dfv Mediengruppe

KI nutzen, aber rechtssicher

KI-Woche 2024, 14. Mai 2024, Session 2

Dr. Anja Keller und Andreas Stock

KI nutzen, aber rechtssicher

01

Urheberrecht und andere Schutzrechte

02

Datenschutz und Vertraulichkeit

03

Ausblick: Kennzeichnungspflichten nach dem AI-Act

04

Fragen

Urheberrecht und andere Schutzrechte

dfv Mediengruppe

Urheberrecht und andere Schutzrechte

Grundlage: Was schützt das Urheberrecht?

Kann der Output von KI die Rechte von Dritten verletzen?

- Verletzung von Urheberrechten
- Verletzung von Marken
- Verletzung von Persönlichkeitsrechten

Wie ist der Output der KI geschützt?

- Rechte am Output der KI
- Rechte an den Prompts

Grundlage: Was schützt das Urheberrecht?

Das Urheberrecht schützt "**Werke**". Werke sind "**persönliche geistige Schöpfungen**" (§ 2 UrhG).

Voraussetzungen:

Schöpferische Leistung eines **Menschen**

- von Tieren oder Maschinen / Algorithmen Geschaffenes ist nicht geschützt.
- Ausnahme: Agieren nur als Werkzeug

Erreichen einer gewissen **Gestaltungshöhe**

- Gewisses Maß an Individualität und geistiger Leistung erforderlich

Für menschliche Sinne **wahrnehmbare Form**

- Kein Schutz von bloßen Ideen/Gedanken, nur deren konkrete Umsetzung

Erfüllt eine Leistung diese Voraussetzungen nicht, kann kein urheberrechtlicher Schutz bestehen. Sie ist "**gemeinfrei**".

"Der Output der KI ist rechtefrei und kann ohne Bedenken genutzt werden“

dfv Mediengruppe

Kann der Output von KI die Rechte von Dritten verletzen?

Urheberrechte

Markenrechte

Persönlichkeitsrechte

Kann der Output von KI die Rechte von Dritten verletzen?

Urheberrechte

Output auf Grundlage urheberrechtlich geschützter Werke erstellt

KI wurde mit urheberrechtlich geschützten Werken (Texten, Bildern, Grafiken, etc.) trainiert → Wir können nicht erkennen, womit und wie nah der Output an diesen Trainingsdaten ist.

- **Texte**: Gefahr der Übernahme ganzer Textstellen (11 Wörter können ausreichen)
- **Bilder**: Gefahr der Übereinstimmung oder zu enger Anlehnung an existierendes Bild

ⓘ Keine Identität erforderlich, enge Anlehnung ausreichend! → Abgrenzung zwischen (zulässiger) Anlehnung/Inspiration und (zustimmungsbedürftiger) Bearbeitung schwierig.

Kann der Output von KI die Rechte von Dritten verletzen?

Urheberrechte

Wer haftet dafür, wenn der Output der KI Urheberrechte verletzt?

- Wir nutzen die generierten Inhalte: Wir haften verschuldensunabhängig auf Unterlassung
- Schadensersatz: Vorsätzliches oder fahrlässiges Handeln erforderlich.

ⓘ Rechtlich noch ungeklärt, wahrscheinlich: Ungeprüftes Übernehmen der Inhalte stellt eigene Sorgfaltspflichtverletzung dar, sodass wir auch auf Schadensersatz haften.

Muss uns der KI-Anbieter (z.B. OpenAI) den entstehenden Schaden ersetzen?

- Haftung in den Nutzungsbedingungen ausgeschlossen → Pflicht zur Sicherstellung, dass Inhalte keine Rechte verletzen, wird Nutzern auferlegt.
- Haftung von OpenAI gegenüber dem Urheber aber möglich.

Kann der Output von KI die Rechte von Dritten verletzen?

Urheberrechte

Was können wir tun, um Risiken zu minimieren?

- „Provozieren“ von Urheberrechtsverletzungen vermeiden, z.B.
  - **Textgenerierung**: Keine Prompts, die darauf angelegt sind, einen bestimmten (fremden) Text aufzugreifen oder sich an diesen anzulehnen
  - **Bildgenerierung**: Keine Prompts, die darauf angelegt sind, ein bestimmtes Bild oder eine berühmte Figur nachzuahmen (Figurenschutz). Imitierung von „Stilen“ grds. zulässig, bringt aber erhöhtes Risiko.
- Bei fremden Texten, die von der KI gezielt umgeschrieben oder zusammengefasst werden sollen: Darauf achten, dass keine zusammenhängenden Passagen übernommen wurden (11 Wörter!)
  - ⓘ Theoretisch kann in diesen Fällen schon der Input rechtswidrig sein, da Vervielfältigung ohne Lizenz.
- Bei Bedenken / Zweifeln: Weitestmöglich ausräumen (Google-Bildersuche, Suche nach Textteilen, Plagiatssoftware)

Kann der Output von KI die Rechte von Dritten verletzen?

Markenrechte

Output enthält Marken von Dritten

- **Redaktionelle Nutzung**: Kaum Risiken, da keine „markenmäßige Benutzung“
- **Nutzung zur Kennzeichnung von eigenen Angeboten**: Nutzung eines identischen oder verwechslungsfähig ähnlichen Zeichens kann Markenrechtsverletzung darstellen.

  ⓘ Wie immer bei neuen Titeln / Logos für Angebote ist vorab zu prüfen, dass Dritten keine älteren Rechte hieran zustehen.

Kann der Output von KI die Rechte von Dritten verletzen?

Persönlichkeitsrechte

Output verletzt Persönlichkeitsrechte Dritter

Veröffentlichen wir KI-generierte Inhalte, gilt nichts anderes als bei der Veröffentlichung von Inhalten menschlicher Urheber. Insbesondere ist daher auch hier vor Veröffentlichung insbesondere sicherzustellen, dass

- veröffentlichte Inhalte allen **presserechtlichen Anforderungen** genügen (nur korrekte Tatsachenbehauptungen, Einhalten der Voraussetzungen der Verdachtsberichterstattung, Zulässigkeit einer Namensnennung, etc.)
- Bilder von Personen nur in Übereinstimmung mit den gesetzlichen Vorgaben zum **Recht am eigenen Bild** veröffentlicht werden

Als Verlag haften wir für Verletzungen des Persönlichkeitsrechts in den von uns publizierten Angeboten und können uns nicht darauf zurückziehen, dass Inhalte KI-geniert sind.

Wie ist der Output der KI geschützt?

Wie ist der Output der KI geschützt?

Grundsätzlich kein urheberrechtlicher Schutz

Urheberrechtlicher Schutz setzt Schöpfung eines Menschen voraus: Texte und Bilder und andere Inhalte (mit Hilfe von KI generierter Sourcecode, etc.) können von jedermann frei kopiert und verwendet werden.

- Dritter können sich frei an Inhalten bedienen, diese übernehmen und auch kommerziell verwerten.
- Mangels eigener Rechte kann Dritten auch keine Lizenz erteilen.
- Inhalte dürfen nicht an die VG Wort oder andere Verwertungsgesellschaften gemeldet werden.

Wie ist der Output der KI geschützt?

dfv Mediengruppe

Wie ist der Output der KI geschützt?

Wann kann ausnahmsweise urheberrechtlicher Schutz bestehen

- Output wurde durch einen Menschen umfangreich überarbeitet, sodass ein eigenes Werk entstanden ist
- KI als Werkzeug:
  - Prompt so detailliert, dass nur ein Ergebnis dabei herauskommen kann: Kaum der Fall, da immer gewisser Gestaltungsspielraum der KI besteht.
  - Schrittweises Prompting (eigene kreative Leistung durch schrittweise Veränderung des Outputs): Urheberrechtlicher Schutz wohl denkbar, aber ohne Dokumentation schwer nachweisbar.

Datenschutz und Vertraulichkeit

Vorsicht
bei Input, Output und
bei der Wahl der KI des Vertrauens

dfv Mediengruppe

Agenda

- Allgemeines zu KI, Datenschutz und Vertraulichkeit
- Datenschutz bei Eingabe von Daten in KI-Anwendungen
- Datenschutz bei Verwendung KI-generierter Inhalte

Allgemeines

- Datenschutz hat 2 Komponenten:
  - Schutz der **Menschen** davor, dass sie durch eine Datenverarbeitung in ihren Rechten verletzt werden (Datenschutz).
  - Schutz der **Daten** selbst vor negativen Einflüssen wie Manipulation, Offenlegung gegenüber Unbefugten oder Löschung (Datensicherheit).
- DSGVO-Vorgaben müssen in der EU beachtet werden, auch wenn KI-Anbieter in anderen Ländern sitzen. EU-Datenschutz gilt für alle, die **in der EU** ihren Firmensitz haben und für alle **außerhalb der EU**, die gezielt Daten von EU-Bürgern verarbeiten.
- Auch **Daten ohne Bezug zu Menschen** können rechtlich relevanten Vertraulichkeitsverpflichtungen unterliegen, z.B. wegen gesetzlicher Vorgaben (Geheimnisschutz) oder vertraglicher Vereinbarungen (NDA).

KI-Anbieter vs. DSGVO

Konfliktpotential besteht insbesondere in folgenden Bereichen:

- Grundsatz der **Transparenz**: Woher kommen die generierten Daten? Was passiert mit eingegebenen Daten?
- Grundsatz der **Datenminimierung**: Nur so viel wie nötig, nach Ende der Erforderlichkeit wird gelöscht.
- **Speicherdauer**: Wann werden Daten gelöscht?
- Erforderlichkeit einer **Rechtsgrundlage**: Einwilligung? Vertrag? Überwiegende berechtigte Interessen?
- **Betroffenenrechte**:
  - Auskunftsanspruch,
  - Anspruch auf Berichtigung und Löschung falscher Daten,
  - Löschungsanspruch.
- Als Verwender einschlägiger Programme sind wir für die Einhaltung der Rechtsvorgaben **verantwortlich**.

Datenschutz-Beschwerde gegen OpenAI

ChatGPT verstößt gegen DSGVO-Grundsätze und setzt Betroffenenrechte nicht um. - Meldung vom 29.04.2024

Eine Person des öffentlichen Lebens fragt ChatGPT nach seinem Geburtsdatum und erhält mehrfach falsche Antworten. Hinweise, dass die Antworten nur vermutet sind und falsch sein können, fehlen. Genauso wie die Möglichkeit, falsche Angaben zu korrigieren oder zu löschen. Angaben zur Datenherkunft kann der KI-Anbieter nicht beantworten. Daher hat der Betroffene Beschwerde bei der Datenschutzaufsicht in Österreich eingereicht.

Auswahl von KI-Anbietern

Worauf muss bei der Auswahl von KI-Anbietern geachtet werden?

Es gelten dieselben Grundsätze, wie bei der Auswahl von anderen Software-Anbietern, es muss ein **seriöses Angebot** sein:

- **Wo** ist der Unternehmenssitz, in welchem Land werden die Daten verarbeitet?

  → Im besten Fall in der EU. Falls nicht:

- Kann die Datenübermittlung in Länder außerhalb der EU **rechtlich abgesichert** werden?

  → Absicherung entweder durch ein Länder-Abkommen zu angemessenem Schutzniveau oder durch spezielle Verträge (Standardvertragsklauseln).

- Können mit dem Anbieter **Verträge** zum Datenschutz geschlossen werden?

  → Vertrag zur Auftragsverarbeitung (AVV) oder Vertraulichkeitsvereinbarung (NDA)?

- Ist die Datenverarbeitung **transparent** und datenschutzkonform nach EU-Recht? Werden Daten an Dritte weitergegeben?

  → Informiert der Anbieter ausreichend über Datenverarbeitungsvorgänge und bietet er Möglichkeiten, eingegebene Daten isoliert zu verarbeiten und wieder zu löschen?

Auswahl von KI-Anbietern

Die DSK gibt Orientierung im KI-Dschungel

Aktuelle Orientierungshilfe der Datenschutzkonferenz (DSK) vom 6. Mai 2024

- Konzeption des Einsatzes und Auswahl von KI-Anwendungen bestimmen.
- dfv-intern: IT, Rechtsabteilung und Datenschutz frühzeitig einbinden.
- Einsatzfelder und Zwecke festlegen: Rechtmäßiges Verarbeiten möglich, ggf. ohne personenbezogene Daten?
- Datenschutzkonformes Training von KI-Anwendungen; geschlossenes oder offenes System?
- Transparenz und Wahlmöglichkeit bei KI-Training und Eingabe-Historie?
- Umsetzung von Betroffenenrechten: Berichtigung, Löschung und Auskunftserteilung möglich?

Datenschutz bei Eingabe

- Wer Daten in ein System eingibt, muss damit rechnen, dass diese Daten an Dritte weitergegeben werden.
- Daher sollte weiterhin **auf die Eingabe sensibler Daten verzichtet werden**.
- **Sensible Daten** i.S.d. DSGVO sind u.a. personenbezogener Daten, aus denen die
  - ethnische Herkunft,
  - politische Meinungen,
  - religiöse oder weltanschauliche Überzeugungen oder
  - die Gewerkschaftszugehörigkeit hervorgehen,
  - sowie die Verarbeitung von biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person,
  - Gesundheitsdaten einer natürlichen Person.
- Bei diesen Daten muss im Zweifel immer eine **Einwilligung** zur Verarbeitung eingeholt werden.

Datenschutz/Vertraulichkeit bei Eingabe

Sensible Daten außerhalb der DSGVO sind u.a. Daten mit oder ohne Personenbezug, die bei unbefugter Verwendung ein **hohes Schadenspotential** bieten:

- Passwörter,
- Kreditkartendaten,
- sonstige Finanzdaten wie z.B. IBAN,
- Angaben, die einer Vertraulichkeitsvereinbarung unterliegen,
- Daten, die dem Geheimnisschutz oder sonstigen vertraglichen oder gesetzlichen Vertraulichkeitsverpflichtungen unterliegen.

Datenpanne bei ChatGPT

User können fremde Chatverläufe einsehen - Meldung vom 24.03.2023

Nutzer von ChatGPT bemerkten, dass in ihren Chatverläufen zwar frühere Unterhaltungen angezeigt wurden. Diese gehörten jedoch nicht zu ihren eigenen und stammten offenbar von anderen Nutzern.

OpenAI-CEO Sam Altman hat die Datenpanne bei Twitter/X bestätigt. Demnach habe ein „kleiner Prozentsatz“ der User den Chatverlauf anderer einsehen können. Dies sei ein „bedeutendes Problem“ gewesen, das aufgrund eines Fehlers in einer Open-Source-Bibliothek aufgetreten sei.

Datenpanne bei Bauhaus

Bing-KI plaudert detaillierte Kundendaten aus - Meldung vom 24.11.2023

Durch den KI-Chatbot von Bing haben die personenbezogenen Informationen ein Eigenleben entwickelt. Wer bei Bing wusste, wonach er suchen muss, konnte Bauhaus-Rechnungen seit 2021 und darin Adressen, Telefonnummern und Rechnungen einsehen. Die Informationen waren im Zwischenspeicher (Cache) des KI-Chats gelandet und dadurch weiter abrufbar. Über den Deep Link – also die genaue Verknüpfung auf die Unterseite einer Website – sind sie immer noch auffindbar. Das lässt darauf schließen, dass Microsoft den Cache von Bing nicht gelöscht, sondern die fraglichen Datensätze nur gesperrt hat.

dfv Mediengruppe

Datenschutz bei Ausgabe

Daten, die KI-Anwendungen ausgeben, müssen kritisch geprüft werden.

- Unklar ist meist, aus welcher **Datenquelle** sie stammen.
- Bei Daten mit Personenbezug ist selten klar, ob sie mit Zustimmung der Personen im Netz gelandet sind oder ob die **Einwilligung** in eine KI-Verarbeitung gegeben wurde.
- Vorsicht ist bei allen **Medienformen** angebracht: Sowohl generierte Texte als auch Bilder oder Videos können Angaben enthalten, die ohne Legitimation verarbeitet wurden.
- KI-generierte Texte müssen immer auf **Wahrheitsgehalt** geprüft werden: KI-Anwendungen sind nicht darauf programmiert, Faktenwissen auszugeben (→ Grundsatz der Richtigkeit personenbezogener Daten).

KI-generierte Inhalte

Auch **KI-generierte Inhalte** können unter die DSGVO fallen.

Werden **Fotos von real existierenden Personen** generiert, greift die DSGVO.

Vorsicht, das kann auch der Fall sein, wenn **Fotos von unbekannten Personen** generiert werden, die tatsächlich existieren.

→ Immer prüfen, was die KI ausgibt, egal ob Text- oder Bildinhalte. Die KI wurde mit „echten“ Fotos trainiert und verwendet diese für neue Kreationen.

Private Bilder aus Krankenakte in Datenbank für KI-Trainings gefunden

Eine Betroffene hat Bilder aus ihrer privaten Krankenakte in einer Bilddatenbank gefunden, die zum Training von KI-Modellen dient. - Meldung vom 23.09.2022

Eine Betroffene stellte fest, dass private Bilder aus ihrer Krankenakte in einem KI-Trainingssatz von LAION verwendet werden. Mit solchen Daten trainieren Text-zu-Bild-Generatoren wie Stable Diffusion und DALL-E. Bei einer Recherche über das Portal https://haveibeentrained.com, über das jeder die Trainingsdatenbank selbst durchsuchen kann, fand die Betroffene Bilder, die ihr Arzt 2013 in ihrer Krankenakte hinterlegt hatte.

KI verleumdet Unschuldigen als Kinderschänder

MSN berichtete über einen Missbrauchs-Prozess gegen einen Radiomoderator, samt Foto eines unbeteiligten Kollegen. - Meldung vom 16.01.2024

Schwere Vorwürfe wegen Verleumdung erhebt der irische Radiomoderator Dave Fanning gegen den Hongkonger Verlag BNN sowie Microsoft. Ein Bericht über einen Strafprozess gegen einen irischen Radiomoderator wegen sexuellen Missbrauchs nannte zwar nicht den Namen des Beschuldigten, zeigte aber ein Foto Fannings. Der hatte mit dem Prozess nichts zu tun. Künstliche Intelligenz soll das Bild ausgewählt haben.

Ausblick: Kennzeichnungs-pflichten nach dem AI-Act

dfv Mediengruppe

Kennzeichnungspflichten nach dem AI-Act

AI Act (KI-Verordnung der EU)

Der AI-Act wurde am 13. März 2024 vom Europaparlament gebilligt. Als eines der ersten Gesetze weltweit soll es den Einsatz künstlicher Intelligenz regulieren.

- Richtet sich vor allem an Anbieter von KI-Systemen: Einteilung in verschiedene Risikoklassen ein; Verbot unakzeptabel riskanter KI-Systeme, Regulierung sonstiger KI-Systeme).
- AI-Act enthält aber auch einige Regelungen, die sich an Nutzer von KI richten. Unter anderem werden Kennzeichnungspflichten für KI-generierte Inhalte (erstmals) geregelt.
- Im Detail noch vieles unklar und auslegungsbedürftig.
- Regelung zur Kennzeichnungspflichten treten erst **Mitte 2026 in Kraft**.

Kennzeichnungspflichten nach dem AI-Act

Welche Kennzeichnungspflichten werden bestehen?

Kennzeichnung von Texten, die mithilfe von KI generiert oder verändert wurden

- Kennzeichnungspflicht nur, wenn **Texte zum Zwecke der Information der Öffentlichkeit über Angelegenheiten von öffentlichem Interesse** publiziert werden

  „[…] text which is published with the purpose of informing the public on matters of public interest […]".

  - Schon Grundvoraussetzung auslegungsbedürftig, Nachrichten- /Wirtschaftsberichterstattung wohl erfasst.
- Selbst wenn Grundvoraussetzung vorliegt: Kennzeichnung entfällt, wenn
  - Der Text vor Veröffentlichung einer **menschlichen / redaktionellen Prüfung** unterzogen wurde; **und**
  - jemand die ***Verantwortung* für den Inhalt** trägt.

Kennzeichnungspflichten nach dem AI-Act

Welche Kennzeichnungspflichten werden bestehen?

Kennzeichnungspflicht für „Deepfakes“

- Pflicht zur Kennzeichnung von Bild-, Audio- und Videoveröffentlichungen, die KI-generiert oder KI-verändert sind, aber dennoch authentisch erscheinen (sog. Deep-Fakes).
- Gesetzliche Definition von Deep-Fakes:

  „AI generated or manipulated image, audio or video content that
  resembles existing persons, objects, places or other entities or events
  and would falsely appear to a person to be authentic or truthful“

- Weiter, als man beim Thema „Deep-Fakes“ erwartet: Nicht nur die Nachahmung echter Personen, sondern auch Nachahmung z.B. echter „Objekte“ und „Orte“.

Fragen

dfv Mediengruppe

Vielen Dank für Ihr/Euer Interesse

- Dr. Anja Keller (anja.keller@dfv.de)
- Andreas Stock (andreas.stock@dfv.de)

#Qualitätsmacher:innen

dfv Mediengruppe