Vergütungsfreiheit des § 60d UrhG⁵⁰⁴ und des Nutzungsumfangs.⁵⁰⁵

Eine andere Bewertung mag zwar mit entsprechender Argumentation noch vertretbar sein, geht aber jedenfalls mit einem erheblichen Risiko von Urheberrechtsverstößen einher. Nach der vorzugswürdigen Ansicht ist eine solche andere Bewertung allerdings gerechtfertigt, wenn und soweit Text und Data Mining auf ein (Teil-)Produkt gerichtet ist, das gerade nicht von vornherein nur dem operativen Einsatz dient (z.B. ein eigenständiges, zunächst trainiertes KI-Modell zur Klärung von Forschungsfragen für die spätere Entwicklung eines KI-Modells für den Produktiveinsatz). Wenn und soweit (privilegierte) wissenschaftliche Forschungszwecke in solchen Konstellationen belastbar angenommen werden können, werden sich zugleich strengere Anonymisierungsanforderungen aus dem Datenschutzrecht ergeben, sprich aus Art. 89 Abs. 1 DSGVO i.V.m. Art. 25 Abs. 2 BayDSG, § 17 Abs. 3 DSG NRW.

d. Vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)

Nach § 44a Nr. 2 UrhG sind „vorübergehende Vervielfältigungshandlungen [zulässig], die flüchtig oder begleitend sind und einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen und deren alleiniger Zweck es ist, eine rechtmäßige Nutzung eines Werkes oder sonstigen Schutzgegenstands zu ermöglichen, und die keine eigenständige wirtschaft-

503 Vgl. zur engen Auslegung der Ausnahmen mit Blick auf den Dreistufentest *EuGH*, GRUR 2017, 610 (Rn. 63) – Stichting Brein/Jack Frederik Wullems; GRUR 2009, 1041 (Rn. 58) – Infopaq.

504 *Ott*, ZUM 2009, 345 (353) in Fn. 77; *Senfileben*, GRUR Int. 2004, 200 (211); vgl. auch *Hofmann*, ZUM 2024, 166 (172).

505 Vgl. *EuGH*, EuZW 2014, 868 (Rn. 56) – TU Darmstadt/Eugen Ulmer KG.

liche Bedeutung haben“. Die rechtmäßige Nutzung umfasst sowohl den Werkgenuss als auch eine von den Schrankenbestimmungen gedeckte Nutzung.⁵⁰⁶ Die Regelung aus § 44a Nr. 2 UrhG adressiert insbesondere das sog. Caching.⁵⁰⁷

Daher kann die Vorschrift für temporäre Zwischenspeicherungen im Zusammenhang mit dem Training eines Sprachmodells Bedeutung erlangen, z.B. wenn temporäre (digitalisierte) Kopien auf den Trainingsservern angelegt werden, die der Ausgangspunkt für abgeleitete Textformate sind. Diese Kopie muss als integraler Teil eines technischen Verfahrens automatisiert nach dessen Abschluss gelöscht werden,⁵⁰⁸ wobei eine entsprechende Programmierung der Software genügt.⁵⁰⁹ Für die wirtschaftliche Bedeutung ist unter Berücksichtigung des Schutzzwecks des Urheberrechts darauf abzustellen, ob spezifisch die urheberrechtlich geschützten Bestandteile eine wirtschaftliche Bedeutung haben.⁵¹⁰ Eine entsprechende wirtschaftliche Bedeutung ist also nicht bereits deshalb anzunehmen, weil auf Grundlage der vorübergehenden Vervielfältigung eine Verwertung von nicht urheberrechtlich geschützten Bestandteilen möglich ist (wie etwa in Form eines Sprachmodells, das nicht einzelne Werke vervielfältigt und Urheber imitiert).⁵¹¹ Im untersuchungsgegenständlichen Projekt können also gestützt auf § 44a UrhG digitale Aktenauszüge auf einem Server abgelegt werden, wenn diese automatisiert als Übergangsstadium zur Übertragung in ein abgeleitetes Textformat oder sonstigen Aufbereitung vor dem Training angelegt und unmittelbar nach der Übertragung automatisiert gelöscht werden.

Grundsätzlich können auch vorübergehende Vervielfältigungen zu Zwecken des Text und Data Mining oder flüchtige Vervielfältigungen im Rah-

506 Schulz, in: Götting/Lauber-Rönsberg/Rauer, § 44a UrhG Rn. 13; Dreier, in: Dreier/Schulze, § 44a UrhG Rn. 8.

507 ErwGr. 33 S. 3 InfoSoc-RL.

508 EuGH, GRUR 2009, 1041 (Rn. 64) – Infopaq.

509 Zu eng daher LG Hamburg, GRUR 2024, 1710 (Rn. 35); wie hier auch Leistner, GRUR 2024, 1665 (1666).

510 EuGH, GRUR 2009, 1041 (Rn. 50) – Infopaq; MMR 2011, 817 (Rn. 175) – Football Association Premier League Ltd. u.a.

511 Vgl. T. Radtke, ZGE 17 (2025), 1 (25 f.); ebenfalls ablehnend, da nicht die Werke, sondern deren Datenpunkte Gegenstand des Trainings seien La Durantaye, AfP 2024, 9 (Rn. 22); wegen Änderung des Vervielfältigungsstücks Hofmann, ZUM 2024, 166 (169); von einer solchen Bedeutung aber regelmäßig im Rahmen des KI-Trainings ausgehend Wissenschaftliche Dienste des Bundestags, Künstliche Intelligenz und Machine Learning - Eine urheberrechtliche Betrachtung, S. 12; Pesch/Böhme, GRUR 2023, 997 (1006). S. auch ErwGr. 9 DSM-RL.

men eines Trainingsprozesses⁵¹² auf § 44a Nr. 2 UrhG gestützt werden. Der europäische Gesetzgeber erkennt in ErwGr. 9 S. 2 DSM-RL ausdrücklich an, dass es auch „Fälle des Text und Data Mining geben [kann], in denen [...] die Vervielfältigungen unter die in Artikel 5 Absatz 1 der Richtlinie 2001/29/EG vorgesehene verbindliche Ausnahme für vorübergehende Vervielfältigungshandlungen fallen, die auch künftig auf Verfahren des Text und Data Mining angewandt werden sollte [...]“. Allerdings dürfte die Aufbewahrung des Trainingskorpus zur weiteren Verbesserung des Sprachmodells und gegebenenfalls des Trainings einer weiteren Version gewünscht sein, sodass es eines Rekurses auf die speziellen Text und Data Mining-Ausnahmen in §§ 44b, 60d UrhG bedarf.⁵¹³

e. Zitate (§ 51 UrhG)

§ 51 UrhG erlaubt Vervielfältigungen zum Zweck des Zitats. Vereinzelte KI-Ausgaben mit geschützten Werkbestandteilen könnten grundsätzlich der geistigen Auseinandersetzung⁵¹⁴ mit dem jeweiligen Werk dienen und daher unter den § 51 UrhG fallen. Allerdings setzt die Vorschrift, wie sich aus einer Gesamtschau der § 51 S. 2 Nr. 1-3 UrhG ergibt, grundsätzlich die Auseinandersetzung im Rahmen eines Werks eines menschlichen Urhebers voraus und dürfte daher nicht die Aufnahme in einen KI-generierten Text umfassen. Ferner wird in den Use-Cases des GSJ-Projekts keine abgrenzende Auseinandersetzung mit dem Werk, sondern eine bloße Vervielfältigung eines zudem noch nicht veröffentlichten Werks erfolgen.⁵¹⁵ Die Vervielfältigung dürfte insbesondere auch nicht die nach § 63 Abs. 1 S. 1 UrhG erforderliche Quellenangabe enthalten. Die Zitatschranke kommt daher grundsätzlich nicht zur Rechtfertigung etwaiger Vervielfältigungen in Betracht.

f. Wissenschaftliche Forschung (§ 60c UrhG)

Die Vorschrift des § 60c UrhG stellt im untersuchungsgegenständlichen Projekt keine geeignete Schranke dar. § 60c UrhG umfasst beispielsweise

512 Jager, Artificial Creativity?, S. 348; vgl. auch Stieper/Denga, GRUR 2024, 1473 (1475).

513 S. auch Chiou, JIPITEC 2019, 398 (405 ff.).

514 BGH, NJW 1968, 1875 (1876 f.).

515 Zu den Anforderungen Peifer, ZUM 2020, 342.

„die Unterrichtung über den Stand der Wissenschaft“,⁵¹⁶ ist nach § 60h UrhG vergütungspflichtig und erfordert nach § 63 Abs. 1 S. 1 UrhG eine Quellenangabe.

Selbst wenn man im GSJ-Projekt die Verfolgung wissenschaftlicher Forschungszwecke unterstellt,⁵¹⁷ ist eine quantitative Beschränkung vorgesehen von jeweils 15 Prozent (§ 60c Abs. 1 UrhG) bzw. 75 Prozent eines Werkes (§ 60c Abs. 2 UrhG). Grundsätzlich wäre für das GSJ-Projekt der geringere Wert von 15 Prozent aus § 60c Abs. 1 UrhG zugrunde zu legen, da § 60c Abs. 2 UrhG nicht die Weitergabe über die eigene Einrichtung hinaus umfasst;⁵¹⁸ solche Weitergaben über die eigene Einrichtung hinaus dürften im GSJ-Projekt bereits durch die Zusammenarbeit zwischen den ausführenden Stellen und den Ministerien erfolgen. In Ansehung der speziell(er)en Regelungen der §§ 44b, 60d UrhG zu Vervielfältigungen für Zwecke des Text und Data Mining ist zudem zweifelhaft, ob § 60c UrhG in diesen Konstellationen überhaupt Anwendung finden kann.

Die Verwertung von bloß 15 Prozent der Werke bei verbleibender Rechtsunsicherheit mit Blick auf die spezielleren §§ 44b, 60d UrhG und die Verfolgung wissenschaftlicher Forschungszwecke ermöglicht nicht belastbar die Verwertung im Rahmen des GSJ-Projekts.

3. Einbindung weiterer Stellen

Im GSJ-Projekt wirken die Ministerien und die ausführenden Stellen zusammen, was die Frage nach dem berechtigten Personenkreis mit Blick auf die jeweiligen Schranken aufwirft. Im Ausgangspunkt bedarf diejenige Person einer Rechtfertigung, die selbst eine Nutzungshandlung begeht bzw. an dieser mitwirkt und ohne eine einschlägige Schrankenbestimmung als Täter, Teilnehmer, Störer oder Hilfsdienst⁵¹⁹ für eine Urheberrechtsverletzung passiv legitimiert wäre.⁵²⁰

516 Grübler, in: Götting/Lauber-Rönsberg/Rauer, § 60c UrhG Rn. 1.

517 S. aber unter E.II.2.c.dd.

518 Grübler, in: Götting/Lauber-Rönsberg/Rauer, § 60c UrhG Rn. 13; Stieper, in: Schricker/Loewenheim, § 60c UrhG Rn. 19; BT-Drs. 18/12329, 40: „Die so hergestellten Kopien dürfen in keiner Form weitergegeben werden“.

519 Reber, in: Loewenheim, § 86 Ansprüche aus Verletzung des Urheber- oder Leistungsschutzrechts, Rn. 52-142.

520 Vgl. Wolff/Wandtke, in: Wandtke/Bullinger, § 97 UrhG Rn. 38.

Die urheberrechtlichen Schranken sind grundsätzlich neutral mit Blick auf die berechtigten Personen und ermöglichen demzufolge die Verwendung von Hilfsmitteln sowie die Einbindung Dritter im Auftrag eines Berechtigten,⁵²¹ verlangen allerdings beispielsweise in § 45 UrhG einen engen Zusammenhang mit einem Gerichtsverfahren, in §§ 44b, 60d UrhG den rechtmäßigen Zugang, wobei § 60d UrhG zusätzliche Anforderungen an die berechtigte Person stellt (i.e. das Erfordernis einer Forschungsorganisation). Dieser rechtmäßige Zugang unterliegt allerdings keinen hohen Anforderungen,⁵²² sondern ist bereits gewahrt, wenn die Werke den ausführenden Stellen durch die Ministerien zugänglich gemacht werden.

Der Rekurs auf § 45 Abs. 1 UrhG scheidet für das untersuchungsgegenständliche Projekt aus, da sich das Projekt – anders als der Einsatz des entwickelten Tools in den einzelnen Use-Cases – nicht auf ein konkretes gerichtliches Verfahren bezieht.⁵²³ Auf §§ 44a, 44b UrhG wirkt sich die Einbindung der ausführenden Stellen zur Durchführung des Sprachmodell-Trainings für die Ministerien bzw. die Justiz nicht aus. Denn trotz Auslagerung bleibt es bei der gleichen Nutzungshandlung, die auch bei eigenständiger Ausführung durch die ausführenden Stellen gerechtfertigt wäre.

§ 60d UrhG scheidet im GSJ-Projekt aus und bedarf daher keiner weiteren Prüfung mit Blick auf die Beschränkung auf Forschungsorganisationen nach § 60d Abs. 2 UrhG.

4. Ableitung von Einsatzbedingungen

Aus dem Vorherigen lassen sich generelle und wesentliche Einsatzbedingungen für das untersuchungsgegenständliche Projekt ableiten.

Das eigentliche Training eines Sprachmodells ist urheberrechtlich grundsätzlich neutral. Allerdings bedürfen Vervielfältigungen und gegebenenfalls Bearbeitungen vor, während oder nach dem Trainingsprozess eines eingeräumten Nutzungsrechts oder einer einschlägigen Schrankenbestimmung.

Soweit nicht schutzfähige Bestandteile von einzelnen Aktenauszügen wahrnehmbar verarbeitet werden, ist das Projekt urheberrechtlich neutral.

521 T. Radtke, ZGE 17 (2025), 1 (24) m.w.N. in Fn. 109; s. insb. am Beispiel der Privatkopie BGH, MMR 2020, 538 (Rn. 22); MMR 2024, 1037 (Rn. 15 ff.); EuGH, MMR 2018, 80 (Rn. 25) – VCAST Limited/RTI SpA.

522 S. oben unter E.II.2.c.cc(2)(c).

523 S. oben unter E.II.2.b.

Es empfiehlt sich daher auf erster Stufe sowohl (i) die Filterung von wahrscheinlich urheberrechtlich geschützten Werken, die zugleich von geringem Mehrwert für den Trainingsprozess sein dürften (z.B. ausgewählte Gutachtenkategorien) als auch (ii) die Vermeidung von Vervielfältigungen dadurch, dass bereits digitalisierte Aktenauszüge unmittelbar in abgeleitete Textformate übertragen und zum Training des Sprachmodells verwendet werden. Nicht vermeidbare Vervielfältigungen können auf zweiter Stufe für eine automatisierte Zwischenspeicherung zur Weiterverarbeitung auf § 44a Nr. 2 UrhG gestützt werden. Vervielfältigungen, die nicht nur vorübergehend gespeichert werden sollen, können auf dritter Stufe vorzugswürdig auf § 44b Abs. 2 S. 1 UrhG gestützt werden. Die Forschungsausnahme nach § 60d UrhG wird mangels Bezugs zu einem konkreten Forschungsprojekt vorliegend nicht in Anspruch genommen werden können. Für eine Berufung auf § 44b UrhG sind die schutzfähigen Bestandteile der Aktenauszüge auf etwaige Vorbehalte (gegebenenfalls automatisiert) zu prüfen, die sodann zu berücksichtigen sind.

Einzelne Werkausschnitte, die während des Trainingsprozesses oder der Ausführung eines LLM verarbeitet werden, sind nach vorzugswürdiger Auffassung mangels Wahrnehmbarkeit bereits nicht als urheberrechtliche Vervielfältigungen zu behandeln. Sofern man dieser Ansicht nicht folgt, sind die betreffenden Vervielfältigungen ebenfalls als Text und Data Mining über § 44b UrhG zulässig, indem sie die Ausgangsbasis für die weitere Analyse auf dem Weg zu einer Ausgabe im konkreten Fall darstellen. Entsprechende Überlegungen gelten für Bearbeitungen nach § 23 Abs. 3 UrhG.

Eine Ablehnung der Wahrnehmbarkeit setzt insbesondere voraus, dass nicht einzelne Werkbestandteile über die Ausgabe des KI-Systems vervielfältigt werden. Solche Vervielfältigungen sind daher durch geeignete Maßnahmen zu verhindern, die analog zu den datenschutzrechtlich gebotenen technischen und organisatorischen Maßnahmen neben Privacy-Konzepten auf Trainingsebene auch Filtermaßnahmen, einen entsprechenden System Prompt und Benutzungsanweisungen sowie gegebenenfalls vertragliche Benutzungseinschränkungen umfassen. Zudem wirkt sich die organisatorische und institutionelle Einbindung zugunsten der Zulässigkeit aus.⁵²⁴ Sollten unter Verstoß gegen diese umfangreichen Maßnahmen durch eine Prompt Injection oder vergleichbare Angriffe gezielt Werkbestandteile im Rahmen der Ausgabe extrahiert werden, kommt im Regelfall nur noch der

524 S. schon oben unter DV.4.c.bb.

jeweilige Anwender als passivlegitimierter Täter dieser Urheberrechtsverletzung in Betracht.

III. Besonderheiten bei der Veröffentlichung des Sprachmodells

Die Veröffentlichung des Sprachmodells kann durch die Eröffnung eines öffentlichen Zugriffs mittels eines KI-Systems, durch die Veröffentlichung auch der Gewichte des Sprachmodells oder durch die Veröffentlichung auch des Trainingskorpus erfolgen. Analog zur datenschutzrechtlichen Betrachtung⁵²⁵ ergeben sich aufsteigend in der dort beschriebenen Reihenfolge (i.e. KI-System, Sprachmodell und Trainingskorpus) verschiedene urheberrechtliche Implikationen.

Die Bereitstellung eines öffentlich zugänglichen KI-Systems ist urheberrechtlich grundsätzlich vertretbar. Zwar erhöht diese Form der Veröffentlichung das Risiko missbräuchlicher Eingabeaufforderungen durch den nicht beschränkten Nutzerkreis bereits nicht unerheblich; zugleich bleiben aber System Prompts und Ausgabefilter erhalten und es dürfte mangels Kennungen schwerfallen, auf urheberrechtlich geschützte Ausdrucksweisen aufzubauen (z.B. durch die Aufforderung: „Schreibe einen Satzesatz im Stil von x“).

Falls es im (seltenen) Einzelfall zu einer Urheberrechtsverletzung kommt, bleibt eine Passivlegitimation der Stelle denkbar, die das KI-System öffentlich zur Verfügung stellt.⁵²⁶ Umfangreiche Vorsorgemaßnahmen können allerdings dieser Passivlegitimation in Form einer Täterschaft, Teilnahme oder sonstiger Haftung aufgrund von Verkehrspflichtverletzungen die Grundlage entziehen.

Die Veröffentlichung auch der Gewichte oder gar des Trainingskorpus steigert das Risiko einer Urheberrechtsverletzung erheblich. In diesem Fall entfallen weitere Schutzmechanismen, wie z.B. die Filtermaßnahmen und der System Prompt. Eine solche Veröffentlichung führt daher regelmäßig zu Urheberrechtsverletzungen, für die eine Passivlegitimation der Projektanbieter denkbar ist. Etwas anderes würde nur gelten, wenn sich die berechtigten Interessen an einer Open Source-Veröffentlichung in den urheberrechtlichen Schrankenbestimmungen niederschlagen würden. Zu-

525 S. oben unter D.IV.3.

526 S. zur Haftung einer Plattform bei der Bereitstellung von geschützten Trainingsdaten *Siglmüller/Gassner*, RDi 2023, 124 (Rn. 32).

mindest auf geschriebene Schrankenbestimmungen wird sich die erhöhte Transparenz durch Open Source, soweit ersichtlich, nicht stützen können.

IV. Exkurs: Eigene Forschungszwecke der ausführenden Stellen

Sofern im hypothetischen Szenario⁵²⁷ die ausführenden Stellen eigene Forschungszwecke verfolgen würden, blieben die grundlegenden urheberrechtlichen Erwägungen mit Blick auf abgeleitete Textformate und § 44b UrhG grundsätzlich unverändert. Unter der Prämisse der Verfolgung eigener Forschungszwecke liegt allerdings eine Anwendbarkeit des Forschungsprivilegs für Text und Data Mining aus § 60d UrhG näher. Dieser Prämisse mit besonderer Relevanz für die Zusammenarbeit von Praxis und Wissenschaft soll nachfolgend ebenfalls mit Blick auf die urheberrechtlichen Implikationen in Form eines Exkurses nachgegangen werden.

1. Vervielfältigungen im Rahmen des GSJ-Projekts

Abhängig von den Einzelheiten der Verwendung der Vervielfältigungen für das Training des Sprachmodells bleibt es bei dem primären Nichtforschungszweck in Form der Entwicklung eines in der Justiz einsetzbaren Sprachmodells und der Nichtanwendbarkeit des § 60d UrhG (s. oben unter E.II.2.c.dd). Die Forschungszwecke könnten aber im Einzelfall durchaus auch überwiegen sowie die Anwendbarkeit des § 60d UrhG für Vervielfältigungen zu Zwecken des Text und Data Mining eröffnen, womit es nicht auf erklärte Vorbehalte nach § 44b UrhG ankommen würde.

Die Anwendbarkeit des § 60d UrhG dürfte insbesondere für Vervielfältigungen anzunehmen sein, die nicht zwingend und unmittelbar über das Text und Data Mining in das Endprodukt eines GSJ-Sprachmodells für die Justiz münden, sondern lediglich Zwischenschritte für die Produktentwicklung, aber auch die (Grundlagen-)Forschung darstellen (z.B. Vervielfältigungen für das Training eines kleinen Testmodells, das selbst nicht in der Justiz operativ zum Einsatz gebracht werden soll).

⁵²⁷ S. zu den datenschutzrechtlichen Implikationen unter D.VII.1.

2. Vervielfältigungen über das GSJ-Projekt hinaus

Vervielfältigungen der ausführenden Stellen zu Zwecken des Text und Data Mining, die von vornherein über die Entwicklung des GSJ-Sprachmodells hinausgehen und der eigenen (Grundlagen-)Forschung der ausführenden Stellen dienen, können grundsätzlich ebenfalls auf § 60d UrhG gestützt werden.

V. Schlussbetrachtung aus urheberrechtlicher Perspektive

Das Urheberrecht erfasst das KI-Training insbesondere über Vervielfältigungen im Vorfeld des KI-Trainings. Diese Vervielfältigungen können im Kontext des GSJ-Projekts gestützt werden auf § 44b UrhG bzw. § 44a UrhG, soweit die Vervielfältigungen temporär als integraler Bestandteil des automatisierten Prozesses angelegt werden. Das GSJ-Sprachmodell selbst enthält (nach allerdings umstrittener Auffassung) grundsätzlich keine Vervielfältigungen.

Vervielfältigungen im Rahmen der Ausgabe sind durch angemessene urheberrechtliche Schutzmaßnahmen zu verhindern, die im Grundsatz den datenschutzrechtlich erforderlichen Schutzmaßnahmen entsprechen. Derartige urheberrechtliche Schutzmaßnahmen verlieren an Wirksamkeit bzw. entfallen sogar, wenn ein KI-System oder das gesamte trainierte Sprachmodell der Öffentlichkeit zugänglich gemacht werden. Eine Veröffentlichung des Sprachmodells ist aus urheberrechtlicher Sicht daher bedenklich.

Daraus lassen sich die folgenden Anforderungen und hierauf bezogenen Handlungsempfehlungen ableiten.

1. Anforderungen an die Zusammenstellung der Trainingsdaten

Durch den Einsatz abgeleiteter Textformate und das KI-Training an der Quelle sind Vervielfältigungen möglichst zu vermeiden.

Wenn und soweit sich Vervielfältigungen nicht vermeiden lassen, sind die folgenden Einsatzbedingungen mit Blick auf die Aktenauszüge zu beachten:

- Sichtung und Clusterung der typischen Aktenbestandteile im Hinblick auf gegebenenfalls auszusortierende Werkkategorien;

- Aussortierung und gegebenenfalls Prüfung von Werken mit besonders hohem urheberrechtlichen Schutzstandard (z.B. umfangreiche, vor allem textbasierte Gutachten);
- Automatisierte Löschung von notwendigen, zwischengespeicherten Kopien (z.B. zur Übermittlung an einen Trainingsserver);
- Prüfung der Aktenauszüge auf Rechtsvorbehalte i.S.d. § 44b Abs. 3 UrhG (z.B. „Kein Text oder Data Mining“, „Kein Text und Data Mining“, „Kein TDM“, „Kein KI-Training“, „Kein LLM-Training“, gegebenenfalls auch „Alle Rechte vorbehalten“ oder „Vervielfältigung vorbehalten“) und Aussortierung der entsprechenden Inhalte.

2. Anforderungen an das KI-System und dessen Einsatz

Im Hinblick auf den Einsatz des KI-Systems sind weitere Bedingungen zu beachten, die zugleich auch auf die Frage zurückwirken, ob das KI-Modell Vervielfältigungen enthält. Diese weiteren Bedingungen stimmen teilweise mit den datenschutzrechtlich erforderlichen Maßnahmen überein (siehe unter D.VIII.4: insbesondere Login-Mechanismus, Vorgabe der Eingabeaufforderung).

Darüber hinaus sind die notwendigen Maßnahmen vor allem im Fall einer frei wählbaren Eingabeaufforderung auch auf die Anforderungen des Urheberrechts auszurichten:

- Gegebenenfalls Einrichtung eines System Prompts, wonach der Stil ausgewählter Dokumentenkategorien aus den Aktenauszügen nicht nachgeahmt werden darf (z.B. Verhinderung von Prompts, einen Schriftsatz im Stil von Aktenauszügen aus einem bestimmten Themengebiet des Verkehrsrechts zu schreiben, der maßgeblich von einem bestimmten Rechtsanwalt dominiert wird);
- Gegebenenfalls Implementierung von Mechanismen zur Prüfung der Ausgabe (z.B. ein separates Sprachmodell zur Überprüfung, ob die Ausgabe vor allem einen konkreten Stil initiiert bzw. imitieren soll).

In Ansehung der bereits datenschutzrechtlich gebotenen Entfernung von Kennungen dürfte es ohnehin grundsätzlich nicht möglich sein, den Stil einer einzelnen Person durch eine entsprechende Eingabeaufforderung gezielt zu imitieren.

3. Anforderungen an die Veröffentlichung des Sprachmodells

Von einer Veröffentlichung des Sprachmodells gehen für das Urheberrecht weniger und geringere urheberrechtliche Implikationen aus als es für das Datenschutzrecht der Fall ist.

Die Veröffentlichung in der Form eines öffentlich zugänglichen KI-Systems erhöht geringfügig die Wahrscheinlichkeit von Urheberrechtsverletzungen wegen des unbeschränkten Nutzerkreises. In Abhängigkeit davon, ob die Eingabeaufforderung frei gewählt werden kann, sind gezielte urheberrechtsverletzende Ausgaben gegebenenfalls unwahrscheinlicher. Dieser Befund gilt insbesondere in Ansehung der fehlenden Kennungen, sodass nicht ohne Weiteres der Bezug zu dem gegebenenfalls urheberrechtlich geschützten Stil einer Person hergestellt werden kann.

Die Veröffentlichung der Gewichte oder gar des Trainingsdatensatzes machen eine Urheberrechtsverletzung allerdings erheblich wahrscheinlicher, da insoweit urheberrechtsrelevante Werkbestandteile mit gewisser Wahrscheinlichkeit extrahiert werden können. In diesen Fällen ist auch eine Passivlegitimation und damit eine Haftung der Ministerien denkbar, sodass dem Grunde nach von einer Veröffentlichung des Modells (Open Source) abzuraten ist.

