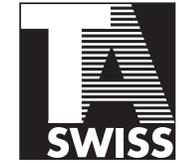


# Conférence suisse des bibliothèques cantonales

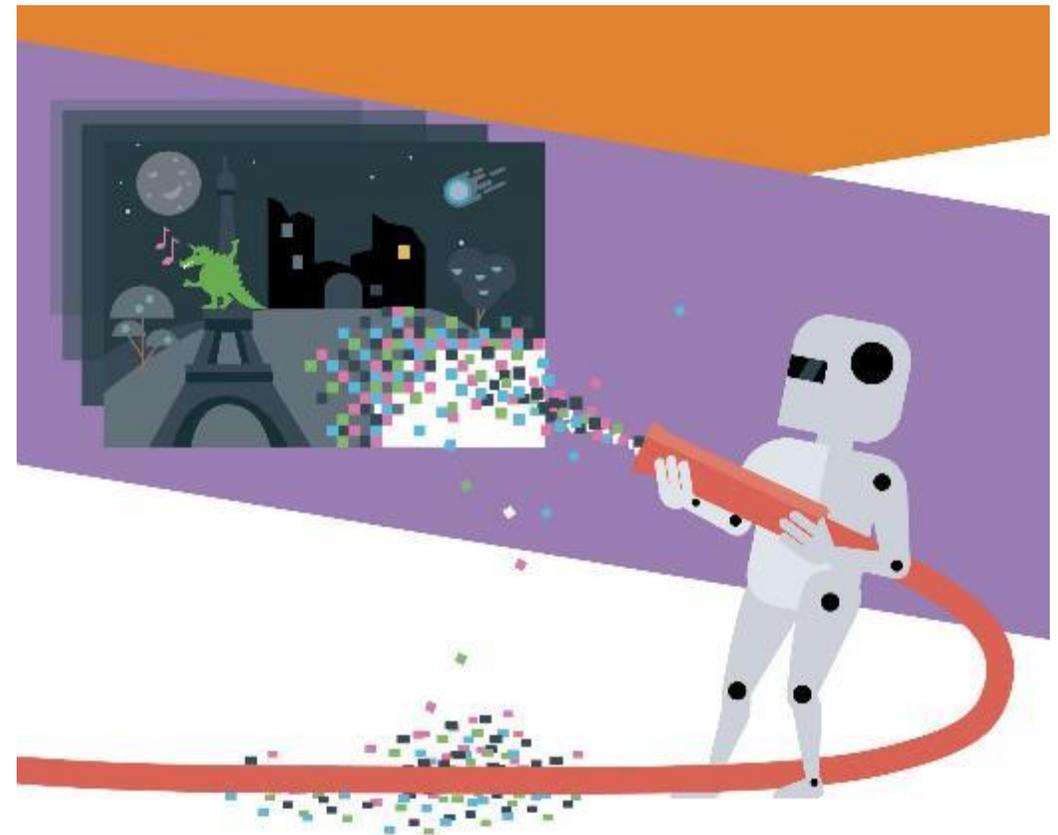


## IA générative : opportunités et risques pour la société, questions d'actualité pour les institutions patrimoniales

Dr. Laetitia Ramelet  
Cheffe de projet et suppléante de la  
directrice

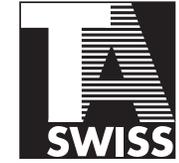
TA-SWISS

mitglied der

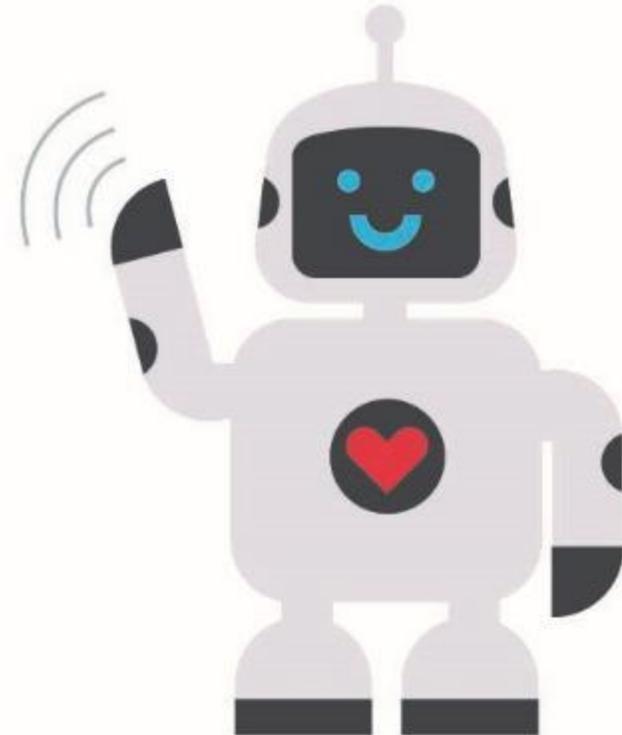


Illustrationen: Hannes Saxer, Bern

# La Fondation TA-SWISS



- **Évaluation des choix technologiques**  
(*Technologiefolgen-Abschätzung*, =TA)
- **Etudes sur les opportunités et risques de nouvelles technologies**
- **Approche interdisciplinaire et participative**
- **Mandat ancré dans la loi fédérale sur l'encouragement de la recherche et de l'innovation** (art. 11)
- **Étude dédiées au Parlement, au Conseil fédéral, aux administrations suisses, à l'économie, la société civile et l'ensemble de la population**



# « Deepfakes et réalités manipulées » (2024)



## «Deepfakes und manipulierte Realitäten.

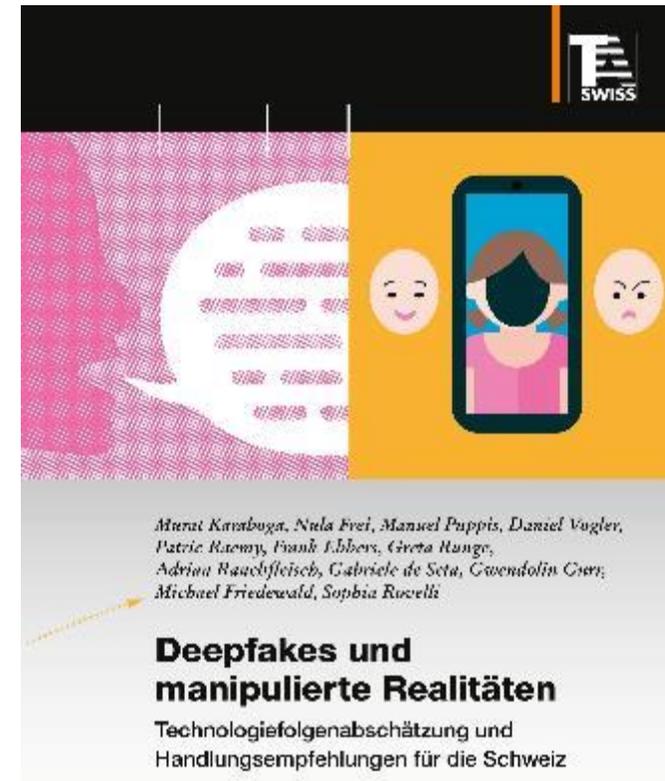
Technologiefolgenabschätzung und  
Handlungsempfehlungen für die Schweiz»

### Auteur(e)s:

**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), Université de Fribourg, Université de Zurich

Télécharger l'étude et sa synthèse: <https://www.ta-swiss.ch/fr/deepfakes>



# Définition FR

«Un **deepfake**, ou hypertrucage, est un contenu **audio ou visuel** synthétique ou manipulé à l'aide de techniques d'IA, qui semble **authentique** et qui, la plupart du temps, fait faire ou dire à un individu quelque chose qu'il **n'a jamais fait ou dit.** »

(Karaboga et al., TA-SWISS 2024)

/ DE

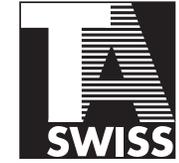


«*Ein Deepfake ist ein **mithilfe von KI-Techniken synthetisierter oder manipulierter Audio-, Bild- bzw. Videoinhalt,** der **authentisch** wirkt, es aber nicht ist.*»

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



# Risques et abus...



- Deepfakes pornographiques et cybermobbing
- Vol d'identité
- « Appels chocs »
- Cyberattaques
- **Diffusion d'informations fausses et manipulations**

## ... et opportunités

- Industrie du divertissement
- Publicité et communication
- **Formats innovateurs pour l'enseignement**
- Protection d'identité



# Arrivons-nous encore à voir la différence ?



- Expérience en ligne en septembre 2023 avec 1361 personnes en Suisse
- Astuces données à la moitié (*Literacy-Intervention*)
- Deepfakes de personnalités connues, « bien faits »

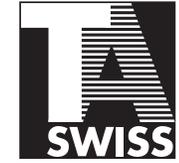
-> **Distinction déjà très difficile**

-> **y compris avec la *Literacy-Intervention***

-> **influence positive des compétences en matière de réseaux sociaux**



# Deepfakes et journalisme



- Un défi pour l'information
  - Crainte d'une perte de confiance dans les médias
  - Une opportunité pour le journalisme
- > Normes de base du journalisme à soigner en pratique et dans la formation
- Besoin d'une sensibilisation du public à la manipulation et à l'importance du travail de vérification des sources
- Aussi une opportunité pour les bibliothèques en tant que lieux de savoir ?



# Quelques recommandations de l'étude TA - concernant l'information



- **Responsabilité personnelle et formation des citoyennes et citoyens:**  
**promotion des compétences médiatiques** et sensibilisation aux enjeux du partage en ligne
- **Mesures de prévention dans les organisations suisses:**

Evaluation des risques, **formation continue**, plans de crise, systèmes d'authentification avancés

# Quelques recommandations de l'étude TA



- **Réglementation des plateformes en ligne:**

Coopération avec les autorités de poursuites pénales, système de notification, blocage des deepfakes illicites, directives de transparence et mécanismes de recours

- **Soutien aux centres de consultation pour les victimes de cybercriminalité**
- **Maintien de normes journalistiques élevées**

# Modèles de langage : des IAs qui produisent et traitent des textes



# Exemples d'applications des modèles de langage



- **Chatbots**

Chatbots généraux, pour questions, recherches et "conversations"

Chatbots spécifiques, par ex. pour demande de citoyens ou de clients

- **Génération de textes** (mails, lettres, rapports, poèmes, publicités, articles, examens ...)
- **Génération de code**
- **Améliorations de textes et de code**
- **Résumés de textes**
- **Simplifications de textes**
- **Traductions**
- **Analyse de grandes quantités de textes**
- **De plus en plus: modèles multimodaux et assistants IA ou "agents IA"**

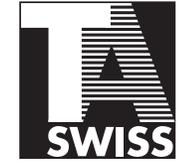
# Risques et limites, du point de vue de l'information



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

- **Erreurs, fausses informations et sources inexistantes**
- **«Biais d'automatisation»**
- **Biais et stéréotypes repris des données d'entraînement**
- **Standardisation des opinions et du savoir**
- **Coût énergétique élevé**

# Modèle de langage et transformations de compétences : un nouveau projet TA



- **Compétences fondamentales:**  
Écriture, lecture, recherche d'informations et d'idées
- **Quelle influence pourrait avoir une utilisation fréquente des LLM sur nos compétences - avec quelles conséquences?**
- **Quelles compétences faudra-t-il maîtriser à l'avenir ?**
- **Quelles mesures pour que chacune et chacun puisse acquérir ces compétences?**
- **Quel rôle pour les différents acteurs et institutions dans ce processus?**



- **Publication fin 2027**
- **Informations sur le thème (DE/FR):**  
<https://www.ta-swiss.ch/fr/modeles-de-langage>

# Opportunités et risques des LLM pour nos compétences



- **Risques** (*deskilling*):

- Perte de compétences
- Dépendance des LLM
- Perte de processus importants
- Baisse de la motivation d'apprentissage
- Dévalorisation de l'expertise
- Pertes de qualité en cas d'augmentation des textes IA

- **Opportunités** (*upskilling*):

- Progrès grâce aux suggestions d'amélioration des LLM
- Plus de temps pour des tâches plus difficiles ou créatives
- Pression sur la qualité de nos performances (!)
- Nouvelles applications pour améliorer nos compétences
- Accès simplifié ou personnalisé à l'information et aux offres de formation

# IA et compétences de demain : quel rôle pour les bibliothèques ?



## **Les bibliothèques en tant que lieux de formation aux compétences médiatiques et numériques:**

- Cours sur les outils d'IA et «AI literacy»
- Evénements sur l'esprit critique à l'ère de l'IA
- Mise à disposition d'outils IA et ateliers créatifs
- ....

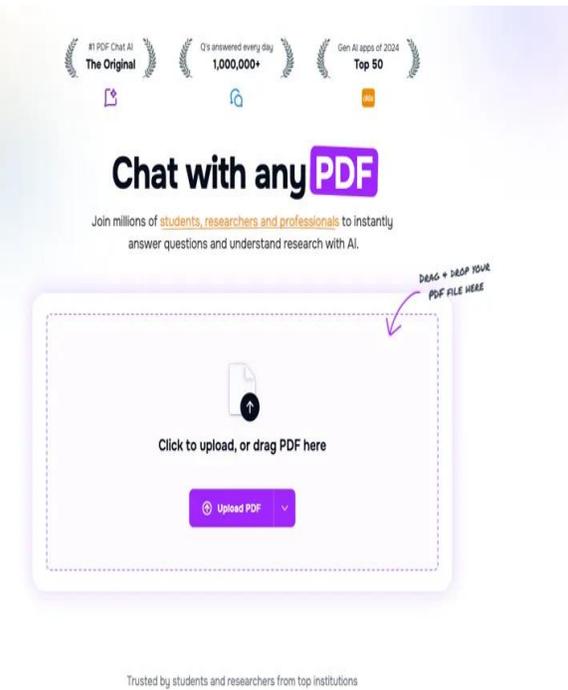
## **Pour les bibliothèques patrimoniales : Quelles possibilités et quelles particularités dans ce contexte ?**

# Des outils d'IA pour valoriser les sources et collections des bibliothèques ?



## Les bibliothèques pourraient-elles mettre à disposition de leur public des outils IA connectés à leur collections ?

- Chatter avec un livre
- ... par écrit ou oralement
- Résumés et simplifications
- Langage simplifié
- Accès personnalisé aux sources
- Traduction en direct de sources
- Combinaison de fonctions possible



## Quid des opportunités pour les bibliothèques patrimoniales d'offrir des nouveaux canaux d'accès à leurs sources et collections ?