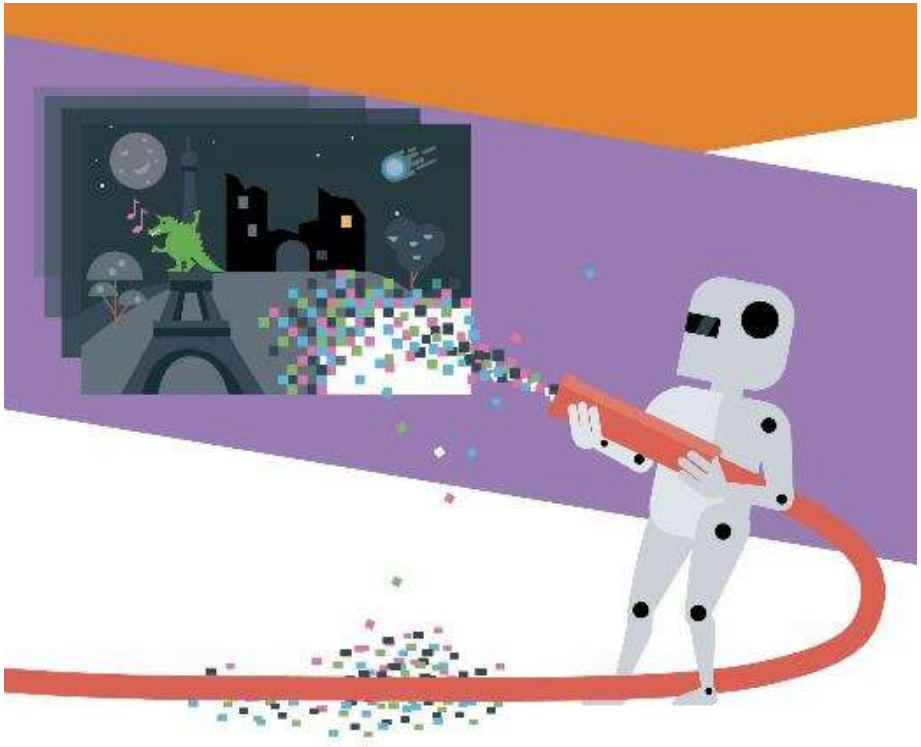
Generative KI und Kulturerbe: Chancen, Risiken, aktuelle Fragestellungen für patrimoniale Institutionen

SKKB-Jahresversammlung 2025

Dr. Laetitia Ramelet
Projektleiterin und Stv. der Geschäftsführerin
TA-SWISS

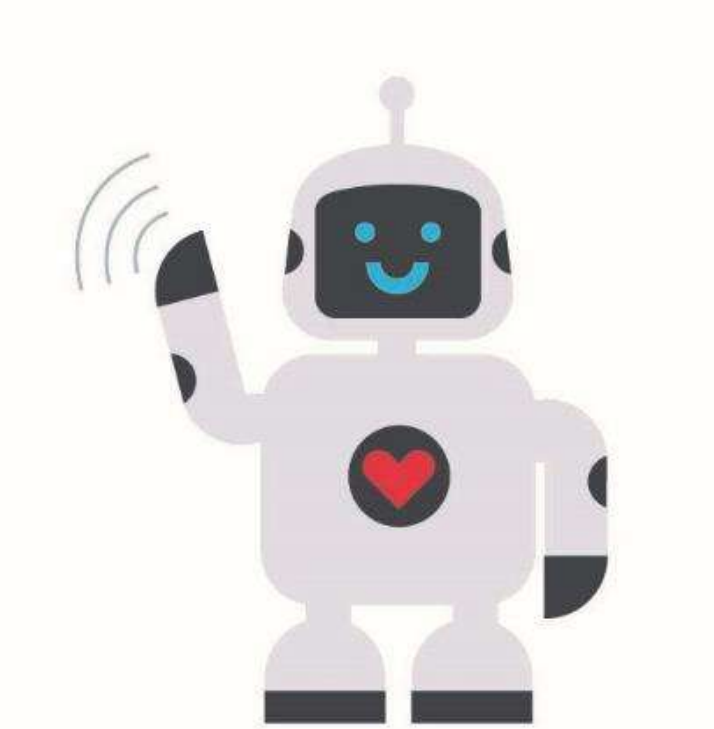


mitglied der
akademien der wissenschaften schweiz

Illustrationen: Hannes Saxer, Bern

TA-SWISS
Stiftung für Technologiefolgen-Abschätzung

- **Abschätzung der Chancen und Risiken neuer Technologien für die Gesellschaft**
- **Sachliche und ausgewogene Studien** für Parlament, Bundesrat, Verwaltung und Stimmbevölkerung
- **Interdisziplinäre Perspektive**
- **Auftrag im Bundesgesetz über die Förderung der Forschung und der Innovation (Art. 11)**



«Deepfakes und manipulierte Realitäten»

Studie im Auftrag von TA-SWISS (2024)

«**Deepfakes und manipulierte Realitäten.**
Technologiefolgenabschätzung und
Handlungsempfehlungen für die Schweiz»

AutorInnen: Murat Karaboga, Nula Frei, Manuel Puppis, Daniel Vogler, Patric Raemy, Frank Ebbers, Greta Runge, Adrian Rauchfleisch, Gabriele de Seta, Gwendolyn Gurr, Michael Friedewald, Sophia Rovelli

Fraunhofer- Institut für System- und Innovationsforschung ISI (Karlsruhe), Universität Freiburg i.Ue., Universität Zürich

Zur Studie: https://www.ta-swiss.ch/deepfakes

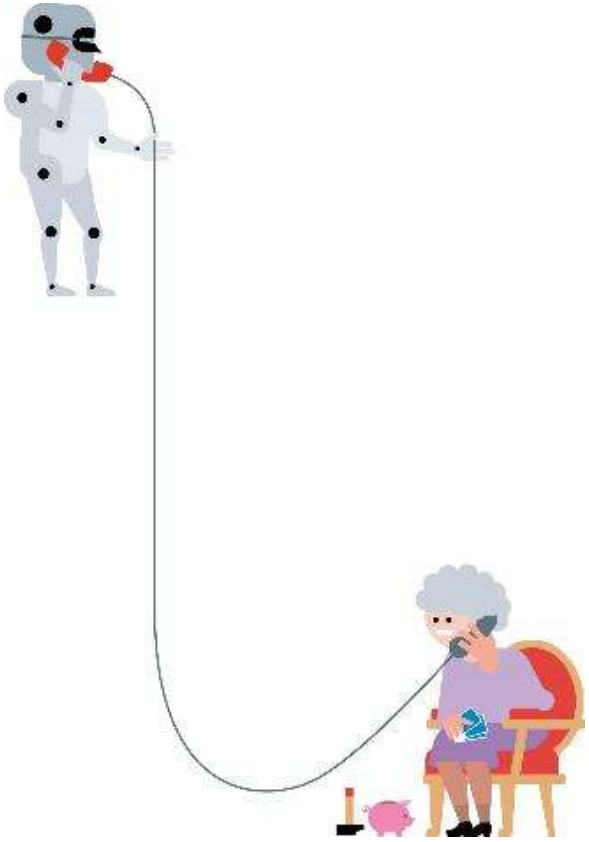
«Ein Deepfake ist ein **mithilfe von KI-Techniken synthetisierter oder manipulierter Audio-, Bild- bzw. Videoinhalt**, der authentisch wirkt, es aber nicht ist.»

Oft wird ein Mensch gezeigt, «**der etwas sagt oder tut, was er nie gesagt oder getan hat.»**

(Karaboga et al., TA-SWISS)

Sources: X Blick.ch. Aww.tokyo

Risiken und Missbräuche…



- Pornographische Deepfakes und Cybermobbing
- Identitätsdiebstahl
- Schockanrufe
- **Verbreitung von falschen Informationen und Manipulation**
- Cyberattacken

… und Chancen

- Unterhaltungsindustrie
- Werbung und Kommunikation
- **Innovative Formate für den Unterricht**
- Schutz von Identitäten

Wie gut können wir Deepfakes von realen Videos unterscheiden?

- Online-Experiment im September 2023 mit 1361 Personen in der Schweiz
- Davon die Hälfte mit einer *Literacy-Intervention* (Tipps), die andere Hälfte ohne
- «Gut gemachte » Deepfakes von bekannten Persönlichkeiten

➔ **Unterscheidung schon jetzt sehr schwierig**

➔ **auch mit *Literacy-Intervention***

➔ **positiver Einfluss von Social-Media-Kompetenzen**

Deepfakes und Journalismus

- **Herausforderung für die Informationsqualität**

- **Müssen Medien einen Vertrauensverlust befürchten?**

- **Chance für den Journalismus**

- **-> Grundlegende journalistische Standards in Praxis pflegen und in Ausbildung vermitteln**

- **Öffentlichkeit für Manipulationen sensibilisieren und auf Wichtigkeit sorgfältiger Überprüfung von Quellen hinweisen**

→ **Auch für Wissensorte wie Bibliotheken eine Chance?**

Einige Empfehlungen der TA-Studie – betr. Informationswelt

- **Bildung und Selbstverantwortung der Bürgerinnen und Bürger fördern**:

Medienkompetenzen und Sensibilisierung beim Teilen von Daten und Deepfakes

- **Präventionsmassnahmen in Schweizer Institutionen und Unternehmen** :

-> interne Beurteilung der Risiken, **Weiterbildung**, Plan für den Umgang mit allfälligen schädlichen Deepfakes, fortschrittliche Authentifizierungsmassnahmen

Einige Empfehlungen der TA-Studie

- **Onlineplattformen regulieren**:

Zusammenarbeit mit den Strafverfolgungsbehörden, Meldesystem, Sperrung gemeldeter Deepfakes (bei Verdacht), Transparenz und Widerspruchsmöglichkeiten

- **Beratungszentren für Opfer von Cyberkriminalität unterstützen**

- **Journalistische Standards hochhalten**

Sprachmodelle: KIs, die Texte verarbeiten und produzieren

Anwendungen von Sprachmodellen

- **Chatbots**
  - allgemeine Chatbots für Fragen, Recherchen und «Gespräche»
  - spezifische Chatbots: z.B. zur Orientierung auf einer Webseite; für Bürger- oder Kundenanfragen; ...
- **Generierung von Texten** (Mails, Briefe, Berichte, Gedichte, Werbeslogans, Artikeln, Prüfungen ...)
- **Generierung von Code**
- **Verbesserungen** bei Texten und Code
- **Zusammenfassungen und Vereinfachung von Texten**
- **Übersetzungen**
- **Analyse** grosser Mengen an Texten
- ...

Zunehmend: multimodale Modelle und KI-Assistenten oder "AI agents"

Grenzen und Risiken (insb. betr. Informationswelt)

Bild: https://www.teepublic.com/

- **Fehler, falsche Informationen und erfundene Quellen**
- **„Automation bias“**
- **Vorurteile und Stereotypen (Trainingsdaten)**
- **Standardisierung von Meinungen und Wissen**
- **Hohe Energiekosten**

Sprachmodelle und Kompetenzverschiebungen: ein neues TA-Projekt

- Grundlegende Kompetenzen: Schreiben, Lesen, Informationssuche, Ideensuche
- Welche Auswirkungen könnte ein häufiger Einsatz von Sprachmodellen auf diese Kompetenzen haben - mit welchen Konsequenzen?
- Welche Kompetenzen müssen künftig in einem Umfeld beherrscht werden, in dem LLMs häufig eingesetzt werden?
- Welche Massnahmen sind notwendig, damit jede und jeder diese Kompetenzen erwerben kann?
- Welche Rolle sollen die verschiedenen Akteure und Institutionen in diesem Prozess spielen?

- **Publikation Ende 2027**
- **Informationen zum Thema:**
  https://www.ta-swiss.ch/grosse-sprachmodelle

Chancen und Risiken für unsere Kompetenzen

Risiken:

- Verlust von Kompetenzen
- Abhängigkeit von den Sprachmodellen
- Wichtige Prozesse verloren
- Weniger Lernmotivation
- Abwertung von Expertise
- Qualitätsverluste im Falle einer Zunahme von KI-Texten

Chancen:

- Fortschritte dank Verbesserungsvorschlägen der Sprachmodelle
- Mehr Zeit für schwierigere oder kreativere Aufgaben
- Druck auf die Qualität unserer Leistungen (!)
- Neue Anwendungen zur Verbesserung unserer Kompetenzen
- Vereinfachter Zugang zu Information und Bildungsangeboten

KI und zukünftige Kompetenzen: Welche Rolle können Bibliotheken spielen?

Bibliotheken als Vermittlerinnen von digitalen Kompetenzen,

Medienkompetenz und «information literacy»:

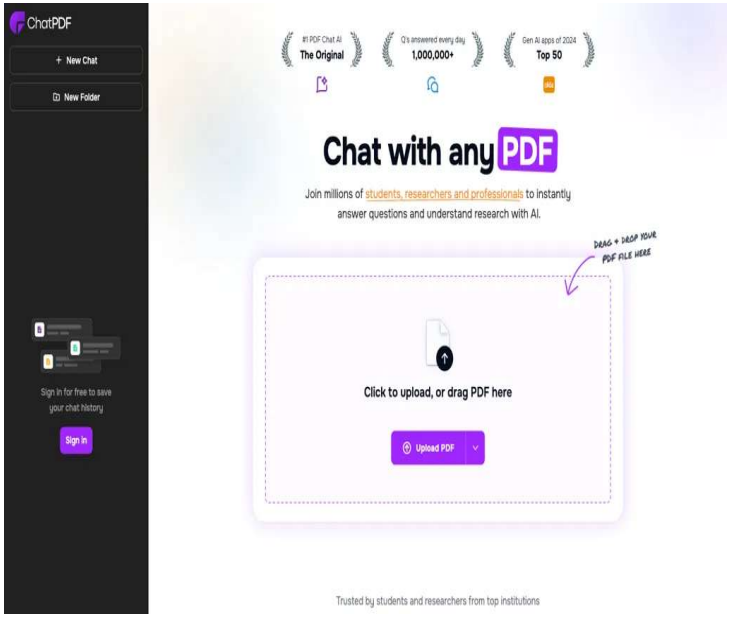
- Kurse über KI-Werkzeuge und «AI literacy»
- Veranstaltungen zu kritischem Umgang mit KI
- KI-Werkzeuge bereitstellen, Kreativworkshops durchführen
- ….

Möglichkeiten und Chancen für patrimoniale Bibliotheken?

KI-Werkzeuge zur Aufwertung bibliothekarischer Sammlungen?

Denkbare Funktionen:

- Mit einem Buch chatten… schriftlich oder mündlich
- Zusammenfassung/Vereinfachung komplexer Inhalte
- Umwandlung von Texten in einfache Sprache
- Personalisierte Zugänge zu den Inhalten
- Direktübersetzungen
- Ev. Kombination dieser Funktionen



Weitere Ideen für neue Zugänge mit KI-Werkzeugen zu den Inhalten patrimonialer Bibliotheken?