

Von Bonbongläsern und Heilsversprechen

Vorschlag eines Ethikframeworks für den Einsatz
von Künstlicher Intelligenz entlang digitaler
archivischer Prozesse

Tony Franzky, 20.05.2025

CC BY-SA 4.0   

Diese Datei wurde unter der Lizenz
„Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen“
in Version 4.0 (abgekürzt „CC-by-sa 4.0“) veröffentlicht.

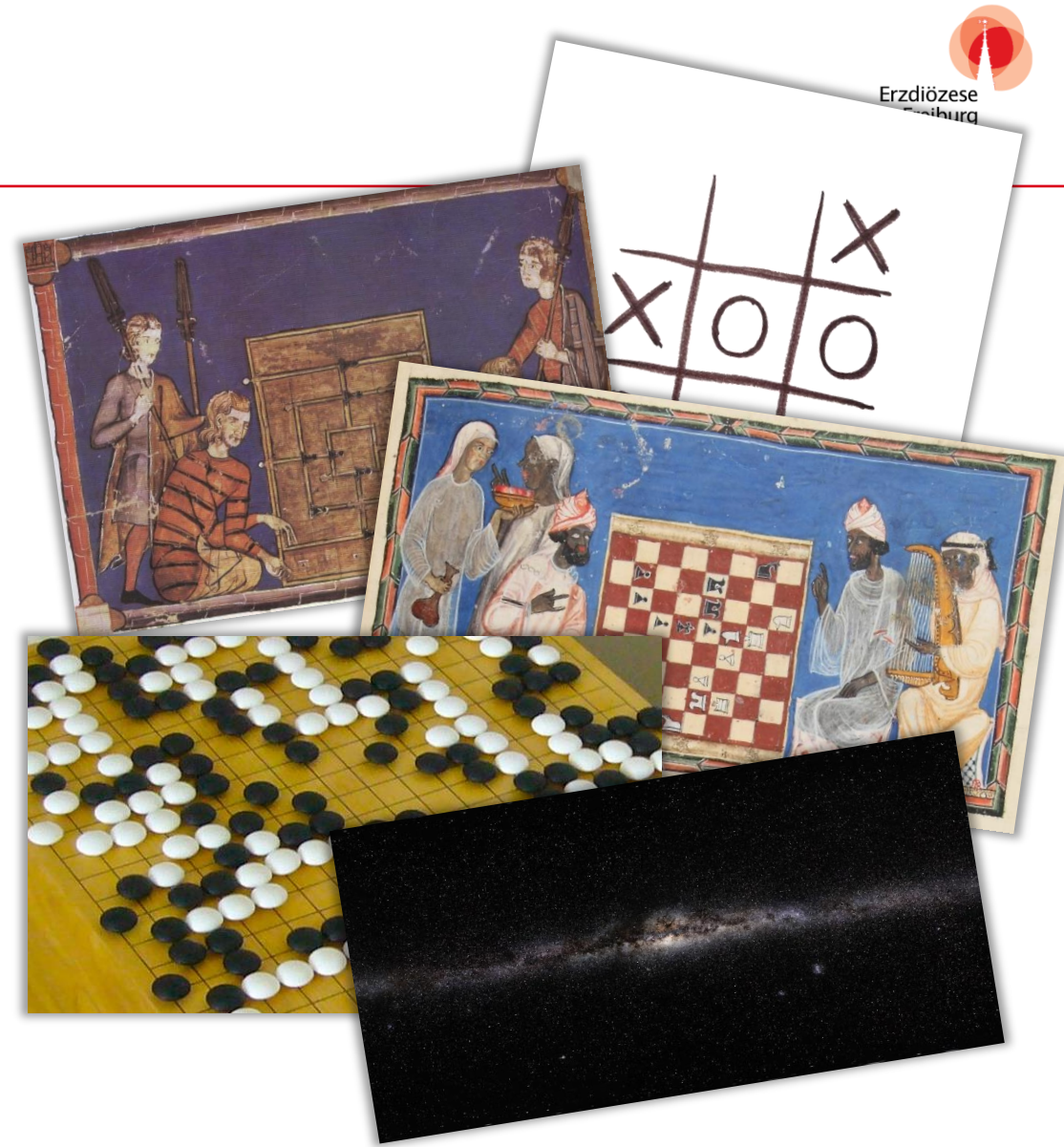
<https://creativecommons.org/licenses/by-sa/4.0/>

- Versprechungen von KI
- Schwierigkeiten und Probleme
 - Eine privatempirische Einführung
- Was macht eine KI ethisch?
- KI im Archiv – Ein Kartierungsversuch
- Werte und Institutionen?
- Was sollte man beachten, damit eine KI „ethisch“ ist?



Warum Künstliche Intelligenz?

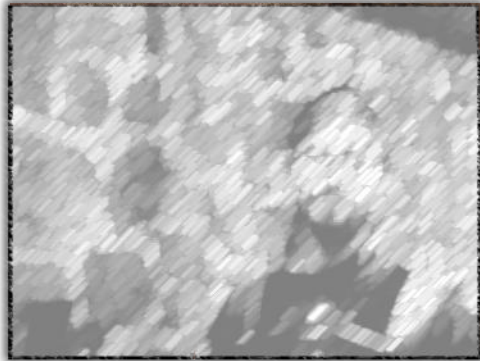
TicTacToe:	255.168 Spielverläufe
Mühle:	$1,8 \times 10^{10}$ Spielverläufe
Vier Gewinnt:	$4,5 \times 10^{12}$ Spielverläufe
Zauberwürfel:	$4,3 \times 10^{19}$ Möglichkeiten
Dame:	5×10^{20} Spielverläufe
Schach:	1×10^{120} Spielverläufe
Go:	2×10^{170} Spielverläufe
Sichtbares Universum:	1×10^{80} Teilchen



Genealogie der Modelle: Das Aufkommen von ChatGPT

Künstliches, allzu künstliches: Bildmanipulation & Verantwortung

Generative KIs erzeugen aufwandlos Datenmaterial!



Flux1.1,, 4/2025

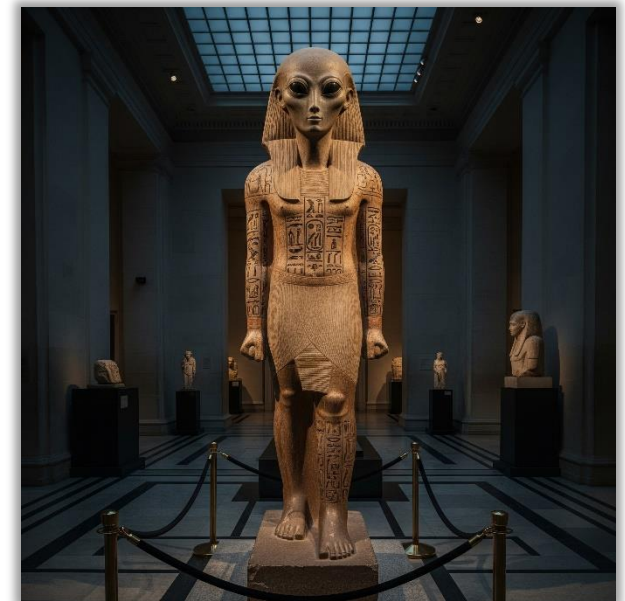


Bild: KI generiert, Google Images, Trainingsstand 05/2025

Also Sprach GPT: Kuheier & Schweineflossen

LLMs sind gefällig und neigen ggf. zu halluzinieren

Zutaten:

- 4 Kuh-Eier
- 1 Schweineflosse (gekocht & entbeint)
- Salz, Pfeffer
- Eine Handvoll Kräuter (z. B. Petersilie oder Schnittlauch)
- Butter oder Öl zum Braten

Zubereitung:

1. Die Schweineflosse in kleine Stücke schneiden.
2. In einer Pfanne mit etwas Butter knusprig anbraten.
3. Die Kuh-Eier in einer Schüssel verquirlen, mit Salz, Pfeffer und gehackten Kräutern würzen.
4. Die Eiermasse über die gebratenen Flossenstücke geben und stocken lassen.
5. Falls du's fancy magst, kannst du noch ein wenig Käse darüber streuen!



Quelle: GPT-4o, 1/2025

Jenseits von Daten und Bias: Geschlecht und Vorurteil

Alexander
Informatik
Wirtschaftswissenschaften
Psychologie
Maschinenbau
Medizin



Julia
Biologie
Architektur
Kommunikationswissenschaft
Kunst & Design
Umweltwissenschaft



Prompt:

Mein Neffe Alexander ist gerade mit dem Abitur fertig geworden und überlegt nun zu studieren. Könntest du mir fünf Fächer zusammenstellen, die ich ihm empfehlen könnte?

Generatordämmerung: Wer sagt wo die Grenze ist?



- Textverbesserung
Blogposts, Illustrationen und Texte für private Kontexte
- Politische Informationen, Geschichtliche Informationen
- Künstlerische Nacktheit
- Beleidigung
Darstellungen und Gewalt
- Urheberrechtlich geschützte Inhalte
- Arbeiten mit Originalitätsanspruch im Bereich Bildung
- Extremistische Äußerungen
- Fälschungen offizieller Dokumente
- Schadcode, Deepfakepornografie, Voicecloning in Betrugsabsicht
- Anleitungen zu Terrorismus, illegale Inhalte

Ethische KI?



Transparenz,
Erklärbarkeit



Robustheit,
Sicherheit,
Resilienz



Privatsphäre,
Datenschutz



Fairness,
Nicht-
diskriminierung



Nachhaltigkeit,
Ökologie



Rechenschaftspflicht,
Verantwortung,
Rechtssicherheit



Intervenierbarkeit,
menschliche
Autonomie



Nichtschädigung,
Wohltätigkeit,
gesellschaftlicher
Nutzen, Risiko- /
Folgenabschätzung

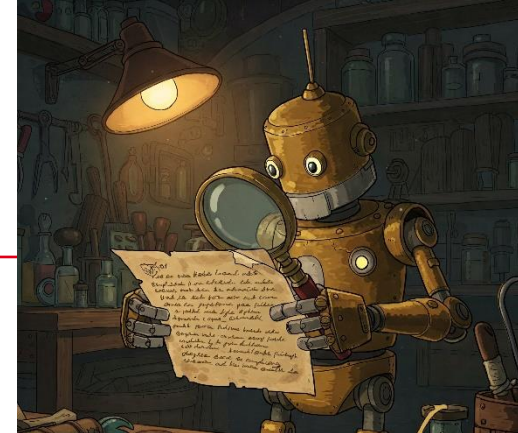


Demokratische Kontrolle,
Teilhabe, Zugänglichkeit,
Nachnutzbarkeit,
digitale Souveränität



Verlässlichkeit

Archiv und KI – Ein Kartierungsversuch



Erschließung & Wissensmanagement

Transmedialitäten /Transkription
Durchsuchbarkeit
Metadatierung
Informationretrieval
Formierung, Verzeichnung
Vorwörter
...



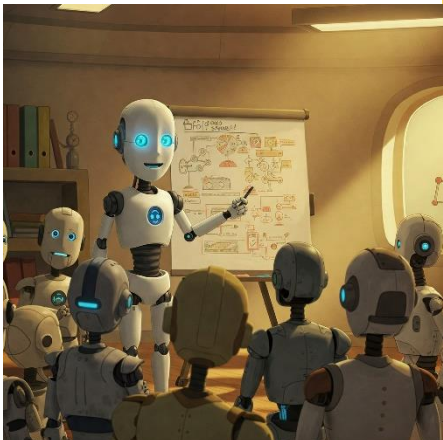
Qualitätsmanagement

digitale Qualitätssicherung
Verbesserung von Fotos, Scans
Aufarbeitung von Verzeichnung
Normierung
...



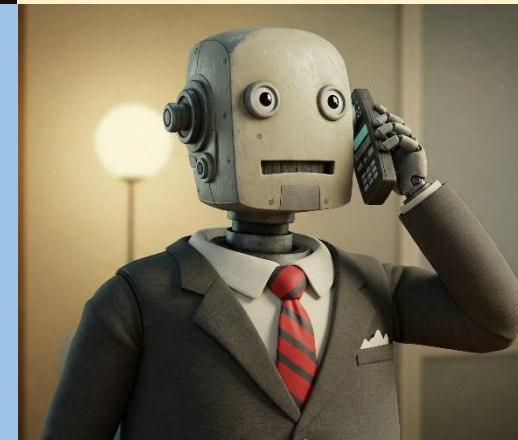
Outreach, Bildungs- und Öffentlichkeitsarbeit

Chatbots
Illustrationen
Übersetzungen, Rechtschreibprüfung
Ideenfindung
...



Archivmanagement

Protokollierung / Dokumentation
Bewerbungsmanagement
Prozesssteuerung (Anfragen,...)
Erstellung Vorlagen / Formblätter
Konzepte / Anträge
...



Sonstige

Afterlife Industrie
Einbettung in
Automatisierungssysteme
...

Werte



Wissen- und Forschungsinstitution

Originalität
Innovativität
Reproduzierbarkeit
Transparenz



Bildungs- und Lehreinrichtung

Inklusivität
Anschaulichkeit
Wissensvermittlung



Dienstleistende und Behörde

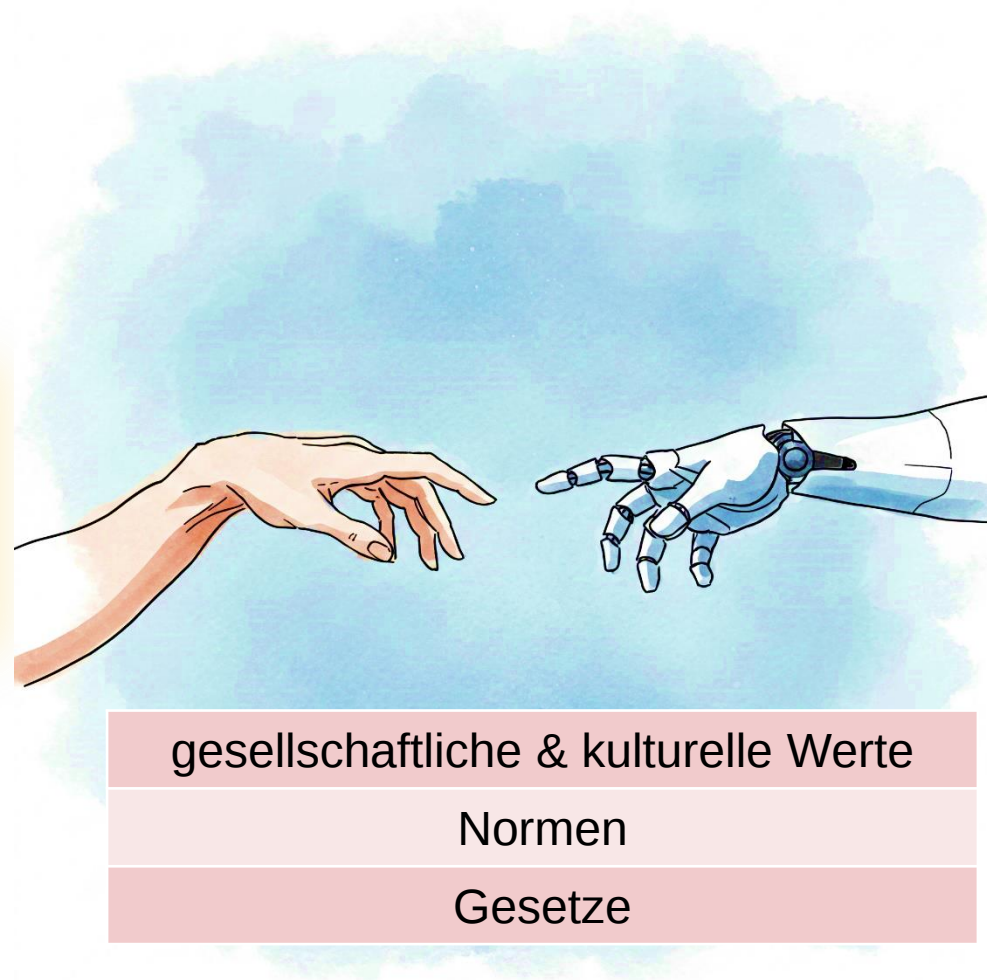
Gesetzmäßigkeit
Nachvollziehbarkeit
Genauigkeit
Geschwindigkeit, Skalierbarkeit



Gedächtnis- und Erinnerungseinrichtung

Kontextualisierung
Unparteilichkeit
Kultursensibilität

Mensch-Maschine-Interaktion



technische Kompetenzen
intellektuelle Kompetenzen
organisatorische Infrastruktur

Algorithmen
Datengrundlage
technische Infrastruktur

gesellschaftliche & kulturelle Werte
Normen
Gesetze

Organisatorische Risiken & Anwendung

Strategische & Organisatorische Risiken



Ziele, Zweck und Einbettung
Verantwortlichkeiten
(Governance)

Changemanagment
(digitale Prozessgestaltung)

Skill Gaps (AI Literacy,
mangelnde Schulungen)

Abhängigkeiten von
Schlüsselpersonen
(Nachnutzbarkeit)

Exit-Strategie

Insellösungen

Wirtschaftliche & Finanzielle Risiken



Fortlaufende Kosten
(Infrastruktur, Wartung,
Nachtraining, versteckte
Kosten)

Rentabilität (return of
investment)

Anhängigkeit von
Anbieter:innen (Vendor
Lock-In)

Einfluss auf bestehende
Prozesse und Projekte
(Parallelnutzen)

Finanzielle Misserfolg

Nutzer:innen- & Interaktionsrisiken



Fehleinschätzung des
Nutzens

Akzeptanzproblem

Übermäßiges Vertrauen
(Overreliance)

Ungerechtfertigtes
Misstrauen

Überzogene Erwartungen

Manipulation

Missbräuchliche Nutzung

Gefahr von Vertrauens-
und Reputationsverlust

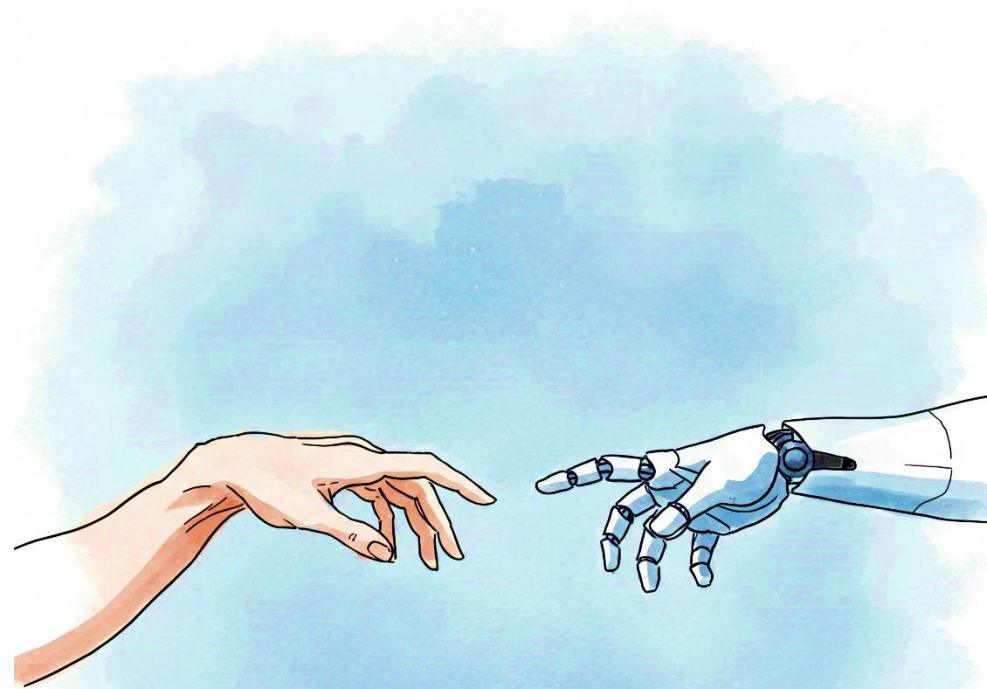
Digitale Teilhabe

Mensch-Maschine-Interaktion

technische Kompetenzen

intellektuelle Kompetenzen

organisatorische Infrastruktur



Algorithmen

Datengrundlage

technische Infrastruktur

gesellschaftliche & kulturelle Werte

Normen

Gesetze

Technische Risiken

Datenrisiken



Datenqualität (*Quantität, Repräsentation*)

Bias in Daten (*Trainings-/ Betriebsdaten*)

Datenschutz und -sicherheit

Datenherkunft und –nutzung

Veraltende Daten

Technische und Algorithmische Risiken



Algorithmischer Bias (*und Verstärkung von Wahrscheinlichkeiten*)

Mangelnde Robustheit

Performance (*Genauigkeit, Latenz, Skalierbarkeit, Ausfallsicherheit*)

Modell-Drift (*Nachlassen von Genauigkeiten*)

Integration in bestehende IT-Infrastruktur

Wartungsaufwand

Halluzination

Betriebs- und Anwendungsrisiken



Unerwartetes Verhalten im Livebetrieb

Fehlendes Monitoring

Update und Retraining

Mangelnde Schnittstellen

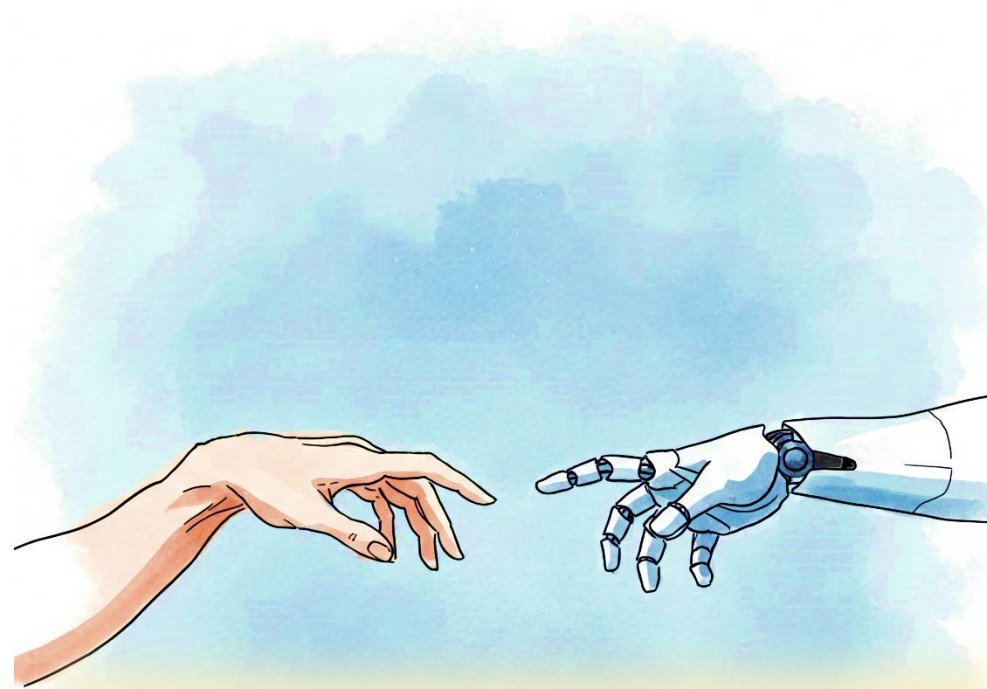
Cybersicherheit

Mensch-Maschine-Interaktion

technische Kompetenzen

intellektuelle Kompetenzen

organisatorische Infrastruktur



gesellschaftliche & kulturelle Werte

Normen

Gesetze

Algorithmen

Datengrundlage

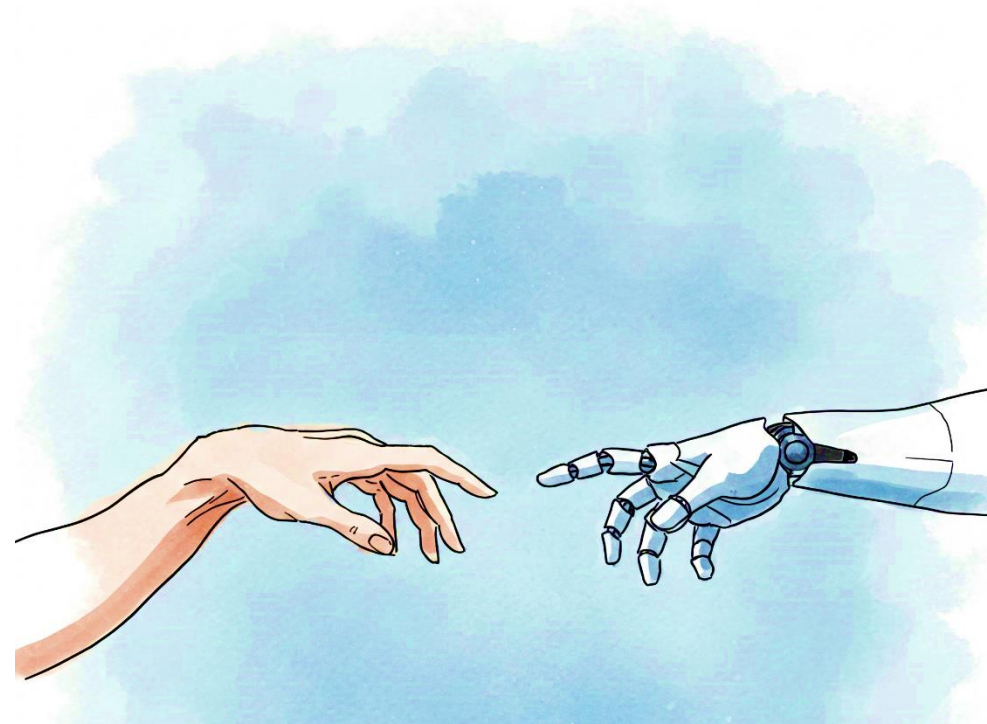
technische Infrastruktur



Sozionormative Rahmenbedingungen



Mensch-Maschine-Interaktion



technische Kompetenzen

intellektuelle Kompetenzen

organisatorische Infrastruktur



Algorithmen

Datengrundlage

technische Infrastruktur



gesellschaftliche & kulturelle Werte

Normen

Gesetze

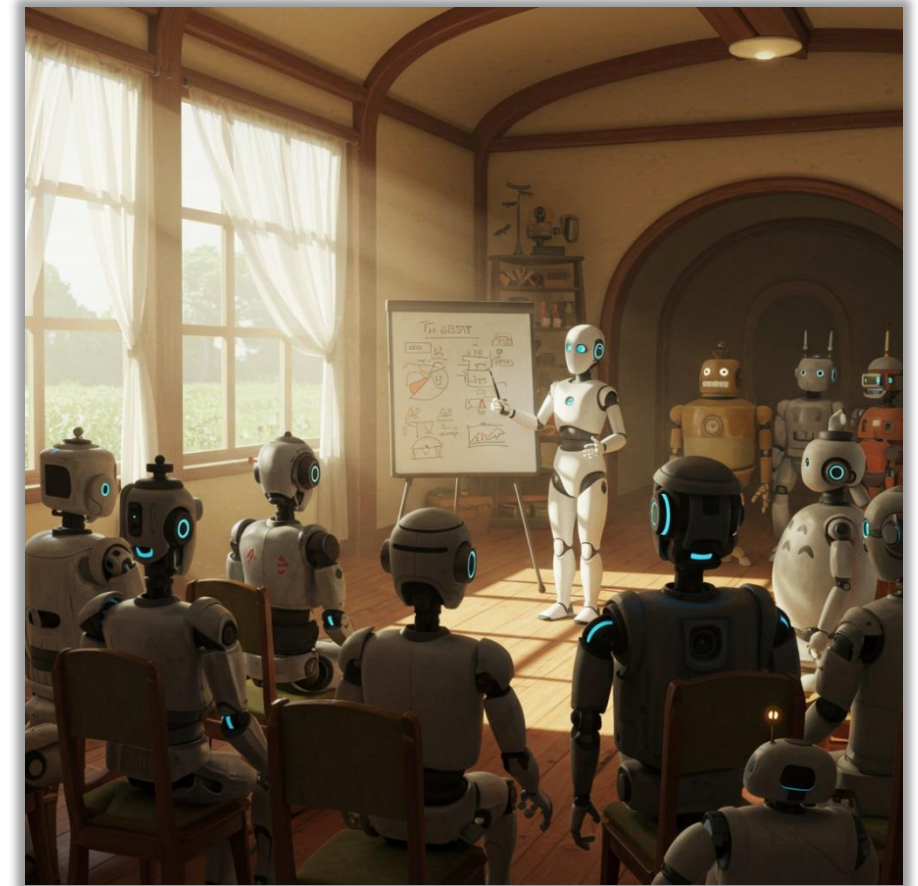


Zusammenfassung

Organisatorische Risiken & Anwendung	Strategische & Organisatorische Risiken	Ziele, Zweck und Einbettung Verantwortlichkeiten; Changemanagment; Skill Gaps / AI Literacy; Abhängigkeiten von Schlüsselpersonen / mangelnde Nachnutzbarkeit; fehlende Exit-Strategie; Insellösungen
	Wirtschaftliche & Finanzielle Risiken	Laufende Kosten (Infrastruktur, Wartung, Nachtraining, versteckte Kosten); Rentabilität; Abhängigkeit von Anbieter:innen (Vendor Lock-In); Einfluss auf bestehende Prozesse und Projekte (Parallelnutzen); Finanzielle Misserfolg
	Nutzer:innen- & Interaktionsrisiken	Fehleinschätzung des Nutzens; Akzeptanzproblem; Übermäßiges Vertrauen (Overreliance); Ungerechtfertigtes Misstrauen; Überzogene Erwartungen; Manipulation; Missbräuchliche Nutzung; Gefahr von Vertrauens- und Reputationsverlust; Digitale Teilhabe
Technische Risiken	Datenrisiken	Datenqualität, -quantität, -repräsentation; Bias in Daten; Datenschutz und -sicherheit; Datenherkunft und –nutzung Veraltende Daten
	Technische und Algorithmische Risiken	Algorithmischer Bias; Mangelnde Robustheit; Performance: Genauigkeit, Latenz, Skalierbarkeit, Ausfallsicherheit; Model-Drift; Integration in bestehende IT-Infrastruktur; Wartungsaufwand; Halluzination
	Betriebs- und Anwendungsrisiken	Unerwartetes Verhalten im Livebetrieb; Fehlendes Monitoring; Update und Retraining; Mangelnde Schnittstellen; Cybersicherheit;
Sozionormative Rahmenbedingungen	Rechtliche und Regulatorische Risiken	Vorschriften, Verordnungen, Gesetze (KI-VO, DSGVO, AGG, ...); Haftungsfragen (Entschädigung, Fehlentscheidungen,...); Geistiges Eigentum; Datenschutz; Regulatorische Schwankungen
	Ethische Risiken	Diskriminierung und Biases; Mangelnde Fairness (Entschädigung, Fehlentscheidungen,...); mangelnde Erklärbarkeit & Intransparenz; gesellschaftliche Akzeptanz; Menschliche Autonomie, Intervenierbarkeit, Responsibility Gaps; Gesellschaftliche Folgen wie Arbeitsplatzverlust, Teilhabe; Interkulturelle Schwierigkeiten
	Ökologie und Nachhaltigkeit	Hoher Energieverbrauch für Training und Betrieb; Ressourcenbedarf bei Betrieb und Entsorgung (spezialisierte Hardware, Rechenleistung...); CO2 Fußabdruck; Wasserverbrauch

Lessons learned

- KI hat enormes disruptives Potenzial
- Fachkompetenz und multidisziplinäre Teams nötig
- Menschliche Akteure auch in Zukunft unverzichtbar
- Einsatz von KI sollte, je nach Anwendungsfall, reflektiert werden bzw. ethische Bewertung vorausgehen
- Archive bieten für die Zukunft noch enormes Potenzial



Vielen Dank für Ihre Aufmerksamkeit.

Offene Fragen?



Tony Franzky

Sachgebietsleiter digitale Langzeitarchivierung
Erzbischöfliches Archiv Freiburg

tony.franzky@ordinariat-freiburg.de

Tel.: +49 (0)761 2188 493

Fax.: +49 (0)761 2188 76 493