
Pitfalls in the Field – Rechtliche Herausforderungen beim Einsatz von ML in der Agrosystemforschung

Constantin Breß ^{1,*}, Lea Sophie Singson ¹

¹ FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Karlsruhe, Germany

* Kontakt: constantin.bress@fiz-karlsruhe.de

Beim Einsatz von Machine Learning in der Agrosystemforschung zeigen sich diverse rechtliche „Pitfalls“, welche hier vorgestellt werden. Zusätzlich wird eine Lösungsmöglichkeit präsentiert.

Die rechtlichen Probleme lassen sich auf verschiedene „Lebensphasen“ eines ML-Modells aufteilen: Training, die Existenz des Modells selbst und sein schlussendlicher Einsatz. Rechtliche Probleme ergeben sich vor allem im Zusammenhang mit dem Training: Daten können rechtlich geschützt sein. Gleichzeitig sieht das Recht gewisse Ausnahmen vor, welche fruchtbar gemacht werden können.

Ein zusätzliches und noch im Wachstum befindliches Feld ist der rechtliche Schutz der beim Training entstehenden Modelle. Bei deren Schaffung bedarf es nicht unerheblicher menschlicher Leistung, sodass die Beteiligten ein Interesse daran haben, ihre Urheberschaft anerkannt zu bekommen und eventuell auch daran, das weitere Schicksal der Modelle zu beeinflussen. Rechtlich bestehen dabei noch massive Unsicherheiten.

Den rechtlichen Problemen gemeinsam ist ihre hohe Komplexität, die sich teilweise erst im Laufe des Vorhabens zeigt. Daher können Rechtsfragen vor Projektbeginn nicht erschöpfend abgehandelt werden, sondern es ist eine rechtliche Begleitung notwendig, welche zudem auf hohem Niveau arbeitet – das auch aufgrund der Herausforderungen durch das neue EU-Datenrecht. FAIRagro setzt dabei auf das Modell des Data Steward Service Center (DSSC). Dieses unterstützt Forschende aus der Agrosystemcommunity bei auftretenden Problemen in allen Bereichen des Forschungsdatenmanagements – inklusive

Publiziert in: Vincent Heuveline, Philipp Kling, Florian Heuschkel, Sophie G. Habinger und Cora F. Krömer (Hrsg.): E-Science-Tage 2025. Research Data Management: Challenges in a Changing World. Heidelberg: heiBOOKS, 2025. DOI: <https://doi.org/10.11588/heibooks.1652.c23925> (CC BY-SA 4.0).

rechtlichen und ethischen Themen. Dieses Modell kann auch Vorbild für andere Forschungsdatencommunities sein.

Keywords: Machine Learning, Data Law, Copyright Law, Data Protection Law, Artificial Intelligence Law

1 Einleitung

Künstliche Intelligenz (KI) sorgt nicht nur in Techniker:innenkreisen, sondern auch in der Rechtswissenschaft und -praxis für Diskussionen. Dies nicht zuletzt auch aufgrund der KI-Verordnung der EU¹, welche am 2. Februar 2025 auch schon teilweise Geltung erlangte (Art. 113). Diese Rechtsfragen stellen sich darüber hinaus nicht nur allgemein, sondern sind teilweise auch sehr domain-spezifisch und einzelfallbezogen.

Dieser Beitrag soll einen Überblick über die verschiedenen rechtlichen Probleme im Zusammenhang mit künstlicher Intelligenz innerhalb und außerhalb der KI-Verordnung im Kontext der Agrarforschung geben. Dafür wird teilweise vereinfachend zunächst das Verständnis von KI dargestellt, welches diesem Paper zu Grunde liegt. Da sich die rechtlichen Fragestellungen je nach Phase in Erstellung und Nutzung eines Modells unterscheiden, orientiert sich die Darstellung an diesen Phasen. Daran anschließend folgt ein Überblick über das Datenrecht, bevor auf die eigentlichen rechtlichen Problemstellungen im Zusammenhang mit künstlicher Intelligenz eingegangen wird. Der Untersuchung wird die Rechtslage in Deutschland zu Grunde gelegt und orientiert sich an den Fragestellungen aus dem NFDI-Projekt „FAIRagro“. Das Geschäftsgeheimnisrecht wird bei diesem Beitrag weitestgehend nicht behandelt.

2 Künstliche Intelligenz

KI oder Machine Learning wird hier in Anlehnung an die Definition der KI-Verordnung für ein KI-System verstanden. In Art. 3 Nr. 1 KI-VO ist ein *KI-System* definiert als

ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Emp-

¹ Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz [...] (Verordnung über künstliche Intelligenz), ABl. L v. 12. Jun. 2024.

fehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können[.]

Indem die KI-Verordnung sich auf ein System bezieht, behandelt sie v.a. ein konkretes Modell, aber auch seine Einbettung in eine bestimmte Infrastruktur. Dieses Paper beschäftigt sich mit den einzelnen Modellen, sodass der Begriff der KI hier enger verstanden wird als in der KI-Verordnung. Dennoch erfasst der Begriff sowohl „einfache“ Modelle wie Klassifikatoren als auch *Large Language Models*. Ein solches Modell lässt sich wiederum (mindestens) in zwei rechtlich relevante Elemente aufspalten: Die grundsätzliche Architektur des Modells und die Weights, die innerhalb dieser Architektur eingebunden werden (so auch Schumacher 2023, S. 36; Übersicht bei Glauner 2024; synonym zu Architektur wird auch „Topologie“ verwendet, etwa bei Schröler und Kuß 2022, Rn. 40–42).

Weights bezeichnet vereinfacht die Gewichtung der verschiedenen Elemente innerhalb der Architektur eines bestimmten Modells. Dadurch entsteht nach einem mehrschrittigen Prozess die Ausgabe des Modells. Die jeweiligen Weights hängen von den zum Training verwendeten Daten ab (weiterführend Schumacher 2023, S. 36 f.). Das Verhältnis von Architektur und Weights ist ungefähr so wie das Verhältnis zwischen einem Apfelkern und dem daraus wachsenden Baum: Der Kern gibt prinzipiell vor, wie der Baum aussehen soll; die konkrete Ausformung hängt aber von den Umgebungsfaktoren wie etwa der Bodenbeschaffenheit ab.

3 Lebensphasen einer KI

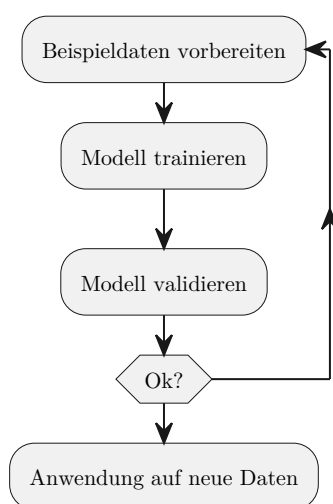


Abbildung 1: Lebenszyklus eines Machine Learning-Modells.

Abbildung 1 zeigt exemplarisch den Lebenszyklus eines Machine-Learning-Modells (in Anlehnung an Kleinkopf 2022, S. 48–62; Lee, Cooper und Grimmelmann 2024, S. 49–52). Dieser wird der weiteren rechtlichen Analyse zu Grunde gelegt.

4 Mögliche Schutzrechte an Daten

Als Rechte an Daten kommen vor allem das Urheberrecht und verwandte Rechte in Betracht. Je nach Zusammenstellung der Daten kann auch das Datenschutzrecht relevant sein.

4.1 Urheberrecht

Das Urheberrechtsgesetz (UrhG) schützt in erster Linie Werke, also konkrete Ausformungen einer Idee. Dieser Schutz setzt mindestens eine persönliche geistige Schöpfung voraus, das Werk muss die sogenannte *Schöpfungshöhe* erreichen (§ 2 Abs. 2 UrhG, für Datenbankwerke § 4 UrhG). Eine persönliche Schöpfung liegt vor, wenn das Werk auf einer menschlich-gestalterischen Tätigkeit beruht (Leistner 2020, Rn. 38; Schulze 2022, Rn.8). Soweit eine Maschine als Hilfsmittel herangezogen wird, kann eine solche Schöpfung gegeben sein, wenn ein Mensch darauf gestaltenden Einfluss hat (Leistner 2020, Rn. 40).

Zentrale Ergebnisse in der Agrarforschung sind Daten(-sammlungen). Zumindest Einzeldaten werden nach (fachspezifischen) Vorgaben erhoben, sodass der forschenden Person kein hinreichender Gestaltungsspielraum zukommt. Damit fehlt die nach § 2 Abs. 2 bzw. § 4 UrhG notwendige Schöpfungshöhe und die Daten sind folglich nicht urheberrechtlich geschützt (Bomhard und Siglmüller 2024, Rn. 11; Schulze 2022, Rn. 26). Für die verschiedenen Bestandteile eines Datensatzes kann sich aber eine abweichende Bewertung ergeben (ausführlich mit Beispielen siehe Hartmann 2022). Das kann auch etwa für eine vorgenommene Annotierung von Daten der Fall sein (Hacker 2020, S. 1027).

Dasselbe gilt in der Regel auch, wenn Daten aus der Verwendung von Computerprogrammen resultieren. Ein vom Schutz für Computerprogramme abgeleiteter Schutz für Daten ergibt sich nur, wenn die Daten im Programm selbst angelegt sind (Schulze 2022, Rn. 14).

Urheberrechtlicher Schutz kann sich in der Agrosystemforschung vor allem bei einer Zusammenstellung mehrerer Datensätze ergeben, sofern die Zusammenstellung Ergebnis eines kreativen, originellen Schaffens war. Dann würde sich ein Datenbankwerk nach § 4 Abs. 2 UrhG ergeben; allerdings sind dabei nicht die Daten als solche, sondern nur die Zusammenstellung als Datenbank geschützt (Weidert und Bug 2023, Rn. 34).

Ob die Zusammenstellung der Datensätze urheberrechtlichen Schutz erlangt, ist aber eine Frage des jeweiligen Einzelfalls.

4.2 Datenbankschutz

Erfolgt die Zusammenstellung des Datensatzes nicht in kreativer, origineller Weise, kann der Datensatz dennoch als Datenbank nach §§ 87a–87e UrhG geschützt sein. Im Unterschied zum Datenbankwerk handelt es sich dabei um ein Leistungsschutzrecht, da es nicht an die schöpferische Tätigkeit einer Person anknüpft, sondern an investierter Leistung. Daher steht dieses Recht auch nicht der:dem Schöpfer:in des Werkes zu, sondern der:dem Datenbankhersteller:in. Da dieses Recht im Vergleich zum sonstigen Urheberrecht andere Voraussetzungen aufweist, handelt es sich dabei um ein Recht eigener Art, es wird daher auch als *sui-generis*-Recht bezeichnet. Ein weiterer Unterschied zum Urheberrecht besteht darin, dass dem:der Datenbankhersteller:in nur das ausschließliche Recht zur Vervielfältigung *wesentlicher* Teile der Datenbank zukommt. Im mkehrschluss kann damit jede Person unwesentliche Teile vervielfältigen, zusätzlich zu den Ausnahmen in § 87c UrhG.

Ein Datensatz ist als Datenbank i.S. des § 87a UrhG geschützt, wenn er sechs Bedingungen erfüllt: Es muss sich um eine Sammlung von Werken, Daten oder anderen unabhängigen Elementen handeln, diese müssen berücksichtigungsfähig für das Recht aus § 87a UrhG sein, systematisch oder methodisch angeordnet, einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sein und schließlich muss die Beschaffung, Überprüfung oder Darstellung eine nach Art oder Umfang wesentliche Investition erfordern.

Sammlung iS des § 87a UrhG ist die Zusammenstellung einer Vielzahl von Elementen (Vohwinkel 2024, Rn. 28 mwN).

Unabhängige Elemente iS des § 87a UrhG liegen vor, wenn sie aus der Datenbank herausgelöst werden können und ihnen trotzdem ein hinreichender Informationswert verbleibt.² Ob ein solcher Informationswert verbleibt, ist aus Perspektive eines interessierten Dritten zu beurteilen.³ Damit reicht es aus, wenn irgendwer etwas mit den herausgelösten Elementen anfangen kann. Zwar muss dies nicht unbedingt auf ein einzelnes Datum zutreffen (wenn es nicht schon den Begriff des Datums iS des UrhG erfüllt), vielmehr kann das auch bei Datenabschnitten hinreichender Länge der Fall sein (Hacker 2020, S. 1029 f.).

Die Elemente müssen überhaupt berücksichtigungsfähig sein, sie dürfen damit keiner Bereichsausnahme unterliegen. Eine solche ist Art. 43 der Daten-Verordnung (D-VO)⁴ (Richter 2024, Rn. 19 f.). Danach sind Daten vom *sui-generis*-Schutz nach § 87a ff. UrhG ausgenommen, wenn sie der Datenverordnung unterfallen und durch vernetzte Produkte

² *EuGH*, Urt. v. 29. Okt. 2015 (C-490/14) – Verlag Esterbauer, Rn. 17; *EuGH*, Urt. v. 9. Nov. 2004 (C-444/02) – Fixtures Marketing, Rn. 29–32.

³ *EuGH*, Urt. v. 29. Okt. 2015 (C-490/14) – Verlag Esterbauer, Rn. 27.

⁴ Verordnung (EU) 2023/2854 des Europäischen Parlaments und des Rates vom 13. Dezember 2023 über harmonisierte Vorschriften für einen fairen Datenzugang und eine faire Datennutzung sowie zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie (EU) 2020/1828 (Datenverordnung), ABl. L v. 22. Dez. 2024.

(Art. 2 Nr. 5 D-VO) oder verbundene Dienstleistungen (Art. 2 Nr. 6 D-VO) erlangt oder erzeugt wurden. Aus ErwG 15 S. 8–10 D-VO ergibt sich, dass Sensordaten, auch in aufbereiteter Form, nicht nach §§ 87a ff. UrhG als Teil einer Datenbank geschützt sein können (Richter 2024, Rn. 17; Antoine 2024, Rn. 25 f.). Geschützt werden können aber Daten, welche aus solchen „Rohdaten“ abgeleitet wurden, etwa durch weitere Auswertung oder „Sensorfusion“ (ErwG 15 S. 13 f. D-VO). Daran werden die meisten Daten in Kontext der Agrarforschung wohl scheitern, sodass hier das *sui-generis*-Recht nicht greift.

Die Anordnung ist systematisch, wenn sich die Sammlung selbst an einem System, einer Klassifizierung oder einem Ordnungsschema orientiert (Dreier 2025b, Rn. 17). Sie ist methodisch, wenn die Sammlung selber einer bestimmten ordnenden Handlungsanweisung oder Plan folgt (Dreier 2025b, Rn. 17). Bei elektronischen Datenbanken kommt es vor allem auf die Durchsuchbarkeit an, wofür eine Aufbereitung der Daten mit Ordnungskriterien notwendig ist (Dreier 2025d, Rn. 7).

Ein Zugang zu den einzelnen Elementen wird dann als möglich angesehen, wenn die Sammlung ein geeignetes Zugangssystem umfasst, etwa ein Abfragesystem oder einen Index, welcher einen Einzelzugriff ermöglicht (Leistner 2020, Rn. 52). Ein solcher Index kann z. B. ein Inhaltsverzeichnis sein.

Die wesentliche Investition i.S. des § 87a UrhG muss sich erstens auf Beschaffung, Überprüfung oder Darstellung der Elemente beziehen. Damit kommt nicht jegliche Investition im Zusammenhang mit der Datenbank in Betracht. Nicht erfasst ist insbesondere die Generierung von Daten, sondern erst die nachgeordneten Schritte, da eben die Datenbank und nicht die Daten selbst geschützt wird (Zech 2015, S. 142 f.).⁵

Zu diesen nachgeordneten Schritten zählt etwa die Sammlung bereits anderweitig vorhandener Daten.⁶ In Bezug auf die Aufzeichnung von Umweltdaten kann auf Grundlage der vom BGH angeführten Differenzierung sowohl für als auch gegen eine Berücksichtigungsfähigkeit argumentiert werden. Denn diese können einerseits als bereits vorhanden angesehen werden (Wiebe 2017, S. 341), andererseits auch als abgeleitet von natürlichen Phänomenen und damit erst neu erzeugt. Im Ergebnis ist das aber aufgrund der Bereichsausnahme in Art. 43 D-VO meist bedeutungslos.

Zweitens muss eine Investition vorliegen und diese wesentlich sein. Investition sind nicht nur finanzielle Aufwendungen, sondern jeder Einsatz von Zeit, Arbeit und Energie.⁷ Die Investition ist wesentlich, wenn sie nicht ganz unbedeutend ist und nicht von jedermann leicht erbracht werden könnte.⁸ Realistisch kann das allerdings nur im Einzelfall bewertet werden.

⁵ BGH, Urt. v. 1. Dez. 2010 (I ZR 196/08) – Zweite Zahnarztmeinung II, Rn. 19 mwN auch zur Rechtsprechung des EuGH.

⁶ BGH, Urt. v. 25. Mär. 2010 (I ZR 47/08) – Autobahnmaut, Rn. 19.

⁷ EuGH, Urt. v. 9. Nov. 2004 (C-444/02) – Fixtures Marketing, Rn. 44.

⁸ BGH, Urt. v. 1. Dez. 2010 (I ZR 196/08) – Zweite Zahnarztmeinung II, Rn. 23.

4.3 Lichtbildschutz

Auch ohne die für den Werkschutz nach UrhG eigentlich notwendige Schöpfungshöhe (dazu Kapitel 4.1) können Lichtbilder nach § 72 UrhG geschützt sein. Voraussetzung dafür ist einerseits die Schaffung eines „Urbilds“ durch Abbildung von etwas. Anders formuliert: Nicht ausreichend ist die bloße Reproduktion, etwa durch Kopie (Vogel 2020, Rn. 33; Lauber-Rönsberg 2024, Rn. 15-17). Andererseits muss die Erzeugung eines Mindestmaßes an Fertigkeiten bedürfen, beispielsweise durch die Bedienung der jeweiligen Vorrichtung (Vogel 2020, Rn. 34). Automatisierte Aufnahmen – beispielsweise durch Feldroboter – fallen damit nicht per se aus dem Schutzbereich heraus (Lauber-Rönsberg 2024, Rn. 34). Der Schutzzumfang entspricht dann weitestgehend dem „normalen“ Urheberrecht (Lauber-Rönsberg 2024, Rn. 27).

In der Agrosystemforschung kann sich ein solcher Schutz v.a. bei Ablichtungen zwecks Bestimmung von Phänotypen ergeben. Ähnliches gilt für Mittel der Fernerkundung, seien es Aufnahmen per Satellit, Flugzeug oder Drohne (Katzenberger 1989, S. 118 f.).⁹ Dabei kommt es nicht darauf an, ob ein Lichtbild im engeren Sinne entsteht, sondern nur, ob die „Aufnahme“ durch Aufzeichnung strahlender Energie erfolgt (Vogel 2020, Rn. 25 f.), sodass z. B. auch Radaraufnahmen erfasst sind.

4.4 Datenschutzrecht

Das Datenschutzrecht greift, wenn ein personenbezogenes Datum i.S. von Art. 4 Nr. 1 Datenschutz-Grundverordnung (DSGVO) vorliegt. Dafür braucht es eine Information, die sich auf eine natürliche Person bezieht. Entscheidend ist dabei, ob diese Person identifizierbar ist. Das ist dann der Fall, wenn sich eine Person zuverlässig aus einer Gruppe absondern lässt (Art.-29-Gruppe 2007, S. 14–24; Europäischer Datenschutzausschuss 2025, Rn. 16–25). Dafür sind Informationen zu berücksichtigen, welche die verarbeitende Stelle wahrscheinlich und vernünftigerweise nutzen würde (und könnte), um eine solche Absonderung vorzunehmen.¹⁰

Das Datenschutzrecht wird im Kontext von FAIRagro neben der Durchführung von Interviews vor allem im Falle von georeferenzierten Daten relevant, oder wenn Personen bei abbildenden Verfahren als „Beifang“ aufgenommen werden. Dafür kommt es aber wiederum auf den jeweiligen Einzelfall an. Auch ist zu betonen, dass auch bei einer Einordnung als personenbezogenes Datum die Verarbeitung der Daten möglich bleibt und dass die DSGVO die Verarbeitung für Zwecke der Forschung (erheblich) erleichtert.

⁹ *LG Berlin*, Urt. v. 30. Mai 1989 (16 O 33/89) – Satellitenfoto, GRUR 1990, S. 270.

¹⁰ *EuGH*, Urt. v. 9. Nov. 2023 (C-319/22) – Gesamtverband Autoteile-Handel, Rn. 45; *EuGH*, Urt. v. 19. Okt. 2016 (C-582/14) – Breyer, Rn. 44–48.

5 Rechtliche Fallstricke im Lebenszyklus

Im Folgenden werden die verschiedenen Fallstricke im Lebenszyklus eines Machine-Learning-Modells dargestellt. Dabei werden Urheberrecht, Datenbank- und Lichtbildschutz gemeinsam abgehandelt, da sie sich im Wesentlichen gleichen (§ 87b UrhG einerseits, § 72 Abs. 1 UrhG andererseits). Sofern sich Abweichungen ergeben sollten, werden diese kenntlich gemacht. Ein eigener Abschnitt ist der Frage gewidmet, ob das resultierende Modell eine Kopie der verwendeten Daten aus rechtlicher Sicht darstellt. Gleiches gilt für die Frage, ob sonst irgendwelche Rechte an dem Modell bestehen.

Da die verschiedenen Fallstricke an den Umgang mit den verschiedenen Datenarten anknüpfen, ergibt sich kein Unterschied aufgrund der Weise, auf die auf das Modell zugegriffen wird. Die Ausführungen gelten also sowohl für lokale als auch gehostete Modelle und auch für das Finetuning von bestehenden Modellen.

5.1 Urheberrecht

Für das Urheberrecht sind nur bestimmte Handlungen relevant, welche mit den Phasen im Lebenszyklus einhergehen. Diese Handlungen sind einerseits die Vervielfältigung, also die Kopie, möglicherweise urheberrechtlich geschützter Werke bei Datensammlung, Training und ggf. Validierung. Andererseits die Verbreitung, falls der verwendete Korpus weitergegeben werden sollte. Mit der Verbreitung verwandt ist die öffentliche Zugänglichmachung; diese betrifft grob umschrieben die Weitergabe über „neue“ Medien. Sollte eine dieser Handlungen vorgenommen werden, kann die:der Rechteinhaber:in Unterlassung (§ 97 Abs. 1 UrhG) und unter Umständen auch Schadensersatz (§ 97 Abs. 2 UrhG) verlangen. Dabei ist aber zu beachten, dass an den meisten Daten kein Urheber- oder verwandtes Recht besteht (s. Kapitel 4).

Vervielfältigung

Eine Vervielfältigung kann auf drei Weisen gerechtfertigt sein: Erstens durch eine Lizenz (etwa CC-BY 4.0), zweitens durch eine (konkludente) Einwilligung oder drittens durch gesetzliche Schranken. Bei der Lizenz kommt es auf die jeweiligen Bedingungen an.

Eine konkludente Einwilligung liegt dann vor, wenn die Zustimmung zu einer bestimmten Handlung nicht ausdrücklich erklärt wird, sich aus dem Verhalten der zustimmenden Person aber eindeutig und unmissverständlich ergibt, dass der Handlung zugestimmt wird. Dies kann beispielweise durch Zugänglichmachung des Werkes der:des Rechtein-

haber:in für bestimmte Nutzungen angenommen werden.¹¹ Auf eine Einwilligung wird für Text- und Datamining (TDM) aber meist nicht zurückgegriffen werden können. Denn Voraussetzung für die Einwilligung ist u. a., dass ein bestimmter Umgang mit einem Werk zu erwarten ist. Das ist für das Training von Modellen (bisher) nicht der Fall (Pesch und Böhme 2023a, S. 1007).

Falls diese beiden Möglichkeiten nicht in Betracht kommen, verbleiben die gesetzlichen Schranken. Gesetzliche Schranken ermöglichen eine Nutzung der geschützten Güter auch gegen den Willen der Rechteinhaber:innen. Sie haben zur Folge, dass bestimmte Nutzungen auch ohne Einwilligung des:der Urhebers:in erfolgen dürfen. Damit dies nicht zu einer extensiven Entwertung der Verwertungsrechte des:der Urhebers:in führt, sind den Schranken wiederum materielle Grenzen gesetzt (Kraetzig 2024, Rn. 46). Diese Grenzen ergeben sich etwa durch den *Dreistufentest*.

Danach sind Ausnahmen von den Verwertungsrechten der:des Urhebers:in nur gestattet, wenn sie (von Welser 2022, Rn. 22):

1. nur in bestimmten Sonderfällen angewandt werden,
2. die normale Verwertung des Werkes nicht beeinträchtigt wird und
3. die berechtigten Interessen des:der Urhebers:in nicht unzumutbar verletzt werden.

Dieser Dreistufentest wurde, zuletzt durch Art. 5 Abs. 5 der InfoSoc-RL¹², mehrfach verbindlich vorgeschrieben. Dennoch wurde er nicht explizit in nationales Recht aufgenommen, da die gesetzlich vorgeschriebenen Schranken an sich bereits die Voraussetzungen des Dreistufentests erfüllen (Kraetzig 2024, Rn. 45).

Im Folgenden sollen zwei der gesetzlichen Schranken hervorgehoben werden: § 44b und § 60d UrhG. Beide ermöglichen unter teilweise abweichenden Voraussetzungen TDM. Das Gesetz versteht darunter die Gewinnung von Informationen über einzelne Daten hinaus (Hofmann 2024, S. 170 f.). Damit ist auch das Training von KI-Modellen inkl. vorbereitender Handlungen erfasst, was sich auch mittelbar aus Art. 53 Abs. 1 lit. c KI-VO und deren ErwG 105 S. 3–7 ergibt (de la Durantaye 2024, Rn. 28; Pesch und Böhme 2023a, S. 1006).¹³

Dem Dreistufentest entsprechend ist TDM nicht allgemein zulässig, sondern an besondere Voraussetzungen geknüpft. Bei § 44b UrhG ist diese Voraussetzung die Flüchtigkeit der Vervielfältigung; bei § 60d UrhG ist die zusätzliche Voraussetzung die Beschränkung auf wissenschaftliche Forschungszwecke. Gemeinsame Voraussetzung beider Schranken ist

11 *BGH*, Urt. v. 11. Sept. 2024 (I ZR 140/23) – *Fototapete*, Rn. 36; *BGH*, Urt. v. 29. Apr. 2010 (I ZR 69/08) – *Vorschaubilder*, Rn. 86; *LG Hamburg*, Urt. v. 27. Sept. 2024 (310 O 227/23) – *LAION*, Rn. 86.

12 Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft, ABl. L 167/10 vom 22. Juni 2001.

13 So auch *LG Hamburg*, Urt. v. 27. Sept. 2024 (310 O 227/23) – *LAION*, Rn. 79.

aber zunächst, dass der Zugang zu den Daten rechtmäßig erfolgt (Bomhard 2024b, Rn. 15). Dieser kann etwa auf einer entsprechenden Lizenz beruhen (welche aber etwa kein TDM zulässt) oder auch auf einer Einwilligung der:des Berechtigten. Eine solche kann sich (konkulent) durch öffentliche Zugänglichmachung im Internet ergeben.¹⁴ Sofern ein Datensatz öffentlich abrufbar ist, kann also meist von einer entsprechenden Einwilligung ausgegangen werden.

§ 44b UrhG Daneben hat § 44b UrhG zwei weitere Voraussetzungen: Für die Daten darf kein *Nutzungsvorbehalt* existieren und die Vervielfältigungen sind zu *löschen*, wenn sie nicht mehr für das TDM benötigt werden.

Rechteinhaber:innen können Nutzungsvorbehalte für Daten bzw. Werke festlegen. Sofern die Daten online zugänglich sind, muss dieser Vorbehalt „maschinenlesbar“ sein, § 44b Abs. 3 UrhG. Welche Anforderungen im Detail daran zu stellen sind, ist umstritten. Vorgeschlagen werden etwa ein Eintrag in der robots.txt, eine eigene ai.txt oder etwa Metadaten-Erweiterungen auf Grundlage von ODRL¹⁵ (Hamann 2024). In der Rechtsprechung wurde bereits ein Vorbehalt in natürlicher Sprache als ausreichend maschinenlesbar angesehen.¹⁶

Ein einheitlicher Standard hat sich bisher nicht herausgebildet, weshalb etwa auch die Wartung und Aktualisierung entsprechender Vorbehalte praktische Probleme zeitigt (Hamann 2024, S. 164 f.). Möglicherweise erarbeitet aber das auf der KI-VO beruhende Amt für künstliche Intelligenz der EU einen Standard, der eine gewisse Verbreitung erfahren könnte (Hamann 2024, S. 166).

Die Vervielfältigungen sind zu löschen, wenn sie nicht mehr für den Zweck des TDM benötigt werden. Darunter kann zwar auch die Validierung des Ergebnisses fallen, eine beliebig lange Speicherung ist davon aber nicht gedeckt (Bomhard 2024a, Rn. 20–23).

§ 60d UrhG § 60d UrhG regelt die Vervielfältigungen für TDM für Forschungszwecke. Der Paragraph kennt im Gegensatz zu § 44b UrhG weder einen Nutzungsvorbehalt noch müssen die Kopien gelöscht werden, wenn sie nicht mehr benötigt werden. Dafür ist der Kreis der Berechtigten auf Forschungsorganisationen (§ 60d Abs. 2 UrhG), Kulturerbe-Einrichtungen (§ 60d Abs. 3 Nr. 1 UrhG) und einzelne Forschende (§ 60d Abs. 3 Nr. 2 UrhG) beschränkt. Wissenschaftliche Forschung i.S. des ist jede methodische und systematische Tätigkeit mit dem Ziel des Erkenntnisgewinns in nachprüfbarer Weise (Dreier 2025c, Rn. 4). Hintergrund ist der besondere Schutz der Wissenschaft in der Rechtsord-

14 BGH, Urt. v. 11. Sept. 2024 (I ZR 140/23) – *Fototapete*, Rn. 36; BGH, Urt. v. 29. Apr. 2010 (I ZR 69/08) – *Vorschaubilder*, Rn. 86; so auch LG Hamburg, Urt. v. 27. Sept. 2024 (310 O 227/23) – *LAION*, Rn. 86.

15 <https://w3c.github.io/tdm-reservation-protocol/spec/>; Besucht am 27. Mai 2025.

16 LG Hamburg, Urt. v. 27. Sept. 2024 (310 O 227/23) – *LAION*.

nung, etwa in Art. 5 Abs. 3 S. 1 GG oder Art. 13 GRCh, weshalb der Begriff tendenziell weit ausgelegt wird.

Verbindendes Merkmal der verschiedenen Berechtigten ist eine nicht-kommerzielle Ausrichtung. Das Merkmal *nicht-kommerziell* wird im UrhG nicht genauer definiert. Die Frage, ob nicht kommerzielle Forschung vorliegt, wird stets anhand der konkreten wissenschaftlichen Tätigkeit beurteilt, nicht dagegen anhand der Organisation und Finanzierung der Einrichtung, in der die Forschung erfolgt (Bullinger 2022, Rn. 16).¹⁷ Damit bleibt § 60d UrhG auch anwendbar, wenn kommerzielle Beteiligte Forschungshandlungen vornehmen oder von ihnen profitieren.¹⁸ Die Einschränkung auf nicht-kommerzielle Forschung bedeutet auch nicht, dass keine Gewinnerzielung oder Kostendeckung mit den Forschungsergebnissen erfolgen darf. Erstere ist dann erlaubt, wenn die Gewinne in die Forschung reinvestiert werden (Bomhard 2024b, Rn. 17). Auch sind Kooperationen mit Angehörigen der Wirtschaft in sog. Public-Private-Partnerships möglich. Dies regelt explizit § 60d Abs. 2 S. 3 UrhG. Dieser erlaubt derartige Kooperationen, solange kein bestimmender Einfluss durch die Wirtschaftsakteure auf die Forschungsorganisation ausgeübt wird und der Wirtschaftspartner keine bevorzugten Zugang zu den Forschungsergebnissen bekommt (Bomhard 2024b, Rn. 22).

Über die Vervielfältigung hinaus ermöglicht § 60d Abs. 4 UrhG auch noch die *öffentliche Zugänglichmachung*. Sie ist sowohl hinsichtlich des Empfängerkreises als auch zeitlich begrenzt. Erlaubt ist die Zugänglichmachung entweder einem bestimmten abgegrenzten Personenkreis für die gemeinsame Forschung (Nr. 1) oder einzelnen Dritten zur Überprüfung der Forschungsqualität (Nr. 2). Die Zugänglichmachung ist zu beenden, sobald die gemeinsame wissenschaftliche Forschung, beziehungsweise die wissenschaftliche Überprüfung, abgeschlossen ist.

Verbreitung

Eine rechtlich relevante Verbreitung kann sich durch Weitergabe einer Datensammlung oder Verlinkung auf Daten ergeben (Linksammlung). Genauso wie die Vervielfältigung bedarf die Verbreitung einer Erlaubnis.

Weitergabe der Daten: Wie auch bei der Kopie der Daten kommen für die Rechtfertigung der Datenweitergabe entweder entsprechende Lizenzbestimmungen, eine Einwilligung oder gesetzliche Schranken in Betracht. Von den behandelten Schranken ermöglicht allerdings nur § 60d Abs. 4 UrhG die Weitergabe der Daten und dies auch nur unter engen Bedingungen.

¹⁷ *LG Hamburg*, Urt. v. 27. Sept. 2024 (310 O 227/23) – *LAION*, Rn. 118.

¹⁸ *LG Hamburg*, Urt. v. 27. Sept. 2024 (310 O 227/23) – *LAION*, Rn. 118.

So muss die Weitergabe entweder der gemeinsamen Forschung dienen oder aber der Überprüfung der Ergebnisse eines Forschungsvorhabens. Danach ist der Zugang zu beenden. Unter dem Vorzeichen der gemeinsamen Forschung kann ein Datensatz aber nicht allgemein öffentlich zugänglich gemacht werden. Vielmehr ist nur die Weitergabe innerhalb eines bestimmten und abgegrenzten Personenkreises zulässig (Bomhard 2024b, Rn. 28).

Verlinkung der Daten: Ein Korpus kann nicht nur durch Zusammenstellung der Daten geschaffen werden, sondern auch durch Verlinkung der zugrundeliegenden Daten. Darin liegt auch eine Möglichkeit, Datensätze, die auf Grundlage von § 44b UrhG genutzt werden (sollen), dauerhaft abrufbar zu halten.

Urheberrechtlich relevant (und damit potentiell erlaubnispflichtig) werden solche Linksammlungen, wenn durch die Verlinkung eine (weitere) öffentliche Zugänglichmachung i.S. des § 19a UrhG erfolgt. Sofern Inhalte nicht eingebettet, sondern nur verlinkt werden, handelt es sich meist nicht um eine erlaubnispflichtige Handlung. Denn durch eine solche Verlinkung wird nach Ansicht der Rechtsprechung kein neues Publikum erreicht. Etwas anderes gilt, wenn die Verlinkung auf ein rechtswidrig veröffentlichtes Werk zeigt. Eine (erneute) öffentliche Zugänglichmachung liegt nach Ansicht des EuGH in diese Falle vor, wenn der verlinkenden Person zumindest bewusst sein könnte, dass der Inhalt rechtswidrig eingestellt wurde. Ein solches Bewusstsein wird dabei vermutet, wenn die Verlinkung mit Gewinnerzielungsabsicht erfolgte (Dreier 2025a, Rn. 6b). Eine solche ist im Rahmen der Agrarforschung meist nicht anzunehmen.

Datenbanken

Besonderheiten ergeben sich für Datenbanken: Bei diesen sind die Vervielfältigung, Verbreitung oder öffentliche Wiedergabe nur dann erlaubnispflichtig, wenn sie einen wesentlichen Teil einer Datenbank betreffen. Ob ein Teil wesentlich ist, kann sich nach Umfang und Art des Teils bestimmen. Die Wesentlichkeit nach Umfang bestimmt sich nach der Größe der Datenbank und dem Verhältnis des entnommenen Teils dazu (Vohwinkel 2025, Rn. 13). Jedenfalls handelt es sich bei einer Entnahme von nur 10 % nicht um einen wesentlichen Teil hinsichtlich des Umfangs (Vohwinkel 2025, Rn. 13).

5.2 Datenschutzrecht

Wenn überhaupt das Datenschutzrecht in Bezug auf die Daten greift (dazu Kapitel 4.4), kann ein Training möglicherweise auf Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO, also ein berechtigtes Interesse gestützt werden (Europäischer Datenschutzausschuss 2024b, Rn. 59–135). Dafür kommt es aber auf den jeweiligen Einzelfall an (Übersicht zu den relevanten Kriterien bei Europäischer Datenschutzausschuss 2024a; Schantz 2025, Rn. 125). Dabei sollte

darauf geachtet werden, die Auswirkungen auf betroffene Personen möglichst zu reduzieren. Als eine solche Maßnahme schlägt der EDSA unter anderem Pseudonymisierung, die Berücksichtigung des urheberrechtlichen Opt-Outs (dazu Kapitel 5.1) oder auch eine Risikoabschätzung dahingehend vor, ob Datenquellen möglicherweise personenbezogene Daten enthalten (Europäischer Datenschutzausschuss 2024b, Rn. 96–108).

Neben der Rechtmäßigkeit muss die Verarbeitung personenbezogener Daten weiteren Prinzipien gehorchen, welche in Art. 5 DSGVO näher beschrieben sind. Hervorzuheben ist dabei das Prinzip der Zweckbindung, welches sich auch auf das Training von Modellen auswirkt: Daten müssen für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nur dann weiterverarbeitet werden, wenn die Zwecke dieser Weiterverarbeitung kompatibel zu den Erhebungszwecken sind. Für die wissenschaftliche Forschung ist diese Zweckbindung gelockert (Art. 5 Abs. 1 lit. b HS. 2 Var. 1 DSGVO), wie auch andere Pflichten innerhalb der DSGVO. Allerdings gilt das nur, wenn die Zwecke der Verarbeitung nicht auch mit pseudonymisierten Daten erreicht werden können, außerdem müssen zusätzliche Maßnahmen getroffen werden, um die Daten zu schützen (Art. 89 DSGVO, ausführlich Weiss und Schuh 2024; Hüger 2024).

Ein kompatibler Zweck ersetzt dabei nach vorherrschender Auffassung nicht die Rechtsgrundlage für die Weiterverarbeitung; dabei kann es sich aber auch um die gleiche wie für die Erhebung handeln (Herbst 2024, Rn. 54). Idealerweise sollten allerdings keine personenbezogene Daten verarbeitet werden; sofern dies nicht ausgeschlossen ist, sollte das Risiko für die betroffenen Personen möglichst reduziert werden.

5.3 Modell als Kopie?

Sowohl aus urheber- als auch datenschutzrechtlicher Perspektive ist umstritten, ob ein Modell eine Kopie der zum Training verwendeten Daten darstellen kann. Urheberrechtlich ist das etwa bei Nutzung der Schranke in § 44b UrhG relevant (dazu Kapitel 5.1), da die Kopien nach Abschluss des TDM zu löschen sind, § 44b Abs. 2 S. 2 UrhG. Datenschutzrechtlich ist vor allem das Prinzip der Speicherbegrenzung zu beachten. Danach dürfen personenbezogene Daten nur so lange gespeichert werden, wie dies für die Verarbeitungszwecke notwendig ist, Art. 5 Abs. 1 lit. e DSGVO.

Verbindend ist dabei für beide Bereiche, wie weit die verwendeten Daten aus dem Modell extrahierbar sind. Entweder, weil so ein urheberrechtlich relevantes Werk extrahiert werden kann oder Informationen auf eine Person zurückgeführt werden können, diese also identifizierbar wird, wodurch die DSGVO anwendbar sein würde.

Dabei werden etwa Vergleiche mit (verlustbehafteter) Kompression angestellt, weshalb auch ein (generatives) Modell eine Kopie der verwendeten Daten darstellen würde (Pesch und Böhme 2023a, S. 1002–1005; ausführlich Dornis und Stober 2024, S. 74–80). Diese

Probleme ergeben sich nicht nur für generative Modelle, sondern auch für „klassische“ KI, da u. a. über sog. *Membership Inference Attacks* die verwendeten Trainingsdaten teilweise wiederhergestellt werden können (Käde und Maltzan 2020).

Gegen solche Ansätze wird zwar vorgebracht, dass Modelle nur Ableitungen und nicht die eigentlichen Daten speichern würden (de la Durantaye 2024, Rn. 19), sowohl Datenschutzrecht als auch Urheberrecht sind in dieser Hinsicht allerdings technikneutral: Solange ein Werk (irgendwie) rekonstruiert werden kann, liegt eine Vervielfältigung vor (Raue 2025, Rn. 18–21), solange eine Person aufgrund von Daten bzw. Rückgaben eines Modells identifiziert werden kann, ist es ein personenbezogenes Datum (ablehnend allerdings Hamburgische Beauftragter für Datenschutz und Informationsfreiheit 2024; differenzierend Europäischer Datenschutzausschuss 2024b).

Nicht nur aus rechtlicher Perspektive lässt sich noch nicht abschließend beurteilen, inwieweit Modelle als Kopien der verwendeten Daten betrachtet werden könnten. Auch Versuche, technische Benchmarks zu schaffen, stoßen aktuell noch an Grenzen (Guldimann u. a. 2025, S. 3 f., 8, 15). Sofern urheber- oder datenschutzrechtlich geschützte Daten zum Training verwendet wurden, sollte also im Zweifel davon ausgegangen werden, dass das Modell eine Kopie dieser Daten darstellen *könnte*, zumal die Entfernung von Informationen aus einem Modell (*machine unlearning*) quasi unmöglich ist (Pesch und Böhme 2023b, 919).

5.4 Rechte am trainierten Modell?

Neben Beschränkungen des faktischen Zugangs kann in der Wissenschaft wohl kein Schutz eines konkreten Modells (besser: der *Weights*) angenommen werden. Ein Schutz als Computerprogramm scheitert oftmals am fehlenden relevanten menschlichen Einfluss auf die konkrete Ausgestaltung (Schumacher 2023, S. 166–169), ein Schutz als Datenbank (§§ 87a–87e UrhG) scheitert an der fehlenden Unabhängigkeit der Elemente eines Modells; anders gesagt: Sie lassen sich nicht herauslösen, ohne, dass sie ihre Bedeutung verlieren (Kaulartz 2020, Rn. 14–18; 25; Schumacher 2023, S. 170). Vielmehr hängen sie von ihrem Kontext innerhalb des Modells ab. Damit verbleibt der Schutz als Geschäftsgeheimnis (Kaulartz 2020, Rn. 33 f. Kritisch dazu Schumacher 2023, S. 179–204). Dieser verträgt sich aber meist nicht mit der angestrebten möglichst hohen Transparenz innerhalb der Wissenschaft (Flöter und Cordes 2024, Rn. 10 f.).

6 Zusammenfassung und Ausblick

- An den meisten Daten besteht meist kein Urheberrecht.
- Einfache Messdaten sind auch in der Zusammenstellung als Datenbank nicht schutzfähig.

- Rechte können sich allerdings an Abbildungen ergeben.
- Datenschutzrecht greift meist nicht, kann aber in Einzelfällen relevant werden.
- Für die verwendeten Daten sollte eine Risikoabschätzung hinsichtlich Schutzrechten vorgenommen werden.
- Das Training von Machine-Learning-Modellen kann durch gesetzliche Ausnahmen („Schränken“) gerechtfertigt werden.
- Ob Modelle Kopien der verwendeten Daten darstellen, ist gegenwärtig stark umstritten und kann hier nicht abschließend beantwortet werden.
- An trainierten Modellen bestehen keine für die Forschung relevanten Schutzrechte.

Die rechtlichen Fallstricke sind auch in der Agrosystemforschung zahlreich. Den rechtlichen Sachverhalte inhärent ist auch, dass eine pauschale Antwort selten möglich ist und dass es stets auf den konkreten Einzelfall ankommt. Gleichzeitig erfordern rechtliche Beratungen einen signifikanten Kosten- und Zeitaufwand, der für Forschungsprojekte nicht stemmbar ist. Ein möglicher Lösungsansatz ist die Einrichtung eines Helpdesks, welcher mit Expertise in den jeweiligen Einzelfallbewertungen unterstützt. Daher umfasst der Helpdesk des FAIRagro-Konsortiums, das Data Steward Service Center (DSSC) auch eine Stelle, welche mit einer Volljuristin besetzt ist.

Danksagungen

Dieses Paper entstand im Rahmen des Projekts FAIRagro, welches von der Deutschen Forschungsgemeinschaft unter der Projektnummer 501899475 gefördert wird.

Beiträge

Constantin Breß und Lea Sophie Singson haben gleichermaßen zur Vorbereitung und Formulierung dieses Papers und des zugrundeliegenden Vortrags beigetragen.

Interessenkonflikte

Die Autor:innen erklären, dass in Bezug auf diese Arbeit keine potenziellen Interessenkonflikte bestehen.

Literaturverzeichnis

- Antoine, Lucie. 2024. „Datenzugang im Spannungsfeld zwischen DSGVO, Geschäftsgeheimnisschutz und Datenbankherstellerrrecht – Achillesferse des Data Act in der Praxis?“ *Computer und Recht*, 73–80. <https://doi.org/10.9785/cr-2024-400204>.
- Art.-29-Gruppe. 2007. *Opinion 4/2007 on the concept of personal data*, Juni.
- Bomhard, David. 2024a. „§ 44b UrhG“. In *BeckOK Urheberrecht*, 45. Aufl., herausgegeben von Horst-Peter Götting, Anne Lauber-Rönsberg und Nils Rauer.
- . 2024b. „§ 60d UrhG“. In *BeckOK Urheberrecht*, 44. Aufl., herausgegeben von Horst-Peter Götting, Anne Lauber-Rönsberg und Nils Rauer.
- Bomhard, David und Jonas Siglmüller. 2024. „9.6 Datenlizenzvertrag“. In *Handbuch der IT-Verträge*, herausgegeben von Helmut Redeker. ISBN: 978-3-504-56008-9.
- Bullinger, Winfried. 2022. „§ 60d UrhG“. In *Urheberrecht*, 6. Aufl., herausgegeben von Artur-Axel Wandtke und Winfried Bullinger. ISBN: 978-3-406-76200-0.
- De la Durantaye, Katharina. 2024. „Nutzung urheberrechtlich geschützter Inhalte zum Training generativer künstlicher Intelligenz – ein Lagebericht“. *Zeitschrift für das gesamte Medienrecht*, 9–22. <https://doi.org/10.9785/afp-2024-550103>.
- Dornis, Tim W. und Sebastian Stober. 2024. *Urheberrecht und Training generativer KI-Modelle: Technologische und juristische Grundlagen*. Nomos. ISBN: 978-3-7489-4955-8. <https://doi.org/10.5771/9783748949558>.
- Dreier, Thomas. 2025a. „§ 19a UrhG“. In *Urheberrechtsgesetz*, 8. Aufl., herausgegeben von Thomas Dreier und Gernot Schulze. ISBN: 978-3-406-79302-8.
- . 2025b. „§ 4 UrhG“. In *Urheberrechtsgesetz*, 8. Aufl., herausgegeben von Thomas Dreier und Gernot Schulze. ISBN: 978-3-406-79302-8.
- . 2025c. „§ 60d UrhG“. In *Urheberrechtsgesetz*, 8. Aufl., herausgegeben von Thomas Dreier und Gernot Schulze. ISBN: 978-3-406-79302-8.
- . 2025d. „§ 87a UrhG“. In *Urheberrechtsgesetz*, 8. Aufl., herausgegeben von Thomas Dreier und Gernot Schulze. ISBN: 978-3-406-79302-8.
- Europäischer Datenschutzausschuss. 2024a. *Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR*.
- . 2024b. *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*.
- . 2025. *Guidelines 01/2025 on Pseudonymisation*.
- Flöter, Benedikt und Christoph Cordes. 2024. „KI und Geistiges Eigentum: Neuronale Netze, Trainingsdaten und Output“. *Gewerblicher Rechtsschutz und Urheberrecht in der Praxis*, 668–670.

- Glauner, Patrick. 2024. „Technical foundations of generative AI models“. *Legal-Tech-Zeitschrift*, 24–34.
- Guldimann, Philipp, Alexander Spiridonov, Robin Staab, Nikola Jovanović, Mark Vero, Velko Vechev, Anna-Maria Gueorguieva u. a. 2025. „COMPL-AI Framework: A Technical Interpretation and LLM Benchmarking Suite for the EU Artificial Intelligence Act“. Vorveröffentlichung, besucht am 16. Juni 2025. <https://doi.org/10.48550/arXiv.2410.07959>.
- Hacker, Philipp. 2020. *Datenprivatrecht: Neue Technologien im Spannungsfeld von Datenschutzrecht und BGB*. Jus privatum, Band 244. Tübingen: Mohr Siebeck. ISBN: 978-3-16-159617-9.
- Hamann, Hanjo. 2024. „Nutzungsvorbehalte für KI-Training in der Rechtsgeschäftslehre der Maschinenkommunikation“. *Zeitschrift für geistiges Eigentum* 16:113–168. <https://doi.org/10.1628/zge-2024-0012>.
- Hamburgische Beauftragter für Datenschutz und Informationsfreiheit. 2024. *Diskussionspapier: Large Language Models und personenbezogene Daten*. Besucht am 17. Juni 2025. https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/240715_Diskussionspapier_HmbBfDI_KI_Modelle.pdf.
- Hartmann, Thomas. 2022. „Forschungsdaten in den Naturwissenschaften: Eine urheberrechtliche Bestandsaufnahme mit ihren Implikationen für universitäres FDM“. In *E-Science-Tage 2021. Share your Research Data*, herausgegeben von Vincent Heuveline und Nina Bisheh. heiBOOKS. <https://doi.org/10.11588/HEIBOOKS.979.C13728>.
- Herbst, Tobias. 2024. „Art. 5 DSGVO“. In *DS-GVO/BDSG*, 4. Aufl., herausgegeben von Jürgen Kühling und Benedikt Buchner. ISBN: 978-3-406-80263-8.
- Hofmann, Franz. 2024. „Retten Schranken Geschäftsmodelle generativer KI-Systeme? Vortrag im Rahmen des Symposions »Generative K.I. und das Urheberrecht – Eine komplizierte Beziehung« des Instituts für Urheber- und Medienrecht am 10.11.2023 in München“. *Zeitschrift für Urheber- und Medienrecht*, 166–174.
- Hüger, Jakob. 2024. „Die Rechtmäßigkeit von Datenverarbeitungen im Lebenszyklus von KI-Systemen“. *Zeitschrift für Digitalisierung und Recht*, 263–291.
- Käde, Lisa und Stephanie von Maltzan. 2020. „Algorithmen, die nicht vergessen – Sicherheitslücken in Machine-Learning-Algorithmen und deren Bedeutung für den Schutz der Daten und der Urheberrechte“. *Zeitschrift zum Innovations- und Technikrecht*, 201–208.
- Katzenberger, Paul. 1989. „Neue Urheberrechtsprobleme der Photographie – Reproduktionsphotographie, Luftbild- und Satellitenaufnahmen“. *Gewerblicher Rechtsschutz und Urheberrecht Internationaler Teil*, 116–119.
- Kaulartz, Simon Apel Markus. 2020. „Rechtlicher Schutz von Machine Learning-Modellen“. *Recht Digital*, 24–34.

- Kleinkopf, Felicitas. 2022. *Text- und Data-Mining: die Anforderungen digitaler Forschungsmethoden an ein innovations- und wissenschaftsfreundliches Urheberrecht*. 1. Auflage. Schriftenreihe des Archivs für Urheber- und Medienrecht, Band 300. Baden-Baden: Nomos. ISBN: 978-3-8487-7561-3.
- Kraetzig, Viktoria. 2024. „vor § 44a UrhG“. In *Urheberrecht*, 13. Aufl., herausgegeben von Axel Nordemann, Jan Bernd Nordemann und Christian Czychowski. ISBN: 978-3-17-043860-6.
- Lauber-Rönsberg, Anne. 2024. „§ 72 UrhG“. In *BeckOK Urheberrecht*, 44. Aufl., herausgegeben von Horst-Peter Götting, Anne Lauber-Rönsberg und Nils Rauer.
- Lee, Katherine, A. Feder Cooper und James Grimmelmann. 2024. „Talkin’ ’Bout AI Generation: Copyright and the Generative-AI Supply Chain (The Short Version)“. In *Proceedings of the Symposium on Computer Science and Law*, 48–63. CSLAW ’24: Symposium on Computer Science and Law. Boston MA USA: ACM. Besucht am 25. Mai 2025. <https://doi.org/10.1145/3614407.3643696>.
- Leistner, Matthias. 2020. „§ 4 UrhG“. In *Urheberrecht*, 6. Aufl., herausgegeben von Ulrich Loewenheim, Matthias Leistner und Ansgar Ohly. ISBN: 978-3-406-72096-3.
- Pesch, Paulina Jo und Rainer Böhme. 2023a. „Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern“. *Gewerblicher Rechtsschutz und Urheberrecht*, 997–1007.
- . 2023b. „Verarbeitung personenbezogener Daten und Datenrichtigkeit bei großen Sprachmodellen“. *Multimedia und Recht*, 917–923.
- Raue. 2025. „§ 16 UrhG“. In *Urheberrechtsgesetz*, 8. Aufl., herausgegeben von Thomas Dreier und Gernot Schulze. ISBN: 978-3-406-79302-8.
- Richter, Heiko. 2024. „Art. 43 DA“. In *BeckOK Datenschutzrecht*, 50. Aufl., herausgegeben von Heinrich Amadeus Wolff, Stefan Brink und Antje v. Ungern-Sternberg.
- Schantz, Peter. 2025. „Art. 6 Abs. 1 DS-GVO“. In *Datenschutzrecht*, 2. Aufl., herausgegeben von Spiros Simitis, Gerrit Hornung und Indra Spiecker gen. Döhmman. ISBN: 978-3-8487-8958-0.
- Schröler, Roman und Christian Kuß. 2022. „E. Urheber-, Patent- und Markenrecht“. In *Künstliche Intelligenz*, 1. Aufl., herausgegeben von Kuuya Chibanguza, Christian Kuß und Hans Steege. ISBN: 978-3-8487-7161-5.
- Schulze, Gernot. 2022. „§ 2 UrhG“. In *Urheberrechtsgesetz*, 7. Aufl., herausgegeben von Thomas Dreier und Gernot Schulze. ISBN: 978-3-406-77832-2.
- Schumacher, Pablo. 2023. *Schutzzfähigkeit Künstlicher Intelligenz: Die immaterialgüterrechtliche Schutzzfähigkeit von KI-Technologien*. Trier University. ISBN: 978-3-758-43933-9.
- Vogel, Martin. 2020. „§ 72 UrhG“. In *Urheberrecht*, 6. Aufl., herausgegeben von Ulrich Loewenheim, Matthias Leistner und Ansgar Ohly. ISBN: 978-3-406-72096-3.

- Vohwinkel, Moritz. 2024. „§ 87a UrhG“. In *BeckOK Urheberrecht*, 42. Aufl., herausgegeben von Horst-Peter Götting, Anne Lauber-Rönsberg und Nils Rauer.
- . 2025. „§ 87b UrhG“. In *BeckOK Urheberrecht*, 45. Aufl., herausgegeben von Horst-Peter Götting, Anne Lauber-Rönsberg und Nils Rauer.
- Von Welser, Marcus. 2022. „§ 44a UrhG“. In *Urheberrecht*, 6. Aufl., herausgegeben von Artur-Axel Wandtke und Winfried Bullinger. ISBN: 978-3-406-76200-0.
- Weidert, Stefan und Hannah Bug. 2023. „c) Rechte an Daten“. In *Handelsgesetzbuch*, 6. Aufl., herausgegeben von Volker Röhrich, Friedrich Graf von Westfalen, Ulrich Haas, Sebastian Mock und Heinz Wöstmann. ISBN: 978-3-504-45523-1.
- Weiss, Mathias und Erik Schuh. 2024. „Die Zweckbestimmung und Zweckbindung als Weichenstellung für die DSGVO-konforme Nutzung von Daten für KI-Systeme“. *Zeitschrift für Digitalisierung und Recht*, 225–262.
- Wiebe, Andreas. 2017. „Schutz von Maschinendaten durch das sui-generis-Schutzrecht für Datenbanken“. *Gewerblicher Rechtsschutz und Urheberrecht*, 338–345.
- Zech, Herbert. 2015. „Daten als Wirtschaftsgut – Überlegungen zu einem „Recht des Datenerzeugers““. *Computer und Recht*, 137–146. <https://doi.org/10.9785/9783504384579-002>.