

KI im Mediensektor: Regeln für verdienten Vertrauen

Ruth Meyer, M.A., ist Direktorin der Landesmedienanstalt Saarland (LMS). Federführend koordiniert sie für die Medienanstalten das Thema KI. 2012-2020 gehörte sie dem Landtag des Saarlandes an.

KIR.() 2026, **89** Die Wahrung einer jeden Demokratie ist Vertrauen: Vertrauen in ihre Verfasstheit, in ihre Organe und in die Prozesse der demokratischen Willensbildung. Im Zentrum dieser Willensbildung steht die medienvermittelte Information über Weltwissen und Tagesgeschehen sowie die diskursive Einordnung dieser Informationen anhand von Werten und Aspekten in Rundfunk und Telemedien. Dass die digitalen Medien uns eine schier endlose Vielfalt an Content und Perspektiven eröffnen, dürfen wir nicht hoch genug schätzen. Hochpotente und allseits verfügbare KI-Anwendungen fordern in diesem Zusammenhang allerdings nicht nur die Redaktionen und noch mehr „den mündigen Bürger“ heraus, sie machen auch ein auf die Potenziale der neuen Technologie angepasstes Recht erforderlich.

I. KI im medialen Alltag– eine Standortbestimmung

Wir schreiben Jahr 4 nach dem Launch von ChatGPT im November 2022 und stellen fest, dass die einst überwiegend mit Staunen registrierten, weil plötzlich leicht verfügbaren KI-Dienste inzwischen mit großer Selbstverständlichkeit in mediale Produktions- und Rezeptionsprozesse Einzug gehalten haben. In unsere digitalen Arbeitstools sind sie als dienstbare Anwendungen integriert, bieten uns mehr oder weniger zutreffende Formulierungsvorschläge, Zusammenfassungen oder Übersetzungen an und präsentieren – unsere Absichten scheinbar vorausahnend – eigenständige Vorschläge für den nächsten Arbeitsschritt oder führen sie gleich selbst aus. In der Individualkommunikation können wir uns virtuelle KI-Freunde als Kommunikationspartner konfigurieren, die selbstlernend Informationen und Diskurse offerieren – je nach Gusto mit fachlichem, politischem, psychologischem oder auch sexuellem Impact.

KI-generierten Content finden wir nicht nur zuhauf in den Feeds unserer Social-Media-Accounts. Auch manche Rundfunkangebote wie bigGPT, Absolut Radio AI oder Studio 47, nähren sich maßgeblich von maschinengenerierten Inhalten. In den USA zeigte kürzlich eine Studie auf, dass fast jeder zehnte Nachrichtenartikel von einer KI geschrieben, nicht jedoch entsprechend gekennzeichnet ist (University of Maryland, AI Use in American Newspapers, 2025). In Deutschland haben sich Verlage, Rundfunkveranstalter und Journalisten und Journalistinnen in ihren Kodizes fast durchgehend dem Human-in-the-loop-Prinzip verpflichtet (Hofeditz/Jung/Mirbabaie, Digital Society 4 2025) und sagen einen verantwortungsvollen transparenten Umgang mit der Technologie zu – ganz iSd Maßgabe der Medienanstalten, Verantwortung zu regeln, Vielfalt zu schützen und Vertrauen zu wahren (die medienanstalten, KI und Medien, Positionspapier v. 20.2. 2024, abrufbar unter: www.die-medienanstalten.de/service/positionspapiere/#c3729). Doch dieser Umgang verlangt uns und den Redaktionen reichlich Disziplin ab, wie nicht zuletzt der ZDF-Beitrag über die ICE-Behörde in den USA gezeigt hat (ZDF, heute-journal v. 15.2.2026, abrufbar unter: <https://www.zdfheute.de/video/heute-journal/heute-journal-clip-6-8310.html>).

Selbst Medien-Profis können Fälschungen aufsitzen, die oft kaum noch als solche zu enttarnen sind, mit Inhalten zum Zwecke der persönlichen Diffamierung Einzelner oder der gezielten Desinformation Vieler. Das Netz ist voll von KI generierten Videos und erfundenen Nachrichten, schon, weil es so leicht funktioniert. Daher müssen wir uns auch rechtlich um „Deepfakes“ kümmern und um den Fakt, dass wir unseren eigenen Augen längst nicht mehr trauen können.

II. Neuere Entwicklungen: AI Search und Agentic AI

Nicht zuletzt hat sich im vergangenen Jahr die Suche nach Informationen im Internet massiv gewandelt. Durch das Zusammenwachsen von KI-Chatbots und klassischen Search Engines, sei es in Form von Suchmaschinen, die – wie Google mit AI Overviews oder Bing mit Copilot Search – KI-Anwendungen in ihre Suche integrieren, oder in Form von KI-Diensten wie Perplexity oder ChatGPT, die wie Suchmaschinen verwendet werden können, da sie auf offene Internetinhalte zugreifen. Wenn sich noch jemand demütig auf die Suche nach belegten Quellen macht, wird er oder sie auf dem Weg jedenfalls permanent der Versuchung widerstehen müssen, KI-Inhalte ungeprüft zu übernehmen und muss stattdessen die Mühen auf sich nehmen, in den Tiefen des Netzes oder hinter Bezahlschranken die tatsächlichen Originale zwecks Re- und Double-Check aufzufinden.

Die Medienanstalten haben ein Gutachten beauftragt, das sich mit der Frage beschäftigt, wie diese Dienste medienrechtlich und regulatorisch einzuordnen sind (Oster/Busch, Integration von KI-Anwendungen in Suchmaschinen – eine medienrechtliche und regulatorische Einordnung, 2026). Ein Fokus dieses Rechtsgutachtens wird auf der Haftung für falsche oder rechtswidrige Inhalte liegen. Dass es sich bei den Ergebnissen einer KI-Suche um „eigene“, den jeweiligen KI-Diensten zuzuordnende Informationsobjekte handelt, wurde bereits im abgeschlossenen ersten Teil des Gutachtens konstatiert (Lewandowski, Integration von KI-Anwendungen in Suchmaschinen und ihre Auswirkungen auf die Meinungsvielfalt, 2025).

Und aktuell stehen wir alle vor einer neuen sprunghaften Entwicklung, die selbst manchem Tech-Insider den Atem stocken lässt: der Integration von Agentic AI in nahezu alle vorstellbaren Prozesse. Quasi wie am Fließband werden hochkomplexe Aufgaben gelöst, indem verschiedene Prompts und Rechenergebnisse aufeinander aufbauend abgearbeitet und an andere Dienste übergeben werden, orchestriert durch selbsttätige spezialisierte Agenten. Die Rückkopplung mit dem User erfolgt über Kommunikationsschnittstellen, zB Messengerdiensten, oder in Form eines allgegenwärtigen und selbstlernenden persönlichen Assistenten. Einmal konfiguriert, arbeiten diese Dienste rund um die Uhr autonom. Sie können Daten auswerten, graphisch aufarbeiten, in zielgruppengerechte Beiträge fassen und diese telefonisch oder via Mail weitergeben, Geld transferieren – die Anwendungsszenarien sind grenzenlos. Open Claw, Manus oder Kimi Agent heißen die Bot-Tools, die in den letzten Monaten für Furore gesorgt haben. Die Möglichkeiten und Risiken liegen auf der Hand – manche reden „von der gefährlichsten Software der Welt“ (Janssen, OpenClaw ausprobiert: Die gefährlichste Software der Welt?, heise online v. 31.1.2026, abrufbar unter: <https://www.heise.de/news/OpenClaw-ausprobiert-Die-gefaehrlichste-Software-der-Welt-11161203.html>), andere sprechen Open Claw gar eine Art Br... sein

zu und nennen es einen „Geist, der vor dem Computer sitzt“ (Knees, Entwickler nennt Open Claw einen „Geist, der vor dem Computer sitzt“, Handelsblatt online v. 13.2.2026, abrufbar unter: <https://www.handelsblatt.com/technik/ki/ki-entwickler-nennt-open-claw-einen-geist-der-vor-dem-computer-sitzt/100197905.html>).

Hinzu kommt, dass solche KI-Systeme mittels agentenbasierter Entwicklungsplattformen wie Claude Code oder Google Antigravity vergleichsweise leicht selbst gebaut werden können. Der bereits zu OpenAI gewechselte Österreicher Peter Steinberger hat das geleistet und viele werden es ihm nachtun.

Den Impact, den KI auf die Reichweite und Wirkkraft maschinell erzeugter Inhalte und damit auf Medienfreiheit und Meinungsvielfalt hat, werden diese neuen Funktionalitäten fraglos potenzieren; und somit auch die Herausforderung, Verantwortung, Vielfalt und Vertrauen in Medien zu sichern. Soweit die kurze Standortbeschreibung.

III. Wir können KI nicht nicht regulieren

Während KI-Systeme und -Modelle rasant in unseren Alltag, die Arbeitswelt und in Medien eingreifen, versucht die Gesetzgebung, mit neuen Regularien nachzusteuern, um Risiken zu minimieren, ohne die Innovation zu ersticken. Das ist erkennbar mühsam, aber unverzichtbar.

Da wir nicht wollen können, dass KI volatile Biases oder gar antrainierte demokratiefeindliche Instruktionen reproduziert und gezielt an ausgewählte Personengruppen adressiert, ist der Weg, den Europa mit der KI-VO iVm DSA, EMFA, Data Act oder DMA geht, uneingeschränkt zu befürworten. Für den Medienbereich bilden KI-VO, MStV und JMStV sowie die jeweiligen Landesmediengesetze den maßgeblichen Rechtsrahmen, um auch unter Verwendung von KI die Medienfreiheit und den Vielfaltsgedanken zu achten und gleichzeitig die Schranken der Persönlichkeitsrechte, des Jugendmedienschutzes und der allgemeinen Gesetze zu wahren. Diesen Schutz mit Zensur gleichzusetzen, wird auch durch permanente Wiederholung – namentlich von dort, wo die großen Digitalkonzerne Milliarden verdienen und verfassungsseits der Meinungsfreiheit quasi keine Grenzen gesetzt sind – nicht zutreffender.

Längst erscheinen diese Konzerne und ihre Verflechtungen „too big to fail“ und ihre Produkte immer unverzichtbarer. Diese werden gelauncht und skrupellos über die Welt verteilt, obwohl ihre nicht abschätzbaren Auswirkungen absehbar disruptiv sind und auch wenn sie gegen geltendes Recht verstoßen. Das Zeitfenster für die Umsetzung einer europarechtskonformen Regulierung schließt sich gerade genauso schnell, wie das für die Entwicklung konkurrenzfähiger europäischer Produkte. Transparenzpflichten, Konformitätsprüfungen und Verbote müssen daher nun mutig und zügig umgesetzt und dürfen nicht in Bussen geparkt werden, bis der Zug endgültig abgefahren ist.

IV. Potenziale und Grenzen von Kennzeichnungspflicht & Co.

Medien- und KI-Regulierung erfolgt grundsätzlich ex post, also erst dann, wenn veröffentlichte Inhalte bestehende Gesetze augenscheinlich verletzt haben. In Zusammenhang mit dem Einsatz von KI im Medienprivileg steht Art. 50 KI-VO mit seinen Transparenzvorgaben im Mittelpunkt. Gerade wird der Code of Practice finalisiert, der die gesetzlichen Anforderungen an eine Kennzeichnung von KI-generierten oder-manipulierten Inhalten sowie Deepfakes näher konkretisiert. So wichtig diese Transparenz für Vertrauen und Kontrollprozesse ist, so klar ist auch bereits jetzt, dass sie von den Usern voraussichtlich genauso mäßig goutiert werden, wie AGBs oder Cookie-Richtlinien. Von den Kontrollinstanzen jedoch werden Verstöße in jahrelangen Prozessen widerlegt werden müssen.

Da KI-Output nur bedingt reproduzierbar und daher auch schwer kontrollierbar ist, erscheint eine ex-ante Regulierung zielführend. So ist es zB notwendig, KI-Praktiken, die ein inakzeptables Risiko für die Grundrechte darstellen, bereits vor deren Veröffentlichung zu verbieten. Für KI-Systeme, die mit einem hohen Risiko verbunden sind, gelten spezifische Anforderungen, etwa, dass sich Anbieter für diese Systeme vorab einer Konformitätsbewertung unterziehen müssen (Art. 43 KI-VO). General Purpose AI – und damit wohl für die oben beschriebenen mächtigen neuen Anwendungen werden vom KI-Büro der EU-Kommission überwacht (Art. 75 Abs. 1 KI-VO). Eine Herausforderung, die weit mehr brauchen wird als Praxisleitfäden.

V. Eine Verfassung für KI?

Wenn mit den neueren Entwicklungen Entitäten entstehen, die mehr sind als abgrenzbare Dienste, braucht es mehr als Gesetze, Richtlinien und Kontrollen. In den Ethik-Sektionen der Digital-Konzerne entstehen sog. „Fair-Use“-Doktrinen. Bei Anthropic ist die Philosophin Amanda Askell für den Aufbau eines moralischen Rahmens für Claude verantwortlich und der Mitgründer und CEO des Unternehmens, Dario Amodei, warnt vor den Potenzialen dessen, was er da gerade baut (Killian, Er wäre gern einer der Guten, DIE ZEIT v. 14.1.2026, abrufbar unter: <https://www.zeit.de/2026/03/dario-amodei-anthropic-openai-kuenstliche-intelligenz-claude>). Inzwischen hat man Claude eine umfängliche Verfassung verpasst (Anthropic, Claude's new constitution, Ankündigung v. 22.1.2026, abrufbar unter: www.anthropic.com/news/claude-new-constitution), die jedoch schon bei oberflächlicher Analyse den Verdacht nährt, von einer mäßig gebabten KI geschrieben worden zu sein.

Dennoch besticht die Einfachheit des Gedankens. Müssten nicht alle KIs zuvorderst und bedingungslos auf die Verfassungen und die Gesetze ihrer Staaten trainiert sein? Anbieter und Betreiber müssten sich von der Konzeption bis zur täglichen Nutzung den rechtskonformen Einsatz und die Erklärbarkeit ihrer Systeme und Modelle obligatorisch zur obersten Maßgabe machen. Medienrechtlich wären etwa Diskriminierungsfreiheit, Quellenschutz oder Sorgfalt hinsichtlich Jugendschutz und Desinformation zu beachten. Wie man die Einhaltung dieser Normen überprüfen könnte, wird derzeit in einem gemeinsamen Forschungsprojekt der Landesmedienanstalt Saarland und des DFKI untersucht (PABeLA – Prüflabor zur Automatisierten Bewertung LLM basierter KI-Systeme). Ziel muss sein, dass generierter Content per se geltendem Recht genügt, höchste Reversibilität gegen Prompt Injection aufweist und Trainings, AGBs und Community-Richtlinien vor geltendem Recht genauso wenig Anwendungsvorrang haben, wie selbsterlernte Biases der KI.

Wenn wir uns einig sind, dass die Währung einer jeden Demokratie Vertrauen ist, dann umfasst dies mehr als je zuvor auch das Vertrauen in KI. Die Verfasstheit von KI-Systemen muss sich schon deshalb streng an der Verfasstheit unseres Rechtsstaats und der der EU orientieren. Und das 24/7.

KI kann Masse, KI kann Analyse, KI kann Kontext. KI kann daher kontextbezogen lernen, vorhersagen und korrigieren. Der oberste Prompt in unserem Rechtsraum muss demnach das geltende Recht und die Maßgabe sein, sich hiernach zu richten. Wenn KI, die in Europa zum Einsatz kommt, so trainiert würde, hätte sie deutlich mehr Vertrauen verdient.