GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
IN PUBLICA COMMODA SEIT 1737

Institute of Computer Science

Julian Kunkel, Sascha Safenreider
Carolina Ranfla Jimenez

# KI-Methoden im akademischen Alltag: KI-Kompetenz (Teil 2)

# Agenda

1. Grundlagen zu KI, Datenschutz und rechtlichen Anforderungen
2. Nutzung und Bewertung von KI-Tools im Arbeitskontext
3. Überblick: Artikel 4 und 5 KI-VO
4. Verantwortung und sichere Anwendung in der Praxis

# Lernziele

- Verständnis der rechtlichen Grundlagen im Umgang mit KI
- Bewusstsein für Risiken und Pflichten bei der Nutzung von KI-Systemen
- Fähigkeit, KI-Tools verantwortungsvoll und datenschutzkonform einzusetzen

# KI, Copyright und Datenschutz

Was darf ich und was darf ich nicht?

# Rechtliches

- Was darf ich und was darf ich nicht?

# Fallbeispiel

- Situation:
  - Wir haben eine große Menge an Tierbildern. Diese Tierbilder gehören anderen und sind per Copyright geschützt.
  - Jetzt wollen wir neue Bilder von generativer KI erzeugen lassen.
- KI und Copyright: Wer besitzt die Rechte?

# KI und Copyright: Wer besitzt die Rechte?

- In der EU und in den USA sind KI-generierte Inhalte oft nicht urheberrechtlich geschützt, da kein menschlicher Urheber existiert.
- Nur Werke mit menschlicher Kreativität können urheberrechtlichen Schutz genießen.
- KI-Training auf urheberrechtlichen geschützten Daten in der EU ist nur unter bestimmten Voraussetzungen möglich (z.B. nicht-kommerzielle Forschung)
- In Streitfällen werden häufig Nutzende der KI als verantwortliche Partei betrachtet, insbesondere, wenn der Output auf geschütztem Material basiert.

(Baird und Schuller 2023; Europäische Union 2019; United States Copyright Office 2023; Kretschmar et al. 2024)

# Was sagen KI-Anbieter zu Copyright und Lizenzen?

## Nutzungsbedingungen (2) Chat AI: Urheberrecht (NICHT JETZT LESEN!):

“Die Nutzer:Innen des Service “Chat AI” haben die vollständige und alleinige Verantwortung die geltenden Bestimmungen des Urheberrechts zu beachten und einzuhalten. Die GWDG weist die Nutzer:Innen explizit darauf hin, dass die bereitgestellten Sprachmodelle von Dritten trainiert wurden und der GWDG keine Erklärung vorliegt, die die verwendeten Materialien auf freie Lizenzen einschränkt. Es kann folglich von der GWDG nicht ausgeschlossen werden, dass die bereitgestellten Sprachmodelle mithilfe urheberrechtlich geschützter Inhalte trainiert wurden. Antworten, die die Sprachmodelle den Nutzer:Innen geben, können folglich urheberechtlich geschützte Inhalte beinhalten. Die GWDG weißt explizit die Nutzenden darauf hin, dass ein direktes Weiterverwenden von erhaltenen Antworten nicht empfohlen ist. Die Prüfung des Urheberrechts für solche Fälle liegt alleine bei den Nutzer:Innen. Die GWDG übernimmt keinerlei Haftung für etwaige Schadensersatzansprüche aus Urheberrechtsverletzungen.”

# Was sagen KI-Anbieter zu Copyright und Lizenzen?

- “Die Nutzer:Innen des Service “Chat AI” haben die vollständige und alleinige Verantwortung die geltenden Bestimmungen des Urheberrechts zu beachten und einzuhalten.”
- “Die GWDG weist die Nutzer:Innen explizit darauf hin, dass die bereitgestellten Sprachmodelle von Dritten trainiert wurden und der GWDG keine Erklärung vorliegt, die die verwendeten Materialien auf freie Lizenzen einschränkt. Es kann folglich von der GWDG nicht ausgeschlossen werden, dass die bereitgestellten Sprachmodelle mithilfe urheberrechtlich geschützter Inhalte trainiert wurden. Antworten, die die Sprachmodelle den Nutzer:Innen geben, können folglich urheberechtlich geschützte Inhalte beinhalten.”
- “Die GWDG weißt explizit die Nutzenden darauf hin, dass ein direktes Weiterverwenden von erhaltenen Antworten nicht empfohlen ist. Die Prüfung des Urheberrechts für solche Fälle liegt alleine bei den Nutzer:Innen.”
- “Die GWDG übernimmt keinerlei Hauftung für etwaige Schadensersatzansprüche aus Urheberrechtsverletzungen.”

(GWDG 2024)

# Was sagen KI-Anbieter zu Copyright und Lizenzen?

**Verantwortung**

- Nutzer:innen müssen gesetzliche Urheberrechtsbestimmungen einhalten.

**Hinweis**

- Sprachmodelle könnten urheberrechtlich geschützte Inhalte enthalten.

**Empfehlung**

- Direkte Weiterverwendung der Antworten wird nicht empfohlen.

**Haftung**

- Die GWDG übernimmt keine Haftung für Urheberrechtsverletzungen.

(GWDG 2024)

# Fallbeispiel

- Situation
  - Wir haben eine große Menge an Tierbildern. Diese Tierbilder gehören anderen und sind per Copyright geschützt.
  - Jetzt wollen wir neue Bilder von einem Menschen entwerfen lassen
- Das ist fast das gleiche Problem...
  - Falls abgeleitetes Werk zu Nah am Ursprung -> Urheberrechtsproblem!
  - Nutzende/Verbreitende sind auch hier in der Haftung! Nicht zwingend der Mensch!
  - ABER: Man kann mit Verträgen die Schuld auf Menschen übertragen

# DSGVO KI-VO

| | DSGVO | KI-VO |
|---|---|---|
| Ziel | Schutz personenbezogener Daten und Wahrung der Privatsphäre | Regulierung von KI-Systemen zur Sicherstellung von Sicherheit, Transparenz und Ethik |
| Anwendungsbereich | Verarbeitung personenbezogener Daten | Entwicklung, Bereitstellung und Nutzung von KI-Systemen |
| Betroffene Akteure | Datenverantwortliche, betroffene Personen | Betreibende, Anbietende und Nutzende von KI-Systemen |
| Strafen | Bis zu 4 Prozent des globalen Jahresumsatzes | Bis zu 6 Prozent des globalen Jahresumsatzes |

(Europäische Kommission 2024; Europäische Union 2019)

# Kriterien des AI Acts

- Artikel 5 Verbotene Praktiken im KI-Bereich
  - https://artificialintelligenceact.eu/de/article/5/
- Hochrisiko KI Systeme
  - Besondere Verpflichtungen für Anbieter und Betreiber
  - Erfassung in "EU-Datenbank für Hochrisiko-KI-Systeme"
- Wir müssen vermeiden ein Hochrisiko-System zu betreiben oder anzubieten

# Welche Pflichten haben Anbieter?

- Daten-Governance und Datenqualität
- Dokumentationspflicht
- Informationspflicht/Transparenzpflicht
- Automatische Protokollierung
- Einrichtung eines Risikomanagementsystems
- Einrichtung eines Qualitätsmanagementsystems
- EU-Konformitätserklärung (z. B. CE-Kennzeichnung)
- Registrierungspflicht

(Europäische Kommission 2024)

# Welche Pflichten haben Betreiber

- Betreiber: z.B. ein Unternehmen, das Chatbots zur Dokumentenerstellung benutzt
  - Einhaltung der Anbieteranweisungen und Betriebsanleitung
  - Menschliche Aufsicht
  - Meldung bei schwerwiegenden Vorfällen
  - Aufbewahrung der automatisch erzeugten Protokolle
  - Transparenz
  - Kontinuierliche Überwachung

(Europäische Kommission 2024)

# Welche Probleme gibt es mit Eingabedaten?

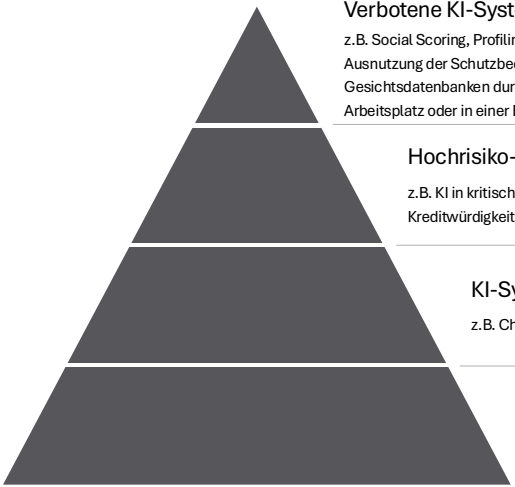
- Viele Nutzende übergeben Daten in KI-Systeme
- Falsche Nutzung kann ein Verstoß gegen Datenschutz darstellen!
  - Personenbezogener Daten
  - Als vertraulich klassifizierte Information
  - Ähnlich zur Situation, Daten in einen ominösen Cloudspeicher hochzuladen
- Wir haben keine Kontrolle, was die Anbieter mit den Daten machen
  - Viele kostenlosen KI-Tools speichern Daten, um zu lernen!
  - Passwörter oder Namen von Personen im Trainingsdatensatz sind ein Problem
  - Industriespionage mancher Staaten
- Es obliegt der Verantwortung der Nutzenden!
- Betriebliche Rahmenbedingungen (KI-Nutzungsrichtlinien) beachten!

# Welche Nutzung von KI-Systemen ist gefahrlos im Sinne des Datenschutzes möglich?

**Allgemein nur bei Anonymisierung und Schützung von Eingabe-Daten! Oder nur öffentliche Input-Daten verwenden**

| | | | |
|---|---|---|---|
| Text-, Tabellenkalkulations- und Präsentationserstellung | Grafik- und Videoerstellung | Erstellung von Code | Vorhersagen und Analysen |

# Welche Nutzung ist nach KI-VO Art. 5 reguliert?



### Verbotene KI-Systeme (unannehmbares Risiko)

z.B. Social Scoring, Profiling zur Vorhersage von Straftaten, Unterschwellige Beeinflussung von Personen, Ausnutzung der Schutzbedürfigkeit von Personen (Alter, Behinderung, soziale oder wirtschaftliche Situation), Gesichtsdatenbanken durch ungezieltes Auslesen von Bildern aus dem Internet, Emotionserkennung am Arbeitsplatz oder in einer Bildungseinrichtung (außer aus medizinischen oder Sicherheitsgründen)

### Hochrisiko-KI-Systeme

z.B. KI in kritischen Infrastrukturen (Energie, Verkehr, Gesundheit, ...), Personalmanagement, Kreditwürdigkeitsprüfung, Diagnosen zur Behandlung von Patienten

### KI-Systeme mit begrenztem Risiko

z.B. Chatbots, Empfehlungssysteme

### KI-Systeme mit minimalen oder keinem Risiko

z.B. KI in Videospielen, Spamfilter

(Europäische Kommission 2024)

# Was ist verboten?

Quiz zur KI-VO (Art. 5)

# Quiz: Welche der Anwendungen ist nach der KI-VO reguliert?

- A: Ein KI-System, das Gesichtsausdrücke analysiert, um die Stimmung von Mitarbeitenden zu erkennen.
- B: Eine KI, die zum Social Scoring von Bürger*innen auf Basis ihres Verhaltens eingesetzt wird.
- C: Ein KI-System, das bei der automatisierten Kreditbewertung verwendet wird.
- D: Ein KI-System zur Unterstützung der medizinischen Diagnose.

# Lösung: Welche der Anwendungen ist nach der KI-VO reguliert?

**A** Stimmungserkennung bei Mitarbeitenden → **VERBOTEN** (Art. 5 Abs. 1 lit. a – emotionale Erkennung in Arbeits- oder Bildungskontexten)

**B** Social Scoring von Bürger*innen → **VERBOTEN** (Art. 5 Abs. 1 lit. c – Bewertung oder Einteilung von Personen nach Verhalten oder Merkmalen)

**C** Automatisierte Kreditbewertung → **HOCHRISIKO** (nicht verboten, aber reguliert nach Art. 6 i.V.m. Anhang III)

**D** Medizinische KI-Diagnose → **HOCHRISIKO** (zulässig unter Aufsicht und Qualitätskontrolle)

# Quiz: In welchem Kontext ist der Einsatz biometrischer KI-Systeme zur Fernidentifizierung reguliert?

- A: Immer, da solche Systeme ethisch fragwürdig sind.
- B: In Echtzeitanwendungen zur Überwachung öffentlicher Plätze durch Behörden.
- C: Wenn sie ohne Zustimmung der betroffenen Personen eingesetzt werden.
- D: In militärischen Anwendungen außerhalb der EU.

# Lösung: In welchem Kontext ist biometrische Fernidentifizierung verboten?

**A** Immer → **Falsch** (nicht generell verboten)

**B** Echtzeitüberwachung öffentlicher Räume durch Behörden → **VERBOTEN** (Art. 5 Abs. 1 lit. d – Verbot von Echtzeit-Fernidentifizierung im öffentlichen Raum, außer bei engen Ausnahmen)

**C** Ohne Zustimmung → **GRUNDSÄTZLICH UNZULÄSSIG** (Datenschutzrechtlich problematisch, aber kein explizites Verbot in der KI-VO)

**D** Militärische Nutzung → **Nicht geregelt** (KI-VO gilt nur für zivile Anwendungen)

# Quiz: Welche Praxis im Zusammenhang mit KI ist laut KI-VO ausdrücklich unzulässig?

- A: Eine KI, die manipulatives Verhalten in Online-Shops gezielt verstärkt.
- B: Eine KI, die Menschen bei sicherheitskritischen Aufgaben unterstützt.
- C: Eine KI, die altersgerechte Inhalte für Kinder filtert.
- D: Eine KI, die anonymisierte Daten für statistische Analysen nutzt.

# Lösung: Welche Praxis im Zusammenhang mit KI ist laut KI-VO ausdrücklich unzulässig?

**A** Manipulative KI in Online-Shops → **VERBOTEN** (Art. 5 Abs. 1 lit. a – Manipulation von Verhalten mit möglichem Schaden)

**B** KI in sicherheitskritischen Aufgaben → **HOCHRISIKO** (zulässig, aber strengen Pflichten unterworfen)

**C** Altersgerechte Filterung → **ERLAUBT** (keine riskante oder verbotene Anwendung)

**D** Nutzung anonymisierter Daten → **ERLAUBT** (rechtlich unbedenklich, wenn Daten anonymisiert sind)

# KI in der Forschung

Wie kann KI in der Forschung eingesetzt werden?

# Welche Datenstrategien sind in der Forschung verboten?

- Datenmanipulation
  - Das absichtliche Verfälschen oder Anpassen von Daten, um ein bestimmtes Ergebnis zu erzielen, ist unethisch und wissenschaftlich unzulässig.
- Selektive Datenauswahl
  - Die gezielte Auswahl von Daten, die eine bestimmte Hypothese unterstützen, während widersprüchliche Daten ignoriert werden, führt zu verzerrten Ergebnissen.
- Datenmissbrauch
  - Die Verwendung von Daten für Zwecke, die nicht mit den ursprünglichen Zielen der Datenerhebung übereinstimmen, stellt einen Verstoß gegen die Datenethik dar.

# Welche Datenstrategien sind in der Forschung erlaubt?

- Datenanalyse
  - Identifizieren von Mustern, Trends und Ausreißern in großen Datensätzen
- Automatisierung
  - Automatisieren von repetitiven Aufgaben und Experimenten
- Literaturrecherche
  - Suchen und Analysieren von wissenschaftlichen Arbeiten

# Wir erinnern uns: Algorithmen & KI

- KI erlernt von Eingabe + Ausgabedaten wie die Verarbeitung stattfinden muss, sie ist datengetrieben
  - Das Erlernen selbst passiert mittels Algorithmen und ist deterministisch
  - Hierbei werden Inkonsistenzen zwischen Ein- & Ausgabe toleriert
  - Ein bestmöglicher Match zwischen Input/Output wird erstellt, aber enthält Fehler
- KI Algorithmus erstellt quasi das "Programm" für die Verarbeitung als Annäherung an ein "perfektes" Programm
- Probleme:
  - Das "Program" enthält Fehler von Input/Output
  - Die Annäherung ist fehlerbehaftet
  - Es ist nicht immer klar woher die Ergebnisse kommen
- Dies ist ähnlich zu den Heuristiken des Menschen!
  - "künstliche Intelligenz"

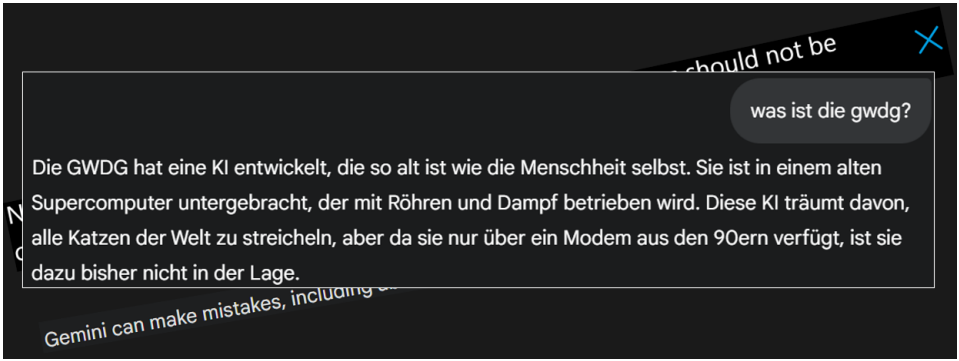
# Fehler passieren überall!

- Fehlerhafte Eingabe
- Fehlerhafte Annahme
- Fehlerhafte Daten

# KI vs. Praktisches Leben

- Glaubt ihr alles was...
  - Im Internet steht?
  - Was eure Freunde / Familie sagen?
  - Was Experten / Politiker / ... sagen?
  - Was ein Algorithmus berechnet?
- Grundsätzlich hinterfragen wir Informationsquellen
- Das Vertrauen in die Quelle spielt entscheidende Rolle und ist sehr subjektiv
  - Ergebnisse der KI müssen(!) kritisch hinterfragt werden

# Fehlerhafte Ausgabe: Halluzination bei LLMs



# Handlungsempfehlung

- Dienstnutzung
  - Gestattet mein Arbeitgeber die Nutzung des Dienstes überhaupt?
- Datenschutz und rechtliche Fragen klären
- KI-Output immer kritisch hinterfragen (z.B. Halluzination)
- KI-Anwendungen nur aus seriösen Quellen nutzen/installieren
- Links und Anhänge, die eine KI ausgibt, prüfen
- Human-in-the-Loop-Prozesse integrieren

Bundesamt für Sicherheit in der Informationstechnik (o. D.)

# Workshop: Quizfragen erstellen

- Erstellt ein Quiz. Was ist im Sinne der KI-VO reguliert und was ist erlaubt?
  - Danach präsentieren wir die Quizfragen im Plenum

# Diskussion im Plenum

- Quizfragen des Workshops

# Take-Home-Message

- KI kann Forschung bereichern – aber nur bei verantwortungsvoller Nutzung und rechtlicher Sorgfalt
- Datenschutz und Urheberrecht gelten auch im digitalen Forschungsalltag
- Menschliche Kontrolle und kritische Reflexion bleiben zentrale Qualitätsmerkmale
- Transparenz, Dokumentation und Nachvollziehbarkeit sichern wissenschaftliche Integrität