

KI-Rechtsrahmen neu gedacht

Generative KI: Innovation und Recht in Arbeitsprozessen

Umgesetzt durch den
Think Tank iRights.Lab

Juni 2026

iRights.Lab 
Lösungen für Recht, Technologie,
Politik & Gesellschaft

**Gen
KI** 

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

Der GenKI-IR-Report

Analysen, Erkenntnisse und Thesen zu den rechtlichen und gesellschaftlichen Rahmenbedingungen beim Einsatz und der Nutzung von generativer Künstlicher Intelligenz (KI).

Ein Bericht des Begleitforschungsvorhabens
„Generative KI – Innovation und Recht in Arbeitsprozessen“
Umgesetzt durch den Think Tank iRights.Lab, Juni 2026

Das Begleitforschungsvorhaben „Generative KI-Innovation und Recht in Arbeitsprozessen“ wird gefördert durch das Bundesministerium für Forschung, Technologie und Raumfahrt.



Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

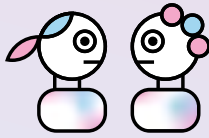
Inhalt



1 Einleitung	4
2 Für ein faires KI-Recht: Innovative Thesenpapiere zu Datenschutz, Vergütung und Kennzeichnung	6
Zwischen Hype und Hybris: Zehn Thesen zu KI-kompatiblen Datenschutz	7
Disruption und Kompensation: Zehn Thesen zu einer KI-Vergütung	10
Sticker drauf und fertig: Zehn Thesen zu der Frage, wieso Kennzeichnung alleine nicht reicht	13
3 Methodenmix für innovative Begleitforschung: Unser Weg zu zentralen Erkenntnissen	16
Zukunftsszenarien: KI-Rechtsrahmen neu gedacht	20
prompt/-Newsletter: Expert*innen-Interviews und Beiträge unter anderem zu Haftungsfragen, KI-Verordnung, Outputs und Urheberrecht	37
Lunch-Talk: Forschungsbefunde und Deep Dive zum Thema KI und Haftung	63
Unser Working Paper – mit zehn Thesen zu KI und Urheberrecht, Datenschutz und Kennzeichnung	66
Recherchen und Hintergrundgespräche: Das Fundament für die gesamte Projektarbeit	76
4 Selbstbeobachtung: So bewerten wir unser Vorgehen	84
5 KI-Rechtsfragen zu Datenschutz, Vergütungen, Kennzeichnungen und Governance – damit wollen wir uns bei GenKI-IR weiter auseinandersetzen	92
6 Anhang	96
Glossar/Terminologie	96
Quellen	97
Über dieses Projekt	98
Das Team	99
Danksagungen	100
Impressum	101

1

Einleitung



Generative KI ist in vielen Berufsfeldern längst Teil des Arbeitsalltags. Sie verändert Arbeitsprozesse, kreative Praktiken und Formen der Wissensarbeit: Chatbots, die journalistische Texte generieren, zusammenfassen oder redigieren; KI-gestützte Werkzeuge, die Anschreiben, Anträge oder Bescheide formulieren und überprüfen; Bildgeneratoren, die Illustrationen, Infografiken oder Animationen erstellen; KI-Umgebungen, in denen Unterrichtsstunden vorbereitet und Klassenarbeiten korrigiert werden können.

Mit der wachsenden Verbreitung generativer KI im Arbeitskontext entstehen neue rechtliche Fragen. Bestehende Rechtsfragen gewinnen an Dringlichkeit. Wer KI-Anwendungen im Berufsalltag einsetzt, muss prüfen, welche Regeln gelten: etwa beim Umgang mit personenbezogenen Daten, bei der Frage nach Urheberrechten an KI-generierten Inhalten oder deren Kennzeichnung. Besondere Sorgfalt ist dort geboten, wo KI in sensiblen Bereichen eingesetzt wird – etwa wenn Entscheidungen die Gesundheit, Sicherheit oder Grundrechte von Menschen berühren. Hinzu kommen arbeitsrechtliche Aspekte, Haftungsfragen oder auch Fragen der Mitbestimmung, etwa wenn KI von Beschäftigten als Arbeitswerkzeug eingesetzt werden soll.

Viele dieser Fragen sind bislang nicht abschließend geklärt. Bestehende Regelungen lassen sich nicht immer ohne Weiteres auf neue KI-Anwendungen übertragen und neue Regeln – wie die Europäische KI-Verordnung – werfen in der praktischen Umsetzung ihrerseits neue Fragen auf. Das erzeugt Unsicherheiten: bei Unternehmen und Behörden, bei Beschäftigten und bei den Menschen, die mit KI-gestützten Diensten in Berührung kommen.

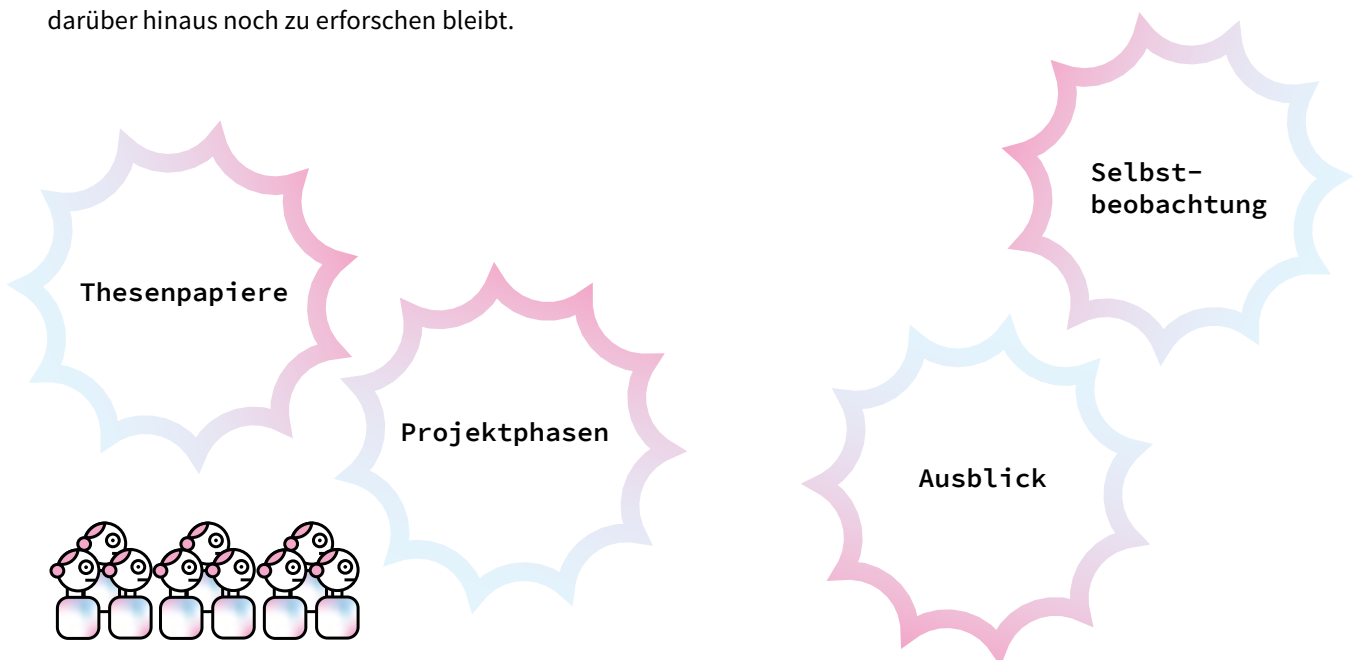
Neben der rechtlichen Dimension kommt dem Einsatz generativer KI aber auch eine gesellschaftspolitische Dimension zu: Welchen Stellenwert haben menschengemachte Kunst und Kultur in unserer Gesellschaft? Wie können wir gewährleisten, dass Kreativschaffende auch in Zeiten von KI fair vergütet werden? Wie kann verhindert werden, dass KI-generierte Inhalte in ohnehin schon herausfordernden Informationsökosystemen nicht für weitere Unsicherheit sorgen – gerade dort, wo Deepfakes, Desinformation und Fälschungen realen Schaden anrichten können?

Auf Basis dieser Überlegungen haben wir die folgenden drei Themen als Kernfelder unserer Forschung identifiziert:

- 1. **Datenschutz und die Frage, wie bestehende Datenschutzregeln angepasst werden können, um personenbezogene Daten in Zeiten von KI vor unerlaubter Verarbeitung zu schützen.**
- 2. **Urheberrecht und die Frage, wie Kreativ- und Medien-schaffende weiterhin fair für ihre Arbeit vergütet werden können.**
- 3. **Kennzeichnung und die Frage, ob und wie KI-generierte Inhalte kenntlich gemacht werden können, um Desinformation und einem fortschreitenden Vertrauensverlust in Informationsangebote zu begegnen.**

Das diesem Report zugrunde liegende Forschungsprojekt nimmt vier Bereiche in den Blick, die exemplarisch für die Breite des Themas stehen: Journalismus, Kreativwirtschaft, Verwaltung und Bildung. Sie wurden ausgewählt, weil sie gesellschaftlich besonders relevant sind, weil generative KI dort bereits intensiv genutzt wird und weil sich die oben beschriebenen Spannungsfelder in ihnen besonders deutlich zeigen. Die gewonnenen Erkenntnisse sind jedoch nicht auf diese Bereiche beschränkt: Viele der identifizierten Rechtsfragen stellen sich in ähnlicher Form auch anderswo.

Den vorliegenden Report haben wir wie folgt strukturiert: Kapitel 2 vertieft die drei Kernthemen jeweils in einem kompakten Thesenpapier mit innovativen Ansätzen. Kapitel 3 dokumentiert die Ergebnisse der einzelnen Projektphasen: Dazu gehören die Erkenntnisse aus Zukunftslaboren, in denen wir gemeinsam mit Expertinnen und Experten aus Wissenschaft, Praxis und Verwaltung rechtliche Gestaltungsoptionen erarbeitet haben sowie Analysen und Interviews aus dem projektbegleitenden Newsletter „prompt/“. Kapitel 4 reflektiert selbstkritisch die gewählten Forschungsansätze und -methoden. Kapitel 5 schließlich benennt offene Fragen und skizziert, was aus unserer Sicht in den drei Kernthemen und darüber hinaus noch zu erforschen bleibt.



2

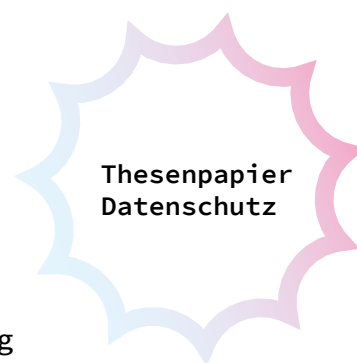
Für ein faires KI-Recht: Innovative Thesenpapiere zu Datenschutz, Vergütung und Kennzeichnung

Wie kann die Regulierung generativer KI zukunftsfähig gestaltet werden? Wie sieht eine gerechte Entschädigung für die ökonomischen und gesellschaftlichen Verwerfungen durch KI-Systeme aus? Wie kann ein modernes Datenschutzrecht aussehen, das technologischen Fortschritt und die Individualinteressen von Bürger*innen gut ausbalanciert? Und welche Zielsetzung sollte einer verpflichtenden KI-Kennzeichnung im Idealfall zugrunde liegen?

Diese und weitere Fragen standen im Mittelpunkt von Diskussionen, die wir im Zuge unserer Zukunftsszenarien mit Fachleuten, Stakeholdern und KI-Nutzer*innen aus unterschiedlichen Branchen geführt haben. Das Ergebnis sind drei Thesenpapiere zu den Themen Datenschutz, Vergütung und Kennzeichnung.

Sie enthalten grundsätzliche Überlegungen aus gesellschaftspolitischer und juristischer Perspektive, wagen Gedankenexperimente und bieten innovative Lösungswege an.

Zwischen Hype und Hybris: Zehn Thesen zu KI-kompatiblen Datenschutz



Wer zukunftstauglichen Datenschutz will, darf nicht nur Felder wie Bürokratieabbau oder Innovationsförderung betrachten. Es braucht einen mutigen Blick für das große Ganze – gerade auf generative KI.

(1) Eine Neufassung des EU-Datenschutzrechts – weil wir dazu gelernt haben

Seit Einführung der DSGVO im Jahr 2018 hat der europäische Datenschutz große Aufmerksamkeit erhalten. Meist jedoch stand er in der Kritik. Häufig wurde eine Reform gefordert, entweder zur Innovationsförderung oder als Beitrag zum Abbau vermeintlich überflüssiger Bürokratie. Überzeugender ist, das bestehende Datenschutzrecht im Sinne guten rechtsstaatlichen Handelns („good governance“) weiterzuentwickeln: Dafür ist einerseits die Einbeziehung aller gewonnenen Erfahrungen aus Unternehmens-, Aufsichts- und Gerichtspraxis wichtig, um bisherige Problemstellungen, Unklarheiten und Nachschärfungsbedarfe gemeinsam anzugehen. Andererseits müssen aber auch jene Personen zu Wort kommen, die sich intensiv mit dem zukünftigen Verhältnis von generativer KI und Datenschutz auseinandersetzen.

(2) Maschinelles Lernen und Big Data stellen den Wert einzelner Datenpunkte grundsätzlich in Frage

Wer einen zukunftstauglichen und KI-kompatiblen Datenschutz möchte, muss sich der Größe der Aufgabe bewusst sein. Die realen und vermeintlichen Potenziale generativer KI-Systeme haben das Grundkonzept der DSGVO unter existenziellen Druck gesetzt. Darüber, ob einzelne Datenpunkte auch in Zukunft noch Bedeutung und auch Wert zugeschrieben bekommen sollten, gibt es gegensätzliche Ansichten. Wo die DSGVO noch Lösch- und Korrekturpflichten in Bezug auf einzelne Daten aufstellt, setzt die KI-Entwicklung auf heuristische Auswertung und stochastische Zusammenhänge in kaum mehr zu erfassenden Datenmengen. Es scheint, als entzögen Qualität und Quantität der für die KI-Entwicklung erforderlichen Daten dem datenschutzrechtlichen Konzept zunehmend den Boden.

(3) Bei einem KI-kompatiblen Datenschutz muss das gesamte Datenrecht mitgedacht werden

Dass der Wert einzelner Datenpunkte unklar geworden ist, ist kein reines Datenschutzproblem: Die urheberrechtliche TDM-Schranke – eine gesetzliche Ausnahmeregelung zum „Text und Data Mining“ – wurde als Innovationsförderung angekündigt und von KI-Anbieter*innen als Erlaubnis zur genehmigungsfreien Nutzung urheberrechtlicher Werke beim KI-Training verstanden. Schließlich komme es für die KI-Entwicklung auf das einzelne Werk gar nicht an. Auch das Marken- und Designrecht verhindert KI-Training mit geschützten Bild- oder Wortmarken nicht. Eine Angleichung dieser und weiterer gesetzlichen Rahmenbedingungen erscheint sinnvoll und zweckmäßig, um Umsetzungsprobleme durch fragmentierte Rechtsrahmen zu verhindern.

2

(4) Politik und Gesetzgebung können sich von innovationsfreundlichen Konzepten aus Wissenschaft, Wirtschaft und Zivilgesellschaft inspirieren lassen

In Wissenschaft, Wirtschaft und Zivilgesellschaft finden sich längst Ansätze, die den vermeintlich verringerten Wert und die womöglich abnehmende Bedeutung einzelner Datenpunkte ins Datenschutzrecht übertragen wollen. So könnte man die Rechtmäßigkeit einer großen Datenverarbeitung nicht mehr von jedem Teildatum abhängig machen. Und auch Betroffenenrechte könnten in Bezug auf solche Verarbeitungsschritte ausgenommen werden, wenn es bei ihnen gar nicht um spezifische Einzeldaten gehen soll. Datenschutz wäre also als Sorgfaltspflicht gegenüber einer Datenmehrheit zu verstehen und nicht als Schutz individuell zuzuordnender Informationen. Diese Ansätze schließen an verschiedene etablierte Fachdiskurse an – zum Beispiel an die vom EuGH angestoßene Diskussion zum Begriff des „personenbezogenen Datums“.

(5) Der Wert einzelner Datenpunkte hängt mit dem Verständnis von Datenschutz als Menschenrechtsschutz in einer digitalisierten Welt zusammen

Ein Datenschutzrecht, das zwar systematische Rechtsgutsverletzungen in den Blick nimmt, aber die Rechtsschutzmöglichkeiten von Individuen beschränkt, etwa im Sinne eines digitalen Verbraucherschutz- und Ordnungsrechts, ist durchaus denkbar. Es lässt sich sogar argumentieren, dass ein solches System vergleichsweise leicht zu handhaben und womöglich sogar effizienter wäre. Über eine entsprechende Rechtsanpassung müsste dann aber auch im Bereich von zum Beispiel Hate Speech, Deep Fakes und Social Bots diskutiert werden: Schließlich ist auch der individuelle Schutz von Meinungsfreiheit, Religionsfreiheit und Identität vom Schutz und der Bedeutung individualisierbarer Datenpunkte abhängig. Bei ganzheitlicher Betrachtung lassen sich damit in der aktuellen datenschutzrechtlichen Reformdebatte Tendenzen wahrnehmen, die den Charakter des europäischen Grund- und Menschenrechtsschutz insgesamt (unbewusst) in Frage zu stellen scheinen.

(6) Ein KI-kompatibler Datenschutz der Zukunft muss als Teil digitaler Menschenrechte verstanden werden

Gerade in einer Gegenwart, in der Smartphones zentrale Hilfsmittel für die Bewältigung des Alltags sind, KI-Systeme im beruflichen Kontext bei wichtigen Arbeitsschritten unterstützen und einem demografischen Wandel mit mehr Technik und Automatisierung begegnet werden soll, muss ein zukunftsfähiges, KI-kompatibles Datenschutzrecht individueller Menschenrechtsschutz bleiben. Politik, Wissenschaft, Wirtschaft und Zivilgesellschaft müssen den Wert, die Bedeutung und die Vorteile dieses Grundansatzes hervorheben und auf diese Weise zu einem neuen, positiveren Image beitragen – etwa durch gemeinsame Werbekampagnen von Bundesministerien und der Industrie, durch aufeinander Bezug nehmende Interviews und Pressemitteilungen, oder durch die Hervorhebung von Best Practices aus Wissenschaft und Wirtschaft. Dazu gehören aber auch Reformen, die die datenschutzrechtlichen Regelungen risikobasiert weiterentwickelt: Vereine und Unternehmen, die keine risikorelevanten Verarbeitungen vornehmen, sollten einerseits entlastet werden, und global agierende Digitalunternehmen andererseits höheren Auflagen entsprechen müssen.

(7) Es braucht klare Rechtsgrundlagen für Entwicklung und Einsatz von KI

Das gilt gerade für generative KI: Ihre Entwicklung und ihr Einsatz müssen rechtskonform möglich sein, ohne hierfür übermäßigen Haftungsrisiken ausgeliefert zu sein. Bisher müssen Organisationen und Unternehmen prüfen und sicherstellen, dass in jedem KI-Verarbeitungskontext die eigenen Verarbeitungsinteressen überwiegen. Das überfordert gerade kleine Organisationen und Unternehmen. Statt den einzelnen Verantwortlichen eine komplizierte Interessensabwägung aufzuerlegen, muss der europäische Gesetzgeber hier selbst vorangehen. Er muss die einzustellenden Interessen und Argumente abwägen und eine präzise gefasste, neue Rechtsgrundlage schaffen – auch und gerade in Bezug auf die Verarbeitung besonderer Kategorien personenbezogener Daten. Das schafft Rechtssicherheit, von der vor allem nicht-digitalspezifische Organisationen und Unternehmen profitieren.

(8) Keine Chatbot-DSGVO: Agentic & Public AI als relevante Szenarien für zukunftstauglichen Datenschutz berücksichtigen

Wer das Datenschutzrecht reformieren will, darf nicht bei heutigen KI-Modellen wie LLMs oder generativen KI-Systemen (wie Chatbots) stehen bleiben. Daten, auf die KI-Agenten zugreifen müssen, um zum Beispiel als persönliche Assistenz von Privatpersonen zu funktionieren, sollten schon heute mitgedacht werden. Dasselbe gilt für gemeinschaftlich organisierte KI-Systeme des öffentlichen Informationsmanagements, die in ganz Europa unter dem Begriff „Public AI“ immer öfter gefordert werden. Beide Entwicklungen sind bei datenschutzrechtlichen Reformüberlegungen zu bedenken und einzubeziehen, und könnten durch passgenaue Erleichterungen und Voraussetzungen im Datenrecht gezielt gefördert oder gesteuert werden.



Thesepapier
Datenschutz

(9) Public AI als Beleg: Datenpunkte haben weiterhin einen Wert – im KI-Output

Gerade Konzepte wie Public AI zeigen zudem: Mag der Wert einzelner Datenpunkte im Rahmen des KI-Trainings in Zweifel stehen – für den KI-Output bleibt er unbestritten. Anders würden Ansätze zu einem KI-basierten Wissens- und Informationsmanagement – sei es proprietär oder gemeinwohlbasiert – nicht tragen. Diese setzen unverändert auf Präzision, Vollständigkeit, Nachvollziehbarkeit und Korrigierbarkeit von Informationen und Daten, die vom System ausgegeben werden. Eine Reform des Daten(schutz)rechts muss daher den Output wesentlich anders behandeln als die KI-Entwicklung.

(10) Handeln statt abwarten – zwischen „best practice“ und „failing forward“

Parallel zu Reformbemühungen in Brüssel sollten auch andere Akteur*innen als EU mit Mitgliedstaaten zu einem zukunftsfähigeren Datenschutz beitragen. Besonders gefördert werden sollten Maßnahmen zum aktiven Wissensaustausch zu best und worst practices in Wissenschaft, Wirtschaft und Zivilgesellschaft – auch unter Einbeziehung der nationalen Aufsichtsstellen. Einen Beitrag könnten hier auch freiwillige Register leisten, in denen staatliche und privatwirtschaftliche KI-Nutzer*innen ihre Dokumentationen, Datenschutzfolgenabschätzungen und Abwägungserläuterungen öffentlich einsehbar hinterlegen. Diese und weitere Maßnahmen können dabei helfen, praxisnahe Anwendungen zu etablieren, Kosten zu senken und gerade durch den Austausch mit Fachbehörden auch Rechtsunsicherheiten zu verringern.

2

Disruption und Kompensation: Zehn Thesen zu einer KI-Vergütung



Thesenpapier
Vergütung

Generative KI entfaltet ihr gesellschaftliches Potenzial auf Kosten jener Informations- und Kreativstrukturen, von denen sie selbst abhängt – ein strukturelles Ungleichgewicht, das weder durch Verbote noch durch das bestehende Urheberrecht aufgelöst werden kann und deshalb einen eigenständigen, zweckgebundenen Kompensationsmechanismus erfordert.

(1) Generative KI ist technische Revolution und wirtschaftliche Disruption zugleich

Generative KI schafft neue Zugangsmöglichkeiten zu Wissen, Analyse und Wissensproduktion. Sie kann dabei helfen, Aufgaben zu lösen, die Menschen allein gar nicht oder nur mit erheblichem Aufwand bewältigen könnten. In vielerlei Hinsicht hat sie erhebliches Potenzial für das Gemeinwohl. Zugleich verursachen ihre Entwicklung und Vermarktung erhebliche Kollateraleffekte mit teils existenzbedrohenden Folgen. Betroffen sind nicht nur einzelne Branchen, sondern zentrale Strukturen menschlicher Informationsproduktion. Da generative KI absehbar nicht wieder verschwinden wird, ist die Frage nach ihrer Verhinderung hinfällig. Stattdessen muss für eine faire Einbettung ihrer wirtschaftlichen Folgen Sorge getragen werden.

(2) Hochwertige KI braucht offene und vielfältige Informationsressourcen

Entwicklung, Angebot und Nutzung von KI sind grundsätzlich weder illegitim noch gesellschaftlich unerwünscht. Um aber ihr Potenzial zu heben, ist es notwendig, dass KI auf viele hochwertige Informationen zugreifen kann. Würde beispielsweise der Zugang zu wissenschaftlichen, journalistischen oder kulturellen Inhalten versperrt, drohten minderwertigere KI-Systeme mit weniger Vielfalt, mehr Bias und größerer Manipulationsanfälligkeit. Die Schließung öffentlich zugänglicher Informationsressourcen ist daher kein Ausweg, sondern ein Irrweg. Das Ziel sollte nicht sein, KI systematisch von Wissen abzuschneiden, sondern die Nutzung öffentlich verfügbarer Informationen fair und nachhaltig in das Informationsökosystem einzubetten.

(3) Die wirtschaftlichen Grundlagen menschlicher Informationsproduktion sind in Gefahr

KI-Systeme nutzen öffentlich verfügbare Daten und Inhalte als Grundlage kommerzieller Produkte. Mit den Erzeugern und Bereitstellern dieser Inhalte gibt es weder routinemäßige Absprachen noch erhalten sie Gegenleistungen. Das ist einerseits nicht fair und erzeugt andererseits gefährliche Substitutionseffekte: Unmittelbare menschliche Leistungen werden weniger nachgefragt. Dadurch sinkt der Anreiz, solche Leistungen zu erbringen, öffentlich zugänglich zu machen und in ihre Qualität zu investieren. Langfristig gefährdet dies die Existenz jener Akteur*innen, auf deren Leistungen das gesamtgesellschaftliche Potenzial von KI-Systemen wesentlich beruht.

(4) Die Disruption beeinträchtigt das gesamte Informations-Ökosystem zugunsten einiger Weniger

Die Folgen der KI-Revolution betreffen nicht nur Urheber*innen, Künstler*innen, Journalist*innen oder andere klassische Kreativberufe. Belastet werden auch Anbieter*innen öffentlich zugänglicher Informationsinfrastrukturen: wissenschaftliche Repositorien, offen zugängliche Bibliotheksbestände, Wikipedia, Open-Source-Repositorien und andere digitale Wissensressourcen. Viele dieser Angebote werden durch KI-Crawler technisch und wirtschaftlich zusätzlich beansprucht, ohne dass hierüber Absprachen getroffen oder Gegenleistungen erbracht werden. Die Disruption verteilt Lasten und Nutzen asymmetrisch: Von der KI-Wertschöpfung profitieren wenige, während viele die Grundlagen dafür bereitstellen und zugleich negative Folgen tragen.

(5) Die KI-Industrie hat ein Interesse an menschlicher Kreativität und Informationsproduktion

An menschlicher Kreativität, journalistischer Arbeit, wissenschaftlicher Qualitätssicherung und offenen Wissensinfrastrukturen besteht ein öffentliches Grundinteresse. Auch die KI-Industrie selbst ist auf eine lebendige menschliche Informations- und Kreativproduktion angewiesen. KI kann sich nicht nur aus KI nähren. Ohne fortlaufend neue, diverse und qualitätsgesicherte menschliche Inhalte drohen Verarmung, Wiederholung und Qualitätsverlust. Eine Beteiligung der KI-Wirtschaft an den gesellschaftlichen Kosten des Umbruchs ist daher nicht nur aus Fairnessgesichtspunkten plausibel, sondern auch funktional sinnvoll.

(6) Rein urheberrechtliche Vergütungsansätze greifen zu kurz

Ein Ansatz, der sich nur auf das bisherige Urheberrecht stützt, kann diese Aufgabe nicht leisten. Denn generative KI-Technologien sind keine Kopiermaschinen. Sie dienen nicht der Reproduktion bestehender Werke, sondern der Generierung neuer Inhalte. Ein urheberrechtlicher Vergütungstatbestand müsste daher mühsam konstruiert werden: Der für das Urheberrecht zentrale Zusammenhang zwischen konkreter Werknutzung und Nutzungsvergütung fehlt. Weder lässt sich der Beitrag eines bestimmten Werkes zu den Einnahmen eines KI-Anbieters verlässlich messen noch ein konkreter Einnahmeverlust einzelner Rechteinhaber. Ein urheberrechtlich verstandener „angemessener Ausgleich“ für die Einbeziehung geschützter Werke in KI-Technologien ist daher dogmatisch und praktisch kaum zu begründen. Selbst, wenn man die dogmatischen Schwierigkeiten irgendwie überwinden könnte, würden urheberrechtliche Vergütungsmodelle nur einen Ausschnitt des Problems erfassen. Viele Betroffene verfügen nicht über Urheberrechte oder sind nicht Teil klassischer Verwertungsstrukturen. Open-Access-Projekte, öffentliche Wissensinfrastrukturen, gemeinfreie Inhalte, staatliche Informationsangebote oder Open-Source-Repositorien blieben weitgehend außen vor.

(7) Eine Ausweitung des Urheberrechts würde den europäischen Interessensausgleich gefährden

Eine Ausweitung des Urheberrechts durch neue Verwertungsrechte oder zusätzliche Vergütungsansprüche würde den sensiblen Ausgleich der DSM-Richtlinie gefährden, auf den sich die EU nach langen Verhandlungen einigen konnte. Den Trade-off zwischen Nutzungsfreiheit und Opt-out-Möglichkeiten aufs Spiel zu setzen, wäre ein riskanter Schritt. Er könnte die Entstehung europäischer KI-Industrie behindern und etablierte Player dazu verleiten, die europäische Jurisdiktion zu vermeiden.

2

(8) Zweckgebundene Disruptionsabgabe zur Herstellung von Reziprozität

Ein neuer, ganzheitlicher Ansatz für den „Value Gap“ zwischen Informationserzeugern und KI-Wirtschaft könnte in einer europaweiten, zweckgebundenen Disruptionsabgabe liegen. Sie würde die Profiteur*innen der KI-Wertschöpfung verpflichten, einen Teil ihrer Erlöse an das Informations-Ökosystem zurückzugeben, aus dem sie ihren wirtschaftlichen Wert schöpfen.

Anders als in der herkömmlichen Internet-Ökonomie sieht das KI-Ökosystem bislang gerade keinen tragfähigen Reziprozitätsmechanismus vor. Zum Vergleich: Klassische Suchmaschinen erhielten indexierbare Inhalte und verschafften Content-Anbieter*innen im Gegenzug Sichtbarkeit, Reichweite, Klicks und mittelbar Werbeeinnahmen. Generative KI verdichtet Inhalte dagegen zu unmittelbar nutzbaren Informationen. Der Aufruf der Ursprungsquelle wird in vielen Nutzungsszenarien entbehrlich. Reichweite, Aufmerksamkeit, Nutzer*innenbeziehung und wirtschaftlicher Wert kumulieren sich bei den KI-Anbieter*innen.

Eine derart einseitig gelagerte Wertschöpfung ist auf Dauer nicht tragfähig. Um diesem Defizit zu begegnen, folgt der Ansatz der Disruptionsabgabe zwei Grundprämissen: Erstens beruht die Wertschöpfung kommerzieller KI-Angebote wesentlich auf öffentlich zugänglichen Inhalten, Informationen, Daten, Code und Wissensinfrastrukturen. Wer hiervon profitiert, sollte an deren Erhalt beteiligt werden (Fairness-Argument). Zweitens verursacht die KI-Disruption gesellschaftliche und wirtschaftliche Folgekosten. Wer diese Disruption maßgeblich vorantreibt und hieran verdient, sollte sich an ihrer Bewältigung beteiligen (Kompensations-Argument).

(9) Eine faire Disruptionsabgabe: zweckgebunden, europäisch, lösungsorientiert

Die vorgeschlagene Disruptionsabgabe fördert diese Ziele in Form eines zweckgebundenen Reziprozitätsmechanismus, um das Informations-Ökosystem zu stabilisieren. Als „Fair-share-Konzept“ orientiert sie sich an den innerhalb der EU erzielten Umsätzen kommerzieller Anbieter*innen von KI-Modellen und -Systemen. Nicht-kommerzielle, wissenschaftliche, gemeinnützige oder öffentliche-rechtliche Angebote sollten ausgenommen werden.

Die Abgabe sollte zweckbezogen erhoben und verwendet werden: Anders als allgemeine Steuern sollte sie nicht im Staatshaushalt aufgehen, sondern gezielt jenen Strukturen zugutekommen, die für die Entstehung, Pflege und Zugänglichmachung öffentlicher Information wesentlich sind. Die Verteilung der Einnahmen sollte über staatlich und zivilgesellschaftlich organisierte Fonds erfolgen, die systemrelevante Bereiche wie Kultur, Journalismus, Bildung, Wissenschaft und digitale Wissensinfrastrukturen fördern. Auch Verwertungsgesellschaften könnten einbezogen werden, soweit Einnahmen an ihre Wahrnehmungsberechtigten verteilt werden sollen. Wünschenswert wäre eine gesamteuropäische Struktur der Erhebung und Verteilung, verbunden mit nationalen Umsetzungs- und Vergabemechanismen.

(10) Die Disruptionsabgabe soll bestehende Schutzmechanismen nur ergänzen

Die Disruptionsabgabe wäre kein Ersatz für urheberrechtliche, medienrechtliche oder KI-regulatorische Instrumente, die es bereits gibt. Sie wäre ein zusätzlicher Baustein eines umfassenderen Ordnungsrahmens. Da die Abgabe nicht an die Nutzung urheberrechtlich geschützten Materials anknüpft, würde sie KI-Anbieter*innen kein Recht auf Zugang zu oder Nutzung von geschützten Inhalten verschaffen. Text-and-Data-Mining-Schranken, Opt-out-Möglichkeiten und Ansprüche bei konkreten Rechtsverletzungen blieben daneben bestehen. Auch die KI-Verordnung und sonstige regulatorische Anforderungen würden weder geändert noch ersetzt.

Sticker drauf und fertig: Zehn Thesen zu der Frage, wieso Kennzeichnung alleine nicht reicht



Wer in einer zunehmend synthetischen Informationswelt Vertrauen in Medieninhalte und die dahinterstehenden Akteure sichern möchte, darf sich nicht auf ein einzelnes Transparenzinstrument verlassen. Kennzeichnung ist wichtig – kann aber nur der Anfang einer größeren Debatte über Erwartungen, Nachvollziehbarkeit, Kontext und Verantwortung beim Einsatz generativer KI sein.

(1) Kennzeichnungspflichten sind ein notwendiger erster Schritt, aber kein vollständiger Schutzmechanismus

Mit Art. 50 der KI-Verordnung verpflichtet die EU Anbieter und Betreiber bestimmter KI-Systeme, KI-generierte Inhalte und die Systeme selbst als solche kenntlich zu machen. Die Vorgabe betrifft also beispielsweise synthetisch erzeugte Bilder und Videos – Deepfakes etwa – genauso wie Chatbots und Systeme zur Emotionserkennung oder biometrischer Kategorisierung. Diese Regelungen schaffen erstmals einen EU-weiten Mindeststandard für Transparenz. Das Problem: Wie kann eine einheitliche Kennzeichnungslösung bei solch einer Vielfalt von Inhalten und Anwendungen praktisch umgesetzt werden? Ist ein für alle KI-Systeme und -Inhalte gleichermaßen gültiges Logo überhaupt denkbar? Es besteht die Gefahr, dass die Debatte sich im Streit um Logos erschöpft, während die tieferliegende Verschiebung von Deutungs-, Verwertungs- und Gestaltungsmacht in den datengestützten Branchen von Journalismus bis Kreativwirtschaft unsichtbar bleibt.

(2) Ein richtiger Ansatz: risikobasierte Kennzeichnung

Die zunehmende Verbreitung KI-gestützter Werkzeuge im Alltag – von Autokorrektur über Bildbearbeitung bis hin zur vollständigen Textgenerierung – zeigt, wie schnell Kennzeichnungspflichten an praktische Grenzen stoßen können: Eine „Überkennzeichnung“ könnte dazu führen, dass die Kennzeichnung ihre Wirkung verliert. Deshalb erscheint der risikobasierte Ansatz der KI-Verordnung sinnvoll: Nicht jeder KI-Einsatz muss gekennzeichnet sein, sondern nur dann, wenn er ein besonderes Risiko birgt. Wie eine angemessene Kennzeichnung in diesen Fällen aussehen sollte, hängt allerdings von unterschiedlichen Aspekten ab.

(3) Die Wirkung von Kennzeichnung hängt stark vom Nutzungskontext ab

Pauschale Kennzeichnungslösungen gehen an der Erwartungshaltung von Nutzer*innen vorbei: Nicht jeder Inhalt wird in jedem Kontext mit denselben Erwartungen wahrgenommen. Wer Werbung betrachtet, geht meist ohnehin davon aus, dass diese inszeniert oder bearbeitet ist. Ob ein hierbei verwendetes Bild von einem Fotografen, aus einer Bilddatenbank oder von

2

einer KI stammt, spielt deshalb eine untergeordnete Rolle. In anderen Bereichen hingegen – etwa im Fotojournalismus – erwarten Rezipient*innen eine authentische Abbildung realer Ereignisse. Die Offenlegung technischer Informationen über verwendete Hilfsmittel ist dabei eher zweitrangig. In solchen Kontexten spielt die Information über den Einsatz von KI eine deutlich größere Rolle – die Anforderungen an die Kennzeichnung steigen.

(4) Wichtig für Kennzeichnung: nicht nur ob, sondern wie KI eingesetzt wurde

Bei der Frage danach, wie KI-generierte Inhalte gekennzeichnet werden sollten, geht es nicht nur darum, kenntlich zu machen, ob generative KI genutzt wurde, sondern auch, wie sie genutzt wurde. Wird KI lediglich unterstützend eingesetzt, etwa zur Rechtschreibkorrektur oder zur Entfernung kleiner Bildstörungen, besteht ein anderes Kennzeichnungsbedürfnis als bei einer vollständig synthetischen Generierung von Texten und Bildern. Eine sinnvolle Transparenzpolitik muss diese Unterschiede berücksichtigen.

(5) Eine robuste Kennzeichnung hilft, manipulierte Inhalte erkennbar zu machen

Mit der zunehmenden Leistungsfähigkeit generativer KI-Systeme verschwimmen die Grenzen zwischen echten und synthetischen Inhalten immer stärker. Bilder, Stimmen oder Texte lassen sich heute täuschend echt erzeugen oder verändern. In dieser Situation kann u. a. eine technisch sichere und dauerhafte Kennzeichnung helfen, echte Inhalte besser von mithilfe von KI generierten oder manipulierten Inhalten zu unterscheiden.

(6) Auch technische Transparenzsysteme haben klare Grenzen

Transparenz-Instrumente können journalistische Sorgfalt oder kritische Medienkompetenz nicht ersetzen. Digitale Wasserzeichen oder Metadaten können entfernt und verändert werden oder verloren gehen, wenn ein Inhalt die Plattformen verlässt, auf der er ursprünglich veröffentlicht wurde. Zudem besteht die Gefahr einer trügerischen Sicherheit: Ein Inhalt, der nicht als KI-generiert gekennzeichnet ist, muss noch lange nicht unverfälscht oder menschlich erstellt sein.

(7) Transparenz braucht mehr Informationen als nur ein Label

Eine einfache Kennzeichnung – etwa ein Symbol oder kurzer Hinweis – liefert meist nur begrenzte Informationen. Für eine wirklich informierte Bewertung von Inhalten sind zusätzliche Details notwendig: Wurde ein Bild vollständig KI-generiert? Wurde ein Text nur übersetzt? Hier setzen sogenannte „Content Credentials“ an. Sie sollen über standardisierte Metadaten nachvollziehbar machen, wie ein Inhalt entstanden ist und welche Bearbeitungsschritte vorgenommen wurden.

(8) Eine Welt voller KI-Inhalte stellt das Konzept der Kennzeichnung selbst infrage

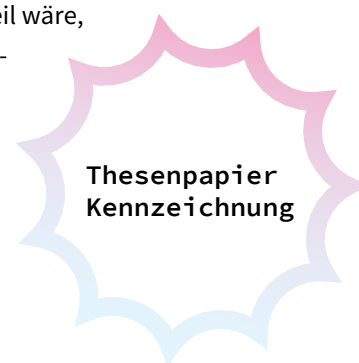
Doch auch eine differenzierte Kennzeichnungs-Politik steht vor absehbaren Herausforderungen: Wenn immer mehr Inhalte KI-gestützt produziert werden, könnte die Kennzeichnung an Aussagekraft verlieren. Diskutiert werden sollte daher auch der umgekehrte Weg: die explizite Kennzeichnung menschlich geschaffener Inhalte, insbesondere in Umgebungen, in denen KI-Outputs zur neuen Norm werden (wie etwa der Grafik oder Content Creation).

(9) Kennzeichnung als Teil einer vertrauensbildenden Compliance-Strategie

Wer Vertrauen in Medieninhalte nachhaltig sichern will, braucht mehr als ein Label in der Ecke eines KI-Outputs. Sinnvoll erscheint stattdessen eine konsequente, niedrighschwellige und überprüfbare Offenlegung von KI-gestützten Arbeitsprozessen, jedenfalls in besonders sensiblen Bereichen. Es braucht interne Leitlinien, wann und wie KI eingesetzt wird, welche Rolle Menschen im Entscheidungs- und Qualitätsprozess behalten und wie diese Rollen gegenüber Nutzer*innen verständlich kommuniziert werden. Kennzeichnung kann hier ein wichtiges Transparenz-Element sein, aber erst im Zusammenspiel mit technischer Nachvollziehbarkeit, organisatorischen Standards und Medienkompetenz entsteht ein vertrauensschaffender Gesamtrahmen.

(10) Staatlich anerkannte Siegel schaffen Vertrauen

Vertrauen entsteht nicht alleine durch Sichtbarkeit, sondern auch durch Glaubhaftigkeit. Gerade deshalb können staatlich oder gesellschaftlich etablierte Stellen eine wichtige Rolle spielen: Sie verleihen einer Kennzeichnung besondere Überzeugungskraft, insbesondere dann, wenn sie auf unabhängigen Prüfungen und klaren Standards beruhen. Der Vorteil wäre, dass ein solches Siegel nicht von einem KI-Anbieter, sondern von einer neutral wahrgenommenen Instanz angebracht würde.



3 Methodenmix für innovative Begleitforschung: Unser Weg zu zentralen Erkenntnissen

Für unser Forschungsvorhaben kombinieren wir mehrere qualitative Forschungsmethoden. Damit verfolgen wir zwei Ziele: Einerseits wollen wir bestehende rechtliche Rahmenbedingungen und die dazu vertretenen Positionen in Wissenschaft und öffentlichen Debatten erfassen. Andererseits wollen wir die Perspektiven handelnder Akteur*innen und die Dynamik bezüglich Herstellung, Einsatz und Nutzung von KI-Systemen und -Anwendungen berücksichtigen.

Wichtiger methodischer Ausgangspunkt ist, dass wir uns zum einen auf vier Arbeitsbereiche fokussieren: Kultur- und Kreativwirtschaft, Journalismus, Bildung und öffentliche Verwaltung. Zum anderen unterscheiden wir bei unseren Bestandsaufnahmen und Ermittlungen drei Phasen im Lebenszyklus eines KI-Systems: 1.) Entwicklung, 2.) Einsatz und 3.) Nutzung des KI-Outputs. Wir nennen diese mehrdimensionale Herangehensweise auch unsere Forschungsmatrix.

Kick-Off und Vor-Recherchen

Quellen-Recherchen & Leitfaden-Interviews



01/24 – 07/25

Projekt-Zeitstrahl

Literaturrecherche

Im ersten Projektschritt beginnen wir mit einer umfangreichen Literaturrecherche in verschiedenen Disziplinen. Das Entwicklungstempo im Bereich KI ist hoch, zugleich wird hierzu sehr viel publiziert. Eine umfassende Auswertung aller einschlägigen Publikationen wäre unmöglich. Für die Basisrecherchen werten wir über 350 wissenschaftliche, fachspezifische und journalistische Aufsätze, Untersuchungen, Stellungnahmen, Positionspapiere und Studien aus, sowie juristische Kommentare und Weiteres – insbesondere aus dem Zeitraum von Januar bis September 2025.

Leitfadeninterviews

Die recherchierten Informationen, Zusammenhänge und Einschätzungen wollen wir einordnen und auf ihre praktische Relevanz für den jeweiligen Arbeitskontext überprüfen. Dazu führen wir insgesamt 30 leitfadengestützte Expert*innen-Interviews mit einer Dauer von jeweils bis zu 60 Minuten durch. Bei der Auswahl der interviewten Fachleute – aus Deutschland und dem europäischen Ausland – achten wir auf eine gleichmäßige Repräsentation aller vier Arbeitsbereiche und der oben beschriebenen drei Phasen des Lebenszyklus von KI.

02/25

Erstausgabe prompt/-Newsletter



E-Mail-Newsletter prompt/

Parallel zu Recherchen und Interviews starten wir im Februar 2025 den E-Mail-Newsletter prompt/. Mit diesem etwa zweimonatlich erscheinenden Format begleiten wir unsere Forschungsarbeit auf journalistische Weise. Ausgehend von einem Leitthema gehen wir in jeder Ausgabe auf aktuelle Ereignisse und Entwicklungen ein, erläutern Hintergründe oder führen Gespräche mit Expert*innen aus der Praxis.

09/25

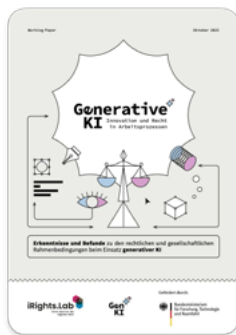
Lunch Talk



Lunch-Talk

Um einen ersten Einblick in die für uns wichtigsten Erkenntnisse und Befunde zu geben, führen wir im September 2025 einen sogenannten Lunch-Talk durch. Bei der Online-Veranstaltung gehen wir schwerpunktmäßig auf das elementare Thema Haftung ein, mit Kurzvorträgen und anschließender Diskussion.

Working Paper



Working Paper

Zum Abschluss der ersten Projektphase und zur Bündelung der bis dahin recherchierten und ermittelten Forschungsergebnisse erarbeiten wir ein Working Paper. Das Paper erscheint im Oktober 2025 (unter der offenen Creative Commons-Lizenz CC BY 4.0) und lässt sich kostenlos herunterladen. Darin arbeiten wir erste Thesen und Erkenntnisse zu unseren drei Kernthemen – Datenschutzgesetze, urheberrechtliche Konflikte und Vergütung sowie Kennzeichnung von KI-Outputs – heraus, identifizieren offene Rechtsfragen und skizzieren dringende Handlungsbedarfe bezüglich der Regulierung rund um den Einsatz generativer KI.

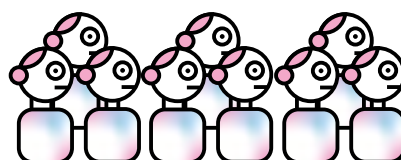
11/25 – 03/26

Zukunftslabore



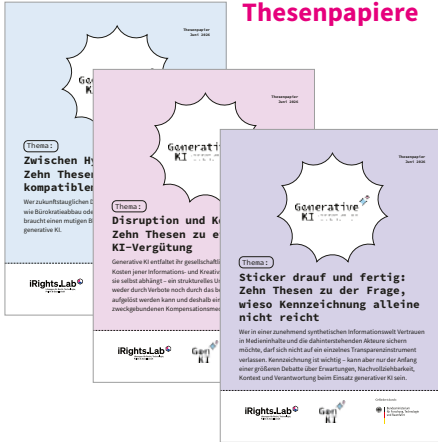
Zukunftsszenarien

Um die Ergebnisse des Working Papers zu vertiefen und offene Fragen zu diskutieren, führen wir in der zweiten Projektphase insgesamt neun Workshops mit KI- und Rechts-Expert*innen, KI-Praktiker*innen und KI-Nutzer*innen aus den von uns untersuchten Arbeitsbereichen durch. Dabei finden zu jedem unserer drei Kernthemen – Datenschutz, Vergütung und Kennzeichnung – je drei Workshops (auch „Zukunftslabore“ genannt) statt.



06/26

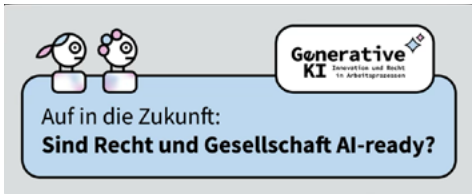
Thesepapiere



Thesepapiere

Die Erkenntnisse aus den Zukunftslaboren fassen wir in Thesepapieren zusammen. Die Schlussfolgerungen und Gedankenexperimente sollen als Denkanstöße für weitere Diskussionen dienen.

Abschlusserevent



Abschlusserevent

Zum Abschluss der Projektlaufzeit bringt eine Präsenz-Veranstaltung Expert*innen aus Politik, Forschung, Kreativbranche, öffentlicher Verwaltung und Zivilgesellschaft zusammen – mit dem Ziel, die Ergebnisse des Projektes zu diskutieren und mit aktuellen, praktischen Perspektiven zu verbinden.

Der GenKI-IR-Report



07/26

12 Ausgaben prompt/



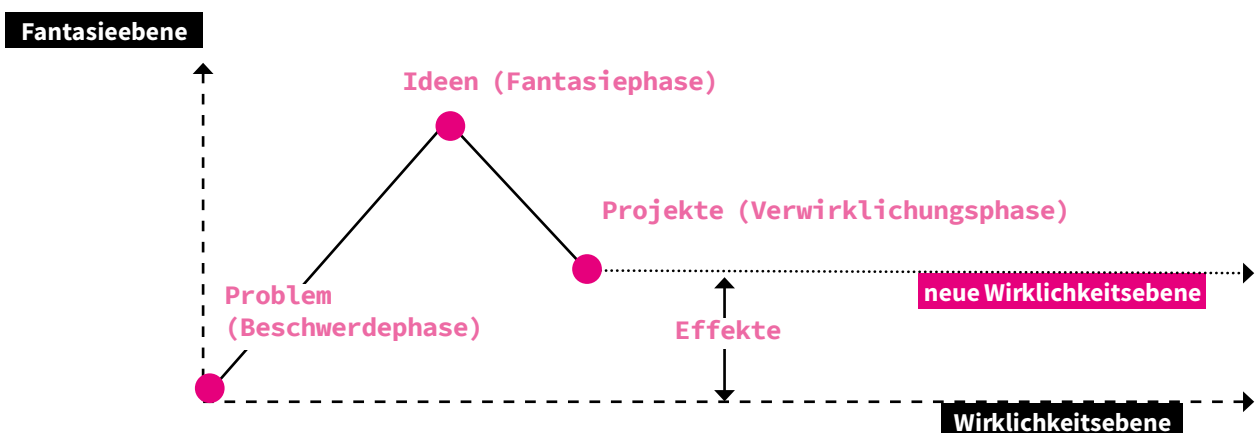
3

Zukunftsszenarien: KI-Rechtsrahmen neu gedacht

Das Format „Zukunftsszenario“

Die Zukunftsszenarien basieren auf dem gleichnamigen methodischen Ansatz, der in den 1970er Jahren von den Zukunftsforschern Robert Jungk, Rüdiger Lutz und Norbert Müllert entwickelt wurde. Er wurde in sehr unterschiedlichen Kontexten erprobt, etwa in der Stadtentwicklung, in Organisationen oder in zivilgesellschaftlichen Initiativen. Im Mittelpunkt dieser Methode steht die Zusammenarbeit sehr heterogener Gruppen, die sich in mehreren Phasen mit einem bestimmten Thema auseinandersetzen, um Problemlagen, Zielvorstellungen und Lösungswege zu erörtern.

Charakteristisch für ein Zukunftsszenario ist, dass nicht allein Expert*innen-Wissen abgefragt wird, sondern auch Praxiserfahrungen und Gestaltungsansprüche von Teilnehmer*innen systematisch mit einbezogen werden. In unserem Fall waren das etwa Personen, die beruflich im Themenbereich des jeweiligen Zukunftsszenarios tätig sind oder in diesem Bereich KI-gestützte Produkte oder Dienste nutzen. Im Rahmen dieser Zukunftsszenarien wurden insgesamt neun Zukunftslabore durchgeführt (drei pro Szenario), die wiederum in drei Phasen gegliedert waren. Die drei Phasen sind im Modell von Jungk, Lutz und Müllert wie folgt definiert:



Grafik unter <https://jungk-bibliothek.org/wp-content/uploads/2016/08/zw-eincf3bchrung-hh-wsp.pdf>
S. 2, Primär-Quelle: Kuhnt/Müllert 1997, S. 13

Kritik- und Beschwerdephase: Hier werden zunächst bestehende Zustände, Ungereimtheiten und Frustrationen systematisch gesammelt. Diese Phase darf „ungehemmt“ sein, um Vorbehalt und Frust abzubauen. In dieser Phase werden oft institutionelle, rechtliche und praktische Probleme sichtbar.

Phantasie- und Utopiephase: In der zweiten Phase wird bewusst ein Schutzraum für ideale Zustände eröffnet. Die Leitfrage ist dabei: „Wie könnte die Realität sein, wenn wir nicht durch Machtverhältnisse, Budgets oder bestehende Institutionen beschränkt wären?“ Hierbei soll insbesondere die Kreativität gefördert und Lösungsmöglichkeiten entwickelt werden. Das Ziel ist, sich von der unmittelbaren Zwangsläufigkeit des Bestehenden zu befreien und Raum für neue – durchaus utopisch anmutende – Ideen zu schaffen.

Realisierungsphase: Schließlich geht es darum, die Wunschvorstellungen und utopischen Entwürfe mit realen Gegebenheiten und Optionen abzugleichen. Jungk, Lutz und Müllert nennen das die Wirklichkeitsebene. Gemeint sind bestehende oder entstehende regulatorische Ansätze sowie institutionelle und praktische Spielräume. Die dritte Phase sucht nach Anschlussmöglichkeiten und realen Veränderungspotentialen. Zwar zeigt sich, dass nicht jedes Ideal und nicht jede Utopie realisierbar sein wird. Doch die entworfenen Lösungswege oder Projekte können dazu beitragen, dass eine neue Wirklichkeitsebene skizziert wird, die zu neuem Denken und Handeln inspiriert und motiviert.

Die Zukunftslabore als Methode schaffen Raum für Austausch und ermöglichen einen Realitätsabgleich, der in vielen Governance-Settings unberücksichtigt bleibt: Denn Regulierung bedeutet nicht nur abstrakte Normsetzung, sondern berührt unmittelbar die Handlungsfähigkeit von Menschen und Organisationen.

Drei Zukunftsszenarien: Themen, Teilnehmende, Ergebnisse

Zwischen Oktober 2025 und Februar 2026 durchlaufen wir mit insgesamt 30 Teilnehmenden drei Zukunftsszenarien: Eines zum Thema Datenschutz, eines zum Thema Urheberrecht sowie eines zum Thema Vergütung und Kennzeichnung. Jedes der Szenarien besteht aus drei individuellen Workshopst – den sogenannten Zukunftslaboren. Im ersten Labor widmen wir uns einer gemeinsamen Problemdefinition. Im zweiten Labor entwickeln die Teilnehmenden eine gemeinsame utopische Version, um diese dann im dritten Labor mit der Realität abzugleichen.

Jedes unserer drei Zukunftsszenarien setzen wir bewusst aus einer diversen Gruppe an Teilnehmenden zusammen. Ziel ist es, in den jeweiligen Gruppen eine gute Balance zwischen Akteur*innen-Vielfalt und diskursfähiger Gruppengröße zu erreichen. Letztlich begrenzen wir die Anzahl der Teilnehmenden deshalb auf jeweils rund zehn Personen.

Die Auswahl der Teilnehmenden zielt nicht darauf ab, bestehende Konflikte zwischen Interessengruppen zu glätten oder konträre Standpunkte zu nivellieren. Vielmehr sollen Spannungen sichtbar werden, um sie als Ausgangspunkt für die Suche nach Lösungsansätzen zu nutzen, die für so viele Teilnehmer*innen wie möglich tragfähig sind. Anders gesagt: Die Zukunftsszenarien sollen produktiven Dissens fördern.



Zukunftsszenarien

Hinweis: Die nachfolgenden Ausführungen geben kurze Zusammenfassungen des Stimmungsbildes sowie der Kernpunkte der jeweiligen Zukunftsszenarien wieder – so, wie sie sich für uns beziehungsweise für unsere hier Bericht erstattenden Teammitglieder dargestellt haben.

3

Zusammenfassung:

In den drei Zukunftsszenarien wird deutlich, dass Fragen des Schutzes, der Regulierung und der ökonomischen Verteilung nicht isoliert verhandelt werden können. Die Beteiligten beschreiben ein Regelungsumfeld, das aus unterschiedlichen Rechtsgebieten und Zuständigkeiten zusammengesetzt ist und in der praktischen Anwendung nur schwer überschaubar scheint.

Für alle drei behandelten Bereiche – Datenschutz, Vergütung und Kennzeichnung – skizzieren die Teilnehmenden eine strukturelle Überforderung, die sich bei Verwaltungen ebenso zeigt wie bei Kreativschaffenden, Medienunternehmen und Aufsichtsstellen. Zugleich wird dieser Befund jedoch nicht als unveränderlicher Zustand wahrgenommen.

Aus den Zukunftsszenarien gehen Vorschläge hervor, die auf institutionelle, rechtliche und technische Gestaltungsmöglichkeiten zielen. Ebenso wird in allen Bereichen KI-Governance als Querschnittsaufgabe verstanden, die Datenschutz, Urheberrecht – und generell Medienrecht – sowie Wettbewerbspolitik miteinander verbindet.



Zukunftsszenario KI und Datenschutz:

Unterschiedliche Schutzlogiken erfordern neue Ausrichtungen

Ausgangsfragen: Wie lässt sich die Datenschutz-Grundverordnung (DSGVO) auf die Praxis des KI-Trainings mit großen, heterogenen Datensätzen anwenden, ohne dass Innovationshemmnisse entstehen oder Schutzstandards aufgeweicht werden? Welche Lücken entstehen in der Regulierungsarchitektur, wenn risikobasierte Ansätze der KI-Verordnung (KI-VO) auf die individualrechtliche Datenschutzlogik der DSGVO treffen?



Zentrale Erkenntnisse:

DSGVO trifft KI-Training: Heterogene Trainingsdaten enthalten oft Personenbezüge, sind aber kaum filterbar. Eine pauschale Rechtsgrundlage fehlt und jede Verarbeitung muss einzeln gerechtfertigt werden.

Ungeklärte Rechtsfragen: Auskunfts- und Löschrechte lassen sich in KI-Systemen kaum umsetzen. Personenbezug und Haftung entlang der Wertschöpfungskette sind rechtlich ungeklärt.

Systemkonflikt: KI-VO und DSGVO folgen unterschiedlichen Schutzlogiken – das erzeugt Rechtsunsicherheit und bedroht entweder Innovation oder Datenschutzstandards.

Teilnehmende: In diesem Szenario bringen wir Vertreter*innen von Datenschutzbehörden, Privatwirtschaft, Stiftungen und NGOs sowie Wissenschaft und öffentlicher Verwaltung zusammen. Diese Mischung soll sowohl die Unternehmens- und Verwaltungssicht auf Innovationshemmnisse nachvollziehbar machen als auch die Sicht von Zivilgesellschaft und Behörden auf Schutzbestrebungen.

Spannungsfeld und Problemlage: Die Anwendung der DSGVO auf generative KI führt zu Unsicherheit bei Verantwortlichen und Aufsichtsstellen sowie zu Konflikten zwischen Innovationszielen und Datenschutzerfordernungen. Gerade in der Trainings- und Entwicklungsphase von großen KI-Modellen entstehen nach bestehendem Recht schwer zu lösende Herausforderungen: Verwendete Trainingsdatensätze sind



3

regelmäßig heterogen – einzelne Datenpunkte können einen Personenbezug enthalten und damit der DSGVO unterliegen. Gleichzeitig sind die Datensätze zu groß, um nach solchen personenbezogenen Daten gefiltert werden zu können. Datenpunkte, die als besonders schützenswert nach Artikel 9 DSGVO einzuordnen sind, stellen die Rechtskonformität des KI-Trainings insgesamt in Frage. Denn eine „pauschale“ Rechtsgrundlage für KI-Training besteht nämlich gerade nicht: Jede einzelne Datenverarbeitung muss für sich genommen gerechtfertigt werden.

Gleichzeitig sind bei Anwendung der datenschutzrechtlichen Grundsätze auf generative KI-Systeme noch nicht alle Fragen zur KI-spezifischen Umsetzung geklärt. Wo und in welcher Form entstehen Kollisionen mit dem Datenschutzrecht? Wie werden der Schutzbereich und der Personenbezug beziehungsweise die Personenbeziehbarkeit definiert? Wie lassen sich Auskunfts- und Löschrechte von Betroffenen umsetzen, wenn ein KI-System zwar keine einzelnen Datenpunkte enthält, diese sich jedoch aufgrund der im Modell angelegten Wahrscheinlichkeits-Zusammenhänge möglicherweise jederzeit reproduzieren lassen? Daraus ergeben sich für die rechtssichere Entwicklung und Anwendung von KI-Technologien – entlang der gesamten KI-Wertschöpfungskette – erhebliche Fragen. Etwa, wer richtigerweise für welche Datenschutzverletzung wie verantwortlich und haftbar zu machen ist.

Zentrale Ergebnisse aus der Kritik- und Beschwerdephase: Die Teilnehmenden teilen das Problembewusstsein darüber, dass das Datenschutzrecht in jetziger Form nicht auf die Entwicklung von KI-Systemen angepasst ist. Dabei wird jedoch auch darauf hingewiesen, dass nicht alle Problemstellungen in gleicher Weise notwendigerweise vom Gesetzgeber zu lösen wären.

Bestätigt wird etwa das Problem heterogener Datensätze – also etwa Daten, die aus unterschiedlichen Quellen stammen, unterschiedliche Schnittstellen aufweisen und nicht

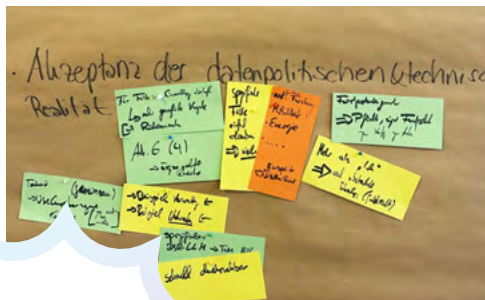
einheitlich aufgebaut beziehungsweise strukturiert sind – und der damit verbundenen Herausforderungen bei der KI-Entwicklung.

Diese betreffen insbesondere die europäischen und nationalen Rechtsprechungen, die den Begriff der personenbezogenen Daten extensiv auslegen und das Ziel „anonymer“ Daten und KI-Systeme damit zusätzlich erschweren. Insgesamt wird die granulare Betrachtungsweise

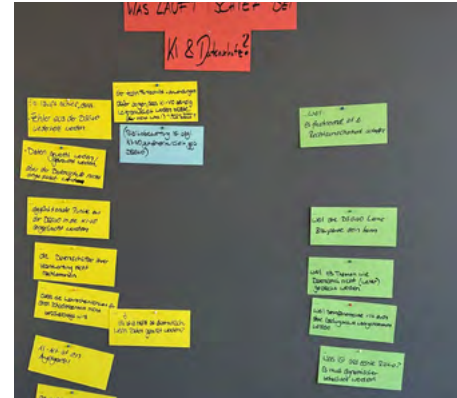
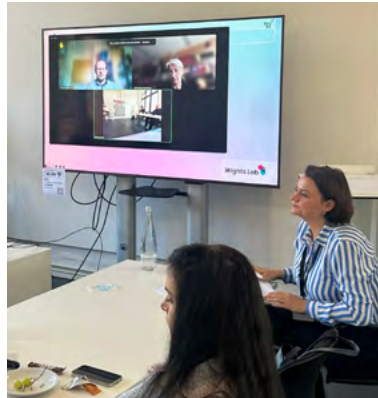
des bestehenden Datenschutzrechts in der Umsetzung als sehr aufwändig eingeschätzt – erst recht im Zusammenhang mit Systemen, die auf massenhafter Datenverarbeitung beruhen.

Kritisiert wird zudem, dass Diskussionen zu häufig normativ geführt würden. Zu selten würden etwa technische Lösungen oder das Potenzial standardisierter Datenschutzlösungen mitgedacht.

In der Berichterstattung und Diskussion über Datenschutz im Zusammenhang mit KI dominierten laut den Teilnehmenden Angst und Sorge. Viel zu selten würde die Wahrscheinlichkeit eines eintretenden Schadens wirklich durchdrungen. Dabei könne man auch sehr gut über Vorzüge und Lösungsansätze der bestehenden Regulierung und Umsetzung berichten. Dennoch dominierten eher negative Berichte. Ein Problem, das Datenschutzdebatten seit eini-



Zukunfts-szenario Datenschutz



gen Jahren anhaftet und das nun auch in die Diskussionen über KI Einzug hält. Insgesamt leide der Datenschutz gerade im Kontext (generativer) KI-Systeme unter einem Image-Problem.

Lösungsimpulse aus der Utopie-Phase: Viele Teilnehmende wünschen sich zuallererst eine neue Perspektive auf die Debatte über den Datenschutz: Ein produktiver, konstruktiver und damit insgesamt positiverer Ton sollte ein neues Datenschutznarrativ prägen, das Datenschutz als „Ermöglicher guter KI“ versteht.

Dafür sollten verbindende und konstruktive Elemente der bestehenden Datenschutzaufsicht deutlicher hervorgehoben und als souveräne, praxisnahe Akteur*innen sichtbar werden – so etwa die Datenschutzkonferenz der Länder.

Einen Beitrag hierzu sollten datenschutzkonforme Best Practices spielen – etwa durch White Lists, differenzierte Handlungsanregungen für unterschiedliche Branchen und Unternehmensgrößen und harmonisierte Leitlinien. Auch Transparenzregister besonders vorbildlicher Datenschutzumsetzungen aus öffentlicher Hand und Privatwirtschaft könnten hier eine Rolle spielen.

Insgesamt müsse aber zugleich das Datenschutzrecht in Theorie und Praxis stärker auf datenpolitische und technische Realitäten angepasst werden. Der Anwendungsbereich des Datenschutzrechts und damit der Begriff der personenbezogenen Daten als solcher müsse risikodifferenziert überdacht werden. Bewährte Methoden der Pseudonymisierung und Anonymisierung müssten weitergedacht werden und im Ergebnis zu einem flexibleren Datenschutz führen. Gleichzeitig könnten auch spezifische Datenverarbeitungskontexte explizit genehmigt werden – etwa im Bereich der medizinischen Forschung, der klimaneutralen Mobilität oder der Energieversorgung. Jedenfalls brauche es angepasste Rechtfertigungsgründe – gerade auch im Bereich der besonders schützenswerten Daten aus Artikel 9 DSGVO.

3

Neue Wirklichkeitsebene aus der Realisierungsphase: In einer differenzierten Auseinandersetzung mit bereits veröffentlichten Reformvorschlägen der DSGVO werden Unterschiede und Parallelen verglichen und mit den bisherigen Ergebnissen der Zukunftswerkstatt abgeglichen. Im Fokus stehen dabei etwa der Entwurf zum Digital-Omnibus der EU-Kommission sowie der Vorschlag des European Law Institute „Simpler, Fairer, More Effective – Towards a Targeted Revision of EU Data Protection Law“.

Mehrere Teilnehmende sprechen sich dafür aus, dass die DSGVO mit einigen der bereits vorhandenen Vorschläge in nachvollziehbarer Weise „KI-fit“ gemacht werden könne – ohne den Datenschutz insgesamt zu opfern. Ein Ansatz hierfür könnte – neben erweiterten Rechtfertigungsgründen – die Etablierung von Risikokategorien nach dem Vorbild der KI-Verordnung sein. Auch sonst müssten Datenschutz und die übrigen Digitalrechtsvorschriften der EU noch besser aufeinander abgestimmt werden. Verwiesen wird hier zudem auf die Problematik einer noch nicht abgeschlossenen Rechtsprechungsentwicklung zur DSGVO.

Auch, weil diese Punkte nicht in allen Vorschlägen und Entwürfen in gleicher Konsequenz bearbeitet werden, wird hier keine Einigkeit erzielt. Deutlich wird, dass die Anpassungsbedarfe im Datenschutz so fundamental sind, dass auch über die Grundlagen des Datenschutzrechts diskutiert wird: Handelt es sich auch in Zukunft um eine Frage der Grundrechtsabwägung?

Fazit:

Bei den Teilnehmenden besteht breiter Konsens, dass die DSGVO in ihrer aktuellen Form nicht KI-fit ist. Daher sollte sie reformiert werden, in Teilen einem risikobasierten Ansatz folgen und eine Differenzierung der Akteur*innen ermöglichen. Insbesondere Anwendungsbereich und Rechtfertigungsgründe bedürfen einer KI-spezifischen Anpassung. Differenzen gibt es, neben praxisrelevanten Detailfragen, vor allem bezüglich der Frage, wie grundlegend die Umbrüche im Datenschutzrecht angegangen werden sollten.



Zukunftsszenario Vergütungen:

Über den Horizont des Urheberrechts hinaus denken

Ausgangsfragen: Was läuft bei der Vergütung kreativer Arbeit im Kontext generativer KI schief – und warum? Lassen sich die Probleme in diesem Bereich allein mit den Hebeln des Urheberrechts lösen? Oder bedarf es ergänzender steuer-, förder- und infrastrukturpolitischer Instrumente?



Zentrale Erkenntnisse:

- - - - - **Rechtsunsicherheiten:** Bezüglich der Verwendung urheberrechtlich geschützter Werke im KI-Training und -Diensten mangelt es an rechtlicher Klarheit und an funktionierenden Kompensationslösungen.
- - - - - **Strukturelle Benachteiligung:** Kreativschaffende sind gegenüber globalen Plattformkonzernen unterlegen und befürchten, neue Vergütungsmodelle kommen primär großen Rechteinhabern zugute.
- - - - - **Neue Ansätze:** Moderne Vergütungslösungen sollten übergreifend wirken, etwa mit Abgabepflichten für Betreiber disruptiver Systeme und Dienste. Auch urheberrechtliche Regeln müssen angepasst werden.

Teilnehmende: Am Vergütungsszenario nehmen Fachjurist*innen, Kultur- und Kreativschaffende sowie Vertreter*innen von Verbänden, Verlagen, Medienunternehmen, Verwertungsgesellschaften und zivilgesellschaftlichen Organisationen teil. Hier ist es besonders wichtig, verschiedene Kultur- und Kreativbranchen einzubeziehen, da sie teils vor sehr unterschiedlichen Herausforderungen stehen.

Spannungsfeld und Problemlagen: Die Nutzung urheberrechtlich geschützter Werke für das Training großer Sprachmodelle und generativer KI-Systeme führt bei Kreativschaffenden, Verwerter*innen und KI-Anbietern zu erheblicher Rechtsunsicherheit – insbesondere in Bezug auf gesetzliche Nutzungsbefugnisse und Vergütungen. Die KI-Hersteller berufen sich auf urheberrechtliche Ausnahmen, bieten jedoch in Einzelfällen vor allem großen Verwertern Kompensationen per vertraglichen Kooperationen an. Zugleich werden in zahlreichen Gerichtsverfahren in Europa und den USA technisch-regulatorische Fragen verhandelt. Dabei geht es auch um mangelnde Transparenz darüber, welche Werke in Trainingsdaten stecken und wie die Teilhabe an Wertschöpfung organisiert ist.



3

Gesellschaftlich betrachtet stehen sich dabei Innovationsziele – Entwicklung und Skalierung generativer KI für allgemeine Einsatzzwecke – und die Schutzinteressen von Urheber*innen und Rechteinhaber*innen gegenüber. KI-Werkzeuge dringen zunehmend in die Erschaffung, Produktion und Distribution künstlerisch-kultureller Werke und Dienstleistungen ein. Parallel dazu drängen synthetisch erzeugte Inhalte auf die Kulturmärkte, wo sie mit menschlichen Kreativleistungen konkurrieren und sie potenziell entwerten.

In diesem Feld treffen mehrere regulatorische Ansätze und Rechtsregime aufeinander: die KI-VO mit ihrem risikobasierten Systemansatz und ihrem Fokus auf Sicherheit, Transparenz und Marktaufsicht; das europäische Urheberrecht mit seinem Fokus auf Schöpfer*in und Werkschutz, exklusive Rechte, sowie individuelle und kollektive Lizenzierungen; und nicht zuletzt sektorspezifische Regelungen, wie urheberrechtliche Leistungsschutzrechte oder das Medienrecht. Angesichts dieser offenen Baustellen ist es fraglich, ob es das Urheberrecht als primäres Instrument leisten kann, die Kreativ- und Kulturwirtschaft wirksam zu schützen. Muss es möglicherweise mit anderen Regimen (wie Steuerrecht, Kartellrecht und Medienrecht) verzahnt werden, um KI-Hersteller und Anbieter, die von den Produkten profitieren, die in diesen Branchen geschaffen werden, zu Zahlungen zu verpflichten?

Zentrale Ergebnisse aus der Kritik- und Beschwerdephase: Kreativschaffende beklagen, dass ihre Werke für KI-Trainings und KI-Dienste genutzt werden, ohne dass sie gefragt oder vergütet werden. Sie erleben kollektiven Frust über die Diskrepanz zwischen der kulturellen Bedeutung ihrer Arbeit und der tatsächlichen Vergütung. Es gibt kaum „Standardwege“, wie die Verwendung urheberrechtlich geschützter Inhalte für das KI-Training rechtssicher lizenziert oder vergütet werden kann. Denkbare Vergütungsmodelle erfordern rechtspolitische Entscheidungen und systemische Neuansätze. Branchen mit starken Verwertungsgesellschaften, etwa die GEMA für den Musikbereich, sind besser aufgestellt als stark fragmentierte Bereiche, beispielsweise freie Journalist*innen und Autor*innen.

KI-Hersteller und Rechteinhaber*innen schließen zwar vereinzelt separate Verträge über Kompensationszahlungen ab, doch im Verhältnis zu den Abermillionen Rechteinhaber*innen, deren Werke in KI-Technologien sang- und klanglos Einzug gehalten haben, sind das bislang marginale Ausnahmen. Außerdem sind viele Kreativschaffende gegenüber globalen Plattformkonzernen strukturell unterlegen – sie haben Sorge, dass neue Vergütungsmodelle wieder primär großen Rechteinhabern zugutekommen. Als reales Risiko mit gesellschaftlichen Folgen nehmen die Teilnehmenden das andauernde „Sterben der Zeitungsverlage“ und die Erosion der Medienvielfalt wahr. Es besteht die Gefahr, dass KI-Systeme insbesondere dem Journalismus die wirtschaftlichen Grundlagen entziehen.

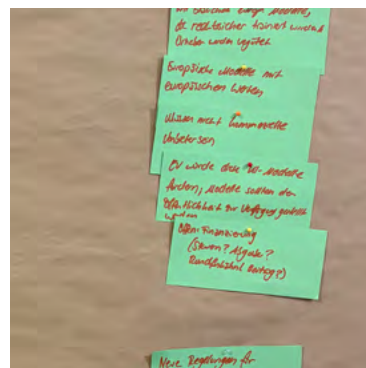
Lösungsimpulse aus der Utopie-Phase: Die Teilnehmenden erarbeiten zunächst drei Leitmotive für ihre Wunschvorstellungen: Angestrebt wurden verlässliche, existenzsichernde Vergütungsmodelle, kollektive Strukturen, die Einzelne entlasten, sowie eine Kombination aus marktfinanzierten und öffentlich getragenen Komponenten.

So entsteht in einer Gruppe die Idee verschiedener Bausteine, mit denen bestehende Vergütungslogiken nicht nur „repariert“, sondern transformiert werden sollen. Ein Ansatz skizziert eine umsatzbasierte Steuer, die kommerzielle Anbieter und Betreiber von KI-Systemen für Vertrieb und Angebot von KI-Technologien auf dem europäischen Markt entrichten sollen. Die daraus resultierenden Einnahmen sollten in den Aufbau einer öffentlichen KI-Infrastruktur fließen (also KI-Technologien, die von der öffentlichen Hand angeboten

**Zukunfts-
szenario
Vergütungen**



werden) und als Kompensation an Autor*innen, Unternehmen und Betreiber gemeinnütziger Informationsinfrastruktur (wie Wikimedia) gezahlt werden. Alternativ oder ergänzend wird eine spezifische, zweckgebundene Abgabe für jene Hersteller/Anbieter/Betreiber disruptiver Technologien vorgeschlagen, die bestehende Märkte nachweislich transformieren und hierbei massiv beschädigen (Disruption Levy). Mit einer solchen Abgabe sollen sich Akteure, die von den disruptiven Effekten ihres Marktverhaltens erheblich profitieren, wirtschaftlich an den von ihnen (mit-)verursachten gesellschaftlichen und ökonomischen Kosten der Disruption beteiligen. Auch eine solche Abgabe sollte bevorzugt betroffenen Branchen, Kreativschaffenden und Akteuren der Informationsinfrastruktur zugutekommen. Hiermit könnten beispielsweise Umschulungen finanziert, wirtschaftliche Einbußen kompensiert und Übergangshilfen beziehungsweise andere Strukturwandelmaßnahmen finanziert werden. Als Vorbilder könnten die Fonds zum Kohleausstieg und regionale Strukturprogramme dienen. Beide Vergütungsmodelle basieren auf einem anderen Ansatz als Urheberrechtsvergütungen. Sie knüpfen nicht an eine – wie auch immer geartete – Werknutzung durch KI-Hersteller/Anbieter/Betreiber an, sondern an unternehmerische Umsätze und/oder den Kosten von kausal zuzuordnenden disruptiven Effekten. In Bezug auf die möglichen Nutznießer*innen gehen sie über urheberrechtliche Vergütungsansätze hinaus, da hiervon nicht nur Inhaber*innen von Urheber- und Verwertungsrechten profitieren könnten, sondern eine bedeutend größere Bandbreite von Akteur*innen. Details wie Abgabehöhe und Verteilungsfragen blieben offen und wären vertiefend zu diskutieren.

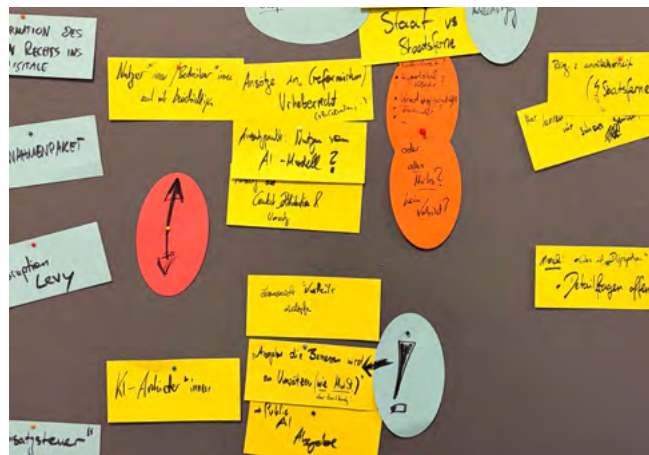


Auch die zweite Gruppe verfolgt die Idee, private KI-Anbieter wirtschaftlich in die Verantwortung zu nehmen. Außerdem sollte als Alternative zu privatwirtschaftlichen Angeboten eine öffentliche, gemeinwohlorientierte KI-Infrastruktur aufgebaut werden. Konkret sollen EU-finanzierte KI-Modelle als Teil der öffentlichen Infrastruktur verfügbar sein, analog zum öffentlich-rechtlichen Rundfunk. Zu diesen öffentlichen KI-Angeboten soll es offenen oder privilegierten Zugang für Bildung, Kultur, Medien und Forschung geben. Sie erfüllen hohe Standards beim Datenschutz, der Transparenz der Trainingsdaten und der Sicherheit. Ziel ist es, sowohl gesellschaftlich wünschenswerte KI-Systeme als auch darauf aufsetzende Projekte mit besonderem gesellschaftlichem Mehrwert, beispielsweise lokale Nachrichten oder Kultur-

3

projekte, zu fördern. Die Finanzierung soll aus Steuermitteln und privaten Beiträgen erfolgen, etwa über Public-Private-Partnerships. Offen blieb unter anderem die Frage, wer definieren würde, was „gesellschaftlich wünschenswert“ wäre.

Neue Wirklichkeitsebene aus der Realisierungsphase: Die Ergebnisse der Utopie-Phase der beiden Gruppen weisen erhebliche Gemeinsamkeiten auf. Die derzeitige Situation, in der einige Wenige erheblich vom Schaffen sehr Vieler einseitig profitieren, ohne etwas zurückzugeben, wurde weitgehend als unfair angesehen. Vergütungslösungen müssten hier für Kompensation und Abschwächung disruptiver Effekte sorgen. Solche können auf dem Urheberrecht basieren, allerdings wäre es damit nicht getan, um die Gesamtproblematik angemessen zu erfassen. Vergütungsfragen im KI-Kontext können nicht über ein einziges Rechtsgebiet gelöst werden, sondern sind als politisch-ökonomische Gestaltungsaufgabe zu betrachten. Auch reichen nationale Alleingänge im Zweifel nicht aus. Vielmehr sind gesamt-europäische und internationale Lösungsansätze erforderlich.




Fazit:

Die drei Zukunftslabore zum Thema „KI, Urheberrecht und Vergütung“ schaffen ein geteiltes, differenziertes Problembewusstsein. Konkrete, teilweise utopisch anmutende Modellvorschläge werden erarbeitet: KI-Steuer, Disruption Levy, öffentliche KI-Infrastruktur. Zugleich wird die Weiterentwicklung urheberrechtlicher Instrumente erörtert, etwa zu Vergütungen für KI-Trainings oder erweiterte kollektive Lizenzen. Wie sich im Verlauf dieses Zukunftsszenarios zeigt, werden hierzu bereits Debatten in Wissenschaft und Zivilgesellschaft geführt. Die erforderliche Prüfung und Ausgestaltung von Steuer- und Abgabinstrumenten zur Finanzierung eines Informations- und Kreativ-Ökosystems ist also nicht undenkbar. Gleiches gilt für den Aufbau öffentlicher beziehungsweise gemeinwohlorientierter KI-Infrastrukturen, samt Governance-Design und Finanzierungsmöglichkeiten.

Zukunfts-
szenario
Vergütungen

Zukunftsszenario Kennzeichnung:**Auf technische Lösungen und sektorale Standards setzen**

Ausgangsfragen: Wie und durch wen sollen KI-generierte Inhalte (KI-Outputs) verpflichtend gekennzeichnet werden? Welche technischen und Governance-bezogenen Anforderungen sind damit verbunden? Wie gut können technische Lösungen funktionieren und etwa vor bewusster Täuschung schützen? Welche Alternativen zu technischen oder regulatorischen Methoden gibt es?



Zukunftsszenario
Kennzeichnung

Zentrale Erkenntnisse:

Notwendig: Kennzeichnungen sollen vor Täuschung, Betrug und Verletzung von Persönlichkeitsrechten durch KI-generierte Inhalte schützen und Transparenz sicherstellen.

Schwierig: Für die Sanktionierung von Verstößen gibt es unterschiedliche Anforderungen je nach Branche, bestehende Kennzeichnungen werden kaum wahrgenommen.

Erforderlich: Robuste technische Lösungen, u. a. nachvollziehbare Metadaten, Wasserzeichen oder Prüfwerkzeuge, Verantwortungsbewusstsein von Medienakteur*innen, Stärkung von Medienkompetenz.

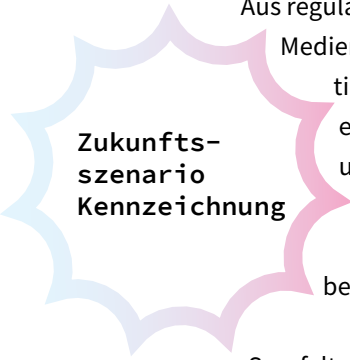
Teilnehmende: An diesem Zukunftsszenario nehmen Vertreter*innen von Medien, Web-Plattformen, Technologieanbietern, Behörden, Wissenschaft und Zivilgesellschaft teil. Die Herausforderungen liegen zum einen darin, in gleichem Maße abzubilden und herauszuarbeiten, was sowohl technisch machbar als auch regulatorisch realisierbar wäre. Zum anderen gilt es, mögliche Alternativen zu bisher debattierten Lösungen aufzuzeigen.

Spannungsfelder und Problemlage: Die Ausgangslage ist klar: Es ist für Nutzer*innen zunehmend schwer bis unmöglich, KI-generierte Inhalte (wie synthetische Bilder, Audios oder Videos) als solche zu erkennen. Aus regulatorischer Sicht scheint es also naheliegend, dass KI-generierte Inhalte gekennzeichnet werden sollen – insbesondere, wenn sie täuschend echt wirken. Auf der anderen Seite gibt es massive Zweifel daran, ob Kennzeichnung tatsächlich die angestrebte Wirkung erzielt. Unklar ist auch, ob Nutzer*innen Kennzeichnungen wahrnehmen und verstehen werden. Nicht zuletzt stellt sich die Frage, wer für das Kennzeichnen von Inhalten verantwortlich ist: Der Hersteller eines KI-Systems, die Betreiber/Anbieter einer KI-Anwendung oder diejenigen, die KI-Outputs veröffentlichen beziehungsweise die Plattformen betreiben, auf denen sie veröffentlicht werden?



3

Bei näherer Beschäftigung mit der Kennzeichnungsfrage ergeben sich weitere Unsicherheiten: So ist unklar, ab welchem Grad der Bearbeitung von Inhalten mittels einer KI-Anwendung eine Kennzeichnungspflicht bestehen soll. Problematisch ist beispielsweise, dass technische Kennzeichnungsmethoden, wie beispielsweise digitale Wasserzeichen oder fest verknüpfte Metadaten, letztlich zu überwinden sind. Menschen, die es darauf anlegen, mithilfe KI-generierter Inhalte zu täuschen oder zu betrügen, sind durch Kennzeichnungen kaum davon abzuhalten.


**Zukunfts-
szenario
Kennzeichnung**

Aus regulatorischer Sicht ist zudem zu berücksichtigen, dass im Bereich Presse und Medien bereits gesetzliche Vorgaben sowie freiwillige Selbstverpflichtungen existieren. So schreibt etwa die KI-Verordnung laut Artikel 50 vor, dass „Betreiber eines KI-Systems, das Text erzeugt oder manipuliert, der veröffentlicht wird, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren, [...] offenlegen [müssen], dass der Text künstlich erzeugt oder manipuliert wurde.“ Journalistische Medien müssen – wie auch andere Betreiber von KI-Systemen – Deepfakes offenlegen.

Zugleich formulieren Pressegesetze und Medienstaatsvertrag bestimmte Sorgfalts- und Transparenzanforderungen, etwa die eindeutige Kennzeichnung von Meinungsbeiträgen und Werbung. Daneben gibt es branchenweite Selbstverpflichtungen, wie den Pressekodex des deutschen Presserats. Zudem stellten Verbände und einzelne Medienhäuser und -unternehmen eigene Leitlinien zum Umgang mit KI-gestützten Werkzeugen und KI-generierten Inhalten auf. Sowohl Transparenzpflichten als auch Medien- und Presserecht sowie die selbstverpflichtende Leitlinien sollen dazu beitragen, die Öffentlichkeit zu informieren, Verbraucher*innen zu schützen und das Vertrauen in journalistische Inhalte zu stärken.

Zentrale Ergebnisse aus der Kritik- und Beschwerdephase: Einige Teilnehmende sehen die Kennzeichnung von KI-generierten und -manipulierten Inhalten als erforderlich an, um gegen Täuschung, Betrug oder Persönlichkeitsrechts-Verletzungen, wie etwa Deepfakes, Identitätsdiebstahl oder auch digitale Gewalt vorzugehen. Kennzeichnungen sollten zudem das Vertrauen in mediale Inhalte stärken und Seriosität signalisieren. Unklar bleibt, ab welchem Bearbeitungsschritt die Kennzeichnungspflicht einsetzen soll. Es wird konstatiert, dass bewusste Verstöße schwer zu sanktionieren sind und bestehende Regelungen (unter anderem Art. 50 KI-VO) nicht ausreichend durchsetzbar oder zu ungenau sind. Zudem zeigt sich, dass sich die Ansprüche an Kennzeichnungen je nach Branche oder Einsatzbereich unterscheiden: Während die Anforderungen an Transparenz im Journalismus tendenziell hoch sind, gibt es bezüglich Werbung weniger klare Erwartungen. Stattdessen gab es teilweise Skepsis, ob Kennzeichnungen hier tatsächlich einen Mehrwert bieten. Auch für die Bereiche Forschung, Bildung oder Unterhaltung sind abweichende Anforderungen denkbar. Die Teilnehmenden stellen zudem fest: Bereits existierende Kennzeichnungen werden nicht wirklich wahrgenommen. Es gibt kein absolut sicheres, „unkaputtbares“ Kennzeichnungssystem.

Lösungsimpulse aus der Utopie-Phase: Aus den Beratungen geht hervor, dass bei der Kennzeichnung KI-generierter oder -bearbeiteter Inhalte differenziert vorgegangen werden sollte: Eine minimale Bearbeitung, wie beispielsweise reine Farbanpassungen bei visuellen Inhalten, sollten in der Regel als nicht kennzeichnungspflichtig gesehen werden. Strukturelle Veränderungen, etwa das Entfernen oder Hinzufügen von Bildelementen, hingegen schon.

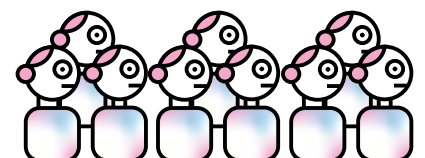
3

Fazit:

Die Teilnehmenden halten es für erforderlich, für KI-Kennzeichnungen robuste technische Lösungen zu schaffen, etwa durch nachvollziehbare Metadaten, Wasserzeichen oder Prüfwerkzeuge. Zugleich müssen Medienkompetenz und kritisches Denken gestärkt werden, um die Achtsamkeit beim Umgang mit medialen Inhalten zu erhöhen. Generell sind KI-Kennzeichnungen differenziert zu betrachten, je nach Branche und Nutzungsszenario.

**Zukunfts-
szenario
Kennzeichnung**

Während im Journalismus auf sehr strikte Kennzeichnungsanforderungen, hohe Transparenzstandards und redaktionelle Kontrolle zu pochen ist, gibt es in der Werbung andere Erwartungshorizonte und den Blick auf Verbraucherschutz. Für Social Media und Plattformen sollte es angesichts geteilter Verantwortlichkeiten - von Betreibern und Nutzer*innen - gegebenenfalls algorithmische Transparenzpflichten geben. Im Bereich Unterhaltung und Kunst können jeweils individuelle Abwägungen zwischen künstlerischer Freiheit und Transparenz sinnvoll sein, während es für Verwaltung und Justiz erhöhte Transparenz-Anforderungen geben sollte, damit dort getroffene Entscheidungen nachvollziehbar sind.



GenKI-IR Zukunftsszenarien:

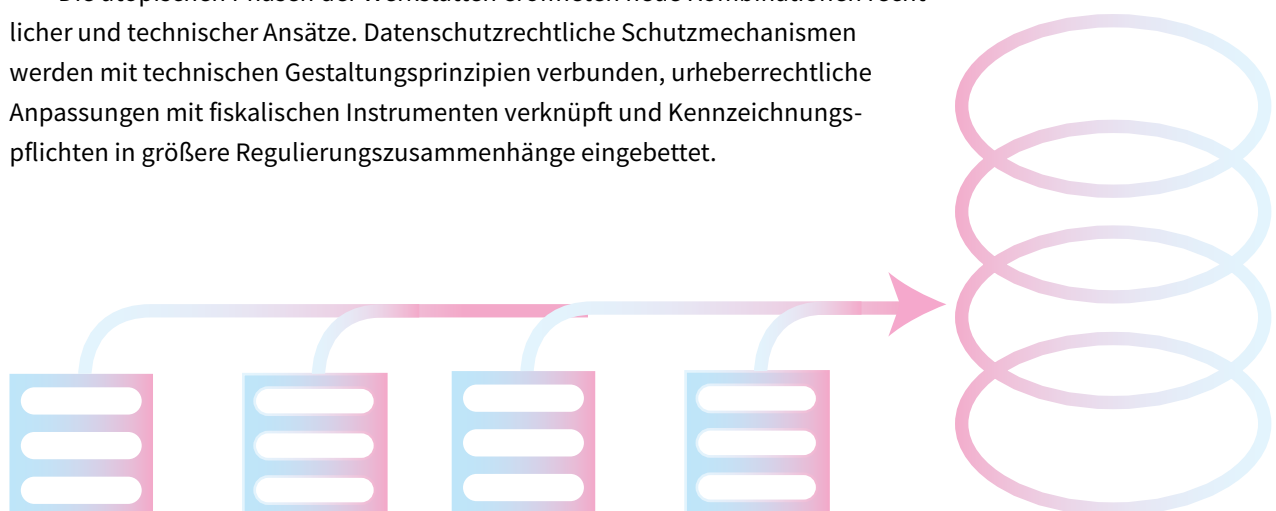
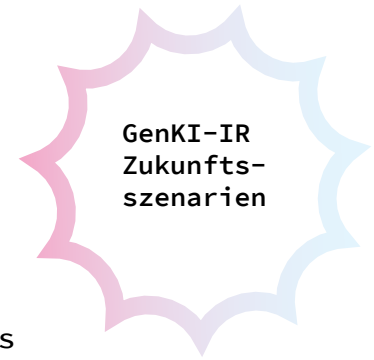
Übergreifende Synthese und Reflexion

Die GenKI-IR-Zukunftsszenarien zeigen, dass sich zentrale Konflikte der KI-Governance in allen drei Themenfeldern wiederfinden. Fragen des Schutzes, der rechtlichen Steuerung und der ökonomischen Verteilung können nicht isoliert verhandelt werden. Die Beteiligten beschreiben ein Regelungsumfeld, das aus unterschiedlichen Rechtsgebieten und Zuständigkeiten zusammengesetzt ist und in der praktischen Anwendung nur schwer überschaubar bleibt. Umsetzungsvorhaben und Initiativen geraten dadurch regelmäßig in langwierige Abstimmungsprozesse, bei denen rechtliche Anforderungen, technische Möglichkeiten und institutionelle Zuständigkeiten aufeinandertreffen, ohne klar ineinanderzugreifen.

In allen Bereichen werden strukturelle Überforderungen beschrieben, in Verwaltungen ebenso wie bei Kreativschaffenden, Medienunternehmen und Aufsichtsstellen. Die bestehenden (vor allem rechtlichen) Instrumente sind auf Konstellationen zugeschnitten, in denen Datenverarbeitung, Nutzung von Werken oder Kennzeichnungspflichten individualisierbaren Sachverhalten, Verarbeitungen oder Situationen zugeordnet werden können. Die Dynamik generativer KI, ihrer Entwicklung, ihres Einsatzes wie auch ihrer Output-vermittelten Wirkung entzieht sich diesen Logiken zunehmend. Einwilligungen verlieren an Steuerungswirkung, klassische Lizenzmodelle geraten an ihre Grenzen und einfache Kennzeichnungspflichten können mit der Verbreitung synthetischer Inhalte kaum Schritt halten.

Die entstandenen Vorschläge zielen auf institutionelle, rechtliche und technische Gestaltungsmöglichkeiten. Eine inhaltliche Schnittmenge ist beispielsweise das Ergänzen individueller Steuerungsinstrumente durch kollektive oder institutionelle Lösungen, die nicht immer den Staat als institutionellen Akteur in den Mittelpunkt stellen. Ebenso wurde in allen Bereichen KI-Governance als Querschnittsaufgabe verstanden, die Datenschutz, Urheberrecht und Medienrecht sowie Wettbewerbspolitik miteinander verbindet.

Die utopischen Phasen der Werkstätten eröffneten neue Kombinationen rechtlicher und technischer Ansätze. Datenschutzrechtliche Schutzmechanismen werden mit technischen Gestaltungsprinzipien verbunden, urheberrechtliche Anpassungen mit fiskalischen Instrumenten verknüpft und Kennzeichnungspflichten in größere Regulierungszusammenhänge eingebettet.



3

Hat sich die Methode „Zukunftsszenario“ bewährt, um damit Rechtsrahmen beim Einsatz generativer KI zu analysieren und neu zu denken? Lesen Sie dazu unsere Einschätzungen in Kapitel 4.

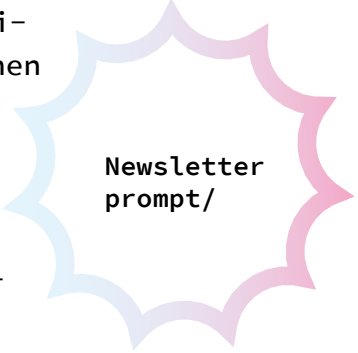
Forschungsbedarf: Die Zukunftsszenarien machen deutlich, dass in allen Bereichen weiterer Forschungsbedarf besteht. Im Datenschutz betrifft dies insbesondere die Umsetzbarkeit datenschutzrechtlicher Vorgaben in Bezug auf große Datenmengen – sei es hinsichtlich der Rechtmäßigkeit der Verarbeitung oder in Bezug auf Betroffenenrechte. Im Urheberrechtskontext bedarf es vertiefter Analysen zu möglichen Vergütungsmodellen und zu internationalen Koordinationsmechanismen. Für die Kennzeichnung von KI-Outputs fehlen belastbare Erkenntnisse zur Wirkung unterschiedlicher Label hinsichtlich des Vertrauens der Nutzer*innen.

Adressat*innen: Die in den Zukunftsszenarien herausgearbeiteten Thesen und Empfehlungen richten sich an verschiedene politische Ebenen und an unterschiedliche Gruppen von Akteur*innen. Nationale und regionale Parlamente, Ministerien und Fachbehörden können durch Leitlinien, Förderprogramme und Koordinierungsstrukturen Einfluss auf die praktische Ausgestaltung von KI-Regulierungen nehmen. Zugleich kommt Praxisakteur*innen eine zentrale Rolle zu, da technische Standards, Kennzeichnungspraktiken und Vergütungsmodelle ohne ihre Mitwirkung nicht realisiert werden können.



prompt/-Newsletter: Expert*innen-Interviews und Beiträge unter anderem zu Haftungsfragen, KI-Verordnung, Outputs und Urheberrecht

Den E-Mail-Newsletter prompt/ richten wir im Februar 2025 als periodisch erscheinendes Format ein, um die extrem dynamischen Debatten rund um generative KI und Recht nachzeichnen zu können. Bis Ende April 2026 entstehen elf Ausgaben zu jeweils einem Leitthema. Unser Anspruch ist, jedes Thema möglichst allgemeinverständlich darzustellen. Dem wissenschaftlich-vermittelnden Ansatz folgend werden wir in prompt/ mit Hilfe von Interviews, Hintergrundartikeln und Analysen unterschiedliche Themen einordnen und vertiefen. Im Folgenden haben wir Auszüge aus den bisher erschienenen prompt/-Ausgaben zusammengestellt, die bis zum Redaktionsschluss dieses Reports erschienen sind. Unter dem Inhaltsverzeichnis finden Sie die entsprechenden Artikel.



Newsletter
prompt/

KI und Urheberrecht

„Jeder Verlag muss für sich entscheiden, ob er klagt oder lizenziert, oder eben beides“ (Interview)

„Dieses GEMA-vs.-OpenAI-Urteil hat erhebliche Signalwirkung“ (Interview)

„Klare Haftungsregeln würden die Akzeptanz von KI-Agenten fördern“ (Interview)

KI und Haftung

KI-Agenten aus rechtlicher Perspektive (Beitrag)

„Eine EU-weite Harmonisierung des KI-Rechts auch außerhalb des Produkthaftungsrechts wäre wünschenswert“ (Interview)

Was ist Schatten-KI? (Beitrag)

Drei Haftungsverfahren (Klagen gegen KI-Anbieter)

KI und Datenschutz

„Aufklärung und Beratung zur Nutzung generativer KI gehören bereits zum Praxisalltag einer Datenschutzaufsichtsbehörde“ (Interview)

APIs bei KI-Anwendungen im Bildungsbereich (Beitrag)

3

KI-Outputs

„KI-Outputs und die Suche nach einer persönlichen geistigen Schöpfung“ (Interview)

Rechtliche Fragen zu den Outputs generativer KI (Beitrag)

KI-Kennzeichnung

„Die KI-Kennzeichnung sollte als einheitliches, nicht entfernbares Symbol über die gesamte Verbreitungskette hinweg bestehen bleiben“ (Interview)

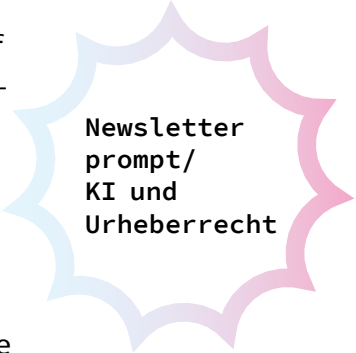
Vorgaben der KI-VO zu Kennzeichnungen von KI-Systemen und KI-Outputs (Beitrag)

KI-Verordnung und Praxisleitfäden

Was Sie über den Praxisleitfaden zur KI-Verordnung wissen müssen (Beitrag)

KI und Urheberrecht

Während Verlage und Verwertungsgesellschaften auf juristischem Wege etwaige Urheberrechtsverletzungen durch KI-Hersteller klären lassen wollen – schließen Medienhäuser mit jenen global agierenden Unternehmen Lizenzverträge ab. Was versprechen sie sich davon? Und was könnten Gerichtsurteile, wie das im Verfahren GEMA gegen Open AI, für beide Seiten verändern? Dazu befragen wir eine Fachanwältin und einen Rechtswissenschaftler.



Newsletter
prompt/
KI und
Urheberrecht



Kerstin Bäcker ist Fachanwältin für Urheber- und Medienrecht und unter anderem Mitorganisatorin der Initiative Verlegerrecht. Sie berät Verlage zu diversen Zukunftsthemen, unter anderem zu urheberrechtlichen und regulatorischen Themen rund um KI.

Foto: Renate Neder, München

Aus „prompt/“, März 2025: „Jeder Verlag muss für sich entscheiden, ob er klagt oder lizenziert, oder eben beides“

Medienrechtsanwältin Kerstin Bäcker zu den Motiven und den Zielen von Medienunternehmen, die Verträge mit KI-Anbietern abschließen

prompt/: „Einige Medienunternehmen verklagen Anbieter generativer KI, andere schließen Lizenzverträge mit ihnen ab, zunehmend auch im deutschsprachigen Raum. Welche Art von Verlagen sind Ihres Wissens nach an diesen Deals interessiert?“

Kerstin Bäcker: „Derzeit werden vor allem im Presse- und Fachverlagsbereich Vertragsabschlüsse beobachtet. Das liegt vor allem daran, dass die KI-Anbieter an solchen Trainingsmaterialien derzeit gesteigertes Interesse haben und insoweit eine höhere Bereitschaft zum Abschluss von Verträgen zu bestehen scheint.“

prompt/: „Was versprechen sich die Verlage von diesen Verträgen?“

Kerstin Bäcker: „Aus meiner übergeordneten Sicht als externe Beraterin geht es erst einmal darum, als Verlag zu kontrollieren, ob und inwieweit die Verlagsinhalte für ein Training genutzt werden – und unter welchen Bedingungen eine Nutzung erlaubt sein soll. Weiter spielt die Vergütung der vergangenen und künftig stattfindenden Nutzungen eine Rolle, und

zwar auf der Trainings-, aber auch auf der Output-Ebene. Zum Teil gewähren die KI-Anbieter – laut Verlautbarungen zu abgeschlossenen Verträgen – den Rechteinhabern auch Einblick, welche ihrer Inhalte in Outputs abgerufen werden. Das kann den Rechteinhabern wichtige Erkenntnisse zum Nutzungsverhalten ihrer potentiellen Kunden verschaffen – und ihnen damit vielleicht für die eigene Produktentwicklung nützen.“

prompt/: „Machen diese Verträge Sinn, oder sollten Rechteinhaber*innen doch lieber klagen?“

Kerstin Bäcker: „Die Ausgangssituationen und die Strategien im Umgang mit KI-Anbietern sind sicherlich sehr unterschiedlich: Manchem Rechteinhaber geht es vorrangig darum, das ungenehmigte KI-Training durch Dritte zu untersagen. Beispielsweise, damit bestehende Produkte nicht über paraphrasierte Zusammenfassungen ersetzt werden dürfen.

Anderen geht es um Entschädigungszahlungen für stattgefundene Nutzungen. Zum Teil stehen die Vergütung und geordnete Lizenzierung künftiger Nutzungen im Fokus. Im Ergebnis muss jeder Verlag für sich selbst entscheiden, ob er klagt oder lizenziert, oder eben beides. Die Strategie wird abhängig sein von der Art seines Katalogs und vor allem auch davon, welche eigenen KI-Nutzungen ein Verlag langfristig plant. Klagen, wie etwa die der ‚New York Times‘, werden unter anderem geführt, um angemessene Lizenzvergütungsansprüche zu realisieren. Denn die KI-Anbieter sind von sich aus entweder gar nicht oder nicht in ausreichender Höhe zu Zahlungen bereit.“

Aus „prompt/“, Dezember 2025: „Dieses GEMA-vs.-OpenAI-Urteil hat erhebliche Signalwirkung“

Mit seiner Entscheidung zur Klage der GEMA gegen den KI-Hersteller OpenAI habe das Landgericht München die Rechte von Urheber*innen gestärkt, sagt Rechtswissenschaftler Prof. Dr. Malte Stieper. Im Interview mit prompt/ erläutert er, dass es Anbieter*innen in die Pflicht nimmt, deren KI-Systeme in der Lage sind, Outputs zu generieren, die mit geschützten Originalwerken im Wesentlichen übereinstimmen.

prompt/: „Was stellt die GEMA-OpenAI-Entscheidung zu urheberrechtlichen Fragen bei KI-Systemen konkret klar?“

Malte Stieper: „Zunächst einmal muss man betonen, womit sich das Urteil nicht befasst. Die bisherige Diskussion im Urheberrecht dreht sich vor allem um die Frage, ob die Speicherung geschützter Werke in einem für das Training generativer KI verwendeten Datensatz unter die gesetzlichen Schranken für das Text- und Data-Mining fällt und damit – unter Berücksichtigung eines Rechtevorbahalts – ohne Zustimmung der Rechteinhaber*innen zulässig ist. [Die Text- und Data-Mining-Schranken beziehungsweise -Ausnahmen sind durch die Paragraphen 44b und 60d des Urheberrechtsgesetzes geregelt, Anmerkung der Redaktion.] Das Landgericht München bejaht das im Urteil zwar beiläufig. Doch solche, dem eigentlichen Training vorgelagerten Speichervorgänge, sind nicht Gegenstand des Rechtsstreits.

Vielmehr sieht das Gericht eine Urheberrechtsverletzung darin, dass ChatGPT von OpenAI die streitgegenständlichen Liedtexte ‚memorisiert‘. Das heißt, dass die Liedtexte auch über das Training hinaus abrufbar in den Parametern des Modells gespeichert bleiben. Es entspricht schon jetzt ganz überwiegender Auffassung in der Rechtswissenschaft, dass eine solche zusätzliche Vervielfältigung der Trainingsdaten im KI-Modell von der Text- und Data-



Prof. Dr. Malte Stieper ist Rechtswissenschaftler und Inhaber der Gundling-Professur für Bürgerliches Recht, Recht des geistigen Eigentums und Wettbewerbsrecht an der Martin-Luther-Universität Halle-Wittenberg. Wir haben die Fragen per E-Mail gestellt, Prof. Dr. Stieper hat sie uns schriftlich beantwortet.

Foto: Nikolaus Brade

3

Mining-Ausnahme nicht gedeckt ist. Offen war bisher aber, wann von einer fort dauernden Speicherung der Trainingsdaten in den Modell-Parametern ausgegangen werden kann. Das Landgericht München lässt für eine Vervielfältigung nun ausreichen, dass das Modell in der Lage ist, auf einfach gehaltene Prompts einen Output zu generieren, der mit dem Originalwerk im Wesentlichen übereinstimmt.

Anders gesagt: Rechteinhaber*innen müssen zum Nachweis einer Urheberrechtsverletzung keinen konkret abgrenzbaren Datensatz im Modell identifizieren, in dem das Originalwerk gespeichert ist. Das Urteil geht aber sogar noch darüber hinaus: Nach Auffassung des Landgerichts München begründet nicht nur die Speicherung der Trainingsdaten in den Modell-Parametern eine Urheberrechtsverletzung durch die KI-Anbieterin. Dies treffe auch auf den jeweiligen Output zu, der die in den Trainingsdaten enthaltenen Originalwerke erkennbar wiedergibt. Obwohl der Output erst auf entsprechende Prompts der Nutzenden generiert wird, rechnet das Gericht diese Verwertungshandlung nicht der einzelnen Nutzerin zu. Vielmehr sei es eine Handlung der ‚öffentlichen Wiedergabe‘ seitens der KI-Anbieterin. Denn sie bestimme die Funktionsweise der Modelle sowie die Auswahl von Trainingsdaten und sei somit für die Architektur der Modelle und die Memorisierung der Trainingsdaten verantwortlich.“

prompt/: „Welche Auswirkungen könnte diese Entscheidung – übergreifend – haben oder nahelegen?“

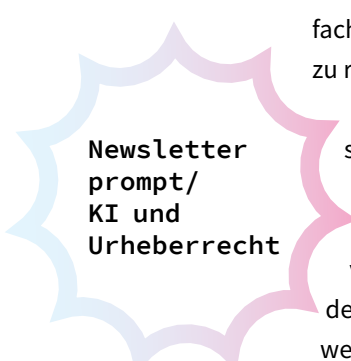
Malte Stieper: „Dieses Urteil hat erhebliche Signalwirkung. Erstmals wird Rechteinhaber*innen ein auch praktisch durchsetzbares Verbotsrecht zugesprochen: Sie können gegen die Verwertung ihrer Werke durch Anbieter*innen generativer KI-Modelle, wie ChatGPT, vorgehen, ohne den konkreten Nachweis führen zu müssen, dass ihre Werke im Inland für das Training des Modells verwendet worden sind. Entscheidend ist, dass das Modell durch einfach gehaltene Prompts dazu gebracht werden kann, relevante Teile der betreffenden Werke zu reproduzieren.

Wenn das Urteil Bestand hat, wird bei der Modellierung generativer KI in Zukunft noch stärker darauf geachtet werden müssen, eine Memorisierung der Trainingsdaten auszuschließen. Gelingt dies, so kann die Verwendung urheberrechtlich geschützter Werke für das KI-Training durchaus rechtlich zulässig sein. Das gilt insbesondere für das Training von KI-Modellen für Zwecke der nichtkommerziellen wissenschaftlichen Forschung, bei dem sogar ein von den Rechteinhaber*innen erklärter Nutzungsvorbehalt nicht beachtet werden muss.

Klar ist aber auch: Die meisten Modelle können durch entsprechendes Prompting dazu gebracht werden, einen urheberrechtsverletzenden Output zu generieren. Umso wichtiger wird es sein, zu definieren, was einen ‚einfach gehaltenen‘ Prompt ausmacht. Wo eine Memorisierung nicht verhindert werden kann, werden KI-Entwickler*innen nicht umhinkommen, für das KI-Training Lizenzen von den betroffenen Rechteinhaber*innen einzuholen.“

prompt/: „Wie wird es in diesem Verfahren in 2026 vermutlich weitergehen?“

Malte Stieper: „Das Urteil ist noch nicht rechtskräftig. Gegen das Urteil kann und wird sicherlich Berufung zum Oberlandesgericht München eingelegt werden, über die Revision muss dann der Bundesgerichtshof entscheiden. Viele in dem Fall aufgeworfene Fragen berühren die Vorgaben des EU-Rechts. Es ist daher damit zu rechnen, dass der Bundesgerichtshof zunächst den Gerichtshof der Europäischen Union (EuGH) anrufen wird, um abschließend über den Fall entscheiden zu können.“



Newsletter
prompt/
KI und
Urheberrecht

KI und Haftung

Entstehen beim Einsatz generativer KI-Chatbots oder -Agenten materielle, finanzielle oder Personen betreffende Schäden, stellen sich haftungsrechtliche Fragen. Wie verantwortlich KI-Hersteller, Anbieter und KI-Nutzer sein könnten, erläutert uns eine KI-Forscherin. Wir sprechen mit Fachanwält*innen über KI-spezifisches und übergreifendes Produkthaftungsrecht. Und wir sehen uns beispielhafte Entscheidungen in gerichtlichen KI-Haftungsverfahren an.

Newsletter
prompt/
KI und
Haftung

Aus „prompt/“, Juli 2025: „Klare Haftungsregeln würden die Akzeptanz von KI-Agenten fördern“

Je autonomer KI-Agenten funktionieren, desto mehr sollen ihre Entwickler*innen für etwaige Folgen verantwortlich sein. So der Vorschlag einiger Forscher*innen vom Ada Lovelace Institute und der Denkfabrik Interface. Wie man KI-Agenten regulieren könnte und was autonome Autos damit zu tun haben, erklärt uns Julia Smakman.

prompt/: „2025 soll das ‚Jahr der KI-Agenten‘ sein. Was bedeutet das eigentlich – und wie realistisch ist das?“

Julia Smakman: „Die Behauptung ist, dass KI-Agenten uns von Chatbots zu etwas führen werden, das in der realen Welt (einigermaßen) autonom handeln kann. Ein Chatbot könnte Ihre Urlaubsplanung erstellen, aber ein Agent könnte auch Ihre Flüge und Hotels buchen. Das klingt großartig, aber es ist nicht klar, wann oder ob KI-Agenten in der Lage sein werden, komplexe, mehrstufige Aufgaben zuverlässig zu erledigen. Wir erleben immer noch, dass Agenten bei relativ einfachen Aufgaben stecken bleiben oder mit unerwarteten Aktionen ‚aus der Reihe tanzen‘. Wir sind noch nicht an dem Punkt angelangt, an dem Sie einem KI-Agenten Ihre Kreditkartendaten anvertrauen können. Es gibt jedoch Agenten, die in der Lage sind, enger definierte Aufgaben in begrenzteren Umgebungen auszuführen. Die Technologie könnte sich rasch weiterentwickeln und zuverlässiger werden, aber es gibt immer noch große Bedenken hinsichtlich Sicherheit, Transparenz und Fairness. Insbesondere besteht die Sorge, dass unsere Rechtssysteme nicht in der Lage sein werden, damit umzugehen, wenn viele KI-Agenten untereinander interagieren.“

prompt/: „Sind wir in Europa rechtlich gut auf den Einsatz von KI-Agenten vorbereitet? Zum Beispiel durch die KI-Verordnung und die kürzlich aktualisierte Produkthaftungsrichtlinie – oder gibt es hier Lücken?“

Julia Smakman: „Leitplanken wie die KI-Verordnung und die aktualisierte Produkthaftungsrichtlinie sind definitiv ein Schritt in die richtige Richtung. Die Richtlinie entlastet betroffene Verbraucher*innen, indem sie die Beweislast in Haftungsfällen etwas erleichtert. Bei der KI-Verordnung wird viel von der tatsächlichen Umsetzung und Durchsetzung abhängen. Und obwohl KI-Agenten laut Verordnung definitiv ‚KI-Systeme‘ sind, kann es vom Kontext, in dem sie eingesetzt werden, abhängen, ob sie den strengeren Anforderungen für ‚KI-Systeme mit hohem Risiko‘ unterliegen.“



Julia Smakman ist wissenschaftliche Mitarbeiterin im Forschungsbereich „Recht und Politik“ des Ada Lovelace Institute. Der Fokus von Julia Smakman liegt auf der wirksamen Regulierung von KI-Systemen einschließlich Foundation Models im Vereinigten Königreich und in der Europäischen Union (EU). Wir haben die Fragen schriftlich gestellt und Julia Smakman hat schriftlich geantwortet. Das Interview wurde mit DeepL aus dem Englischen ins Deutsche übersetzt und redaktionell auf Richtigkeit überprüft.
Foto: Karla Gowlett

3

prompt/: „Sie haben kürzlich zusammen mit Kolleg*innen der Berliner Denkfabrik Interface einen möglichen Regulierungsrahmen für KI-Agenten vorgeschlagen. Als Vorbild diente Ihnen die Regulierung selbstfahrender Autos in Großbritannien. Warum haben Sie gerade dieses Beispiel gewählt, und was konnten Sie davon übernehmen?“

Julia Smakman: „Die britischen Rechtsvorschriften für selbstfahrende Autos sind ein interessantes Beispiel für einen Rechtsrahmen, bei dem sich die Haftung mit zunehmender Autonomie und abnehmender Kontrolle durch den Nutzer immer weiter von ihm weg verlagert. Obwohl selbstfahrende Autos in einem spezifischeren physischen Umfeld operieren – Straßen im Gegensatz zum Internet –, bieten sie eine nützliche Analogie für KI-Agenten, insbesondere, wenn man über Haftung und Verantwortung nachdenkt. Die Analogie hat jedoch auch Grenzen. Nach britischem Recht benötigen selbstfahrende Autos eine vorherige Genehmigung, KI-Agenten nicht. Künftige Nutzer*innen müssen also bei der Auswahl eines Agenten möglicherweise mehr Sorgfalt walten lassen. Auch die Versicherung ist für selbstfahrende Autos sehr wichtig. Der Markt für KI-Versicherungen beginnt sich gerade erst zu entwickeln.“

prompt/: „In Ihrem Ansatz unterscheiden Sie zwischen fünf Autonomiestufen, je nachdem, wie selbstständig ein KI-Agent handelt. Was genau verbirgt sich hinter diesen Stufen – und wie helfen sie bei der Frage, wer im Falle eines Schadens haftet?“


Julia Smakman: „Wir wollten differenzieren und zeigen, dass ‚nicht alle Agenten gleich sind‘. Es gibt verschiedene Autonomiestufen, und wir sollten unterschiedliche Erwartungen an Entwickler*innen und Nutzer*innen stellen, je nachdem, welchen Einfluss sie auf das Verhalten der Agenten haben und ob sie in der Lage sind, Schäden vorherzusehen und zu verhindern.“

Auf niedrigeren Autonomiestufen sind die Benutzer*innen eher in der Lage, den Agenten zu kontrollieren, so dass sie mehr rechtliche Verantwortung für die Handlungen des Agenten übernehmen. Bei einem höheren Grad an Autonomie, bei dem die Benutzer*innen weniger in der Lage sind, den KI-Agenten zu kontrollieren, tragen die Entwickler*innen möglicherweise eine größere rechtliche Verantwortung. Dies könnte ein Anreiz für die Entwickler*innen sein, nur hochgradig autonome Agenten mit angemessenen Sicherheitsvorkehrungen zu veröffentlichen, um vorhersehbare Schäden zu verhindern.“

prompt/: „Welche politischen, rechtlichen oder praktischen Hürden müssten überwunden werden, damit Ihr Modell tatsächlich umgesetzt werden kann?“

Julia Smakman: „Wir sehen nur begrenztes politisches Interesse daran, neue rechtliche Rahmenbedingungen für KI zu schaffen. Die Industrie wiederum könnte zögern, mehr rechtliche Verantwortung für hochgradig autonome Agenten zu übernehmen, mit dem Argument, dass Regulierung die Innovation hemmt. Wir glauben jedoch, dass klare Haftungsregeln die Akzeptanz von KI-Agenten fördern würden. Die Motivation, den Autonomous Vehicles Act in Großbritannien auf den Weg zu bringen, war, sicherzustellen, dass Menschen sich in und mit selbstfahrenden Autos wohl fühlen, wenn diese auf die Straße kommen. Wer möchte schon ein autonomes System nutzen, das er nicht richtig kontrollieren kann, und gleichzeitig rechtlich für alles haften, was schiefgeht – persönlich oder als Teil seines Unternehmens?“

prompt/: „Solange es noch keinen festen rechtlichen Rahmen gibt: Welchen Rat würden Sie jemandem geben, der bereits heute mit KI-Agenten arbeitet?“



Newsletter
prompt/
KI und
Haftung

Julia Smakman: „Gehen Sie mit der gebotenen Sorgfalt vor und vergewissern Sie sich, dass jeder KI-Agent, den Sie einsetzen, ordnungsgemäß für den Zweck getestet wurde, für den Sie ihn verwenden. Verstehen Sie die Nuancen: Die Risiken von weniger autonomen Agenten, die für spezifischere Aufgaben eingesetzt werden, sind geringer als die Risiken, die mit hochgradig autonomen Agenten mit Internetzugang verbunden sind, die ‚open-end‘ Aufgaben ausführen. Wie bei jedem KI-System können Sie haftbar gemacht werden, wenn die KI einen Schaden verursacht. Vergewissern Sie sich also, ob Ihnen wohl dabei ist, dieses Risiko einzugehen. Wenn nicht, sollten Sie überdenken, ob ein KI-Agent die beste Lösung für Sie ist.“

Julia Smakman war auch Referentin bei unserem GenKI-IR-Lunch-Talk, bei dem wir das Thema KI und Haftung weiter vertieften. Mehr dazu lesen Sie weiter unten in diesem Kapitel.

Aus „prompt/“, Juli 2025: KI-Agenten aus rechtlicher Perspektive

Problem of many hands: Im Zusammenhang mit KI-Agenten sprechen manche vom „Problem of many hands“, einem Konzept aus Philosophie und Ethik. Es fragt danach, wie in einem komplexen System mit vielen verschiedenen Akteur*innen – auch rechtliche – Verantwortung zugewiesen werden kann. Im Zusammenhang mit KI-Agenten bezieht sich das „Problem der vielen Hände“ darauf, dass vom Training des zugrunde liegenden Modells über die Entwickler*innen des KI-Agenten bis zu dessen Einsatz viele Akteur*innen an der Wertschöpfungskette beteiligt sind und darin auch miteinander interagieren.

Hinzu kommen mögliche, von KI-Agenten ausgelöste Interaktionen, etwa die Delegation von Aufgaben an andere KI-Agenten oder Menschen. Ein wesentlicher Unterschied zwischen KI-Agenten und anderen Produkten, die ebenfalls das Ergebnis einer komplexen Wertschöpfungskette sind (Autos oder Maschinen zum Beispiel), ist also, dass die Agenten autonome Entscheidungen treffen und ihr Verhalten in Echtzeit ohne menschliche Intervention ändern können. Geht dabei etwas schief, ist es schwer zu bestimmen, wo genau der Fehler lag und wer dafür verantwortlich ist.

Haftungsfragen und mögliche Lösungsansätze: Will man klären, wer im Schadensfall haftet, gibt es gleich mehrere Herausforderungen. Zunächst ist es nicht einfach, angemessene Sorgfaltspflichten festzulegen. Wie entscheidet man beispielsweise, ob ein*e Nutzer*in einem KI-Agenten sorgfältigere Anweisungen hätte geben müssen? Oder ob ein*e Entwickler*in hätte sicherstellen müssen, dass der KI-Agent auch zweideutige Anweisungen interpretieren kann? Wo können unterschiedliche Akteur*innen entlang der Wertschöpfungskette also wirklich Kontrolle ausüben und wo können sie entsprechend auch verantwortlich – und haftbar – gemacht werden?

Eine weitere Herausforderung ist, dass Schäden, die durch den Einsatz von KI-Agenten entstehen können, nicht immer materieller Natur sind. Auch immaterielle Schäden, wie Diskriminierung oder die Verletzung von Rechten, kommen infrage. Denkbar sind auch systemische Schäden, die erst nach längerer Zeit erkennbar werden: Die Folgen von Desinformation oder makroökonomische Schäden wie Marktverzerrungen, zum Beispiel. Wie man damit umgeht, ist bislang nicht ganz klar.

Die KI-Verordnung kann ein Teil der Lösung sein. Sie weist verschiedenen Akteuren in der Wertschöpfungskette, abhängig von ihren individuellen Möglichkeiten, unterschiedliche Pflichten zu: Beispielsweise hinsichtlich erforderlicher Risikobewertungen und Transparenzgebote, oder auch zu technischen Einsatzkontrollen und menschlicher Aufsicht bestimmter

3

Arbeitsgänge. Die Verordnung wurde aber nicht explizit für KI-Agenten konzipiert und weist deshalb einige Schwachstellen auf: Es ist zum Beispiel nicht ganz klar, in welche der drei Risikokategorien, die die KI-VO definiert, KI-Agenten fallen. Bei Haftungsfragen ist die KI-VO zudem wenig konkret und verweist meist auf die Produkthaftungsrichtlinie (siehe dazu auch die Mai-Ausgabe von *prompt/*) und nationales Recht. Erste Ergänzungsvorschläge, etwa ein Regulierungsmodell, das sich an den Regeln für autonome Fahrzeuge orientiert, gibt es bereits (siehe dazu das nachfolgende Interview).

Aus „prompt/“, August 2025: „Eine EU-weite Harmonisierung des KI-Rechts auch außerhalb des Produkthaftungsrechts wäre wünschenswert“

Marieke Mattern und Dr. Jan-Philipp Günther-Burmeister beschäftigen sich als IT- und Datenschutz-Rechtsanwältin und -Rechtsanwalt schon länger mit Haftungsfragen beim Einsatz von KI. Sie würden eine einheitliche europäische Haftungsregelung für KI begrüßen, weil dann die Risiken für die Beteiligten – insbesondere auch für Versicherer – voraussichtlich besser zu kalkulieren wären.


prompt/: „Wie sehr prägen Fragen nach den möglichen Haftungsrisiken beim Einsatz von generativen KI-Systemen Ihren Arbeitsalltag als Anwält*innen?“

Marieke Mattern: „Bislang kommen unsere Mandant*innen primär mit Fragen zum rechtskonformen Einsatz von KI-Systemen in ihren Unternehmen zu uns und wünschen insbesondere die Erstellung von Richtlinien zum KI-Einsatz. Hierbei weisen wir auch auf die Anforderungen zur Vermeidung von Haftungsfällen speziell für Unternehmen hin. Daneben wird die Gestaltung von Nutzungsverträgen über KI-Systeme immer relevanter, bei denen es ebenfalls um Fragen zur Haftung der jeweiligen Vertragspartner geht.“

Dr. Jan-Philipp Günther-Burmeister: „Viele der Anfragen unserer Mandant*innen betreffen dabei nicht nur unmittelbar die mit dem Einsatz eines bestimmten KI-Systems verbundenen Haftungsfragen, sondern auch Aspekte, die darüber hinausgehen: Beispielsweise können sich Schadensersatzansprüche ergeben, die aus Datenschutzverletzungen im Zusammenhang der Verwendung von KI resultieren. Auch droht Leitungsorganen von Gesellschaften das Risiko einer Haftung gegenüber ihrer Gesellschaft oder unter Umständen sogar gegenüber Dritten, wenn eine Verletzung ihrer Sorgfaltspflicht im Zusammenhang mit der Einführung und der Nutzung von KI-Systemen vorliegt.“

prompt/: „Wie groß ist das Haftungsrisiko wirklich – kann ich mir das nicht alles versichern lassen?“

Marieke Mattern: „Hier kommt es ganz darauf an, was mit der Versicherung vereinbart wurde. So können die Berufshaftpflichtversicherungen für Rechtsanwält*innen die Haftung für wissentliche Pflichtverletzungen vom Versicherungsschutz ausschließen. Hiervon wird in der Praxis häufig Gebrauch gemacht. Wenn Anwält*innen daher zum Beispiel wissentlich KI-generierte Inhalte übernehmen, ohne diese noch einmal einer eigenen Prüfung, insbesondere einem Faktencheck, zu unterziehen, müssen sie für hieraus resultierende Schäden in der Regel selbst haften.“



Newsletter
prompt/
KI und
Haftung

Dr. Jan-Philipp Günther-Burmeister: „Und dies gilt auch für andere Versicherungen. Obwohl es seit kurzer Zeit auch einige Versicherungen für ‚fehlerhafte KI‘ gibt, ist zu bezweifeln, dass eine solche Versicherung zahlen wird, wenn der Fehler nicht direkt auf die KI, sondern auf Versäumnisse des Versicherungsnehmers zurückzuführen ist. Weiterhin ist aus versicherungs(-rechtlicher) Perspektive die Frage spannend, inwieweit möglicherweise eine ‚silent coverage‘ bestehen könnte. Gemeint ist ein ‚unbeabsichtigter‘ Versicherungsschutz für KI-Risiken, der sich aus herkömmlichen Sach- und Haftpflichtversicherungen ergeben kann, wenn KI-Risiken nicht ausdrücklich ausgeschlossen sind.“

prompt/: „Welche Auswirkung hat es, dass die EU-Kommission den Entwurf der KI-Haftungsrichtlinie zurückgezogen hat? Bestehen jetzt mehr Unsicherheiten?“

Marieke Mattern: „Die KI-Haftungsrichtlinie sah grundsätzlich ‚nur‘ Regelungen zur Beweiserleichterung für Nutzer vor. Etwa in Form von Offenlegungspflichten und Kausalitätsvermutungen für Pflichtverletzungen der Anbieter. Hiervon waren auch nicht alle Pflichtverletzungen umfasst, aus denen eine mögliche Haftung resultieren könnte. Es wurde allerdings vermehrt diskutiert, ob die KI-Haftungsrichtlinie neben den (neuen) Regelungen des Produkthaftungsrechts überhaupt erforderlich war – meines Erachtens ein berechtigter Einwand. Aus dieser Sicht hat der Rückzug der KI-Haftungsrichtlinie wenig Auswirkungen.“

Generell wäre eine EU-weite Harmonisierung des KI-Rechts auch außerhalb des Produkthaftungsrechts aber wünschenswert, da die Regelungen zum Schadensersatz in den einzelnen EU-Mitgliedsstaaten teils sehr unterschiedlich und unübersichtlich sind und daher hinsichtlich der Haftung außerhalb des deutschen Rechts in jedem Fall noch Rechtsunsicherheit herrscht. Hier wird sich aber erst in der Praxis zeigen müssen, in welchen Bereichen dies von Relevanz sein könnte. Dann könnte man sich auch fragen, ob nicht eine EU-Verordnung der richtige Weg zur Beantwortung von noch offenen Haftungsfragen sein könnte.“

Dr. Jan-Philipp Günther-Burmeister: „Der Vorteil einer einheitlichen europäischen Haftungsregelung für KI wäre, dass die Risiken für die Beteiligten – insbesondere auch für Versicherer – voraussichtlich besser zu kalkulieren wären. Ob dabei Offenlegungspflichten und Kausalitätsvermutungen tatsächlich die richtigen Instrumente sind, sei an dieser Stelle offengelassen.“

Aus „prompt/“, Oktober 2025:

Was ist Schatten-KI?

Was mystisch klingt, ist eigentlich ganz einfach: Wenn am Arbeitsplatz nicht genehmigte KI-Anwendungen genutzt werden, spricht man von Schatten-KI. Ein Beispiel: Eine Schule besitzt keine offizielle Lizenz, die Lehrenden nutzen aber – ohne Erlaubnis – ihren privaten ChatGPT-Account, um Aufgaben zu erstellen oder Arbeiten ihrer Schüler*innen zu korrigieren. Oder: Eine Mitarbeiterin in der Verwaltung muss einen Sachbericht erstellen und gibt dafür Daten in einen privat abonnierten Chatbot ein. Jetzt könnte man fragen: Wo ist das Problem?

Generell geht es Arbeitgebern zunächst darum, dass die von den Beschäftigten genutzten Werkzeuge und Materialien legal erworben, funktional geeignet und geprüft sind, auch, um damit interne Vorgaben und Qualitätsmaßstäbe einhalten zu können. Bei KI-Tools meint das



Marieke Mattern und Dr. Jan-Philipp Günther-Burmeister sind Rechtsanwältin und Rechtsanwalt mit Schwerpunkten auf IT- und Datenschutzrecht bei einer Kanzlei in Hannover.

Fotos: Patrice Kunte

3

unter anderem den Umgang mit sensiblen Firmendaten oder personenbezogenen Daten. Verwenden die Beschäftigten nun eigenmächtig Anwendungen – im Schatten der Kontrollmechanismen – kann dies unerwünschte Folgen haben: Im Fall KI-gestützter Chatbots in der Cloud besteht die Gefahr, dass sensible Daten unkontrolliert an die KI-Anbieter abfließen. Oder dass KI-generierte Outputs fehlerhaft sind und womöglich falsche beziehungsweise erfundene Informationen enthalten. Bei daraus resultierenden Schadensfällen könnte der Arbeitgeber dafür haftbar gemacht werden.

Newsletter prompt/ KI und Haftung

Was kann man dagegen tun? Die einfache Antwort wäre vermutlich: Lizenzen anschaffen, Konfigurationen steuern und die Nutzung möglichst allen zugänglich machen – auch externen, vertraglich eingebundenen Mitarbeiter*innen. Wenn das nicht möglich ist, ließe sich vielleicht per anonymer Umfrage ermitteln, welche KI-Tools von den Beschäftigten genutzt beziehungsweise gefordert werden. Daraufhin können die Verantwortlichen Risikoeinschätzungen vornehmen und Richtlinien für den Einsatz der ausgewählten KI-Anwendungen erstellen. Darin kann auch definiert sein, was ausdrücklich nicht erlaubt ist. Nicht zuletzt sollten Mitarbeitende im Zuge einer organisierten KI-Einführung qualifiziert werden, wie es die KI-Verordnung ohnehin verpflichtend vorsieht.

Aus „prompt“, April 2025:

Drei Haftungsverfahren

Man könnte sie als **Vorboten betrachten für vermutlich kommende juristische Auseinandersetzungen: Verfahren, in denen Haftung und Entschädigung verhandelt werden, die auf falsche Ergebnisse von oder fehlerhaftem Umgang mit KI-Systemen zurückgehen. Zumindest lassen drei Fälle aus Deutschland, den USA und Kanada Rückschlüsse auf diesbezügliche Kernfragen zu.**

Haftung für Falschinformationen

Wer klagt? Ein mittelständisches Unternehmen klagt am Landgericht Kiel gegen einen Wirtschaftsauskunftsdienst.

Was ist passiert? Der Wirtschaftsauskunftsdienst veröffentlichte eine KI-gestützte Analyse, in der fälschlicherweise behauptet wurde, dass ein Unternehmen wegen Vermögenslosigkeit aus dem Handelsregister zu löschen sei. Das betroffene Unternehmen klagte daraufhin auf Unterlassung.

Wie lautet das Urteil? Der Beklagte wurde zur Unterlassung und zur Zahlung vorgerichtlicher Rechtsanwaltskosten verurteilt.

Wie lautet die Begründung? Der Auskunftsdienst haftet für die durch die KI erzeugten Falschinformationen. Laut Landgericht Kiel liegt eine Verletzung des Unternehmenspersönlichkeitsrechts vor. Der Dienst kann sich nicht darauf berufen, nicht an dem automatisierten Verfahren beteiligt gewesen zu sein, da er das KI-System bewusst zur Beantwortung von Suchanfragen eingesetzt hat. Der Beklagte habe sich die KI-generierten Inhalte zu eigen gemacht und nach außen erkennbar die inhaltliche Verantwortung übernommen.

Quelle: BBS Rechtsanwälte

Haftung für seelische und körperliche Schäden

Wer klagt? Das Social Media Victims Law Center und das Tech Justice Law Project klagen in Texas / USA gegen Character Technologies.

Was ist passiert? Die Eltern zweier texanischer Teenager klagen gegen Character.AI, den Anbieter eines Chatbots, der es Usern ermöglicht, mit prominenten oder fiktiven Personas – so genannten Charakteren – zu interagieren. Die Kläger*innen behaupten, ihre Kinder hätten nach längerer Interaktion mit dem Chatbot problematische Verhaltensweisen entwickelt. Eines der Kinder sei etwa gewalttätig gegenüber seinen Eltern geworden, nachdem diese versucht hatten, seine Bildschirmzeit zu verringern. Ein Screenshot von einer Konversation zwischen dem 17-Jährigen und dem Chatbot zeigt, wie dieser dem Jungen vorschlägt, dass er doch seine Eltern töten könne, um die Bildschirmzeit nicht verringern zu müssen. Die Kläger*innen argumentieren, dass das Unternehmen in mehrfacher Hinsicht fahrlässig gehandelt hätte. Zum Beispiel sei es zu fahrlässigen Fehlern im Design des Systems gekommen. Zudem hätten Warnhinweise zu den möglichen Risiken gefehlt, die mit der Nutzung des Chatbots verbunden sind.

Wie lautet das Urteil? Das Verfahren läuft noch.

Quelle: The National Law Review

Schadensersatz für fehlerhaften Chatbot

Wer klagt? Ein Passagier gegen Air Canada

Was ist passiert? Ein KI-Chatbot von Air Canada lieferte einem Passagier falsche Informationen zu einem vermeintlichen Rabatt auf Flugreisen. Der Chatbot hatte behauptet, dass Reisende, die wegen eines Todesfalls in der Familie reisen, bis zu 90 Tage nach ihrer Reise einen sogenannten Trauerrabatt beantragen können. Zwar gibt es diesen Rabatt tatsächlich, allerdings kann er nicht rückwirkend geltend gemacht werden. Air Canada wies den Antrag des Klägers auf Rückerstattung also ab.

Wie lautet das Urteil? Das Gericht entschied zugunsten des Klägers auf Schadensersatz. Air Canada sei für alle Informationen auf seiner Website verantwortlich, unabhängig davon, ob sie von einem Chatbot oder einer statischen Website stammen. Vom Kunden könne nicht erwartet werden, dass dieser die Informationen, die ihm das Unternehmen auf seiner Webseite anzeigt, nochmals eigenständig überprüft.

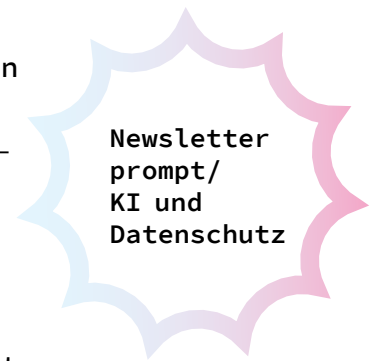
Quelle: American Bar Association



3

KI und Datenschutz

„In der öffentlichen Verwaltung, aber auch in Unternehmen, Verbänden oder im Dienstleistungssektor, sind datenschutzrechtliche Vorschriften und Prozesse etabliert. Ergeben sich daraus Hürden für den Einsatz generativer KI-Werkzeuge – und stehen die Regelungen der KI-Verordnung damit im Widerspruch oder im Einklang? Das erläutern uns zwei Landesdatenschutzbeauftragte. Und wir erklären, wie wichtig hierbei Programmierschnittstellen (APIs) sein können.“



Yvonne Herzog ist Sachbearbeiterin KI und Verwaltung in Abteilung 2 (Inneres, Videoüberwachung und Verkehr), Daniel Maslewski ist Referent und Beauftragter für Künstliche Intelligenz / KMU- und Start-up-Beratung in Abteilung 4 (Datenschutz in der Privatwirtschaft), jeweils beim Landesbeauftragten für Datenschutz und Informationsfreiheit Baden-Württemberg.

Fotos: Privat

Aus „prompt/“, Mai 2025: „Aufklärung und Beratung zur Nutzung generativer KI gehören bereits zum Praxisalltag einer Datenschutzaufsichtsbehörde“

Interview mit Yvonne Herzog und Daniel Maslewski

prompt/: „Der datenschutzkonforme Einsatz von generativen KI-Systemen spielt für die öffentliche Verwaltung, aber auch für die Zivilgesellschaft und für Unternehmen eine große Rolle. Wie stellt sich das im Praxisalltag einer Datenschutzaufsichtsbehörde bisher dar?“

Yvonne Herzog: „Wir versuchen hier ganz bewusst, uns auch proaktiv mit der Materie zu beschäftigen: Beispielsweise veröffentlichen wir Handreichungen und Diskussionspapiere, vermitteln Best Practices, halten und leiten Fachvorträge auf KI-Veranstaltungen oder organisieren Schulungen in unserem eigenen Bildungszentrum (BIDIB). Zur Vorbereitung hierfür gibt es regelmäßig interne Fachkreise und Arbeitsgruppen, an denen unsere KI-Fachreferent*innen aus den verschiedenen Abteilungen teilnehmen.“

Daniel Maslewski: „Unsere Aufgaben als Aufsichtsbehörde sind sehr vielfältig, weshalb uns die Thematik im Praxisalltag an verschiedenen Stellen begegnet. Es kann beispielsweise sein, dass eine betroffene Person eine Beschwerde bei uns einreicht und dabei eine Datenschutzverletzung durch ein Unternehmen rügt, das KI einsetzt. Oder wir haben ganz konkrete Beratungsanfragen von verantwortlichen Unternehmen oder öffentlichen Stellen, die wissen wollen: Was dürfen wir mit KI machen? Wo gibt es beim Einsatz von KI datenschutzrechtliche Probleme? Worauf müssen wir achten?“

Je nachdem, von welcher Stelle die Anfrage kommt, können das auch immer wieder sehr spezifische Problemstellungen sein, die sich auf konkrete KI-Systeme und Verarbeitungsprozesse beziehen. Häufig wird uns aber auch die pauschale Frage gestellt, ob ein System eingesetzt werden darf. Entscheidend ist daher, dass wir uns gezielt mit innovativen Themen wie KI beschäftigen und unsere Expertise weiter ausbauen.“

prompt/: „Warum spielt die schon 2018 in Kraft getretene DSGVO eigentlich für eine neue Technologie, wie generative KI, überhaupt noch eine Rolle?“

Daniel Maslewski: „Der Grund hierfür liegt in der bewusst technologieneutralen und innovationsoffenen Ausgestaltung der Datenschutz-Grundverordnung (DSGVO). Die Vorschriften der Verordnung gelten insoweit grundsätzlich auch bei Datenverarbeitungen im Zusammenhang mit innovativen Technologien, wie etwa Künstlicher Intelligenz. Insbesondere beim Training oder beim Einsatz von generativer KI werden regelmäßig große Mengen an personenbezogenen Daten verarbeitet.“


Dabei kann es sich um ganz verschiedene Arten von Informationen handeln. So können große Trainingsdatensätze beispielsweise unterschiedlichste Inhalte aus dem Internet enthalten – beispielsweise aus Wikipedia oder anderen Plattformen -, in denen Meinungen oder Positionen geäußert werden, die sich auf konkrete Personen zurückführen lassen. Solche Informationen stellen personenbezogene Daten dar. Verantwortliche müssen folglich auch beim Einsatz und Training von KI die datenschutzrechtlichen Anforderungen einhalten und sich mit den KI-spezifischen Problemen auseinandersetzen.“

prompt/: „Manche Behörden und Unternehmen wünschen sich mehr datenschutzrechtliche Absicherung – zum Beispiel durch klare Aussagen, Bewertungen oder Zertifizierungen von KI-Systemen durch die Aufsichtsbehörden. Wie gehen Sie hier vor?“

Daniel Maslewski: „Die datenschutzrechtliche Zulässigkeit des Einsatzes von Künstlicher Intelligenz hängt stets von den konkreten Umständen der jeweiligen Datenverarbeitung ab. Als unabhängige Aufsichtsbehörde prüfen wir die entsprechenden Verarbeitungsvorgänge daher stets einzelfallbezogen. Konkrete Empfehlungen oder pauschale Zertifizierungen für einzelne KI-Anwendungen entsprechen insoweit nicht dem Zweck unserer Aufsichtstätigkeit. Gleichzeitig sehen wir es jedoch als unsere Aufgabe an, verantwortliche Stellen bei der rechtlichen Einordnung zu unterstützen. So haben wir ein Diskussionspapier zu den einschlägigen Rechtsgrundlagen erstellt, das einen Überblick über die datenschutzrechtlichen Anforderungen beim Einsatz von KI bietet und als Grundlage für die weitere Auseinandersetzung dienen soll.“

Daneben verfolgen wir in unserer Beratungspraxis das Ziel, ein grundlegendes Verständnis für den datenschutzkonformen Umgang mit KI zu fördern und verantwortliche Stellen für datenschutzrechtliche Fragestellungen zu sensibilisieren. Dabei stellen sich regelmäßig zentrale Fragen wie: Was macht das System, das ich einführen will, mit den geschützten Daten? Gibt es für die Verarbeitung personenbezogener Daten eine einschlägige Rechtsgrundlage? Werden die Daten zu einem anderen Zweck verarbeitet? Wie können die Betroffenenrechte bei intransparenten KI-Systemen umgesetzt werden? Werden die technischen und organisatorischen Maßnahmen eingehalten und entsprechen diese dem Stand der Technik? Wie sieht es mit regelmäßigen Updates aus? Wohin werden die Daten übermittelt? Werden abgeschottete Systeme verwendet, bei denen der KI-Anbieter nicht dauerhaft auf die Datensätze zugreifen kann?“

Yvonne Herzog: „Wichtig ist auch, dass die von der DSGVO vorgesehene Rollenverteilung klar wird: Arbeitet ein Unternehmen oder eine öffentliche Einrichtung mit personenbezogenen Daten, hat das Unternehmen oder die Einrichtung die Pflicht und Verantwortung, diese zu schützen. Der größte Fehler, den man machen kann, ist, sich nicht aktiv mit den Fragen auseinanderzusetzen, weil mir zum Beispiel mein KI-Anbieter sagt: Es passt schon alles!“



Newsletter
prompt/
KI und
Datenschutz

3

Aus „prompt/“, November 2025: APIs bei KI-Anwendungen im Bildungsbereich

Was ist eine API? Eine Application Programming Intertace (API) ist eine sogenannte Programmierschnittstelle. Sie ermöglicht es, dass bestimmte Daten einer Datenbank beziehungsweise einer Anwendung freigegeben oder frei zugänglich gemacht werden, damit diese von anderen Anwendungen genutzt werden können. Ein typisches Beispiel ist die Google Maps API: Externe Anwendungen können die digitalen Karten via API einbetten und darauf zusätzliche visuelle Informationen platzieren. Im Kontext von KI-Modellen ermöglichen APIs die Einbindung von Drittanbieter-Anwendungen im eigenen (KI-)System. Mehr noch: Ohne APIs wären KI-Technologien von Drittanbietern nicht einsetzbar.

Was meint die „datenschutzkonforme Anbindung via API“? Immer, wenn der Austausch personenbezogener Daten stattfindet, gelten die Regelungen der Datenschutz-Grundverordnung (DSGVO). Diese gelten auch, wenn dieser Austausch beim Zugriff auf ein KI-Modell eines Drittanbieters via API erfolgt. Zunächst muss die Erforderlichkeit (sprich: konkrete notwendige Gründe) für die Verarbeitung personenbezogener Daten sichergestellt werden. Zudem sind die Prinzipien „Datenschutz durch Technikgestaltung“ (Privacy by Design) und „datenschutzfreundliche Voreinstellungen“ (Privacy by Default) einzuhalten. Wenn ein Unternehmen beispielsweise ein Sprachmodell oder ein KI-Tool über eine API in seine Produkte und Services integriert, ist es verantwortlich dafür, „Zwecke und Mittel“ der Datenverarbeitung festzulegen.

Datenschutzkonforme Anbindung via APIs im Bildungsbereich Zu den KI-Anwendungen im Bildungsbereich, die als – vereinfachend ausgedrückt – „datenschutzkonform“ angeboten werden, zählen unter anderem Telli, Sidekick, KAI, DieSchulApp, fAlr Chat, emiGPT/emuDALLE, Fobizz, SchulKI, Cornelsen.ai, IServ und FelloFish. In der Regel wollen die Anbieter die DSGVO-Konformität unter anderem dadurch sicherstellen, dass sie ihre Systeme nur innerhalb der EU betreiben und eine Datenübertragung in unsichere Drittländer verhindern. Zudem wird zusätzlich eine sogenannte Auftragsverarbeitungsvereinbarung (AVV) abgeschlossen, um die Verarbeitung von personenbezogenen Daten festzulegen.

Newsletter
prompt/
KI und
Datenschutz



KI Outputs

Wie die Arbeitsergebnisse generativer KI-Systeme – auch als KI-Outputs bekannt – aus urheberrechtlicher Perspektive zu bewerten sind, diese Frage wird seit längerem debattiert und hat mehrere Ebenen. Wir warfen den Blick auf etwaige Schutzrechte für Prompts und hybride Werke. Und wir sprachen mit einer Wirtschaftsinformatikerin über vermeintliche Plagiate und das Phänomen der Doppelschöpfung.

Newsletter
prompt/
KI Outputs

Aus „prompt/“, Oktober 2025: „KI-Outputs und die Suche nach einer persönlichen geistigen Schöpfung“

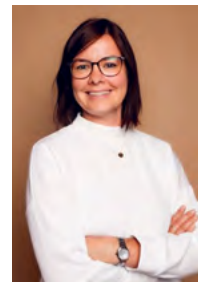
Kopie, Plagiat oder künstlerische Bearbeitung? Wenn generative KI Werke ausgibt, die vorhandenen Originalen ähneln, stellen sich urheberrechtliche Fragen. Lisa Käde über das Konzept der „unabhängigen Doppelschöpfung“, über die „Pastiche“-Regelung und die Frage nach den Intentionen.

prompt/: „In Bezug auf den Output von generativer KI sagen Sie, ‚wenn Mickey rauskommt, muss Mickey auch drin sein‘ – was ist damit gemeint?“

Lisa Käde: „Dahinter steckt die verbreitete Annahme, dass ein Prompt in irgendeiner Weise zwangsläufig zur Reproduktion führt – ähnlich dem Abrufen von Daten aus einer Datenbank. Dabei wird aber verkannt, dass generative KI-Systeme nicht einfach Gelerntes wiedergeben können oder sollen. Diese Systeme lernen vor allem, zu verallgemeinern und mithilfe von Statistik die ‚Essenz‘ des Gelernten wiederzugeben. Außerdem kann ein User, der es darauf anlegt, möglicherweise durch eine extrem präzise und ausführliche Formulierung des Prompts das KI-System zu einem Output leiten, der sehr nah an bestimmten Trainingsdaten liegt – weil es etwa in Bezug auf bestimmte Merkmale nichts zu verallgemeinern gibt. Dementsprechend handelt es sich bei dem generierten Output nicht um eine ‚Kopie‘ im technischen Sinne, denn es wird nichts einfach ‚wiedergegeben‘, sondern es wird ein neuer Output errechnet.“

prompt/: „Ist es urheberrechtlich nicht unerheblich, wie ein Plagiat zustande kommt, denn es ist so oder so eine Urheberrechtsverletzung?“

Lisa Käde: „Jain. Es gibt das Konzept der ‚unabhängigen Doppelschöpfung‘. Wenn zwei Menschen die gleiche Idee haben und sie auf gleiche Weise umsetzen und ein sehr ähnliches oder gleiches Werk dabei herauskommt, ohne dass beide voneinander abgekupfert haben können, kann für beide Urheberrechtsschutz entstehen. Es wurde auch schon für generative KI-Modelle angedacht, ob ein derartiger Ansatz bei der urheberrechtlichen Betrachtung des Outputs zu berücksichtigen ist – insbesondere, wenn das, was scheinbar plagiiert wurde, nicht in den Trainingsdaten enthalten war. Womit ich Ihnen aber Recht gebe: Es kommt nicht darauf an, wie der Output – technisch – zustande gekommen ist. Wenn der Output Trainingswerken stark ähnelt oder gleicht, kann eine Vervielfältigung vorliegen – egal, ob aus einer Datenbank abgerufen oder statistisch aufgrund des Prompts neu erzeugt.“



Lisa Käde ist Wirtschaftsinformatikerin und Rechtsanwältin in der Kanzlei JBB Rechtsanwält:innen in Berlin und berät in allen rechtlichen Fragen rund um Entwicklung und Nutzung von KI-Systemen sowie Open-Source-Compliance.

Foto: JBB Rechtsanwält:innen Partnerschaft

3

prompt/: „Lassen sich diese Überlegungen auf die Klage der GEMA gegen das Unternehmen Suno anwenden? Die GEMA legte Beispiele dafür vor, dass die Suno-KI Werke erzeugt, die sehr nah an Originalen lagen und damit als unerlaubte Vervielfältigungen zu sehen seien.“

Lisa Käde: „In der Pop-Musik kommt vermutlich verschärfend dazu, dass sich einige Muster etabliert haben, die sich in vielen Werken wiederfinden. Beispielsweise bestimmte Akkordfolgen – Am, F, C, G –, bestimmte Rhythmen oder Song-Strukturen, sodass allein in den Trainingsdaten – also in dem Musikbestand – schon vieles ähnlich klingen dürfte. Da auch Musik-KI darauf getrimmt ist, zu produzieren, was gefällt, liegt es nahe, dass die Ergebnisse in diese bekannten und bewährten Muster fallen und dementsprechend ähnlich klingen. Zudem vermute ich, dass der Bestand aktueller Pop-Musik – aus einer bestimmten zeitlichen Epoche – deutlich eingeschränkter sein dürfte als der Bildbestand und dass auch die Parameter deutlich eingeschränkter sind (Takt, Tempo, Tonart, Rhythmus, Dynamik, Instrument, Stimme, Text). Auch deshalb können hier größere Ähnlichkeiten bestehen.“

prompt/: „Im EU-Urheberrecht gibt es die ‚Pastiche‘-Regelung, wonach künstlerisch verfremdende Bearbeitungen geschützter Werke erlaubt sein können. Wäre das ein rechtlicher Weg für solche KI-generierten Werke, in denen Originale erkennbar, aber eben nicht vervielfältigt sind?“

Lisa Käde: „Hier dürfte es unter anderem darauf ankommen, ob die Nutzerin es mit ihrem Prompt darauf anlegt, verschiedene Werke zusammenzuführen oder fremde Werke zu nutzen. So könnte sie sich etwa Kombinationen überlegen wie ‚Mona Lisa spaziert durch Paris‘ oder ‚Mickey und Donald stehen in Monets Garten‘. Wenn eine Übernahme fremder Werkteile bewusst und im Rahmen einer geistigen Auseinandersetzung erfolgt, könnte die Pastiche-Schranke auch in diesem Fall an Bedeutung gewinnen – wobei auch noch abzuwarten ist, wie sich der Pastiche-Begriff in der Rechtsprechung entwickelt. Spannender ist vielleicht sogar, wie es aussieht, wenn eine Übernahme gar nicht beabsichtigt war. Etwa durch den Prompt ‚zwei Enten stehen in einem malerischen Garten‘, aber dieser Prompt dennoch durch das KI-System so interpretiert wird, dass im Ergebnis eine Übernahme fremder Werkteile erfolgt.“

Dann dürfte es überaus fraglich sein, ob eine inhaltliche beziehungsweise geistige Auseinandersetzung mit dem Ausgangswerk erfolgt, die die Anwendung der Pastiche-Schranke rechtfertigt. Auch beobachten wir, dass es nicht ausreicht – wie früher, ‚vor KI‘ – nur auf das Ergebnis zu schauen, sondern verstärkt auch auf den dahinter liegenden Prozess, die ‚Werkerschöpfungskette‘. Im Rahmen der Ermittlung des Urheberrechtsschutzes an KI-Output passiert das jetzt schon auf der Suche nach einer persönlichen geistigen Schöpfung: Es wird untersucht, ob ein Mensch das KI-System so gesteuert hat, dass der Output eine persönliche geistige Schöpfung darstellt. Und auch beim Pastiche wird vermutlich zu erforschen sein, ob es ‚die Idee der KI war‘, bestimmte Elemente zu nutzen und zu kombinieren, oder ob ein Mensch das beabsichtigt hat.“

Aus „prompt/“, Oktober 2025: Rechtliche Fragen zu den Outputs generativer KI

Was bedeuten Input und Output bezüglich generativer KI? Damit ein KI-Modell Ergebnisse generieren kann, muss es zuvor Input bekommen, also mit bestehenden Inhalten trainiert und durch die Eingaben der Nutzer*innen (Prompts) zu einem bestimmten


Auftrag angeregt werden. Zu den Trainingsdaten zählen in der Regel auch urheberrechtlich geschützte Werke. Rein rechtlich muss die Nutzung dieser Werke entweder durch die Urheber beziehungsweise Rechteinhaber*innen erlaubt worden sein oder durch bestimmte Ausnahmeregelungen im Urheberrecht ermöglicht werden (die Text- und Datamining-Schranke beispielsweise). Der Output umfasst hingegen die Inhalte, die eine generative KI auf Basis dieser Trainingsdaten und der Prompts eigenständig erstellt.

Sind durch KI erzeugte Outputs urheberrechtlich geschützt? Für einen wirksamen Urheberrechtsschutz ist der „menschliche Faktor“ entscheidend. Nur persönliche geistige Schöpfungen einer natürlichen Person sind schutzfähig. Der Output von KI ist daher nicht urheberrechtlich schutzfähig. Das heißt, die Eingabe von Prompts – ob einfache Befehle, umfangreiche Anweisungen oder komplexe Aufgabenstellungen – ist nach bisheriger Rechtslage nicht ausreichend, um ein urheberrechtlich geschütztes Werk zu schaffen. Es wird davon ausgegangen, dass der Mensch den kreativen Prozess dabei nicht ausreichend kontrolliert oder steuert.

Wie ist die Rechtslage bei hybriden Werken? Wurde ein Werk teilweise von einer KI generiert und teilweise von einem Menschen geschaffen, entsteht eine „hybride Schöpfung“. Bearbeitet jemand ein KI-generiertes Werk in eigenschöpferischer Arbeit weitgehend genug, kann das Endergebnis urheberrechtlichen Schutz genießen. In solchen Fällen spielt die sogenannte Schöpfungshöhe eine entscheidende Rolle. Danach wird bemessen, wie hoch der Anteil der persönlichen menschlichen Leistung an einem Werk ist. Wer beispielsweise einen von der KI erzeugten Text weiter ausformuliert oder ein KI-Bild grafisch in hinreichender Weise verändert, kann Urheber des neu entstandenen Werkes werden. Die Veränderungen müssen dabei nicht außergewöhnlich umfangreich sein. Geringfügige oder unwesentliche Bearbeitungen, wie das Umschreiben von Wörtern oder das Ändern von Farben, reichen allerdings nicht aus, um urheberrechtlichen Schutz zu erlangen.

Wer ein geschütztes Werk mit Hilfe eines KI-Werkzeugs bearbeitet – etwa bei Übersetzungen oder aufwändigen Adaptionen von Romanen für Hörspiele – kann dadurch nicht das Urheberrecht am Ausgangswerk tilgen. Generell dürfen Bearbeitungen eines Werkes nicht ohne Zustimmung der Urheber veröffentlicht werden (Paragraph 23 Urheberrechtsgesetz). Bei hybriden Outputs ist oft nur schwer oder gar nicht genau messbar, wie hoch der Anteil der KI war und wie hoch der des Menschen. Entscheidend ist hierbei, inwieweit der Mensch das KI-Produkt kontrolliert und beeinflusst hat. Wie so oft bei urheberrechtlichen Fragen, läuft es hier auf Einzelprüfungen beziehungsweise gerichtliche Einzelfallentscheidungen hinaus.

Existiert ein Leistungsrecht für KI-Outputs?: Im Urheberrecht sind auch sogenannte Leistungsschutzrechte verankert. Sie stehen jenen zu, die an der Entstehung oder Verbreitung eines Werkes beteiligt sind. Dazu gehören zum Beispiel ausübende Künstler*innen oder Unternehmen, die technisch oder finanziell für die Produktion sorgen. Beim Leistungsschutzrecht geht es nicht um die kreative Leistung, sondern um den Aufwand und die Investitionen, die in ein Werk gesteckt wurden. Nach derzeitigem Stand wird davon ausgegangen, dass es kein Leistungsrecht für KI-Outputs gibt.




Newsletter
prompt/
KI Outputs

3

KI-Kennzeichnung

Die KI-Verordnung sieht Kennzeichnungen von KI-generierten Inhalten und Werken vor. Doch wer prüft, ob sie auch umgesetzt werden? Und was sanktioniert etwaige Verstöße? Zu diesbezüglichen Pflichten von Rundfunkveranstaltern und Telemedienanbietern und zu den Zuständigkeiten der Landesmedienanstalten befragen wir deren Themenbeauftragte für KI. Zudem geben wir einen Überblick, welche Kennzeichnungsregeln laut KI-VO ab dem 2. August 2026 EU-weit gelten.



Newsletter
prompt/
KI-Kenn-
zeichnung



Ruth Meyer ist Direktorin der Landesmedienanstalt Saarland (LMS) und innerhalb der Gemeinschaft der Landesmedienanstalten die Themenbeauftragte für Künstliche Intelligenz. Wir haben die Fragen per E-Mail gestellt, Ruth Meyer hat sie uns schriftlich beantwortet.

Foto: Carsten Simon

Aus „prompt/“, Februar 2026: „Die KI-Kennzeichnung sollte als einheitliches, nicht entfernbare Symbol über die gesamte Verbreitungskette hinweg bestehen bleiben“

Wenn Rundfunkveranstalter und Telemedienanbieter KI-Systeme betreiben sowie komplett oder teilweise KI-generierte Inhalte veröffentlichen, fällt dies auch in die Aufsichtsbereiche der Landesmedienanstalten. Welche Rolle sie dabei spielen und wie sie auf die kommenden Kennzeichnungspflichten blicken, erfragten wir bei Ruth Meyer.

prompt/: „Die KI-Verordnung der EU enthält Transparenz- und Kennzeichnungspflichten bezüglich des Einsatzes von KI. Was bedeutet das für die Landesmedienanstalten?“

Ruth Meyer: „Diese Kennzeichnungspflichten gemäß Artikel 50 der KI-VO betreffen auch Rundfunkveranstalter wie Telemedienanbieter, soweit diese Anbieter oder Betreiber von KI-Systemen sind – und zwar ab August 2026 beziehungsweise Februar 2027. Die Medienanstalten sind die geborenen zuständigen Behörden für die nationale Umsetzung dieser Pflichten und werden für die Marktüberwachung im Bereich Rundfunk und Telemedien bezüglich der KI-VO zuständig sein. Sie sind bei den Rundfunkveranstaltern und Telemedienanbietern als vertraute Ansprechpartner bekannt, zudem war dies auch aus Gründen der Staatsferne absolut geboten. Aktuell bringen sich die Medienanstalten bereits in den Prozess der Konkretisierung der Kennzeichnungsleitlinien im Praxisleitfaden ein, dem Code of Practice, kurz CoP.“

prompt/: „Was sollten KI-Kennzeichnungen in Medien Ihrer Auffassung nach erfassen – und was nicht?“

Ruth Meyer: „Wir halten eine KI-Kennzeichnung insbesondere dann für erforderlich, wenn für durchschnittlich informierte und aufmerksame Nutzer:innen nicht auf den ersten Blick ersichtlich ist, dass die Inhalte künstlich erzeugt oder manipuliert wurden. Insofern unterstützen wir eine einheitliche Form zur Kennzeichnung von KI. Diese soll dagegen nicht erforderlich sein bei geringfügigen, mittels KI vorgenommenen technischen Optimierungen, wie Rauschunterdrückung, Farbkorrektur oder kleineren Korrekturen, solange diese den Inhalt, Kontext oder Aussagegehalt nicht signifikant verändern. Sinnvoll ist auch eine Art Vertrauensvorschuss für Medien, die nach etablierten professionellen Standards arbeiten.“

prompt/: „Wie könnten KI-Kennzeichnungen idealerweise konkret aussehen und funktionieren?“

Ruth Meyer: „Genau diese Fragestellung soll der erwähnte Praxisleitfaden konkretisieren. Hierin werden konkrete Ausgestaltungen für eine entsprechende KI-Kennzeichnung vorgeschlagen, beispielsweise wie ein grafisches Logo aussehen könnte, wie ein Texthinweis auszugestaltet ist. Oder auch, welche technischen Anforderungen an die Kennzeichnung zu stellen sind, etwa, dass sie durch ein zu entwickelndes System der EU maschinell ausgelesen und überprüft werden kann. Je nach geltender Vorschrift ist die Kennzeichnungspflicht auf Maschinenlesbarkeit beschränkt oder muss darüber hinaus für die Rezipient:innen klar, sofort sichtbar und unmissverständlich sein.“

Die KI-Kennzeichnung sollte als einheitliches, nicht entfernbares Symbol über die gesamte Verbreitungskette hinweg robust bestehen bleiben. Gut wäre ein EU-weit einheitliches Symbol, um die Wiedererkennung zu fördern. Zudem sollten vor der finalen Festlegung empirische Nutzertests erfolgen, da die Wirkung stark vom Kontext und der Zielgruppe abhängt. Die von uns beauftragte Studie ‚Transparenz-Check: Wahrnehmung von KI-Journalismus‘ hat gezeigt, dass implizite Hinweise von den Nutzer:innen oft nicht verstanden werden.“

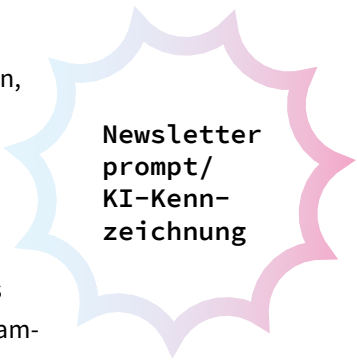
prompt/: „Die Untersuchungen für diese Studie fanden 2024 statt. Wie hat sich aus Sicht der Medienanstalten die Lage zu KI-generierten Inhalten in Medien und zu deren Kennzeichnungen verändert?“

Ruth Meyer: „Quintessenz der Befunde von 2024 war, dass die zunehmende Nutzung KI-generierter Inhalte in den Medien grundsätzlich auf Akzeptanz stößt. Es werden klare Regeln für eine transparente vertrauenswürdige Kennzeichnung von Avataren, KI-generierten Stimmen oder von KI verfassten Artikeln gewünscht. Inzwischen haben sich in den verschiedenen Mediengattungen in Selbstverpflichtung unterschiedlichste Kennzeichnungsarten etabliert: Bei fast allen Medien findet man Bezeichnungen für Programme oder Formate, beispielsweise ‚Absolut AI‘ oder ‚Avatar-Talk‘, aber auch Einblendungen wie ‚erstellt mit KI-Unterstützung‘ oder eingesprochene Hinweise.“

Außerdem finden sich Informationen in Guidelines und Footern darüber, auf welchen Grundlagen journalistische Inhalte automatisiert generiert und distribuiert werden. Entscheidend ist, dass die Verantwortung für KI-Inhalte bei den Telemedienanbietern und Rundfunkveranstaltern liegt, die diese Systeme einsetzen. Sofern dabei medienrechtliche Vorgaben nicht eingehalten werden, tragen die Rechtsfolgen hierfür die Inhalteanbieter beziehungsweise Vermittlungsdienste, insbesondere soziale Netzwerke wie Facebook und X, Online-Marktplätze wie eBay und Amazon, Suchmaschinen wie Google und Bing, oder Internetzugangsdienste wie Telekom und Vodafone – und zwar unabhängig davon, wie die Inhalte erstellt wurden.“

prompt/: „Wenn die KI-VO-Transparenzpflichten ab August dieses Jahres in Kraft treten, wird es dann Sanktionen gegen Verstöße oder Unterlassungen geben? Und wie sehen die aus?“

Ruth Meyer: „Bei Verstößen oder Unterlassungen gegen die Transparenzpflichten für Anbieter und Betreiber gemäß Artikel 50 der KI-VO ergeben sich die möglichen Sanktionen aus Artikel 99 Absatz 4 Buchstabe g der KI-VO. Demnach werden Geldbußen von bis zu 15 Millionen Euro oder – im Falle von Unternehmen – von bis zu drei Prozent des gesam-



Newsletter
prompt/
KI-Kenn-
zeichnung

3

ten weltweiten Jahresumsatzes des vorangegangenen Geschäftsjahres verhängt, je nachdem, welcher Betrag höher ist. Ergänzend hierzu können die Mitgliedstaaten auch Verwarnungen und nichtmonetäre Maßnahmen erlassen und anwenden, sofern diese die von der EU-Kommission erteilten KI-VO-Leitlinien berücksichtigen.

Jenseits der KI-VO enthält aber auch der Medienstaatsvertrag Vorschriften bezüglich Sorgfalt und Transparenz, die beim Einsatz von KI zu beachten sind und bereits jetzt zu Beanstandungen führen können. Beispielhaft können hierzu Verstöße gegen die Werbegrundsätze und die Werbekennzeichnungs- und Trennungspflichten im Rundfunk und in Telemedienangeboten oder Verstöße gegen die journalistischen Sorgfaltspflichten genannt werden.“

Aus „prompt/“, Februar 2026: Vorgaben der KI-VO zu Kennzeichnungen von KI-Systemen und KI-Outputs

Nicht alles, was mit KI-Werkzeugen generiert wird, muss laut KI-Verordnung gekennzeichnet werden, aber doch einiges. Wir fassen zusammen, was ab dem 2. August 2026 EU-weit rechtlich gelten wird.

„Wurden diese Inhalte nur von Menschen erstellt, oder wurde auch generative KI eingesetzt?“ Das fragen sich Nutzer*innen und Verbraucher*innen immer öfter. Denn Texte, Fotos, Grafiken, Videos, Musik und sogar Stimmen lassen sich zunehmend weniger eindeutig zuordnen. Genau deshalb rückt die Kennzeichnung von KI-generierten Inhalten immer stärker in den Fokus.



Eine zentrale Rolle spielt dabei die KI-Verordnung der EU. Sie enthält Vorgaben zu Kennzeichnungspflichten, festgehalten im Artikel 50. Die darin formulierten Regelungen sollen für mehr Transparenz auf dem Markt sorgen und treten für die meisten KI-Anbieter ab dem 2. August 2026 in Kraft. Ziel ist es, Nutzer*innen klar darüber zu informieren, wann sie mit Künstlicher Intelligenz interagieren oder KI-generierte Inhalte wahrnehmen. Allerdings sind nicht alle KI-Anwendungen und nicht alle KI-generierten Inhalte laut KI-VO zu kennzeichnen. Vielmehr trifft dies auf bestimmte, im Folgenden erläuterte Anwendungsfälle zu.

Wer muss kennzeichnen? Die Kennzeichnungsvorgaben der KI-Verordnung betreffen je nach konkreter Pflicht die Anbietenden und die Betreibenden der KI-Systeme. Das kann – je nach konkretem Inhalt – also auch jene betreffen, die KI-gestützte Inhalte veröffentlichen: Verlage, Sender, Medienhäuser, Podcaster und Blogger sowie – nicht zuletzt – auch Social-Media-Plattformen, doch ebenso auch Handelsunternehmen, Behörden, Dienstleister, Verbände und so weiter. Wichtig ist aber auch, zu differenzieren: Es gibt zudem Kennzeichnungspflichten außerhalb der KI-Verordnung – gerade für all jene, die öffentlich kommunizieren.

Was muss gekennzeichnet werden? Zum einen gelten die Vorschriften für jene KI-gestützten Systeme und -Dienste, die eigenständig mit natürlichen Personen interagieren (beispielsweise Chatbots) und mit denen sich synthetische Inhalte erzeugen lassen, die Emotionen und biometrische Merkmale erkennen und kategorisieren, etwa zur Stimmenanalyse. Zu kennzeichnen sind zum anderen KI-generierte oder KI-manipulierte Inhalte, die echt wirken und Personen täuschen sollen – die KI-Verordnung spricht dann von Deepfakes. Aber auch veröffentlichte Texte, mit denen die Öffentlichkeit über gesellschaftlich relevante The-

men informiert wird, fallen darunter, also Nachrichten, Berichte, Meldungen oder politische Kommentare.

Welche Ausnahmen gibt es, bei denen keine Kennzeichnung gefordert wird? Etwa, wenn KI-generierte oder -manipulierte Medieninhalte von Menschen redaktionell geprüft und verantwortet wurden. Das ist bei etablierten, professionell und seriös vorgehenden Großverlagen, Sendern und Medienhäusern Standard. Zugleich agieren zunehmend neue, kleine und kleinste Anbieter in Online-Medien und -Kanälen, die ihre „News“ fast vollständig automatisiert erstellen lassen und ungeprüft veröffentlichen. Auf Social-Media-Plattformen oder bei KI-gestützten Suchmaschinen finden sich ebenfalls vermehrt maschinell erzeugte und nicht von Menschen geprüfte Beiträge.

Auch für Werke der Kunst, der Fiktion oder der Satire, die klar als solche erkennbar sind, sieht die Verordnung Ausnahmen von der Kennzeichnungspflicht vor. Dazu heißt es in Artikel 50, Absatz 4 der KI-Verordnung: „Ist der Inhalt Teil eines offensichtlich künstlerischen, kreativen, satirischen, fiktionalen oder analogen Werks oder Programms, so beschränken sich die in diesem Absatz festgelegten Transparenzpflichten darauf, das Vorhandensein solcher erzeugten oder manipulierten Inhalte in geeigneter Weise offenzulegen, die die Darstellung oder den Genuss des Werks nicht beeinträchtigt.“

Wie muss gekennzeichnet werden? Generell muss die Kennzeichnung zu dem Zeitpunkt wahrzunehmen sein, zu dem eine Person mit einem KI-System interagiert oder einen KI-generierten oder -manipulierten Output wahrnimmt. Sprich, sie muss direkt mit dem Inhalt verknüpft sein. Die Kennzeichnung muss außerdem klar, verständlich, gut sichtbar und barrierefrei sein. Wie genau das auszusehen hat, steht nicht in der KI-Verordnung. Muster und Vorlagen wurden bislang noch nicht festgelegt. Aktuell arbeitet die EU-Kommission an einem sogenannten Code of Practice (CoP), der konkrete Umsetzungshilfen liefern soll. Ein erster Entwurf dieses Praxisleitfadens zeigt unter anderem, wie ein Icon zur Kennzeichnung aussehen könnte. Der finale Code of Practice soll zwischen Mai und Juni 2026 veröffentlicht werden.

Auch ohne Logo werden gängige Formulierungen wie „(Vollständig/teilweise) KI-generiert“ oder „Dieser Text/diese Abbildung/dieser Film/die Stimme wurde vollständig/teilweise mit KI-Anwendungen erzeugt oder manipuliert“ als ausreichend erachtet. Wichtig ist zudem, dass die Kennzeichnung auch technisch eindeutig und unumgänglich an die Inhalte gekoppelt werden soll. Beispielsweise mit Wasserzeichen, maschinenlesbaren Metadaten oder digitalen Markierungen, die sich nicht entfernen lassen und auch nach Bearbeitungen den Ursprung sowie die Merkmale des Inhalts enthalten.

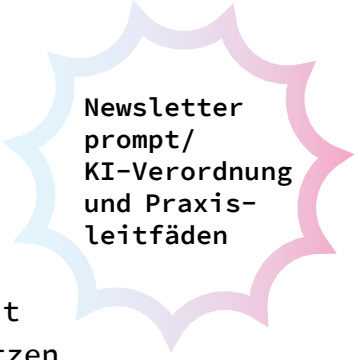
Was, wenn jemand seiner Kennzeichnungspflicht nicht nachkommt? Wer sich nicht an die ab August 2026 geltenden Pflichten hält, hat mit Bußgeldern zu rechnen. Für die Aufsicht und Kontrolle dieser Regelungen der KI-Verordnung werden die entsprechenden Behörden in den EU-Mitgliedsstaaten zuständig sein. In Deutschland soll diese Aufgaben zu großen Teilen die Bundesnetzagentur übernehmen. Im Bereich der Medienhäuser und Telemedienanbieter soll auch den Landesmedienanstalten eine Aufsichtsfunktion zukommen (siehe hierzu das Interview mit Ruth Meyer).



3

KI-Verordnung und Praxisleitfäden

Als Handlungshilfe sieht die EU-Kommission die „Praxisleitfäden“, die sie erstellen ließ, um die KI-VO zu ergänzen. Wir erläutern, wie der erste Praxisleitfaden für Allzweck-KI-Modellen zu Transparenz, Urheberrecht sowie Sicherheit und Schutzmaßnahmen die Anwender*innen unterstützen soll, die Vorgaben der KI-VO in die Praxis umzusetzen. Und wir erklären, wie sich vorherige Praxisleitfäden, etwa zu Desinformation, bewährt haben.



Newsletter
prompt/
KI-Verordnung
und Praxis-
leitfäden

Aus „prompt“, August 2025: Was Sie über den Praxisleitfaden zur KI-Verordnung wissen müssen

Nach einem etwa elfmonatigen Entwicklungsprozess hat das European AI Office (Europäisches Büro für Künstliche Intelligenz) den sogenannten General-Purpose AI Code of Practice an die EU-Kommission übergeben. Der Praxisleitfaden für Allzweck-KI-Modelle (General Purpose AI, kurz: GPAI) besteht aus drei Kapiteln – Transparenz, Urheberrecht sowie Sicherheit und Schutzmaßnahmen – und soll dabei helfen, die Vorgaben der KI-Verordnung in die Praxis umzusetzen.

Wie ist der Praxisleitfaden entstanden? Das Büro für Künstliche Intelligenz hat in einem Multi-Stakeholder-Prozess etwa 1.000 Teilnehmende aus Industrie, Wissenschaft, Zivilgesellschaft sowie internationalen und EU-Regierungen in vier Arbeitsgruppen zusammengebracht. Geleitet wurden die Gruppen von 13 ausgewählten Expert*innen. In einem iterativen Prozess wurden mehrere Entwürfe erarbeitet, bis man sich schließlich auf die aktuelle Fassung einigte.

Kritik am Prozess gibt es vor allem von zivilgesellschaftlichen Akteur*innen, die bemängeln, dass Industrie und Lobbyverbände zu großen Einfluss auf die Entwicklung des Leitfadens gehabt hätten.

Worum geht es konkret? Der Leitfaden ist in erster Linie ein sogenanntes Compliance Tool – ein Werkzeug, um bindende Selbstverpflichtungen aufzustellen. Er soll Anbietenden von Allzweck-KI-Modellen, wie beispielsweise GPT-4, Gemini oder Mistral, dabei helfen, die Vorgaben der KI-Verordnung in die Praxis umzusetzen. Unternehmen oder Organisationen können ihn freiwillig unterzeichnen und sich damit selbst dazu verpflichten, die Vorgaben einzuhalten. Ihn zu unterschreiben, soll Anbietenden vor allem zwei Vorteile bringen: Erstens, weniger administrativen Aufwand, um nachzuweisen, dass sie die Vorgaben der KI-VO einhalten. Für die Unterschreibenden gilt nämlich zunächst pauschal die Vermutung, dass sie dies tun. Für alle, die nicht unterschreiben, dürfte der Nachweis mit mehr bürokratischem Aufwand verbunden sein. Zweitens verspricht ein Bekenntnis zum Leitfaden eine positive Außenwirkung: Wer mitmacht, signalisiert Vertrauenswürdigkeit.

An wen richtet sich der Praxisleitfaden? Grundsätzlich gilt der Code of Practice für alle, die Allzweck-KI-Anwendungen anbieten. Klassische Anbietende sind etwa große


Player wie Mistral, OpenAI oder Anthropic. Aber auch kleinere Akteure fallen in diese Kategorie, sofern sie eigene KI-Modelle entwickeln und vertreiben oder bestehende KI-Modelle in ihre eigenen Produkte integrieren und sie unter eigenem Namen anbieten. Wer als Anbieter*in gilt und wer nicht, ist nicht immer ganz eindeutig. Der Übergang kann schleichend sein und muss teilweise im Einzelfall beurteilt werden. Für Unternehmen und andere Akteure, die KI-Modelle für eigene Zwecke einsetzen wollen, kann diese Unsicherheit herausfordernd sein. Denn: Für Anbietende gelten laut KI-Verordnung grundsätzlich strengere Pflichten als etwa für Betreibende von KI-Systemen.

Vorgaben sind nicht für alle Anbietenden gleich Besonders zu beachten ist, dass nicht alle Vorgaben des Praxisleitfadens für alle Anbietenden gleichermaßen gelten. Anbietende von Open-Source-Modellen müssen sich beispielsweise nicht an die Vorgaben zur Transparenz halten – außer diese Modelle stellen ein systemisches Risiko dar (siehe obiger Text). Demgegenüber müssen die Vorgaben zum Schutz von Urheberrechten ausnahmslos alle Anbietenden einhalten. Die vorgegebenen Anweisungen zu Sicherheits- und Schutzmaßnahmen wiederum gelten lediglich für Anbietende besonders leistungsfähiger KI-Modelle, von denen „systemische Risiken“ ausgehen können.

Als Grundlage der Risikoeinschätzung wird die Rechenleistung herangezogen, die für das Training eines Modells benötigt wurde. Übersteigt diese einen gewissen Schwellenwert, müssen die Anbietenden zusätzliche Auflagen erfüllen. Aktuell liegt dieser Wert laut Artikel 51 der KI-Verordnung bei 1025 FLOP (Floating Point Operations). Die Zahl markiert, wie viele Rechenschritte es braucht, um ein Modell zu trainieren. Je höher die Zahl, desto rechenintensiver das Modell. 1025 FLOPs bedeutet also, dass mindestens 10.000.000.000.000.000.000.000.000 Rechenschritte nötig waren, um ein Modell zu trainieren. Insgesamt fallen laut einer Untersuchung von Epoch AI vom Juni 2025 derzeit etwa 33 Modelle in diese Kategorie. Die Unterzeichnung des Kapitels zu Sicherheits- und Schutzmaßnahmen ist für alle, die keine Modelle dieser Größenordnung entwickeln und anbieten, optional.

Was legt der Leitfaden fest? Der Leitfaden ist in drei Kapitel unterteilt, die jeweils spezifische Maßnahmen für einen Wirkungsbereich beinhalten:

Transparenz: Anbietende verpflichten sich dabei u. a. dazu, zu dokumentieren, wie Trainings- und Testdaten beschafft wurden, welche Rechenleistung für das Training aufgebracht wurde und wie hoch der Energieverbrauch eines Modells ist. Die Verfasser*innen des Leitfadens haben ein Formular aufgesetzt, das die Unternehmen ausfüllen können und das es dem EU KI-Büro oder nationalen Prüfbehörden erleichtern soll, die Angaben der Anbietenden zu überprüfen. Außerdem gibt es nun ein sogenanntes „Trainingsdaten-Template“. Darin sollen die Anbietenden angeben, welche Trainingsdaten für das Training ihrer Modelle genutzt werden. Das ist vor allem für Urheber*innen und Verwertungsgesellschaften interessant. Fraglich bleibt jedoch, wie detailliert diese Angaben sein werden. Der Leitfaden fordert lediglich eine „hinreichend detaillierte Zusammenfassung“ der Trainingsdaten und bleibt damit recht unkonkret.



Newsletter
prompt/
KI-Verordnung
und Praxis-
leitfaden

3

Urheberrecht: Insgesamt legt der Leitfaden fünf Maßnahmen fest, die Anbietende ergreifen sollen, um die Rechte von Urheber*innen zu schützen. Dazu gehört u. a., keine Piraterie-Webseiten für Trainingsdaten zu crawlen, Rechtebeschränkungen im Sinne der DSA-Verordnung einzuhalten oder Opt-out-Protokolle zu respektieren (zum Beispiel auf Basis einer robot.txt-Datei). Außerdem sollen ausreichende technische Maßnahmen ergriffen werden, um zu vermeiden, dass KI-Modelle Outputs generieren, die Urheberrechte verletzen. Rechteinhaber*innen muss es zudem möglich sein, sich bei den Anbietenden zu beschweren.

Sicherheit und Schutzmaßnahmen: Hier sind die Vorgaben des Leitfadens am umfangreichsten. Anbietende müssen unter anderem fortlaufend protokollieren, wie sie mögliche Risiken und Schäden, die während der Entwicklung oder dem Einsatz eines GPAI-Systems entstehen können, identifizieren und bewerten – aber auch, wie sie den Risiken und Schäden entgegenwirken wollen. Diese Risikobewertung muss dem KI-Büro der EU vor Veröffentlichung eines Modells vorgelegt und alle sechs Monate aktualisiert werden. Die Liste der möglichen Risiken, die dabei beachtet werden müssen, umfasst unter anderem Cyberangriffe, schädliche Manipulationen oder künstlich erzeugte Nacktbilder, die auf Fotos echter Personen basieren.

Außerdem müssen Anbietende interne Governance-Strukturen schaffen, die eine fortlaufende Risikobewertung ermöglichen und eine „gesunde Risikokultur“ fördern.

Ab wann gilt der Leitfaden? Der Leitfaden gilt seit seiner Veröffentlichung am 10. Juli 2025, muss aber noch von den EU-Mitgliedsstaaten und der EU-Kommission gebilligt werden. Außerdem räumt die EU-Kommission den Unterschreibenden eine Schonfrist ein. Erst ab August 2026 sollen Sanktionen gegenüber den Unternehmen erfolgen.

Was bleibt offen? Der Praxisleitfaden sorgt zwar in mancher Hinsicht für mehr Klarheit. Einige Maßnahmen bleiben allerdings ungeklärt, andere haben es nicht in die finale Fassung geschafft. Anbietende müssen zum Beispiel keine Notfallpläne mehr anlegen. Sie müssen also nicht mehr im Voraus festlegen, wie sie im Ernstfall mit Sicherheitsrisiken umgehen. Externe Prüfungen sind nicht vorgesehen. Anbietende überprüfen quasi selbst, ob sie die

Auflagen des Praxisleitfadens wirklich einhalten, ohne dass ihre Angaben von Dritten oder der EU noch einmal begutachtet werden müssten. Gleichwohl muss das KI-Büro der EU die Umsetzung der Vorgaben der KI-Verordnung durch die KI-Anbietenden überwachen und Versäumnisse ahnden.

Wer macht bereits mit? Zu den bekanntesten Unterzeichnenden gehören bisher Unternehmen wie Mistral, OpenAI, Microsoft, Anthropic, Google oder Aleph Alpha. Da Aleph Alpha sich aus der Entwicklung von Allzweck-KI-Modellen zurückgezogen hat, unterzeichnet das Unternehmen laut eigenen Angaben lediglich die Leitfäden zu Transparenz und Urheberrecht. Auf der Webseite zum Code of Practice listet die EU-Kommission die Unterzeichnenden auf, Stand 11. August 2025 waren das 26 Unternehmen. Das Bekenntnis der Unterzeichnenden, insbesondere der bekannten Big Player, könnte ein Signal an weitere Unternehmen sein, sich dem Leitfaden ebenfalls anzuschließen. Es gibt aber auch Verweigerer: Der Facebook-Mutterkonzern Meta hat angekündigt, den Leitfaden nicht zu unterzeichnen, der Anbieter XAI will nur Teile des Codes unterzeichnen.

KI-Anbieter oder -Betreiber: Worin besteht der Unterschied? Anbieter sind laut KI-Verordnung Unternehmen, die ein KI-Modell selbst entwickeln (beziehungsweise an der Entwicklung beteiligt sind) oder ein bestehendes Modell in ein eigenes Produkt integ-

Newsletter
prompt/
KI-Verordnung
und Praxis-
leitfäden

rieren und unter eigener Marke vermarkten. Die reine Einbettung eines fremden KI-Systems, etwa in die eigene Website, reicht in der Regel nicht aus, um als Anbieter zu gelten.

Betreiber hingegen setzen KI-Systeme laut KI-Verordnung lediglich für interne Zwecke ein, ohne sie weiterzuentwickeln oder selbst anzubieten – zum Beispiel zur Prozessoptimierung oder im Kundenservice. Sie müssen die Systeme regelkonform nutzen, unterliegen aber nicht denselben Pflichten wie Anbieter.

Wichtig ist hier außerdem die Unterscheidung zwischen KI-Modell und KI-System. Ein Modell ist die Grundlage, auf dem ein KI-System aufbaut und ist alleine noch nicht funktionsfähig. Ein Beispiel: GPT4 ist das KI-Modell des Unternehmens Open AI, auf dem das KI-System ChatGPT aufsetzt beziehungsweise eingebettet ist. Das Modell kann auch für andere Systeme genutzt werden. Gerade Allzweck-Modelle sind in vielen unterschiedlichen Kontexten einsetzbar.

Aus „prompt/“, August 2025: Vorbild aus dem Jahr 2018:
Der „Code of Practice on Disinformation“

Ein Praxisleitfaden als Umsetzungshilfe – kein neuer Ansatz der EU-Kommission Der Code of Practice (Praxisleitfaden) ist eines von mehreren in der KI-VO vorgesehenen „Hilfsmitteln“, um Unternehmen, Organisationen und Institutionen dabei zu helfen, die rechtlichen Vorgaben der Verordnung umzusetzen. Dazu gehören unter anderem folgende:

- Harmonised Standards (deutsch: harmonisierte Normen) werden bei den europäischen Normungsorganisationen CEN und CENELEC erarbeitet und betreffen vor allem Hochrisiko-KI-Systeme
- Guidelines (deutsch: Leitlinien) sind von der Europäischen Kommission erarbeitete Auslegungshilfen
- Codes of Practice (deutsch: Praxisleitfäden) sind von der Europäischen Kommission zusammen mit Unternehmen und Interessierten erarbeitete Selbstverpflichtungen
- Codes of Conduct (deutsch: Verhaltenskodizes) sind von Unternehmen ohne unmittelbare Mitwirkung der Europäischen Kommission erarbeitete Selbstverpflichtungen

Weniger Überprüfungen, mehr Vertrauenswürdigkeit Sich dem Praxisleitfaden (freiwillig) anzuschließen, hat für die Unterzeichnenden den Vorteil, dass die EU-Kommission sie weniger auf Einhaltung der diesbezüglichen KI-VO-Regelungen kontrolliert. Demgegenüber müssten Anbietende, die ihn nicht unterzeichnen, mit umfänglicheren Prüfungen rechnen. Der Praxisleitfaden ließe sich demnach auch als eine, salopp formuliert, „Beglaubigung“ für gutes, weil gesetzeskonformes „Benehmen“ betrachten. Und das soll sich – so die Absicht der EU-Kommission – positiv auf die Vertrauenswürdigkeit der Anbietenden auswirken. Hingegen könnten jene, die ihn nicht unterzeichnen, an Reputation einbüßen.

Vorbild: Verhaltenskodex zur Bekämpfung von Desinformation Diesen Regelungsansatz wählte die EU-Kommission schon einmal: 2018 ging es auf europäischer Ebene um die Bekämpfung von Desinformation im Netz. Als ein erster Entwurf des Digital Services Act (DSA) vorbereitet wurde, erarbeitete die „Generaldirektion Kommunikationsnetze, Inhalte und Technologien“ – dieselbe Stelle in der EU-Kommission, die jetzt für die KI-VO zuständig

3

ist – zusammen mit großen Onlineanbietern den „Praxisleitfaden zur Bekämpfung von Desinformation“ (englisch: Code of Practice on Disinformation).

In diesem verpflichten sich die Unterzeichnenden auf Vorgehensweisen und Mindeststandards, um der Verbreitung von Desinformation auf ihren Diensten und Plattformen entgegenzuwirken. Beispielsweise, indem die Verantwortlichen für Desinformation keine werbebasierte Finanzierungsmöglichkeit erhalten. Dieser Praxisleitfaden – oft mit dem Begriff „Verhaltenskodex“ übersetzt – wurde zunächst von Unternehmen wie Facebook, Google, X und Mozilla unterzeichnet, bis 2020 traten ihm auch Microsoft und TikTok bei.

2022 und 2025: Überarbeitung und Anerkennung als „Code of Conduct“

2022 – dem Jahr, in dem erste Teile des Digital Services Act in Kraft traten – überarbeiteten 34 Organisationen auf Vorschlag der Kommission den „Code of Practice on Disinformation“. Zu den Zielen gehörte, die auferlegten Verpflichtungen bereichs- und dienstspezifisch zu präzisieren, sowie die Zusammenarbeit mit Forschenden und Faktenprüfer*innen zu erleichtern. Mittlerweile haben 41 Unternehmen den Code of Practice 2022 unterzeichnet, neu hinzu kamen unter anderem Adobe und Twitch.

2025 wurde eine Brücke zwischen Digital Services Act und dem parallel entwickelten Code of Practice geschlagen: Die Europäische Kommission und das beratende Gremium für digitale Dienste erkannten ihn formal als „Code of Conduct“ (der englische Begriff für Verhaltenskodex) im Sinne des Artikel 45 des Digital Services Act an. Spätestens mit diesem Schritt können sich die Unterzeichnenden des „Code of Practice 2022“ der Kommission gegenüber darauf berufen, mittels dessen Einhaltung auch geeignete Risikominderungsmaßnahmen nach dem Digital Services Act vorgenommen zu haben.

Fazit: Regelungs-Instrumente wie Code of Practice und Verhaltenskodex scheinen sich zu bewähren Wie sich anhand mancher Unschärfe bei den Begriffen Praxisleitfaden, Leitfaden und Verhaltenskodex zeigt, agieren unterschiedliche Ebenen in der EU sprachlich nicht immer einheitlich. Gleichwohl lassen sich die Aufnahme von Codes of Practice und Codes of Conduct in die Regelungen der KI-Verordnung – getreu dem Vorgehen beim DSA – durchaus perspektivisch interpretieren: Aus Sicht der EU scheint sich dieses Zusammenspiel aus rechtlichen Vorgaben, unternehmerischen Selbstverpflichtungen und Erleichterungen in der Rechtsdurchsetzung bewährt zu haben.

Wie zufrieden sind wir mit dem prompt/-Newsletter? Lesen Sie dazu unsere Einschätzungen in Kapitel 4.



Lunch-Talk: Forschungsbefunde und Deep Dive zum Thema KI und Haftung

Anfang September 2025 laden wir zu einem GenKI-IR-Lunch-Talk ein. Die Online-Veranstaltung dient zum einen dazu, die wesentlichen Zwischenergebnisse unserer bisherigen Forschungsarbeit vorzustellen, und zum anderen, Einblick in die für uns wichtigsten Erkenntnisse und Befunde zu geben. Hierbei gehen wir zudem schwerpunktmäßig auf Fragen zur Haftung bei KI-verursachten Schäden ein. Nach Kurzvorträgen folgt eine rege Diskussion.

Lunch Talk



»Was, wenn was passiert?«

Generative KI, Haftung
und Innovation

04. Sept. 2025
13:00–14:00 Uhr



GenKI iRights.Lab

In einer Präsentation geben wir einen ersten Einblick in die für uns wichtigsten Erkenntnisse und Befunde, die wir nach dem Lunch-Talk als Zwischenbericht in Form eines Working Papers veröffentlichen. Darin formulieren wir – entlang der drei Phasen des KI-Lebenszyklus – zehn Thesen zu rechtlichen Fragen.

Ergänzend dazu stellen wir die Frage: Wer haftet, wenn beim Einsatz generativer KI ein Schaden entsteht? Unter dem Titel „Was, wenn was passiert? Generative KI, Haftung und Innovation“ erörtern wir gemeinsam mit zwei eingeladenen Expert*innen, welche gesetzlichen Rahmenbedingungen für Haftungsfragen beim Einsatz generativer KI relevant sind, wie die Verantwortung in der Wertschöpfungskette und bei KI-Agenten zugewiesen werden kann und wie sich der regulatorische Rahmen in Zukunft entwickeln könnte. Im Anschluss an die Vorträge erhalten die Teilnehmenden Gelegenheit für Fragen und Raum für eine vertiefende Diskussion.

EU-Rechtsrahmen zur KI-Haftung: Produkthaftungsrichtlinie und KI-Verordnung

In das Schwerpunkt-Thema des Lunch-Talks führt Prof. Dr. Martin Ebers ein, Präsident der Robotics & AI Law Society (RAILS) und Professor für IT-Recht an der Universität Tartu. In seinem Impulsvortrag erläutert er relevante Rechtsrahmen und wichtige Fragen zur Haftung bei KI-verursachten Schäden.

3

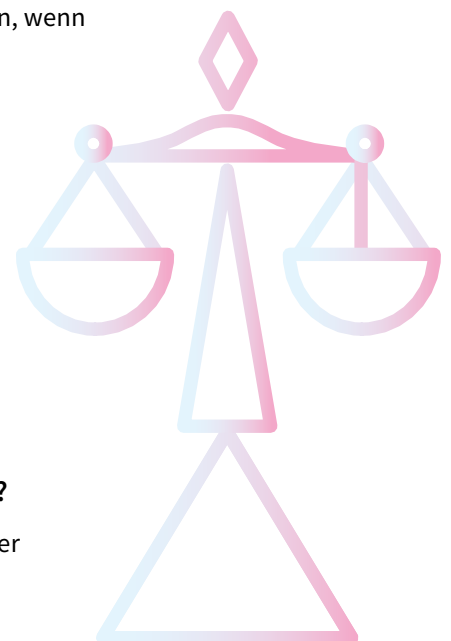
Er zitiert eine Erhebung, wonach die Anzahl von KI-bezogenen, haftungsrelevanten Vorfällen zwischen 2012 und 2023 angestiegen sei sowie Unfallbeispiele aus dem „AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC) Repository“. Dort werden u. a. Vorfälle gesammelt, bei denen KI-gestützte Tools die Kleidung von Personen entfernten, fälschliche Beschuldigungen aufstellten, Deepfakes eine Rolle spielten sowie tödliche Autopiloten. Als relevanter europäischer Rechtsrahmen für solche Haftungsfälle gilt zum einen die EU-Produkthaftungsrichtlinie (Product Liability Directive, kurz „PLD“). Sie wurde 2024 in Kraft gesetzt und muss von den europäischen Mitgliedstaaten bis Ende 2026 in nationales Recht umgesetzt werden. Zum anderen gibt es aber auch die KI-Verordnung der EU, die ebenfalls seit 2024 schrittweise in nationales Recht umgesetzt werden muss und Verantwortlichkeiten bei Entwicklung und Einsatz von KI-Systemen reguliert. Darüber hinaus erläutert Ebers die deutsche Rechtslage, die sowohl vertragliche als auch deliktische Haftung vorsieht.

In seinem Vergleich von KI-Verordnung und Produkthaftungsrichtlinie (PLD) stellt er fest, dass die KI-VO zwar keine spezifische Haftung vorsieht, jedoch die Pflichten des Anbieters festlegt. Demnach handelt es sich um eine sogenannte ex-ante-Produktionssicherheit („von vorher“, also vor einer Handlung), während die PLD eher ex-post-Haftung regelt („von nachher“, sprich nach einer Handlung). Seiner Ansicht nach umfasst die Produkthaftungsrichtlinie auch eine Haftung für Schäden, die durch fehlerhafte Software entstanden sind, in diesem Fall durch generative KI.

Allerdings sei derzeit unklar, ob und wie die rechtliche Lage bei Service-Produkten (AI-as-a-Service) ist, in denen KI-gestützte Tools integriert sind. Weitere solcher „Haftungslücken“ in den genannten Regulierungen betreffen die Verletzung des allgemeinen Persönlichkeitsrechts, Diskriminierung, Beschädigung gewerblich genutzter Sachen und reine Vermögensschäden. In solchen Fällen ist ein Rückgriff auf nationales Recht nötig. Eine weitere Regelung in der PLD betrifft Produktfehler. Bezogen auf KI-Systeme ist jedoch nicht abschließend geklärt, ob bei Haftungsfragen bereits eine einzelne fehlerhafte Information ausreicht oder ob das gesamte KI-System betrachtet werden muss und welche Kriterien dem zugrunde liegen. Laut Ebers erleichtert die 2024 reformierte Produkthaftungsrichtlinie die Durchsetzung eines Schadensersatzanspruchs, es sei eine Verbesserung für Nutzer*innen und Verbraucher*innen. Eine Haftung tritt ein, wenn körperliche oder psychische Schäden entstehen oder private Daten vernichtet oder verfälscht werden. Im Falle eines Schadens gelten zugunsten des Geschädigten Erleichterungen bei der Beweisführung. Zudem wurde die Verantwortlichkeit von Unternehmen erweitert, denn auch Software und KI-Systeme sind nun in die Haftungsrichtlinie einbezogen, auch wenn sie über Cloud-Dienste oder Software-as-a-Service-Modelle bereitgestellt werden. Dennoch wird bei vielen Schadensfällen ein Rückgriff auf das nationale Haftungsrecht nötig sein.

Einstufung von KI-Agenten nach deren Autonomiegrad?

Im zweiten Vortrag beschäftigt sich Julia Smakman, Researcher am Ada Lovelace Institute, mit Haftungsfragen in Bezug auf sogenannte KI-Agenten, die teilweise autonom agieren. Sie



leitet dabei Erfahrungen und Lehren aus dem „Automated Vehicles Act“ des Vereinigten Königreichs ab, der dort 2024 in Kraft trat.

Dieses Gesetz enthält Regulierungen für autonome Fahrzeuge. Es unterscheidet zwischen fünf „Levels“ an Autonomie, die sich nach dem Grad der Eingriffs- und Steuerungsmöglichkeiten richtet, die Menschen bei der Bedienung von Fahrzeugen haben. In Anlehnung an diese Normen schlägt Julia Smakman eine Kategorisierung von KI-Agenten-Systemen vor, also auch einem klar definierten Autonomiegrad entsprechend. Eine Reihe von Fragen seien bei diesem Ansatz noch offen. Etwa, ob es möglich ist, alle relevanten technischen Aspekte zu definieren, mit denen die autonomen Fähigkeiten von Agenten eingeteilt und abgegrenzt werden können, da sie so vielfältig konfigurier- und einsetzbar sind. Unklar ist auch, wie eine effektive menschliche Aufsicht einer Agentenregulierung funktioniert und ab wann ein Agent als nicht mehr kontrollierbar gilt.

Fazit: Impulse für ein immer wichtigeres KI-Rechtsthema

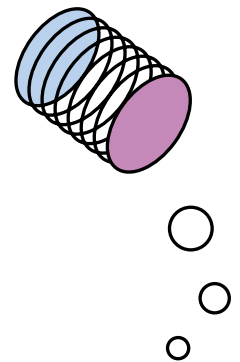
Wie sich an den Vorträgen und der anschließenden Fragerunde zeigt, ist das Thema „Schäden durch KI-Anwendungen, KI-Outputs und agentische KI-Systeme“ schon jetzt bedeutend – und wird an Bedeutung zunehmen. Dafür spricht, dass die KI-Wirtschaft die technischen Entwicklungen mit hohem Tempo vorantreibt, neue Systeme und Dienste oft übereilt in die Märkte bringt und dabei in Kauf zu nehmen scheint, dass mangelhafte Produkte nicht nur fehlerhafte Ergebnisse erzeugen, sondern auch relevante Schäden verursachen können.

Für diese Erkenntnisse exemplarisch mag der Hinweis von Martin Ebers auf bestehende Haftungslücken sein, insbesondere auch bezüglich der Verletzung von Persönlichkeitsrechten mittels KI-Einsatz. Dies scheint sich im März 2026 zu bestätigen, als eine bundesweit geführte Debatte über digitale sexualisierte Gewalt mittels KI-generierter Deepfakes aufflammte. Die darauf reagierende Gesetzesinitiative der Bundesregierung zur Anpassung des Strafrechts hat letztlich mittelbar auch mit der Frage nach Verantwortlichkeit und Haftung seitens der Hersteller, Anbieter und Betreiber von KI-Systemen zu tun. Denn allein sie können ihren Produkten funktionale Grenzen setzen – wenn sie es denn wollen oder eben auch müssen.

Auf eine Möglichkeit, wie man die Unternehmen dazu bringen kann, weist im Lunch-Talk Julia Smakman hin. Sie sagt, dass Schäden durch fehlerhafte KI-Agenten oder mangelhafte KI-Systeme wohl weiterhin passieren werden. Doch strenge Haftungsregularien, die deutliche Schadensersatzzahlungen vorsehen, sollten für Unternehmen Anreiz sein, ihre Produkte so zu verbessern, dass sie eben gar nicht erst in die Haftung geraten.

Lunch Talk

Mehr zum Ansatz von Julia Smakman lesen Sie im Interview mit ihr in Kapitel 3.



Unser Lunch-Talk räumte dem Thema Haftung viel Zeit für Reflexion und Vertiefung ein. Lagen wir mit der Themenwahl und dem Format richtig? Lesen Sie dazu unsere Einschätzungen in Kapitel 4.

3

Unser Working Paper – mit zehn Thesen zu KI und Urheberrecht, Datenschutz und Kennzeichnung

Working Paper

Um die relevanten Rahmenbedingungen für die Entwicklung und den Einsatz von generativen KI-Systemen sowie für deren Output zu identifizieren und einzuordnen, starten wir Ende 2024 unsere umfangreichen Quellenrecherchen, unter anderem in einschlägigen Fachpublikationen aus verschiedenen Disziplinen. Die zusammengetragenen Informationen, Zusammenhänge und Einschätzungen gilt es sodann auszuwerten und einzuordnen. Mit Blick auf deren Relevanz und auf die realen Verhältnisse in den von uns bestimmten, beispielhaften vier Arbeitsbereichen, führen wir leitfadengestützte Expert*innen-Interviews durch.

Lesen Sie mehr zu Quellenrecherchen und Expert*innen-Interviews weiter unten in diesem Kapitel.

Die Ergebnisse der Recherchen und der Interviews führen wir zusammen und werten sie inhaltlich aus, um wesentliche Schlüsse und wichtige Erkenntnisse daraus zu ziehen. Unser im Oktober 2025 erscheinendes Working Paper bündelt diese Überlegungen, und legt dabei einen Schwerpunkt auf niedrigschwellige Vermittlung. Es lässt sich kostenlos herunterladen (unter der offenen Creative Commons-Lizenz CC BY 4.0 veröffentlicht).

Aus dem Working Paper:

Hinweis: Im Folgenden geben wir Auszüge aus dem Working Paper wieder, die wir 1:1 daraus übernommen haben. Dies soll dazu dienen, die Geschichte des Projekts und den Erkenntnisprozess nachzuzeichnen. Aus diesem Grund verzichteten wir darauf, Inhalte des Working Papers nachträglich zu bearbeiten, auch, wenn mitunter aufgrund gesetzgeberischer oder juristischer Entwicklungen Aktualisierungen möglich wären.

Die Publikation erhebt nicht den Anspruch, alle relevanten rechtlichen Aspekte in gleicher Weise und angemessenem Umfang darstellen und wiedergeben zu können. Sie reflektiert eine Reihe von rechtlichen und gesellschaftlichen Rahmenbedingungen, die zum Zeitpunkt der Betrachtung viel diskutiert wurden. Auf dieser Grundlage will das Working Paper zudem bestimmte Klarstellungsbedarfe benennen. Es versteht sich nicht als juristische Fachpublikation, sondern als Informationsressource, die allgemein verständlich formuliert ist und einen Überblick über den Stand der Problematik und Diskussion geben will. Damit richten wir uns insbesondere an Policy Maker auf europäischer und nationaler Ebene, an netz- und digitalpolitische Akteur*innen aus Politik und Zivilgesellschaft sowie an Angehörige und Beobachter*innen praxisorientierter Wissenschaft und Wirtschaft.

[...]

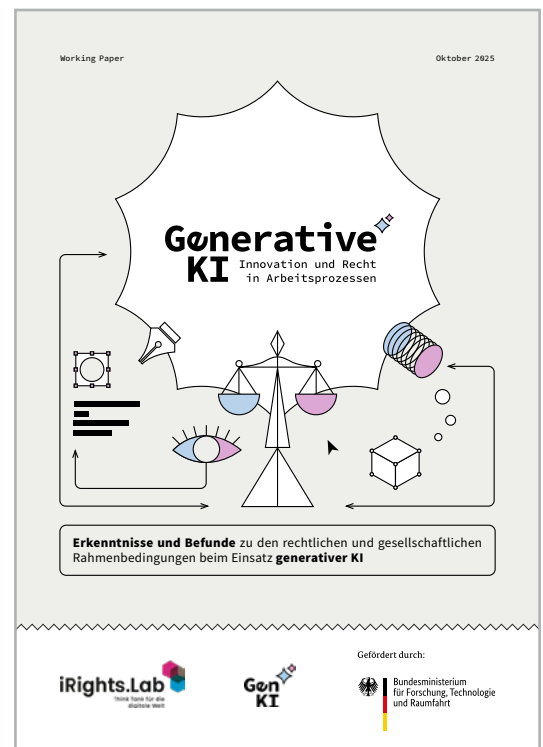
Betrachtet man den Wirtschaftsbereich „KI-Entwicklung und -Modelle“, fällt auf, dass dort einige wenige Unternehmen nicht nur neue Märkte erschaffen haben, sondern sie auch weitgehend beherrschen: OpenAI, Meta, Google, Microsoft, Anthropic, Perplexity, DeepSeek und ein paar weitere. Diese Hersteller und Anbieter von großen Sprachmodellen (Large Language Models, LLM) und GenKI-Systemen sind international agierende Unternehmen aus den USA und China, die man mit Fug und Recht als „Big Player“ kategorisieren kann. Ihre Modelle werden millionenfach genutzt für KI-basierte Systeme, Agenten und Dienste, ihre Produkte und Werkzeuge werden in nahezu allen Branchen und Arbeitsbereichen eingesetzt, und dort jeweils für zahlreiche Tätigkeiten und Aufgaben.

Für manche Arbeitsbereiche könnte die Nutzung KI-gestützter Systeme – wie sie durch Beschäftigte, Mitwirkende, Verbraucher*innen und Kund*innen bereits stattfindet – zu einer wirtschaftlichen Bedrohung werden. Beispielsweise lassen sich mittels KI-Systemen Fotos, Bewegtbilder, Musik oder Sprachwerke erstellen und bearbeiten – was bei Kulturschaffenden bereits zu Auftragsrückgängen führt. KI-gestützte Chatbots und Suchmaschinen können außerdem als primäre Nachrichtenquelle genutzt werden, indem sie Zusammenfassungen, Meldungen oder Berichte zu nachrichtlichen Themen anbieten. Auf diese Weise könnten die KI-Anbieter den journalistischen Medien sukzessive Publikum und Einnahmen entziehen. Das heißt, die KI-Anbieter begeben sich indirekt und direkt in den Wettbewerb mit Medienhäusern, Kulturproduzent*innen und Kreativarbeiter*innen. Sie ziehen Aufmerksamkeit und Mediennutzungszeit des Publikums auf sich. Zudem vereinnahmen die kommerziell aufgestellten KI-Anbieter via Abonnement-Gebühren (für Pro-Accounts) auch Teile der für Medien aufgewendeten Budgets – etwa von Verbraucher*innen, Unternehmen oder Verwaltungen.

Andererseits profitieren die Akteur*innen in den genannten Arbeitsbereichen aber auch selbst davon, die KI-gestützten Werkzeuge und Dienste gezielt und systematisch einzusetzen. Etwa, weil sie sich davon versprechen, einzelne Arbeitsschritte zu optimieren und selbst produktiver und wirtschaftlicher arbeiten zu können. Somit erscheint es insbesondere im Kultur- und Kreativitätsbereich, in Medien und Bildung, Wissenschaft und Forschung eine Ambivalenz oder auch Widersprüchlichkeit zu geben: Denn der Einsatz der KI-Tools verspricht einerseits willkommene Effizienzsteigerungen, führt aber andererseits zu einer teils massiven und bedrohlichen Disruption der eigenen angestammten Geschäftsfelder.

Das vergleichsweise rasche Eindringen von KI-Systemen und -Diensten in den Arbeitsalltag sowie in das Mediennutzungs-Verhalten sehr vieler Menschen wirft die Frage auf, welche Regeln, welche rechtlichen, soziotechnischen und gesellschaftlichen Rahmenbedingungen eigentlich für die Anbieter, deren Kund*innen sowie Nutzer*innen und Anwender*innen gelten – oder gelten sollten: Was soll für die Hersteller großer Sprachmodelle verbindlich sein, wenn sie diese und ihre KI-Systeme entwickeln und vermarkten? Ebenso wichtig ist, auf jene gesetzlichen, rechtlichen und gesellschaftlichen Rahmenbedingungen zu blicken, unter denen Anbieter und Betreiber von KI-Systemen auf den neu entstehenden Märkten agieren. Sowie auf solche, nach denen man sich als Nutzer*in und Anwender*in in Kultur, Medien, Bildung und Verwaltung richten muss oder orientiert, um KI-Systeme bewerten, auswählen, konfigurieren und effizient nutzen zu können.

Bei unseren Untersuchungen stellte sich heraus, dass in der ersten Phase des KI-Lebenszyklus vor allem Vorschriften aus dem Urheberrecht und Datenschutzrecht eine maßgebliche Rolle einnehmen. Beide Rechtsbereiche fußen auf der Überlegung, dass bestimmte Informationen und Daten nicht ohne gesetzliche oder privatrechtliche Grundlage verarbeitet und genutzt werden dürfen. Demgegenüber steht der Bedarf der Anbieter, KI-Modelle und -Systeme mithilfe möglichst vieler Informationen und Daten zu entwickeln. Die Debatte darüber, ob die KI-Entwicklungen teilweise nicht rechtskonform verlaufen sind, ist bei Weitem noch



Mehr zu unserer
Forschungsmatrix
lesen Sie weiter
unten in diesem
Kapitel.

3

Working
Paper

nicht abgeschlossen und dürfte – angesichts zahlreicher laufender Gerichtsverfahren – voraussichtlich noch Jahre andauern. Zugleich artikulieren viele Akteur*innen und Nutzer*innen ihre Forderungen nach rechtlicher Klarheit seitens des Gesetzgebers und mehr Transparenz über die KI-Entwicklung seitens der Hersteller.

Bei der Auswahl und dem Einsatz generativer KI-Systeme – der zweiten Phase des KI-Lebenszyklus – sehen sich alle Branchen vor der Herausforderung, urheberrechts- und datenschutzkonform vorzugehen; Wie kann beispielsweise verhindert werden, dass Mitarbeitende beim Einsatz und der Verwendung von KI gegen Gesetze verstoßen?

Eng damit verbunden sind die in allen Arbeitsbereichen deutlich gewordenen Qualifizierungsbedarfe – sowohl der Leitungsebenen als auch der Mitarbeiter*innen. Dazu kommen die Fragen nach Prüf- und Kontrollobliegenheiten von KI-Betreibern. Bei all diesen Aspekten geht es den Beteiligten darum, möglichst umfassend und ganzheitlich rechtskonform zu handeln. Zugleich wollen sie Maßnahmen ergreifen, um sich selbst zu regulieren und damit Verantwortungsbewusstsein zu zeigen, Stichwort: Compliance.

Für die dritte Phase des KI-Lebenszyklus, der Erstellung und Verwendung von KI-Output, sehen viele in den untersuchten Bereichen zum einen Herausforderungen und Klärungsbedarfe bezüglich der rechtlichen Einordnung und der rechtskonformen Kennzeichnung. Zum anderen fragen sich viele, ob und wie sie womöglich in Haftung geraten, etwa, wenn bestimmter KI-Output Schäden anrichtet, der jedoch auf Fehler im KI-Modell zurückzuführen sein könnte.

Übergreifend über die drei Phasen lässt sich feststellen, dass die komplexen Zusammenhänge zwischen Modell-Entwicklung und KI-Einsatz die beteiligten Akteur*innen häufig vor Herausforderungen oder gar Zerreißproben stellen: KI-Systeme basieren mehrheitlich auf den KI-Modellen Dritter, deren Entstehungsdetails den Systemanbietern und -betreibern in aller Regel nicht bekannt sind und auf die sie keinerlei Einfluss haben. Das muss die Verantwortungsverteilung berücksichtigen, um „chilling effects“ zu vermeiden: Akteur*innen würden sonst für Compliance-Anforderungen verantwortlich gemacht, die sie weder einschätzen noch kontrollieren können. Wenn es beispielsweise zur Nutzung von Output kommt, der zu nah an den Trainingsdaten ist und daher in Urheberrechte eingreifen könnte. Oder wenn ein Output unrechtmäßig erhobene oder falsche personenbezogene Daten enthält. Steuerungsmöglichkeiten, um dies zu verhindern, haben nur die Entwickler von KI-Modellen und gegebenenfalls die Betreiber von KI-Systemen, nicht aber die Anwender*innen. Transparenzpflichten oder -gebote – von vielen gefordert und in der KI-Verordnung in gewisser Weise verankert – können die Verantwortungsallokation durchaus beeinflussen. In konzentrierten oder monopolaritägen Märkten funktioniert das aber in der Regel nicht.

Unsere ersten Erkenntnisse in zehn Thesen, entlang des KI-Lebenszyklus

Erste Phase des KI-Lebenszyklus: Entwicklung, Training und Finetuning

(1) Zwiespältig: Viele sehen das Training von KI als urheberrechtlich zweifelhaft an, zugleich profitieren sie von KI-Systemen

Große KI-Hersteller und -Anbieter sehen sich bei der Nutzung urheberrechtlich geschützter Werke und Inhalte für die Trainings ihrer Modelle im Recht. Sie berufen sich auf die Fair-use-Regelung (in den USA) und auf die gesetzliche Ausnahme für das Text- und Data-Mining (TDM, Urheberrecht der EU).

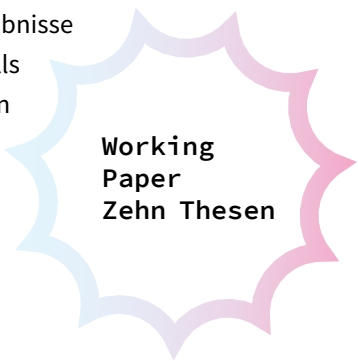
Demgegenüber argumentieren unter anderem Kreative, Urheber*innen und Verwerter*innen aus Kultur, im Journalismus und im Bildungsbereich, dass gesetzliche Erlaubnisse für das KI-Training nicht gelten. KI-Unternehmen müssten Lizenzen einholen, jedenfalls aber für ihre massenhaften Nutzungshandlungen Vergütungen bezahlen. Des Weiteren seien Nutzungsvorbehalte (Opt-outs) missachtet worden. Zahlreiche Klagen sind in den USA, China, der EU und in Deutschland anhängig. Bisherige Gerichtsentscheidungen gehen auf Einzelaspekte ein, eine endgültige Klärung durch die obersten Gerichte liegt jedoch noch in weiter Ferne.

Die Abwehrhaltung gegen KI-Technologie ist in der Kultur- und Kreativwirtschaft ebenso weit verbreitet wie ambivalent: Denn angesichts des vielfältigen Einsatzes KI-gestützter Werkzeuge für die unterschiedlichsten Tätigkeiten sind die kreativen Branchen und Akteur*innen auch Nutznießende der neuen Technologien. Die Tools eröffnen ganz neue Möglichkeiten der Inhaltsproduktion, steigern Effizienz, Produktivität und damit zu einem gewissen Grad auch wiederum die Wirtschaftlichkeit von Kultur und Medien. Dementsprechend profitiert auch die Medien- und Unterhaltungsbranche durchaus davon, dass die Werke Dritter – und womöglich auch ihre eigenen – herangezogen werden, um KI-Systeme zu trainieren beziehungsweise zu optimieren.

(2) Problematisch: Betreiber brauchen rechtssichere KI-Systeme, damit sie ihrer Verantwortung gerecht werden können

Wer generative KI-Systeme und KI-gestützte Werkzeuge, insbesondere für nicht-private Zwecke, nutzen will, muss zum einen wissen, welche Risiken damit verbunden sind. Zum anderen sind Informationen und Kenntnisse erforderlich, ob und wie man mit den Tools rechtskonform agieren kann. Mehr noch: Wer KI-Modelle so konfiguriert und nutzbar beziehungsweise zugänglich macht, dass er*sie selbst als KI-Anbieter agiert, kann verpflichtet sein – je nach KI-System – die zugrundeliegenden Datenverarbeitungen, die vorgenommenen Entwicklungsschritte und die vorgesehenen Einsatzzwecke nachzuweisen. Um diesen Pflichten nachzukommen und um auf dieser Grundlage Risiken und Rechtskonformität einschätzen zu können, braucht es Einblicke oder Offenlegungen seitens der Hersteller in deren KI-Entwicklung und zu den KI-Produkteigenschaften. Beispielsweise darüber, mit welchen Daten und Parametern sie ihre Modelle trainieren und optimieren – oder auch dazu, nach welchen Regeln oder Kriterien ihre KI-Systeme beim Generieren von Output vorgehen.

Hinweis: Im Folgenden geben wir die Erkenntnisse unserer Recherchen und Forschungen so wieder, wie wir sie bereits im Working Paper dokumentiert haben. Daraus leitet sich auch die Unterteilung der zehn Thesen gemäß des von uns genutzten Phasenmodells vom KI-Lebenszyklus ab.



Working
Paper
Zehn Thesen

3

Die Modell-Anbieter verweisen jedoch auf bestehende Geschäftsgeheimnisse sowie eigene wirtschaftliche Interessen. Außerdem tendieren sie dazu, sich den Forderungen und Pflichten nach Transparenz weitgehend zu entziehen. Auch sind diesbezügliche Vorgaben der KI-Verordnung (KI-VO) eher unspezifisch. Die ergänzend zur KI-VO von der EU-Kommission vorgelegten Praxisleitfäden („Codes of Practice“) sowie Leitlinien („Guidelines“) enthalten zwar Präzisierungen und Handlungsanleitungen zur Transparenz. Sie sollen die KI-Hersteller und -Anbieter dazu anhalten und motivieren, mehr Transparenz über ihre Entwicklungen und Systeme zu zeigen. Doch die EU setzt darauf, dass sich die Hersteller den Leitfäden und Guidelines freiwillig unterordnen und sich diesbezüglich quasi selbst prüfen und bewerten. Darunter fällt auch die Prüfung der inhaltlichen Richtigkeit der für die KI-Entwicklung erforderlichen Trainingsdaten und -verfahren. Das heißt, der Ansatz der Praxisleitfäden und Guidelines muss sich erst noch in der Praxis beweisen: nämlich, ob diese tatsächlich eine Wirkungsmacht entfalten können im Sinne von mehr Transparenz und mehr Rechtssicherheit für KI-Betreiber und -Nutzer*innen.

(3) Uneindeutig: Für die Klassifizierung als KI- Anbieter oder -Betreiber fehlt es an Orientierung und Unterstützung

Zu einem der Grundsätze der KI-Verordnung gehört, die Akteur*innen nach deren Tätigkeitsfeld und Umgang mit KI-Systemen in unterschiedliche Kategorien einzuteilen beziehungsweise ihnen Rollen zuzuweisen. Für die Unterscheidung der KI-Modelle und -Systeme selbst enthält die KI-VO recht umfangreiche und konkrete Risikoklassifizierungen. Dennoch bleiben hier viele Fragen offen.

Das schafft Rechtsunsicherheit, denn es liegt in der eigenen Verantwortung der KI-Entwickler und -Nutzer*innen, ihre Rolle und die von den Produkten ausgehenden Risiken einzuschätzen.

Unsicherheiten bestehen vor allem in der öffentlichen Verwaltung und im Bildungsbereich: Welche Anwendung ist nach der Risikoklassifizierung der KI-Verordnung wie einzuordnen? Welche Voraussetzungen muss ich erfüllen, um als Betreiber und nicht als Anbieter eines KI-Systems zu gelten? Welche Risiken bestehen im Falle einer fehlerhaften Selbsteinschätzung?

Zweite Phase des KI-Lebenszyklus: Auswahl, Einführung und Einsatz generativer KI

(4) Neuland: Bei der Auswahl rechtskonformer KI-Systeme fehlt es an Orientierung – und an strategischem Vorgehen

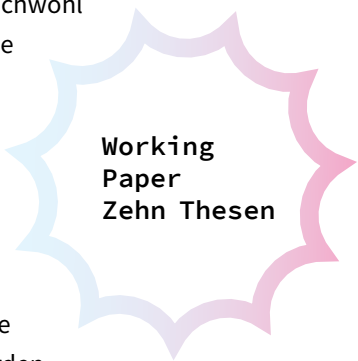
KI-Systeme rechtskonform einzusetzen, ist häufig herausfordernd. Die Fülle der Regularien und Vorgaben ist groß, die rechtlichen Unklarheiten sind, schon aufgrund der Neuheit der Materie, mannigfaltig. Dies beeinträchtigt den Einsatz in allen beleuchteten Arbeitsbereichen, besonders aber die öffentliche Verwaltung und den Bildungsbereich, die besonders auf Rechtsklarheit angewiesen sind, um ihre strengen Konformitätsanforderungen erfüllen zu können. Hier fehlt es den Akteur*innen an Orientierungshilfen, etwa durch erkennbare Qualitätsmerkmale, Best Practices und Zertifizierungen. Zudem erwarten sie mehr Rechtssi-

cherheit darüber, welche Verantwortung KI-Betreiber für etwaige Datenschutzverstöße beim Modelltraining tragen.

Die Recherchen in den Bereichen Journalismus und öffentliche Verwaltung zeigen, dass verantwortliche Stellen selten strategisch vorgehen und bei ihrer Auswahl oft ohne hinreichende Fachexpertise entscheiden müssen. Expert*innen aus dem Vergabe- und Beschaffungsrecht betonen die Haftungsrisiken und drohenden Qualitätsdefizite, die mit einer unzureichenden vertraglichen Vorbereitung der KI-Beschaffung einhergehen können.

(5) Bemüht: KI-Qualifizierungen finden im Arbeitsalltag statt, die Einführung von KI-Systemen erfolgt selten partizipativ

In allen von uns untersuchten Arbeitsbereichen zeigt sich ein gesteigertes Bedürfnis, sich selbst beziehungsweise die Mitarbeiter*innen und Beteiligten für den Umgang mit KI-Systemen und -Anwendungen sowie mit den generierten Ergebnissen (KI-Output) zu sensibilisieren und zu qualifizieren. Dies deckt sich mit Vorgaben aus der KI-Verordnung, wobei auch hier vielen unklar ist, ob und wann sie diese Qualifizierungsverpflichtungen betreffen. Gleichwohl finden Schulungen und Fortbildungen im Arbeitsalltag statt, damit sich möglichst viele Mitarbeiter*innen mit den Herausforderungen und Möglichkeiten der KI-Werkzeuge auseinandersetzen können. Zugleich stellt sich heraus, dass die Einführung der KI-gestützten Technologien und Hilfsmittel nur selten partizipativ erfolgt. Es existieren Unklarheiten beziehungsweise unterschiedliche Auffassungen darüber, ob hierbei die Beteiligung und Mitwirkung von Betriebsräten verpflichtend ist. Darüber hinaus sehen viele, wie bedeutend eine Einbeziehung von Mitarbeiter*innen ist – schon bei der Auswahl und bei der Konfiguration der KI-Systeme –, insbesondere dann, wenn die KI-Werkzeuge mittel- und langfristig in Produktions- und Arbeitsabläufe integriert werden sollen.



**Working
Paper
Zehn Thesen**

(6) Sensibilisierend: Der Umgang mit geschützten Werken für den KI-Input betrifft Urheberrechte und Datenschutz

Als „Prompt Injection“ oder „In-Context Learning“ werden weitverbreitete Methoden bei der Nutzung von GenKI bezeichnet. Hierbei laden Anwender*innen einzelne Materialien beim Prompting in das System, um etwa ein Dokument zu analysieren oder zusammenzufassen oder als Orientierungsmaßstab für den gewünschten Output („Erstelle mir ein Bild in dem Stil der Bilder, die ich hochgeladen habe.“).

Da solche Materialien häufig urheberrechtlich geschützt sind und/oder personenbezogene Daten enthalten, stellen sich allerhand Fragen nach der Zulässigkeit dieser ebenso nützlichen wie beliebten Methode. Derzeit sind hier noch viele Fragen offen, beispielsweise, ob die TDM-Schranken des Urheberrechts auch solche Einzelkopien gestatten oder ob sich aus dem Datenschutzrecht eine Rechtsgrundlage für die jeweiligen Datenverarbeitung ergibt und welche das ist.

Dass solche Fragen bei Aufkommen einer neuen Rechtsmaterie nicht unmittelbar geklärt sind, liegt in der Natur der Sache. Hieraus ergibt sich jedoch Unsicherheit, die gerade bei risikoaversen Akteur*innen dazu führt, neue Technologien und Praktiken abzulehnen und zu vermeiden. Arbeitgeber sind daher gefragt, Orientierung durch interne Richtlinien, Aufklärungsmaterialien und Schulungen zu schaffen. Auch Politik und der Gesetzgeber sind gehalten – insbesondere im Datenschutzrecht –, weitergehende Klarheit zu schaffen, indem

3

Regelungslücken zügig geschlossen oder einheitliche Leitlinien zu den wichtigen Fragen verabschiedet werden.

(7) Pragmatisch: Mit KI-VO-gerechten Leitlinien schaffen sich Betreiber einen eigenen Handlungsrahmen

Working
Paper
Zehn Thesen

Interessanterweise werden Leitfäden, die nicht nur Rechtsfragen betreffen, branchenübergreifend als wichtig erachtet: In Ergänzung zu rechtlichen Vorgaben sollen beispielsweise verbindliche Selbstverpflichtungen oder ethische Handlungsempfehlungen bei der Auswahl, Einführung und Verwendung generativer KI-Systeme helfen. Sie finden sich bei großen Unternehmen, werden von Branchenverbänden oder Interessenvertretungen vorgelegt, entstehen in gemeinsamen Abstimmungsprozessen in Organisationen oder als Vorschrift in Verwaltungen.

Die Umfänge, die Inhalte und der jeweilige Entwicklungsstand dieser Leitfäden und Handlungsempfehlungen unterscheiden sich stark: Sie reichen von den etablierten Grundsätzen journalistischer Sorgfalt und selbst auferlegten Qualitätskodizes im Journalismus über verschiedene Grundsätze und Praxisleitfäden in der öffentlichen Verwaltung bis zu pragmatischen Leitlinien im Bildungsbereich. In der Kultur- und Kreativwirtschaft treffen mitunter abweichende Vorstellungen zu den Herangehensweisen aufeinander. Freiwillige Selbstverpflichtungen, Leitfäden und Compliance-Vorgaben können durchaus eine Wirkmacht entfalten und sich als ebenso pragmatische wie maßgebliche Rahmenbedingungen etablieren. Beispielsweise, wenn sie von etablierten beziehungsweise einflussreichen Verbänden oder Branchen-Allianzen organisiert sind

Dritte Phase des KI-Lebenszyklus: Umgang mit dem Output generativer KI

(8) Unschärf: Für eine rechtssichere Kennzeichnung von KI-Output braucht es Konkretisierung und Klarheit

Für Betreiber und Nutzer*innen stellen sich die Pflichten und Methoden zur Kennzeichnung als Herausforderung dar, zumal es noch an konkreten Standards oder Normen fehlt. Auch hier sind Vorgaben oder Vorlagen für rechtskonformes Handeln gefordert: In welchen Fällen muss KI-Output als solcher gekennzeichnet sein – und wie muss diese Kennzeichnung genau aussehen? Wer ahndet unzureichende oder fehlende Kennzeichnungen? Sind technische Lösungen für eine rechtskonforme Kennzeichnung denkbar und gewollt? Braucht es unabhängige Zertifizierungsstellen beziehungsweise werden diese eingerichtet?

Zugleich hat für viele der verantwortungsbewusste und transparente Umgang mit KI-Output auch eine Bedeutung für die Außenwirkung ihres Unternehmens, ihrer Organisation, ihrer Verwaltung oder ihrer öffentlichen Einrichtung. Mit internen Regeln für Prüfschleifen und Kennzeichnungen, für Metadaten und Dokumentation zu KI-Output können und wollen viele Compliance belegen und vertrauensbildende sowie reputationsfördernde Prozesse unterstützen.

Außerdem stellt sich die Frage, ob angesichts der rasch anschwellenden Flut von Inhalten, die mit mehr oder weniger KI-Unterstützung entstanden sind, die Kennzeichnung überhaupt noch zielführend ist: Denn wenn beim überwiegenden Teil von Inhalten entsprechende

KI-bezogene Angaben zu finden sind, wird dann noch differenziert? Oder steigert das nicht das Misstrauen, weil Inhalte pauschal als „nicht ganz echt“ betrachtet werden? Aus diesen Überlegungen heraus entstehen Ansätze, Technologien und das Bemühen um Standards, mit denen die Authentizität und „Echtheit“ von Werken gekennzeichnet wird beziehungsweise nachzuweisen ist. Obwohl hierfür primär im (foto-)journalistischen und medialen Umfeld besonderes Interesse besteht, hat dieser Aspekt schon jetzt den gesellschaftlichen Diskurs erreicht und verhandelt ihn als bedeutende Rahmenbedingung.

(9) Ambivalent: Urheberrechtlich ungeschützter KI-Output wirft zugleich urheberrechtliche Regulierungsfragen auf

Als eine zentrale Rahmenbedingung für den Umgang mit KI-Output gilt – insbesondere in den schöpferischen Arbeitskontexten von Journalismus, Kultur- und Kreativwirtschaft – dessen Schutzwürdigkeit. Denn häufig entstehen Inhalte auch hybrid, sind also das Ergebnis einer Mischung von maschineller und menschlicher Bearbeitung. Da reiner KI-Output schutzlos ist, alle menschlichen Anteile hingegen schutzfähig sind, stellt sich unter anderem die Frage, wie sich die menschlichen, schutzfähigen Anteile verlässlich oder auch rechtssicher nachweisen lassen.

Neben der vor allem rechtspolitischen Debatte, ob das Urheberrecht auch für KI-generierte Werke gewährt werden sollte, stellt sich die Frage, ob nicht schon die geltenden Leistungsschutzrechte ausreichenden Schutz gewähren. Sie schützen beispielsweise die kreativen Leistungen ausübender Künstler*innen, aber auch den Aufwand und die Investitionen jener, die an der Produktion, Veredelung und Veröffentlichung eines urheberrechtlich geschützten Werks mitwirken. Diese Rechte sind teilweise unabhängig davon, ob Menschen an den Leistungen mitgewirkt haben oder ob sie rein computergeneriert sind. Zumindest denkbar ist, dass an KI-erzeugten Tonaufnahmen Leistungsschutzrechte des Tonträgerherstellers entstehen. Auch das Leistungsschutzrecht für Presseverleger könnte einen gewissen Schutz gegen die Nachnutzung synthetischer Pressebeiträge bieten – doch das ist noch unklar. Leistungsschutzrechte haben einen engeren Anwendungsbereich als das Urheberrecht: Letzteres gilt universell für alle menschengemachten persönlichen Kreationen. Leistungsschutzrechte gelten dagegen nur punktuell für bestimmte Arten von Produkten.

(10) Vielfältig: Weitere Themenfelder zu rechtlichen und gesellschaftlichen Rahmenbedingungen bei KI-Verwendung

Entsteht durch den Einsatz eines KI-Systems ein Schaden, ist die Haftung nach den allgemeinen Grundsätzen zu ermitteln. Ein belastbarer Überblick über relevante Haftungsrisiken fehlt vielen Akteur*innen im Bildungsbereich noch. Gründe hierfür sind einerseits der Regulierungsansatz über das Produktsicherheitsrecht und andererseits die noch jungen Reformen.

Der Einsatz von GenKI-Anwendungen im Bildungsbereich prägt unter anderem die Debatte um neue Rahmensetzungen für die Bearbeitung von Prüfungsaufgaben. Zwar sind KI-Werkzeuge an Schulen wie Hochschulen weder pauschal verboten noch pauschal erlaubt – zentral ist die Frage der Bewertbarkeit einer eigenständigen Leistung der geprüften Personen. Gleichwohl drängen viele darauf, diesbezügliche prüfungsrechtliche Fragen zu klären. Erste rechtliche Beurteilungen und Empfehlungen der Bundesländer liegen bereits vor. Sie gehen



Working
Paper
Zehn Thesen

3

in die Diskussion um den künftig sinnvollen Einsatz von generativen KI-Systemen in Prüfungskontexten ein, die in vollem Gange ist.

Bundes- und landesrechtlich werden die für die Entwicklung von generativen KI-Systemen geltenden Vorschriften der KI-VO vereinzelt durch Gesetze und – teils unveröffentlichte – Verwaltungsvorschriften weiter präzisiert und ausgeführt. Auch der Nutzung des Outputs generativer KI-Systeme sind dienst- und verwaltungsrechtliche Grenzen gesetzt. Akteur*innen aus Verwaltungspraxis und Wissenschaft weisen auf die Hebelwirkungen des Beschaffungs- und Vergaberechts hin. Im Zuge von Beschaffung und Vergabe könnte man – mit Blick auf die Vorschriften – rechtliche und ethische Anforderungen an das betreffende KI-System gegenüber Dienstleistern und Entwicklern einfordern. Beispielsweise, dass es Qualitätsanforderungen zur Produktsicherheit einhält oder dass es bei einer ethischen Folgenabschätzung gut abschneidet.

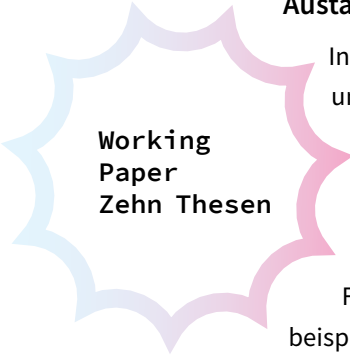
Weitere Rückschlüsse und Thesen

Mangelhafte Transparenz der KI-Hersteller

Die vorliegende Untersuchung verdeutlicht, dass für eine ganze Reihe rechtlicher, ethischer und gesellschaftlicher Rahmenbedingungen ein Aspekt besonders bedeutend ist: Der Einblick in die bei der KI-Entwicklung verwendeten Daten. Welche Datensätze für Training und Finetuning wie genutzt wurden, das betrifft sowohl den Datenschutz und das Urheberrecht als auch Antidiskriminierungsrechte, Produktsicherheit, Neutralität und Weiteres.

Austausch von Leitlinien zwischen den Branchen

Intensive diskursive und reflektierte Auseinandersetzungen innerhalb einzelner Branchen und Arbeitskontexte bringen immer wieder klare Hilfestellungen und Leitlinien zum sinnvollen Einsatz generativer KI-Systeme hervor. Besonders vielversprechend erscheinen kontinuierliche branchenübergreifende Austauschformate: Erfahrungswerte des Journalismus mit selbst auferlegten Pflichten und Qualitätskriterien können etwa in der öffentlichen Verwaltung oder der Kreativwirtschaft von Bedeutung sein. Pädagogische Fachdiskurse zur Entwicklung sinnvoller Leistungsabfragen im Bildungsbereich könnten beispielsweise von den Debatten der Kreativwirtschaft rund um ko-kreative Schaffensprozesse profitieren.



Working
Paper
Zehn Thesen

Unterschiedliche Auffassungen auch innerhalb der Branchen

Wie sich zeigt, gibt es in den untersuchten Arbeitsbereichen jeweils deutlich auseinandergehende Auffassungen zur KI-Nutzung. Im Journalismus und in Teilen der vor allem Musik, Video und Film produzierenden Kreativbranche kommen generative KI-Systeme umfassend zum Einsatz: Große Medienhäuser und Agenturen setzen hier sogar auf eigene Entwicklungen, Anpassungen und KI-Angebote. Auf der anderen Seite wird – besonders in der Kultur- und Kreativwirtschaft – auch grundlegend über die Sinnhaftigkeit des Einsatzes generativer KI-Systeme diskutiert. Ähnlich unterschiedliche Haltungen finden sich auch bei Akteur*innen im Bildungsbereich sowie in der öffentlichen Verwaltung. Sowohl die zögernden Skeptiker*innen als auch die den Einsatz generativer KI-Systeme begrüßenden Befürworter*innen treiben die Diskussion um die disruptiven Potenziale der Technologie an: Während die einen

den stärkeren Schutz menschlicher Kreativität und Schaffensprozesse fordern, sehen die anderen die Chancen der Effizienzsteigerung durch Technologieeinsatz, der dann (humane) Ressourcen für Qualitätsverbesserung freisetzen könnte.

Bestehende Regeln umsetzen statt auf Gesetzesreformen hoffen


Nicht zuletzt ist über alle untersuchten Anwendungsbereiche hinweg festzustellen, dass viele beim Blick auf zentrale Rahmenbedingungen vor allem realitätsorientierte Lösungen anstreben: Im Mittelpunkt steht weniger die Hoffnung auf gesetzgeberische Reformbemühungen als die Suche nach einer praktikablen und interessengerechten Umsetzung der bestehenden Rechtslage. Diese Lösungen sollen aber im Zusammenhang von Auswahl, Einführung und Einsatz generativer KI-Systeme zugleich möglichst mehr als unverbindliche Empfehlungen sein: Angestrebt werden Verbindlichkeit und Handlungssicherheit.

Wie bewerten wir selbst unsere Zwischenbilanz und das Working Paper? Lesen Sie dazu unsere Einschätzungen in Kapitel 4.



3

Recherchen und Hintergrundgespräche: Das Fundament für die gesamte Projektarbeit



Recherchen
und
Hintergrund-
gespräche

Hinweis: Im Folgenden geben wir hier Auszüge aus unserem Working Paper wieder, die wir 1:1 daraus übernommen haben. Dies soll dazu dienen, die Geschichte des Projekts und den Erkenntnisprozess nachzuzeichnen. Aus diesem Grund verzichtet wird darauf, Inhalte des Working Papers nachträglich zu bearbeiten – auch wenn aufgrund gesetzgeberischer oder juristischer Entwicklungen Aktualisierungen möglich wären.

Die unserem Working Paper zugrunde liegenden Informationen stützen sich auf eine umfangreiche Literatur- und Sachrecherche anhand diverser fachbezogener Quellen sowie einschlägiger Fachpublikationen aus verschiedenen Disziplinen:

- Gesetze, Verordnungen, Richtlinien (EU, D) sowie Normen und Standards, Leitlinien, Auslegungsgrundsätze und Praxisleitfäden
- Bücher, wissenschaftliche Aufsätze in Büchern oder in wissenschaftlichen Periodika
- Artikel, Beiträge in Magazinen, Zeitschriften, Zeitungen etc.
- Studien, Untersuchungen, Working Paper
- Stellungnahmen

Aufgrund der hohen Entwicklungsdynamik im Bereich KI – sowohl in technischer Hinsicht, als auch zu politischen, gesellschaftlichen, rechtlichen oder ethischen Aspekten – wird hierzu sehr viel publiziert. Für die Basisrecherchen werten wir eine Vielzahl von Publikationen aus, darunter zahlreiche Veröffentlichungen aus dem Zeitraum von Januar bis Juli 2025 sowie vereinzelte Materialien aus August und September 2025 aus. Die ausführliche Quellen- und Literaturliste finden Sie im Working Paper als Literaturverzeichnis, von Seite 79 bis Seite 109.

Um die recherchierten Informationen, Zusammenhänge und Einschätzungen einordnen und auf ihre praktische Relevanz für den jeweiligen Arbeitskontext überprüfen zu können, führen wir insgesamt 30 leitfadengestützte Interviews mit Expert*innen aus der Wissenschaft, Wirtschaft und dem Bildungssektor mit einer Dauer von jeweils bis zu 60 Minuten durch. Bei der Auswahl der interviewten Expert*innen achten wir darauf, dass sowohl alle untersuchten Arbeitsbereiche gleichmäßig repräsentiert waren als auch alle drei untersuchten Phasen des Lebenszyklus von KI.

Der zugrunde gelegte Interviewleitfaden fußt auf den zentralen Forschungsfragen des Forschungsvorhabens und wird in der Interviewsituation jeweils an die spezifische Expertise der Interviewpartner*innen angepasst. Alle Fragestellungen und Interviews setzen einen explorativen Schwerpunkt auf offene Erzählaufforderungen, um den Interviewpartner*innen die Möglichkeit zu geben, weitergehende Eindrücke und Ideen zu GenKI in ihrem jeweiligen Arbeitskontext zu formulieren. Die Antworten, Aussagen und Reflexionen der Interviewpartner*innen dienen wie die Literaturrecherche dazu, sich ein Gesamtbild über die schwer überschaubare Thematik und den derzeitigen Diskussionsstand zu verschaffen.

Forschungsmatrix: Mehrdimensionaler Blick auf KI-Rechtsrahmen und KI-Realität

Wichtiger methodischer Ausgangspunkt ist, dass wir bei unseren Bestandsaufnahmen und Ermittlungen zum einen drei Phasen im Lebenszyklus eines KI-Systems unterscheiden: Entwicklung, Einsatz und Nutzung des KI-Outputs. Zum anderen fokussieren wir uns auf vier Arbeitsbereiche, die wir als gut geeignet identifizieren, um an ihnen beispielhaft die rechtlichen Rahmenbedingungen beim Einsatz generativer KI zu untersuchen: Kreativitäts- und Kulturwirtschaft, Journalismus und Medien, Bildung und Verwaltung.

Diesen mehrdimensionalen Ansatz bezeichnen wir als Forschungsmatrix. Sie soll es uns ermöglichen, einerseits gleichgeartete rechtliche Fragen und Probleme herauszuarbeiten, die sich branchen- und bereichsübergreifend beim Einsatz generativer KI stellen. Andererseits dient die Forschungsmatrix auch dazu, die spezifischen Aspekte in einzelnen Arbeitsbereichen zu identifizieren, die es in einer bestimmten Lebenszyklusphase beim Umgang mit generativer KI gibt.

Die drei Phasen des Umgangs mit generativer KI

Um die rechtlichen und gesellschaftlichen Rahmenbedingungen für den Umgang mit GenKI verständlich und übersichtlich zu erörtern, identifiziert das Forschungsprojekt drei wesentliche Phasen des Umgangs mit generativer KI: Die erste Phase beschäftigt sich mit der Entwicklung der KI-Modelle durch große Hersteller und Anbieter. In der zweiten Phase geht es um Auswahl, Einführung und Einsatz generativer KI-Modelle und -Systeme, etwa durch Unternehmen, Einrichtungen oder Organisationen. Die dritte Phase bezieht sich auf den KI-Output, also die mittels KI-Modellen, KI-Systemen sowie KI-gestützten Werkzeugen generierten Werke und Ergebnisse.

Erste Phase: Entwicklung, Training und Finetuning

Die Entwicklung von KI-Modellen und -Systemen wird häufig vereinfacht als „Trainingsphase“ bezeichnet. Sie umfasst das Training mittels großer Datenbestände und Millionen Parameter sowie das Finetuning mit aufbereiteten Datenbeständen – oftmals auch das eigentliche Training durch Tausende menschliche Arbeitskräfte.

Die Entwicklung eines KI-Systems ist ein kleinteiliger Prozess mit zahlreichen Schritten. Sie lässt sich als Wertschöpfungskette verstehen, an der in der Regel unterschiedliche Akteur*innen beteiligt sind.

Bei klassischen, anwendungsfallorientierten Systemen steht am Anfang die Festlegung von Zweck und Anwendungskontext. Daran schließen sich Sammlung, Bereinigung und Aufbereitung der erforderlichen Daten sowie Modellierung und Training an. Für die Datenerhebung und -annotation hat sich ein global vernetzter Markt spezialisierter Dienstleister etabliert, während Modellierung und Training in der Regel bei den entwickelnden Institutionen verbleiben.

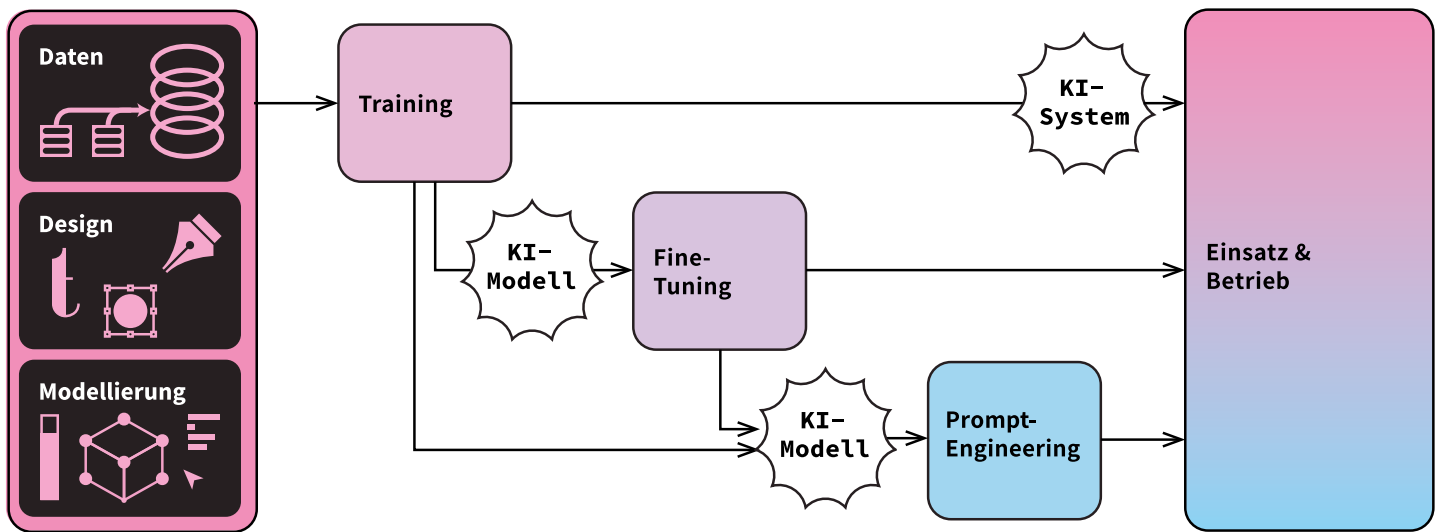


Abbildung 1: Wertschöpfungskette eines generativen KI-Systems (Quelle: iRights.Lab)

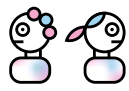
Recherchen und Hintergrundgespräche

Bei generativen KI-Systemen (Foundation Models, General Purpose Artificial Intelligence, kurz: GPAI) ist der Ablauf entkoppelt vom konkreten Zweck: Große Basismodelle werden zunächst auf sehr breiten Datensätzen trainiert. Die Spezifizierung für bestimmte Anwendungsfelder erfolgt erst nachgelagert – insbesondere durch Finetuning oder Prompt-Engineering. „Finetuning“ bezeichnet die Nachjustierung eines großen Basismodells mit zusätzlichen, bereichsbezogenen Datensätzen. Dadurch lassen sich Modelle an spezifische Aufgaben oder Fachbereiche anpassen, etwa an juristische Texte, medizinische Daten oder unternehmensinterne Wissensbestände. „Prompt-Engineering“ umfasst dagegen die gezielte Gestaltung von Eingaben, um den Output eines Modells in die gewünschten Bahnen zu lenken. Dazu zählen strukturierte Prompts, Vorlagen oder sogenannte Meta-Prompts, die von Nutzer*innen nicht verändert werden können. Auf diese Weise lassen sich etwa die Antworten auf bestimmte Themenbereiche fokussieren oder unerwünschte Ergebnisse ausschließen (zum Beispiel urheberrechtsverletzende Inhalte).

Anschließend folgen Testphasen und Qualitätssicherung, bevor das KI-System in Verkehr gebracht und am Markt angeboten wird. Auch in der Einsatzphase ist eine fortlaufende Evaluation notwendig.

Zweite Phase: Auswahl, Einführung und Einsatz generativer KI

Auswahl, Einführung und Einsatz generativer KI-Modelle und -Systeme können für Unternehmen, Einrichtungen oder Organisationen herausfordernd sein. Setzen sie KI-Systeme ein, sind sie als „Betreiber“ (Art. 3 Nr. 4 KI-Verordnung) anzusehen und unterliegen rechtlichen Anforderungen und Verantwortlichkeiten. Rechtskonforme KI-Systeme auszuwählen und einzusetzen – etwa, um darauf eigene Produkte oder Dienste aufzusetzen –, ist regulativ anspruchsvoll. Die Akteur*innen müssen in dieser Phase Kompetenzen aufbauen und Beteiligte, Mitwirkende und Betroffene gezielt qualifizieren. Ein weiterer Aspekt betrifft Qualitätsstandards und Qualitätssicherung, die sich insbesondere auf den Einsatz für interne Produk-



Außenkontakt

- Kund*innen
- Geschäftspartner*innen
- Bürger*innen
- Schüler*innen
- Leser*innen

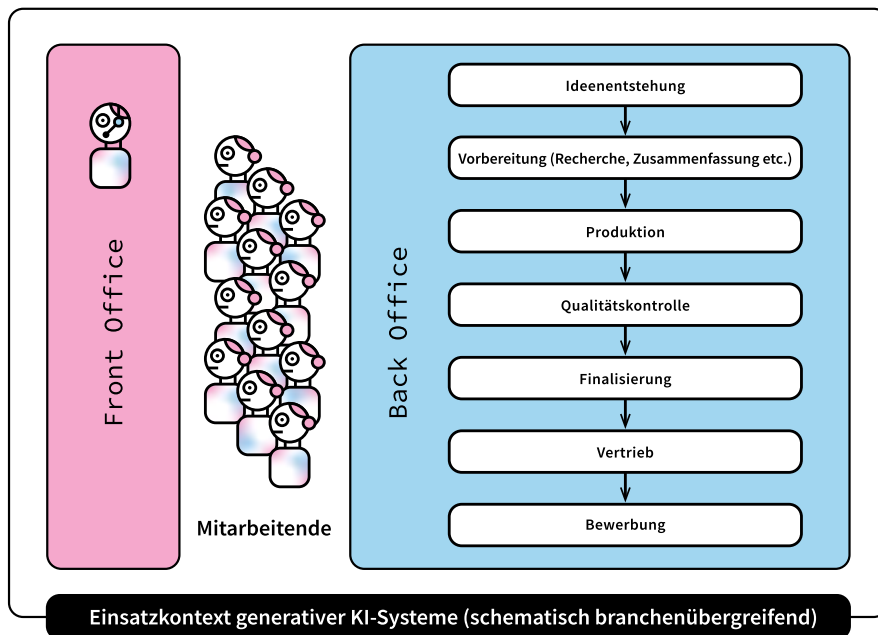


Abbildung 2: Einsatzkontexte generativer KI-Systeme (Quelle: iRights.Lab)

tionsschritte, Tätigkeiten und Prozesse beziehen. Je mehr sich die angebotenen KI-Systeme anhand bestimmter Leistungsmerkmale und -parameter oder transparent gemachter Arbeitsweisen bewerten und vergleichen lassen, desto mehr lassen sich Erkenntnisse gewinnen, ob sie den jeweiligen Qualitätsstandards der Betreiber oder Nutzer*innen genügen oder nicht.

Dritte Phase: Umgang mit dem Output generativer KI

Was KI-Systeme aufgrund bestimmter Eingabebefehle – dem „Input“ – generieren und dann als „Ausgabe“ präsentieren, wird übergreifend als „Output“ bezeichnet. Auf diesen KI-Output bezieht sich die dritte Phase, also die mittels KI-Systemen sowie KI-gestützten Werkzeugen generierten Werke und Ergebnisse: Texte, Bilder, Grafiken, Illustrationen – aber auch komplette audiovisuelle, multimediale Inhalte sowie Programmiercode und vieles mehr.

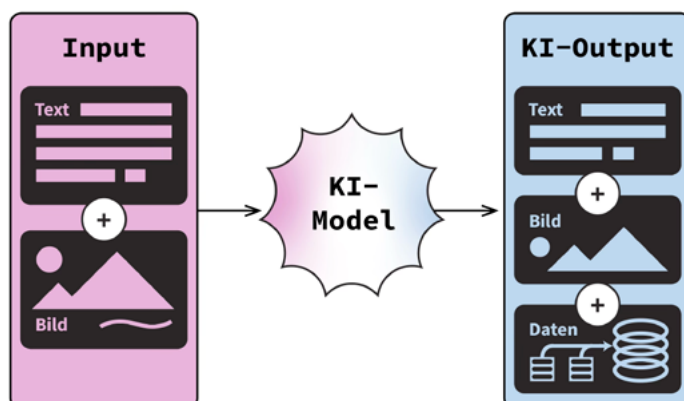


Abbildung 3: Entstehung von KI-Output (Quelle: iRights.Lab)

3

Betrachtungsperspektive Arbeitskontexte

Vier beispielhaft ausgewählte Arbeitsbereiche

Um sich der Beantwortung der Leitfragen in den genannten, umfassenden Perspektiven möglichst ganzheitlich und praxisrelevant zu nähern, werden in einem ersten Schritt vier Arbeitskontexte definiert: Journalismus, Bildung, Kultur- und Kreativwirtschaft und (öffentliche) Verwaltung (siehe hierzu die näheren Beschreibungen aller vier Arbeitsbereiche weiter unten in diesem Kapitel).

Für die Auswahl der untersuchten Arbeitskontexte wird herangezogen, in welchen Arbeitsfeldern generative KI-Systeme bereits in größerem Umfang genutzt, Potenziale und Risiken in der KI-Nutzung identifiziert und mögliche Veränderungen der Arbeitsprozesse diskutiert werden. Hinzu kommen die wirtschaftliche und gesellschaftliche Bedeutung der Tätigkeitsfelder sowie die Übertragbarkeit des untersuchten Arbeitskontexts auf vergleichbare Branchen und Arbeitsbereiche.

Die Fokussierung dieser vier beispielhaften Arbeitskontexte soll dazu beitragen, den Untersuchungsgegenstand „Generative KI-Systeme in Arbeitskontexten“ greifbar zu machen. Bereichsübergreifend relevante Problemlagen werden erkennbar, ohne branchenspezifische Problemstellungen zu übersehen.

Kultur- und Kreativwirtschaft

Schon heute kommen generative KI-Systeme in der Kultur- und Kreativwirtschaft umfassend zum Einsatz. Von der Erstellung von Blogartikeln und Werbekampagnen über die Zusammensetzung von Hintergrundmusik bis hin zur Skizzierung von Logos, Schriftzügen oder Bildmarken: Das Potenzial der neuen Technologien in der kreativen Produktion ist vielfältig. Neben Assistenzaufgaben in der Erstellung kreativer Werke kommen KI-Systeme auch für vollständig neuartige Kreationen und Kunst-Zusammenhänge zum Einsatz. Es scheint daher absehbar, dass in Zukunft nahezu alle Phasen des kreativen Prozesses – von der Ideenfindung über die Produktion und Kuratierung bis hin zur Vermarktung – von generativer KI beeinflusst sein werden.

Dabei lassen sich zwei zentrale Anwendungsziele unterscheiden: Zum einen die Automatisierung, Beschleunigung und Optimierung von Routine- und (Post-)Produktionsprozessen (zum Beispiel Farbkorrektur, Entfernung von Aufnahmeelementen, Musik-Mastering). Zum anderen der (mitunter noch experimentelle) Einsatz als kreatives Werkzeug innerhalb gestalterischer Schaffensprozesse. Nach gängigem Narrativ bedingen sich die Anwendungsziele: Durch die Effizienzsteigerungen im Rahmen von Routineaufgaben erhoffen sich Entscheider*innen und Akteur*innen mehr Raum für Kreativität und Originalität.

In der Tat bieten generative KI-Systeme insbesondere kleineren Unternehmen und Einzelpersonen die Chance, ehemals aufwendige Produktionen kostengünstig durchzuführen. Auf der anderen Seite deutet sich in Form von Auftragsrückgängen und Einkommensverlusten, Aufgabenveränderungen und dem Wegfall von Arbeitsplätzen auch das disruptive Potenzial generativer KI-Systeme an. Dabei werden die Einsatzmöglichkeiten von generativen KI-Systemen und die Nutzung von KI-Output im Allgemeinen noch längst nicht ausgeschöpft – zu

häufig bremsen bislang Implementierungskosten und Anpassungsaufwände die Veränderung bestehender schöpferischer Prozesse.

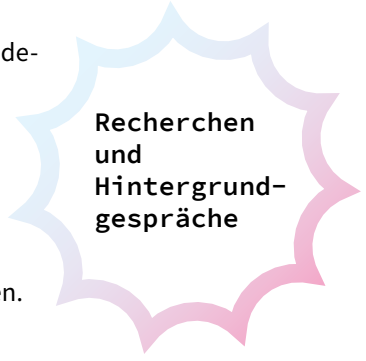
Auf der einen Seite übernehmen generative KI-Systeme vereinzelt bereits ganze Produktionsschritte im musikalischen, bildnerischen oder sonstigen Kreativeprozess – und drängen Kreativschaffende damit in die Rolle, den KI-Output zu beaufsichtigen und zu kontrollieren. Auf der anderen Seite experimentieren Kultur- und Kreativschaffende mit immer neuen kooperativen Funktions- und Einsatzmöglichkeiten. Das heißt: Die Kreativwirtschaft nimmt – trotz der berechtigten Kritik am Umgang der KI-Hersteller mit ihren Werken beim Training und Finetuning – zugleich längst auch die Rolle ein und an, selbst Betreiberin und Entwicklerin von KI zu sein. Sie nutzt die KI-Werkzeuge schöpferisch und produktiv und trägt zu ihrer Optimierung sowie letztlich zur Veränderung der eigenen Märkte durch den Einsatz von KI-Systemen bei.

Allen Akteur*innen ist klar, dass die Kultur- und Kreativwirtschaft derzeit eine beispiellose Disruption erlebt. Die kommerzielle Nutzung von KI-Systemen und KI-Output wirkt wie ein Katalysator für Veränderungsprozesse, die schon seit Längerem aus wirtschaftlichem Druck vorangetrieben werden, vor allem zur Einsparung von menschlicher Arbeit und Personalkosten. Diese Entwicklung polarisiert: Die einen nehmen die durch KI bedingten Veränderungen gegenüber traditionellen Arbeitsweisen offen an, entwickeln spezifische KI-Systeme mit und setzen sich künstlerisch-kreativ mit der Funktionsweise der Technologie auseinander. Die anderen begegnen KI mit Skepsis und Vorbehalten und lehnen Technologie und hierauf basierende Entwicklungen ab, weil sie sich angesichts des Vorgehens der KI-Industrie übergangen und ausgebeutet fühlen. Viele – vor allem Kreativschaffende – stellen sich die Grund-satzfrage: Was ist Kunst wert – und wer sollte für sie zahlen?

Bildung

Generative KI-Systeme gewinnen im Bildungsbereich zunehmend an Bedeutung, bieten sie doch die Möglichkeit, Lerninhalte individuell zu erstellen, Lernprozesse zu unterstützen und neue Formen der Interaktion zwischen Lehrenden und Lernenden zu ermöglichen. Bislang sind die effektive Entwicklung und spätere Integration der Systeme im Alltag jedoch noch häufig mit Herausforderungen verbunden, die über die rein technischen Aspekte hinausgehen. Das vergleichsweise breit aufgestellte Akteur*innenfeld reicht von Lehrenden und Lernenden über Entscheidungsträger*innen und unterschiedlich aufgestellten Bildungseinrichtungen bis zu Schulbuchverlagen und Ed-Tech-Unternehmen. Jede Gruppe bringt unterschiedliche Erwartungen, Interessen und rechtliche wie organisatorische Herausforderungen mit.

Trotz intensiver Diskussionen um ihre Risiken und Herausforderungen kommen generative KI-Systeme im Bildungswesen bereits in vielen unterschiedlichen Anwendungsfällen zum Einsatz. Lernende nutzen generative KI-Tools etwa, um Aufgaben zu bearbeiten, Texte zu erstellen oder Inhalte zu recherchieren. Dafür lassen sie sich von KI-Tools Textbausteine generieren, Inhalte sortieren und zusammenfassen, große Datenmengen strukturieren, zu eigenen Texten Feedback geben oder diese korrigieren. Lehrende nutzen KI-Tools, um Unterrichtsmaterialien vorzubereiten und zu erstellen, wie z. B. Arbeitsblätter, Präsentationen oder Prüfungsaufgaben. Dadurch wird die Unterrichtsvorbereitung effizienter und sie kann besser auf die individuellen Bedürfnisse der Schüler*innen abgestimmt werden. Zudem setzen Lehrende KI-Tools ein, um individualisiertes Feedback zu geben oder Prüfungen auszuwerten. Schulverwaltungen setzen generative KI-Tools vor allem zu organisatorischen Zwecken ein,



Recherchen
und
Hintergrund-
gespräche

3

beispielsweise, um Lehrpläne zu erstellen oder Ressourcen effektiver zu planen. Die für den (schulischen) Bildungsbereich relevanten Rahmenbedingungen für Auswahl, Einführung und Einsatz von generativen KI-Systemen wurden von der Kultusministerkonferenz (KMK) eingekreist – zumindest lässt sich die im Oktober 2024 von der KMK herausgegebene „Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen“ so einordnen.

Die Fragen zur rechtskonformen Entwicklung von generativen KI-Systemen beschäftigt bislang vor allem die Anbieter von Bildungs-KI-Systemen: namentlich etwa private Ed-Tech-Unternehmen wie fobizz und FelloFish oder durch die öffentliche Hand finanzierte Angebote wie das KI-Tool „KAI“ der Technischen Universität Dresden. Dass generative KI-Modelle von Grund auf autonom entwickelt und trainiert werden, kommt im Bildungssektor aufgrund der hohen Kosten für die notwendige technische Infrastruktur, des immensen Energieverbrauchs sowie des fehlenden Know-hows jedoch kaum vor. Stattdessen werden Lizenzen für vortrainierte KI-Modelle kommerzieller Anbieter oder frei verfügbare Open-Source-Modelle erworben, genutzt und für bildungsrelevante Einsatzzwecke insbesondere im Wege des Prompt-Engineerings angepasst.

Journalismus und Medien

Für die Betrachtung der Rahmenbedingungen im Arbeitsbereich Journalismus sind dessen besondere Merkmale zu berücksichtigen. So ist Journalismus unter anderem dadurch gekennzeichnet, dass in vielen Fällen eine Massenproduktion von Inhalten stattfindet: Täglich werden je nach Größe des Medienunternehmens Nachrichten, Berichte, Reportagen und Interviews in gewaltiger Zahl und unter hohem Zeit- und Wettbewerbsdruck erstellt. Ein solcher Produktionsalltag legt nahe, zahlreiche generative KI-Systeme zu testen und einzusetzen. Tatsächlich haben Algorithmen- und KI-gestützte Werkzeuge schon seit längerem Einzug in vielen Produktionsbereichen gehalten und haben sich in vielen Fällen bereits bewährt – von der Recherche über die Produktion bis hin zu Vertrieb und Management.

Recherchen und Hintergrund- gespräche

Neuere Werkzeuge, die auf GenKI basieren, kommen spätestens seit Ende 2022 in vielen Teilbereichen der journalistischen Produktion immer regelmäßiger und systematischer zum Einsatz. Entscheider*innen und Akteur*innen beabsichtigen, die journalistische Produktion mittels leistungsfähiger und verlässlicher KI-Werkzeuge zu rationalisieren. Ebenso streben Verlage danach, mittels KI-gestützter Analysen des Online-Leseverhaltens den Leser*innen personalisierte Inhalte auszuspielen. Hierfür nutzen journalistische Betriebe die KI-Systeme großer internationaler KI-Anbieter, um sie per Prompt-Engineering auf ihre eigenen, spezifischen Bedarfe zuzuschneiden.

Insbesondere große Medienunternehmen investieren zudem in die Optimierung KI-gestützter Prozesse: In Redaktionen entstehen Prompt-Sammlungen (Libraries) zur Vereinfachung des Prompt-Engineerings, das zum festen Bestandteil journalistischer Arbeit wird. Mit KI-Agenten beziehungsweise KI-Assistenten, RAGs (= Retrieval-Augmented Generation) und Metaprompts, Multi-Hop-Fragen und ähnliche Methoden werden halbautomatisierte Arbeitsschritte miteinander verschränkt und auf die spezifischen Datenquellen des Verlags zugeschnitten. Solche und weitere Werkzeuge und Funktionalitäten entwickeln und liefern ihnen die Anbieter großer Content-Management-Systeme als neue und direkt integrierbare Plug-ins und Erweiterungen.

Den Journalist*innen kommt in einer freien, demokratischen Gesellschaft eine besondere Verantwortung zu: Sie sollen als unabhängige, staatsferne Gatekeeper fungieren und mit verlässlichen, korrekten und vertrauenswürdigen Informationen zu Aufklärung und Meinungsbildung beitragen. Zu den generellen rechtlichen Rahmenbedingungen, unter denen der Journalismus und dessen Akteur*innen agieren – auch unter dem Begriff „Presse“ zusammengefasst – gehören sowohl bestimmte Privilegien als auch Pflichten. Beispielsweise Ausnahmerechte hinsichtlich aktueller Berichterstattung, des Informantenschutzes oder der Aufdeckung von gesellschaftlich relevantem Fehlverhalten, zugleich etwa die Verpflichtungen zu Sorgfalt und Transparenz.

Nicht zuletzt ist der Journalismus heutzutage bereits nahezu täglich mit Inhalten konfrontiert, die zu gewissen Teilen oder komplett generativer KI-Output sind oder in irgendeiner Weise mit KI-gestützten Tools erzeugt wurden. Das umfasst Texte, Bilder, Audio-, Video- und audiovisuelle Inhalte gleichermaßen und betrifft sämtliche denkbaren Distributionsformen: Webseiten, Apps, Podcasts, Druck-erzeugnisse, Radio oder TV. Daraus ergibt sich für Journalist*innen einerseits die Frage nach der Richtigkeit der Informationen, Bilder und Berichte – auf die sie mit akribischen Prüfungen reagieren müssen. Zudem werden Kennzeichnungen und Nachweise immer wichtiger, etwa von KI-generierten oder KI-manipulierten Inhalten und Beweise der Authentizität, etwa eines fotojournalistischen Bildes. Denn die inhaltliche Qualität, die Glaubwürdigkeit und die Verlässlichkeit der Arbeit ist entscheidend für das Vertrauen der Gesellschaft in den Journalismus und für den Erhalt seines besonderen Status.

(Öffentliche) Verwaltung

Im Rahmen unserer Untersuchungen zeigte sich, dass Akteur*innen in der öffentlichen Verwaltung gegenüber generativen KI-Systemen eine ambitionierte Haltung einnehmen: In interdisziplinären Teams außerhalb klassischer Hierarchien und Zuständigkeiten wird auf kommunaler bis nationaler Ebene erprobt und reflektiert, welche Aufgaben und Leistungen durch generative KI-Systeme unterstützt oder komplett ausgeführt werden können. Getrieben scheinen diese Initiativen von der Hoffnung auf eine KI-gestützte Effizienzsteigerung, die eine Antwort auf Herausforderungen des demografischen Wandels und eines zunehmenden Fachkräftemangels bieten kann. Bei den KI-Systemen, die bislang in der öffentlichen Verwaltung diskutiert werden, handelt es sich in Deutschland wie in ganz Europa in der Regel um angepasste Standardsysteme, bei denen große KI-Modelle verwaltungsspezifisch nachtrainiert und finegetuned werden.

Es lässt sich festhalten, dass sich öffentliche Verwaltungen in Deutschland um einen ambitionierten Umgang mit generativen KI-Systemen bemühen. Der für die öffentliche Verwaltung relevante Output generativer KI-Systeme ist ebenso vielfältig wie die Einsatzmöglichkeiten von KI-Systemen: Je nach Funktion und Kontext können KI-Anwendungen inhaltliche Zusammenfassungen, Informationstexte, Bilddarstellungen, Skizzierungen, aber auch Bewertungshilfen und Entscheidungsempfehlungen erzeugen.



War die Forschungsmatrix ein sinnvoller Ansatz? Lesen Sie dazu unsere Selbstbeobachtungen in Kapitel 4.

4

Selbstbeobachtung: So bewerten wir unser Vorgehen

„Hinterher ist man immer schlauer“: Mit etwas Abstand werden Entscheidungen klarer, Annahmen überprüfbar und blinde Flecken sichtbar. Genau das ist der Ausgangspunkt dieses Kapitels. Wir werfen einen selbstkritischen Blick auf unsere Methoden, Ansätze und Vorgehensweisen:

Was hat gut funktioniert?

Was nicht ganz so gut?

Und was nehmen wir daraus für die Zukunft mit?



Die Auswahl der Arbeitskontexte Bildung, Journalismus, Kultur- und Kreativwirtschaft, Verwaltung

Warum haben wir uns auf genau diese vier Arbeitskontexte konzentriert? Wir wollten Bereiche untersuchen, in denen sich die Auswirkungen generativer KI bereits konkret beobachten lassen. Außerdem waren wir der Überzeugung, dass eine begrenzte Auswahl an Arbeitskontexten sinnvoll ist, um eine gewisse Vergleichbarkeit zu gewährleisten. Gleichzeitig sollten sie unterschiedlich genug sein, um Redundanzen zu vermeiden. Ein weiterer Faktor: Alle vier Arbeitsbereiche stehen bereits im Fokus von Medien und Forschung.

Rückblickend zeigt sich, dass selbst vier Bereiche eine große Bandbreite darstellen. Überschneidungen und thematische Dopplungen ließen sich nicht vollständig vermeiden. Eine stärkere Fokussierung – etwa auf zwei bewusst kontrastierende Bereiche – hätte möglicherweise noch klarere Vergleichslinien ermöglicht.

Gleichzeitig wurde deutlich, dass viele Herausforderungen bereichsübergreifend auftreten. Urheberrechtsfragen betreffen etwa Bildung und Kreativwirtschaft gleichermaßen. Datenschutz spielt im Journalismus eine ähnliche Rolle wie in der Verwaltung. Diese Paralleltät wurde im Zuge der praktischen Auseinandersetzung nochmals besonders deutlich und hat uns bei der thematischen Zuspitzung für die Zukunftslabore geholfen.

Was war gut? Je gewähltem Arbeitskontext war eine Teilgruppe des Teams verantwortlich und federführend. So waren Zuständigkeiten eindeutig geklärt und jede Person konnte die bestehende Expertise detailliert und passgenau erweitern. Zudem konnten wir durch diese Aufteilung die große und kontinuierlich anwachsende Menge an Quellen und Informationen zu generativer KI und den jeweiligen Bereichen besser bewältigen.

Was war nicht so gut? Die Aufteilung brachte im Umkehrschluss mit sich, dass zunächst Wissenssilos entstanden. Mit dem Ziel, das jeweilige Detailwissen zusammenzuführen, um eine Synthese der Forschungsergebnisse und verbindende Erkenntnisse zu identifizieren, mussten wir diese Herausforderung im Projektverlauf neu angehen.

Was nehmen wir mit? Auch in scheinbar engen Untersuchungsfeldern, wie beispielsweise Kreativwirtschaft oder Verwaltung, bringt die Dynamik der KI-Entwicklung eine sehr große Bandbreite an rechtlichen Fragen mit sich. Das haben wir, wie viele andere auch, in der praktischen Dimension unterschätzt. Hier gilt es künftig, früher nachzujustieren und die Fokussierung zu schärfen. Es bietet sich etwa an, bestimmte Teilbranchen unterschiedlicher Arbeitskontexte gemeinsam zu bearbeiten, wenn diese bei ähnlicher Problemlage vergleichbare Lösungsansätze suchen – zum Beispiel zwischen Verwaltung und Bildung.

Mehr zur Auswahl der Arbeitskontexte und zu unserer Forschungsmatrix lesen Sie in Kapitel 3.



4

Literaturanalyse

Den Ausgangspunkt des Forschungsvorhabens bildete eine umfassende Literaturanalyse. Ziel war es, den bestehenden Wissensstand zu erfassen: Welche Forschung gibt es bereits? Wo und wie werden unsere Fragestellungen medial diskutiert? Welche Akteur*innen sind relevant? Wo können wir anknüpfen?

Unsere Recherche umfasste Gesetze, Gesetzesbegründungen und staatliche Positionierungen, wissenschaftliche Publikationen, internationale Medienberichte sowie Publikationen von NGOs, Stiftungen und anderen zivilgesellschaftlichen Organisationen. Obwohl – oder vielleicht gerade, weil – generative KI ein vergleichsweise junges Phänomen ist, ist das Themenfeld extrem dynamisch. Technologische Fortschritte – beispielsweise Größe, Komplexität und Fähigkeiten großer Sprachmodelle oder die Möglichkeiten darauf aufsetzender KI-Agenten – entwickeln sich schnell und damit oft auch die rechtlichen Rahmenbedingungen ihres Einsatzes. „Neue“ Erkenntnisse können also schnell veralten.

Was war gut? Die fachbereichsübergreifende Literaturanalyse hatte gleich mehrere Vorteile: Sie erbrachte eine fundierte Basis für unsere Leitfadeninterviews, ermöglichte einen guten Überblick über relevante Akteur*innen und wurde zur Grundlage für den Aufbau einer belastbaren, durchsuchbaren Datenbank.

Was war nicht so gut? Die Ergebnisse einer Literaturanalyse in einem Team sinnvoll zu strukturieren und die richtigen Formate zu finden, um Erkenntnisse zu teilen, ist generell eine Herausforderung – und so war es auch für uns. Außerdem: Zum Zeitpunkt der Recherche gab es noch zu wenig empirische Forschung zu den von uns ausgewählten Arbeitsbereichen – die hätten wir allerdings gut gebrauchen können. Gleichzeitig zeigt sich ab einem bestimmten Zeitpunkt im Forschungsprojekt, dass neue umfangreiche Werke nur noch mit unverhältnismäßigem Aufwand in die eigenen Arbeiten und Arbeitsschritte einbezogen werden können.

Was nehmen wir mit? Gerade bei solch diversifizierten Themenfeldern wie KI-Rechtsfragen gilt es, vorab klar zu definieren, nach welchen Vorgaben man Literatur und Quellen archiviert. Zudem scheint es in solchen Fällen geboten, die Kategorien und Strukturen der Literatursammlung im Team immer wieder zu evaluieren und gegebenenfalls anzupassen. Auch hier bleiben Open-Source- und Open-Access-Lösungen eine gute wissenschaftliche Praxis.



Leitfadeninterviews

In der zweiten Phase des Projekts führten wir rund 30 qualitative Interviews mit Expert*innen aus den vier Arbeitskontexten durch. Die Gespräche dauerten zwischen 30 und 60 Minuten und wurden mithilfe von KI-gestützten Tools (Aiko, Amberscript, WhisperAI) transkribiert. Ziel war es, die Erkenntnisse aus der Literaturanalyse sowie erste Annahmen, die sich daraus ergeben hatten, zu überprüfen und durch praktische Perspektiven zu ergänzen. Die Interviews lieferten genau das: konkrete Einblicke in Arbeitsabläufe, neue Fragestellungen und einen wertvollen Realitätscheck – mit teilweise überraschenden Ergebnissen.

Die Auswahl der Interviewpartner*innen erwies sich als Herausforderung: Um ein ausgewogenes Bild zu bekommen, versuchten wir möglichst unterschiedliche Erfahrungswerte abzubilden. Das ist letztlich nur zum Teil gelungen. Viele Interviewpartner*innen sind in ihren jeweiligen Arbeitskontexten überwiegend Expert*innen oder zumindest technikaffine Early-Adopter. Perspektiven von Menschen, die sich eher nicht oder zumindest nicht eigeninitiativ mit jenen Veränderungen beschäftigen, die durch den Einsatz generativer KI in ihren Arbeitsbereichen geschehen, konnten wir nicht ausreichend einfangen. Das lag zum Teil daran, dass Personen mit solchen Profilen ihre Absagen an uns genau damit begründeten, dass sie zu wenig praktische Erfahrungen und Kenntnisse mit beruflich nutzbaren KI-Werkzeugen hätten. Erfreulich war, dass wir zu jedem Arbeitskontext auch Interviewpartner*innen aus dem europäischen Ausland gewinnen konnten, was teilweise zu sehr interessanten Erkenntnissen führte.

Was war gut? Die Leitfadeninterviews erwiesen sich als wertvolle Ergänzung zur Literaturrecherche. Sie erweiterten – wie von uns erhofft – die theoretische Perspektive um konkrete Einblicke in den Arbeitsalltag und lieferten praxisnahe Beobachtungen. Gleichzeitig brachten die Gesprächspartner*innen neue Impulse und Fragestellungen ein, die direkt aus realen Arbeitssituationen entstanden.

Was war nicht so gut? Eine große Herausforderung dieser Methode liegt im dafür erforderlichen Aufwand. Planung, Durchführung und Auswertung erwiesen sich als sehr zeitintensiv – wobei zumindest der Einsatz von KI-gestützten Werkzeugen zum Transkribieren und Übersetzen zählbare Entlastung brachte.

Was nehmen wir mit? In den Expert*innen-Interviews wurde deutlich, dass die Herangehensweisen und Erfahrungshorizonte beim Einsatz generativer KI noch sehr individuell sind. So divers, wie die Fachleute die tatsächlich mit KI-Tools erprobten Arbeitsschritte und ihre spezifischen Wahrnehmungen von den erwarteten und erfolgten KI-„Wirkungen“ beschrieben, so sehr wuchs in uns die Erkenntnis, dass wir bei zukünftigen, qualitativen Interviews zu diesem Thema wohl eher auf einen standardisierten Leitfaden verzichten würden.

4

Zwischenbericht / Working Paper

Als eine Art erste Wasserstandsmeldung sollte nach etwa der Hälfte der Projektlaufzeit ein Zwischenbericht erstellt werden. Das Ziel des Berichts schien zunächst klar: Unsere bisherigen Ergebnisse zusammenzufassen und zu präsentieren. Allerdings warf die Fülle an relevanten Aspekten schnell die Frage auf, welche Form und welche Zielgruppe für diesen Zwischenbericht die geeignetste wäre. Wie sich beispielsweise allein am Thema Datenschutz zeigte, gibt es sehr viele unterschiedliche Perspektiven und Bedarfe zu KI-Rechtsfragen. Etwa von Datenschutzbeauftragten in Behörden oder Unternehmen, die das schwierige Wechselverhältnis von DSGVO und KI-VO interessiert. Demgegenüber fragen Beschäftigte im Bildungsbereich eher danach, welche datenschutzrechtlichen Kenntnisse sie für die berufliche Nutzung von KI-Anwendungen benötigen.

In einem ersten Anlauf entwickelte sich ein Dokument, das am Ende zu viel auf einmal sein wollte: Forschungsbericht für interessierte Expert*innen aus den Rechtswissenschaften, kurzweiliger Input für Medienschaffende und die allgemein interessierte Öffentlichkeit, Ergebnislage für politische und wirtschaftliche Entscheidungsträger*innen. Quasi auf halber Strecke nahmen wir zur Kenntnis, dass das so wohl nicht funktionieren würde. Wir entschieden, den Ansatz zu revidieren: Das neue Konzept sah ein kompaktes Working Paper vor, das im Kern zehn klare Thesen zu offenen Rechtsfragen und rechtlich begründeten Konflikten beim Einsatz generativer KI enthält, sowohl prägnant zusammengefasst als auch inhaltlich vertieft. Umrandet von Erläuterungen zu unserer Methodik und zur genutzten Terminologie, vermieden wir so, dass der Bericht ausuferte oder zu beliebig wurde.

Was war gut? Die Revision, das Nachdenken, Umdenken und Überarbeiten von Konzept und Texten hat zwar Zeit gekostet, doch das hat sich aus unserer Sicht gelohnt: Wir halten unser Working Paper „Erkenntnisse und Befunde zu den rechtlichen und gesellschaftlichen Rahmenbedingungen beim Einsatz generativer KI“ trotz aller Herausforderungen für ein gut strukturiertes und gut lesbares Dokument, das für verschiedene Zielgruppen anschlussfähig ist.

Was nehmen wir mit? „Snackable Content“ hat auch einen Platz in Forschungsberichten: Kurzweilige und niedrigschwellige Einschübe können einen langen und vermeintlich schwer zugänglichen Bericht auflockern, und bieten eine zugängliche Grundlage für die Weiterwertung – etwa für Konferenzbeiträge, Meinungspapiere oder sonstige Publikationen.

Die zehn Kernthesen und weitere Auszüge aus dem Working Paper lesen Sie in Kapitel 3.

prompt/ Newsletter

Ein zentrales Ziel des Projekts ist die öffentlichkeitswirksame Kommunikation der aggregierten Ergebnisse und gewonnenen Erkenntnisse. Daher entschieden wir uns für einen periodisch erscheinenden E-Mail-Newsletter, den wir – passend zum Hauptthema generative KI – „prompt/“ nannten.

Damit verbinden wir den Anspruch, aktuelle Entwicklungen sowie Grundwissen rund um generative KI und Recht praxisnah, verständlich und branchenunabhängig aufzubereiten. Die Ansprache ist bewusst niedrigschwellig, Fachbegriffe werden erklärt. Die Reichweite des Newsletters entstand organisch: Abonnent*innen gewannen wir Ausgabe für Ausgabe durch bestehende Kontakte und Empfehlungen der Leser*innen, Hinweise auf die neuesten prompt/-Inhalte bei Veranstaltungen sowie in den Social-Media-Kanälen des iRights.Lab; Nicht zuletzt auch dank der Übernahme von Beiträgen auf anderen Plattformen, wie beispielsweise [iRights.info](https://www.iriights.info). Da die prompt/-Inhalte unter offenen Creative Commons-Lizenzen veröffentlicht werden, ist diese Übernahme leicht möglich und ausdrücklich erwünscht.

Was war gut? Der prompt/-Newsletter ermöglicht es, auf wichtige Ereignisse und neue Entwicklungen „prompt“ zu reagieren, beispielsweise Klagen und Entscheidungen in Gerichtsverfahren gegen KI-Hersteller oder KI-relevante Lizenzen von Verwertungsgesellschaften im Kontext der Forschung einzuordnen. Gleichzeitig bietet er mit dem Fokus auf Hintergrundwissen zu Rechtsrahmen, wie der KI-VO, sowie Übersichten zu KI-rechtlichen Fragen, immer wieder einen praktischen Mehrwert für die Leserschaft. Zudem schlägt er die Brücke in die Praxis und schafft Vernetzungsmöglichkeiten, insbesondere durch die exklusiven Interviews mit KI-Rechts-Expert*innen.

Was war nicht so gut? Noch besser wäre eine höhere Frequenz und damit mehr Ausgaben gewesen. Zudem ein weiterer Ausbau des Formats, etwa durch mehr Interaktionsmöglichkeiten mit den Leser*Innen. Das war im bisherigen Projektverlauf jedoch noch nicht vorgesehen.

Was nehmen wir mit? Da wir sowohl mit den Inhalten als auch mit dem kontinuierlich ansteigenden Zuspruch sehr zufrieden sind, fällt unsere Bilanz hier in den Bereich Wunsch und Visionen: Wir hätten gerne mehr Ausgaben mit noch mehr Wissensvermittlung und noch mehr Einordnung zu aktuellen und kommenden Rechtsfragen beim Einsatz generativer KI vorgelegt – und streben genau das an.


Auszüge aus Beiträgen und Interviews aus den bisher zehn prompt/-Ausgaben lesen Sie in Kapitel 3.



4

Zukunftsszenarien

Die Bestandsaufnahmen und Analysen der rechtlichen Rahmenbedingungen für den Einsatz generativer KI, die wir in den ersten beiden Projektphasen durchführten, sollten als Basis für ein zentrales Ziel dienen: Innovationen und konkrete Zukunftsperspektiven zu entwickeln. Dafür organisierten wir drei Zukunftsszenarien – moderne Workshopformate für interdisziplinären, konstruktiven Austausch.



Selbstbeobachtung

Mehr zur Methode und den Inhalten der Zukunftswerkstätten lesen Sie in Kapitel 3.

Die Zukunftsszenarien fokussierten jeweils eines der drei Kernthemen, für die wir im Zuge unserer Forschung sowohl Handlungsbedarf als auch die Erfordernisse für neue, innovative Denkansätze identifizierten: Datenschutz, Vergütung und Urheberrecht sowie Kennzeichnung. Diesen Themen widmeten sich jeweils Gruppen von etwa zehn Teilnehmenden. Die bewusst heterogen zusammengesetzten Gruppen brachten Rechts-/Wissenschaftler*innen, Akteur*innen aus der Wirtschaft und Verbänden sowie Einzelaktuer*innen mit Praxisbezug zusammen.

Ziel der Zukunftsszenarien war es, von einer gemeinsamen Problemdefinition über Visionen zu konkreten Handlungsschritten zu gelangen. Diese drei Schritte wurden in jeweils drei kompakten Einzelveranstaltungen umgesetzt, moderiert und angeleitet durch das GenKI-IR-Team.

Was war gut? Die Zukunftsszenarien ermöglichten einen intensiven Austausch mit ausgewählten Stakeholdern. Sie zeigten, dass dieses Format für Fragen der KI-Governance einen Mehrwert bietet. Die Abfolge von Kritik, Entwurf und Konkretisierung schafft einen Raum, in dem Probleme zunächst benannt werden können, ohne sie sofort in konkrete Lösungen zu übersetzen. Die Gruppengröße erwies sich als gut gewählt: vielfältig genug für unterschiedliche Perspektiven, klein genug für fokussierte, konstruktive Kommunikation zwischen den Beteiligten. Vertreter*innen von Aufsichtsbehörden, Ministerien, Unternehmen, Kreativschaffenden und Zivilgesellschaft traten nicht primär als Repräsentant*innen ihrer Institutionen auf, sondern als Workshopteilnehmer*innen mit eigenen Meinungen und Erfahrungen. So konnten Problemdeutungen entstehen, die über sektorale Perspektiven hinausreichen und konkrete Vorschläge ermöglichen, die in klassischen Fachdialogen häufig nicht formuliert werden.

Was war nicht so gut? Es blieben auch Grenzen des Formats sichtbar. Die Zusammensetzung der Gruppen war bewusst selektiv. Dies ermöglicht eine dichte, produktive Arbeitsatmosphäre, gleichwohl begrenzt es die Breite an Perspektiven auf die erörterten Fragen und Lösungswege. Um beispielsweise weitere Nutzer*innengruppen oder internationale Sichtweisen zu berücksichtigen, bräuchte es mehr Zeit und mehr Beteiligte. Hier gilt es abzuwägen und den Rahmen der zeitlich und konzeptionell kompakten Zukunftsszenarien gegebenenfalls anders zu setzen.

Was nehmen wir mit? Die in unseren Zukunftsszenarien entwickelten Modelle und Lösungswege bewegen sich auf einer vergleichsweise hohen Abstraktionsebene. Gleichwohl liefern sie als kollektiv entwickelte Gedankenexperimente handfeste Vorlagen, auf die juristische, technische und organisatorische Ausarbeitungen aufsetzen können. In allen drei Fällen zeigte sich uns, dass das Format Zukunftsszenario dafür geeignet ist, um komplexe Konfliktlagen zu ordnen, normative Spannungen sichtbar zu machen und eine Grundlage für vertiefte rechtspolitische Arbeit zu schaffen. Daran wollen wir anknüpfen und die Methode auch zukünftig einsetzen.

Nähere Beschreibungen zu Inhalten und Ergebnissen der drei Zukunftswerkstätten lesen Sie in Kapitel 3.



5

KI-Rechtsfragen zu Datenschutz, Vergütungen, Kennzeichnungen und Governance – damit wollen wir uns bei GenKI-IR weiter auseinandersetzen

GenKI-IR hat gezeigt, wie rasant generative KI in zentrale gesellschaftliche Bereiche hineinwirkt und wie begrenzt bestehende Rechtsrahmen und Institutionen bislang darauf vorbereitet sind. Besonders deutlich wird das bei den drei Zukunftsszenarien zu den Bereichen Datenschutz, Urheberrecht sowie Kennzeichnung und Transparenz von KI-generierten Inhalten (mehr dazu lesen Sie in Kapitel 2). Die Diskussionen dort zeigen uns: Zwar sind zahlreiche Fragestellungen arbeitskontextübergreifend relevant und lassen sich zu strukturellen rechtlichen Herausforderungen verdichten. Hierzu entstehen in der zweiten Projekthälfte erste Lösungsansätze, Modellierungen und Handlungsspielräume, um den erkannten rechtlichen Unsicherheiten zu begegnen.

Nun gilt es, die Erkenntnisse weiter zu vertiefen, aktuelle rechtliche und technische Entwicklungen systematisch zu analysieren und neue beziehungsweise fortgeschriebene Fragestellungen kontinuierlich einzuordnen. Das Feld generativer KI ist dabei geprägt durch dynamische rechtliche Rahmenbedingungen – etwa neue regulatorische Vorgaben, ethisch-rechtliche Implikationen, Rechtsprechung und Verwaltungspraxis – sowie durch kontinuierliche technologische Veränderungen. Daher braucht es weiterhin eine aufmerksame Begleitforschung, die rechtliche und technische Entwicklungen fortlaufend beobachtet, analysiert und einordnet – und mithilfe bewährter Formate, wie Zukunftslabore, konkrete und belastbare Lösungsansätze für einen rechtssicheren und verantwortungsbewussten Einsatz generativer KI in Arbeitsprozessen entwickelt. Forschungsinteresse und Ziel ist, rechtliche, organisationale und gesellschaftliche Bedingungen zu definieren, wie generative KI rechtssicher und sinnvoll in Arbeitsprozessen eingesetzt werden und diese unterstützen kann.

Offene Baustellen im Datenschutz

Im Themenfeld Datenschutz bestätigt sich, dass die DSGVO in ihrer aktuellen Auslegung nur bedingt auf großskalige Trainingsprozesse und komplexe Modellarchitekturen übertragbar ist. Besonders umstritten bleiben die Abgrenzung sensibler Daten, tragfähige Rechtsgrundlagen für KI-Training sowie die praktische Durchsetzung von Betroffenenrechten entlang mehrstufiger Wertschöpfungsketten. Zugleich wird deutlich, dass es bereits konkrete Reformansätze gibt – von risikobasierten Regulierungsmodellen, die einen Teil der Einschätzungsverantwortung auf den Gesetzgeber übertragen, bis hin zu Neukonzeptionen zentraler Anknüpfungspunkte wie der Datenverarbeitung oder dem Personenbezug. Für die nahe Zukunft ergeben sich für uns mehrere Forschungsfragen:

- Wie lässt sich die DSGVO so justieren, dass sie KI-Trainings und -Fine-Tunings adressiert, ohne den Schutzstandard auszuhöhlen?
- Welche Kombination aus rechtlichen Leitlinien und technischen Mindeststandards (Privacy by Design, datenschutzfreundliche Systemarchitekturen) ist nötig, damit Betroffenenrechte im Umfeld generativer KI mehr sind als ein formaler Anspruch?
- Welche Anwendungsbereiche könnten von der DSGVO ausgenommen werden – und welche Konsequenzen ergäben sich daraus?
- Wie gehen Praxis und Wissenschaft mit den Problemstellungen in der KI-Entwicklung – Qualität der Datensätze, keine einheitliche Anonymisierung – um? Welche Lösungen sollten gefördert werden und warum?



5

Verteilungskonflikte und Vergütung

Im Bereich Urheberrecht und Vergütung machen die Zukunftslabore ein strukturelles Problem sichtbar: Kreativschaffende erleben, dass ihre Werke in Trainingsdaten eingehen und sie zugleich durch KI-gestützte Angebote unter Wettbewerbsdruck geraten, ohne dass daraus für sie verlässliche Einnahmen oder Zukunftsperspektiven entstehen. Das Urheberrecht bleibt als Instrument unverzichtbar, stößt aber angesichts multipolarer Interessenkonflikte, einer Vielzahl von betroffenen Akteur*innen, globaler Plattformökonomien und asymmetrischer Verhandlungsmacht an klare Grenzen.

Die im Rahmen von GenKI-IR diskutierten Ansätze – von KI-Steuern und „Disruption Levies“ über kollektive Vergütungsfonds bis hin zu öffentlich beziehungsweise gemeinwohlorientiert finanzierter KI-Infrastruktur – markieren deshalb keine fertigen Lösungen, sondern Koordinaten für weitere Forschung. Offene Punkte, an die man in Zukunft anknüpfen sollte, sind unter anderem:

- Wie lassen sich kollektive Vergütungsmechanismen so gestalten, dass sie tatsächlich bei Kreativschaffenden ankommen und nicht nur etablierte Rechteinhaber bevorzugen?
- Welche Governance-Modelle sind für Verwertungseinrichtungen und öffentliche KI-Infrastrukturen geeignet (Transparenz, Rechenschaft, Beteiligung)?
- Wie können urheberrechtliche und kulturpolitische Instrumente so kombiniert werden, dass sie international anschlussfähig bleiben und globale Akteure effektiv adressieren?

Kennzeichnung, Vertrauen und Plattformverantwortung

Beim Thema Kennzeichnung von KI-Output zeigt sich, dass einfache Label-Pflichten zu kurz greifen. Technische Lösungen, wie Wasserzeichen oder Metadaten, gelten zwar als prinzipiell geeignet. Doch sie sind oft manipulierbar und entfalten gerade dort geringe Wirkung, wo Akteur*innen gezielt täuschen wollen. Gleichzeitig kann eine Label-Inflation zu einem Wahrnehmungsproblem führen: Kennzeichnungen werden übersehen, missverstanden oder erzeugen unerwünschte Authentizitätseffekte.

Hier verschieben sich die spannenden Fragen deutlich weg vom „Ob“ hin zum „Wie“:

- Welche sektoralen Standards braucht es, etwa für Journalismus, Werbung, Verwaltung oder Justiz, um unterschiedliche Vertrauensanforderungen abzubilden?
- Wie können technische Authentifizierungsverfahren mit klaren Verantwortlichkeiten für Plattformen, Intermediäre und Aufsichtsstellen verzahnt werden?
- Wie können sinnvolle, idealerweise internationale Standards für Kennzeichnungen erreicht werden und welche

Rolle sollten dabei Stakeholdergruppen wie Nutzer*innen, Wissenschaft, Wirtschaft und der Staat einnehmen?

- Welche Rolle spielen kartell und wettbewerbsrechtliche Instrumente, wenn es darum geht, Kennzeichnungs und Transparenzstandards gegenüber großen Plattformen durchzusetzen?

KI-Governance als Daueraufgabe – ab sofort und unverzüglich

Über die drei Kernthemenfelder unserer Forschung hinweg kristallisiert sich ein gemeinsamer Befund heraus: Individuelle Steuerungsinstrumente – Einwilligungen, Einzellizenzen, isolierte Kennzeichnungspflichten – stoßen im Bereich generativer KI an strukturelle Grenzen. Benötigt werden kollektive, institutionell unterlegte Lösungsansätze, die Datenschutz sowie Medien und Wettbewerbsrecht nicht nur nebeneinanderstellen, sondern miteinander verschränken. Gleichzeitig sind die bestehenden Aufsichts, Beratungs- und Förderstrukturen nur begrenzt koordiniert. Viele Akteure berichten von einem Flickenteppich aus Zuständigkeiten und unklaren Schnittstellen.

In der Zukunft gilt es daher, an diesen neuralgischen Punkten anzusetzen:

- **Rechtliche Reformvorschläge – etwa zu DSGVO-Anpassungen, kollektiven Vergütungsmodellen oder Kennzeichnungspflichten – systematisch mit institutionellen und infrastrukturellen Szenarien verknüpfen.**
- **Offene Forschungsfragen multiperspektivisch vertiefen, insbesondere unter Einbeziehung von Praxis, Wissenschaft, Wirtschaft und Aufsicht.**
- **Mit transdisziplinären Formaten (wie den Zukunftsszenarien) fortlaufend Erfahrungen aus Verwaltung, Wirtschaft, Aufsicht, Kreativwirtschaft und Zivilgesellschaft zu konsolidieren, um innovative Ansätze generieren und kontinuierlich in rechtspolitische Prozesse zurückspielen zu können.**



Die Entwicklung der Rechtsrahmen zu generativer KI braucht kontinuierliche Begleitforschung

Die Spannungen, die wir im Rahmen von GenKI-IR zwischen Innovation und Grundrechtsschutz, zwischen Wertschöpfung und Vergütung, zwischen Transparenzanspruch und tatsächlichem Vertrauen identifizieren, werden sich in naher Zukunft nicht auflösen. Entscheidend wird sein, Rechtsrahmen so zu strukturieren, dass Rechtspolitik, Aufsicht und Praxis nicht nur kurzatmig auf Krisen reagieren, sondern handlungsfähig bleiben. Genau dafür braucht es neben einer strukturierten gesamtgesellschaftlichen Debatte eine kontinuierliche Begleitforschung, die Widersprüche und Konflikte sichtbar macht, Optionen durchdekliniert und tragfähige Regulierungspfade erdenkt und vorbereitet.

6

Anhang

Glossar/Terminologie

Um Widersprüche zu den untersuchten rechtlichen Rahmenbedingungen zu vermeiden und gleichzeitig die allgemeine Lesbarkeit zu gewährleisten, orientiert sich das vorliegende Working Paper in seiner Terminologie überwiegend an der im August 2024 in Kraft getretenen KI-Verordnung der Europäischen Union.

„**KI-Modelle**“ lassen sich als Vorprodukte verstehen. Sie dienen als Datenbasis von → KI-Systemen. KI-Modelle basieren wiederum auf spezifischen technischen Verfahren (wie dem maschinellen Lernen) und eröffnen oft eine große Bandbreite von Anwendungsmöglichkeiten. Bekannt sind daher auch die Begriffe „Basismodell“ oder „großes Sprachmodell“ (Large Language Model, kurz: LLM). Der Verwendungszweck ist bei der Entwicklung von KI-Modellen in der Regel noch nicht festgelegt. Er ist abhängig „von der Art und Weise seines Inverkehrbringens“. KI-Modelle werden in der Regel von ihren Anbietern vermarktet. Im Zuge dessen werden Lizenzen an Anbieter von KI-Systemen vergeben.

Ein „KI-System“ wird verstanden als „ein maschinengestütztes System, das für einen [...] autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben [...] explizite oder implizite Ziele ableitet“ (Art. 3 Nr. 1 KI-VO). Maßgeblich ist hier das Kriterium der Ableitung. Darin spiegelt sich die Entwicklung des Systems auf der Basis von Prinzipien wider, vor allem denen des maschinellen Lernens. KI-Systeme sind daher auf → KI-Modellen basierende Anwendungen, die auf unterschiedliche Art in den Verkehr gebracht werden. Beispiele für bekannte KI-Systeme sind ChatGPT, Claude, Gemini, Midjourney oder Dall-E.

„**Generative KI-Systeme**“ ermöglichen eine Erzeugung von Inhalten, etwa in Form von Text, Audio, Bild oder Video. Aufgrund der vielseitigen Einsatzmöglichkeit und Nutzbarkeit insbesondere von textlichem KI-Output weisen generative → KI-Systeme oft ein breites Spektrum an möglichen Einsatzgebieten und Aufgaben auf, die mit ihrer Hilfe erledigt werden können.

„**Anbieter**“ sind jene Unternehmen oder Personen, die ein → KI-Modell oder ein → KI-System entwickeln oder entwickeln lassen und unter eigenem Namen auf den Markt bringen. Je nach Entwicklungsprozess eines KI-Systems können (und werden häufig) die Anbieter von KI-Modellen und von KI-Systemen unterschiedliche Akteure sein.

„**Betreiber**“ sind Personen, Unternehmen oder Institutionen, die → KI-Systeme für berufliche und/oder kommerzielle Zwecke einsetzen.

Der Begriff „Nutzer*innen“ wird in der KI-Verordnung nicht definiert. Verstanden werden hierunter Unternehmen oder Personen, die KI-Systeme verwenden, unabhängig davon, zu welchem Zweck.

Quellen

Die Ergebnisse, Analysen und Erkenntnisse, die diesem Begleitforschungsprojekt als Ausgangspunkte dienten, basieren zu großen Teilen auf umfänglichen Literaturrecherchen in verschiedenen Disziplinen, die insbesondere im ersten Projektschritt stattfanden. Angesichts des hohen Entwicklungstempos im Bereich KI und der enormen Zahl an Veröffentlichungen, wäre eine umfassende Auswertung aller einschlägigen Publikationen unmöglich. Für die Basisrecherchen werteten wir über 350 wissenschaftliche, fachspezifische und journalistische Aufsätze, Untersuchungen, Stellungnahmen, Positionspapiere und Studien aus, sowie juristische Kommentare und Weiteres – insbesondere aus dem Zeitraum von Januar bis September 2025. Ein ausführliches Quellen- und Literaturverzeichnis finden Sie im „Working Paper zu rechtlichen und gesellschaftlichen Rahmenbedingungen beim Einsatz generativer KI“ (ab Seite 70, <https://irights-lab.de/outputs/genki-working-paper>)

6

Über dieses Projekt

Das Begleitforschungsvorhaben „Generative KI – Innovation und Recht in Arbeitsprozessen“ nimmt relevante Aspekte und strittige rechtliche Fragen in den Blick. Es untersucht Anwendungskontexte entlang des Lebenszyklus generativer KI-Systeme und anhand beispielhafter Nutzungen in den Bereichen Kultur- und Kreativwirtschaft, Journalismus, Bildung sowie Verwaltung. Dabei orientiert es sich an der zukunftsgerichteten Leitfrage: **Welchen rechtlichen und gesellschaftlichen Rahmen braucht es, damit Anwendungen generativer künstlicher Intelligenz bestimmte Arbeitsprozesse sinnvoll unterstützen können?**

Hierbei untersuchen und vermitteln wir gegebene Rechtsgrundlagen beim Schaffen und Verwerten von KI-generierten Inhalten für bestimmte Branchen. Zu den Zielen gehört, Handlungsspielräume sowie Zukunftsszenarien des rechtssicheren und sinnvollen Einsatzes von generativer KI in den betrachteten Kontexten aufzuzeigen.

Im Rahmen dieses Forschungsprojekts erscheint periodisch auch der Newsletter „prompt/“ mit kompakten Analysen, Interviews und Tipps zu den wichtigsten Entwicklungen im Bereich generativer KI.

Das Projekt wird gefördert vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR.)

Das Team



Dr. Matthieu Binder



Solvejg Gunkel



Lena Biskup



Merlin Münch



Henry Steinhou



Katja Berg



Ella Jordan



Dr. Till Kreuzer



Yannick Hahn



Kathrin Maurer



Elena Kalogeropoulos



Dr. Wiebke Glässer



Philipp Otto

6

Danksagungen

Das gesamte GenKI-IR-Projekt-Team bedankt sich bei folgenden Personen, die an einzelnen Projektvorhaben mitgewirkt, zu Einschätzungen, Erkenntnissen und Einordnungen beigetragen und uns während der Projektlaufzeit mit ihrem Wissen, ihren Expertisen oder ihrer Tatkraft unterstützt haben:

In alphabetischer
Reihenfolge und nach
Vornamen sortiert.

Prof. Dr. Albert Sanchez-Graells	Marco Urban
Dr. Amit Datta	Marie-Lene Armingeon
Dr. Anke Reich	Marie-Theres Mayer
Dr. Anke Schierholz	Marlis Prinzing
Aurélie Waeterinckx	Prof. Dr. Martin Ebers
Berhard Gmeiner	Martin Kretschmer
Bettina Rehm	Prof. Dr. Martin Steinebach
Boris Eldagsen	Matthias Hornschuh
Chan-Young Ramert	Mattias Wickberg-Hugerth
Dr. Christian Gemmin	Max-R. Ulbricht
Prof. Dr. Christiane Wendehorst	Dr. Moritz Karg
Christoph Brill	Nicole Wrana
Dagmar Schwelle	Nils Pooker
Daniel Maslewski	Nina George
Davy Wang	Nina Miu Graf
Félicien Vallet	Paul Keller
Felix Reda	Ramak Molavi Vasse'i
Florian Dohmann	Rebekka Weiß
Frederick Richter	Prof. Dr. Rolf Schwartmann
Guido Scorza	Ruth Meyer
Hanna Möllers	Sabrina Konrad
Hannes Jakobsen	Prof. Dr. Steffen Eckhard
Hendrik Haverkamp	Steffen Holly
Dr. Jens Keßler	Sven Hilpisch
Dr. Jonas Botta	Dr. Tabea Golgath
Jonas Leschke	Tiger Stangl
Joscha Falck	Tobias List
Julia Smakman	Torsten Zesch
Jürgen Geuter	Yvonne Herzog
Dr. Jyn Schultze-Melling	
Dr. Kai Welp	
Katharina Schell	
Lara Trompeter	
Prof. Dr. Linda Kuschel	
Lisa Käde	
Prof. Dr. Malte Stieper	
Dr. Marco Holtz	

Impressum

Ein Bericht im Rahmen des iRights.Lab-Forschungsprojekts
 Generative KI – Innovation und Recht in Arbeitsprozessen (GenKI-IR)

iRights.Lab GmbH
 Oranienstraße 185, D-10999 Berlin
 Telefon: +49 (0)30 40 36 77 230
 Fax: +49 (0)30 40 36 77 260
 E-Mail: kontakt@irights-lab.de

Geschäftsführer: Philipp Otto
 Projektleitung: Solvejg Gunkel
 Autor*innen und Redaktion: Henry Steinhau, Lena Biskup, Matthieu Binder, Solvejg Gunkel,
 Dr. Till Kreutzer, Ella Jordan, Merlin Münch
 Gestaltung: Beate Stangl / tigerworx; Titelfoto: unsplash.com/ChristianPapaux
 Lektorat: Kathrin Maurer

Juni 2026

Rechtshinweis: Alle Texte und Abbildungen stehen unter der offenen Lizenz CC-BY 4.0, es sei denn, sie sind anders gekennzeichnet.

Registergericht: Amtsgericht Berlin-Charlottenburg
 Registernummer: HRB 185640 B

Finanzamt für Körperschaften II
 USt-IdNr.: DE311181302

Inhaltlich Verantwortlicher i. S. d. § 18 Abs. 2 MStV:
 Philipp Otto (Anschrift siehe oben)

Das Projekt GenKI-IR wird vom unabhängigen Think Tank iRights.Lab verantwortet und durchgeführt. Das iRights.Lab entwickelt Strategien und praktische Lösungen, um die Veränderungen in der digitalen Welt vorteilhaft zu gestalten. Wir unterstützen öffentliche Einrichtungen, Stiftungen, Unternehmen, Wissenschaft und Politik dabei, die Herausforderungen der Digitalisierung zu meistern und die vielschichtigen Potenziale effektiv und positiv zu nutzen.

Weitere Informationen über das iRights.Lab finden Sie unter <https://irights-lab.de/>

Das Begleitforschungsvorhaben „Generative KI – Innovation und Recht in Arbeitsprozessen“ wird gefördert durch das Bundesministerium für Forschung, Technologie und Raumfahrt.

Gefördert durch:



Bundesministerium
 für Forschung, Technologie
 und Raumfahrt

iRights.Lab 
Lösungen für Recht, Technologie,
Politik & Gesellschaft