

04/2023 &gt; Spezial

# Das Ende der Wahrheit

**KÜNSTLICHE INTELLIGENZ** — Wenn plötzlich alles Fake ist, kann das ein echtes Problem für die Stabilität unserer Demokratie werden. Ein Gespräch über den Umgang mit Deepfakes

21. Juni 2023



Der Papst im Daunenmantel und Frankreichs Präsident als Müllsammler – die KI macht's möglich Foto: screenshot Reddit / Twitter

Deepfakes sind Videos, Bilder oder Audioaufnahmen, die mit Hilfe maschineller Lerntechniken, also Künstlicher Intelligenz (KI), erstellt oder manipuliert wurden. Betroffen sind insbesondere Personen des öffentlichen Lebens, aber auch Privatpersonen werden Opfer von Betrugsma-  
schen. Daniel Seibert und sein Team am Institut für Kommunikations- und Medienwissenschaft in Leipzig widmen sich dem Publikum und dessen Kenntnisstand über Deepfakes sowie den Möglichkeiten und Herausforderungen der sich rasant weiterentwickelnden Technologie.

**ver.di publik: Welches ist das krassste Deepfake, das du je gesehen hast?**



Daniel Seibert ist Kommunikations- und Medienwissenschaftler an der Universität Leipzig Foto: Uni Leipzig

**Daniel Seibert:** Das war im Frühjahr 2020. Da kam mein Professor Alexander Godulla auf mich zu und meinte, es gibt ein mögliches neues Thema, welches wir in einem größeren Projekt verfolgen wollen. Er hat mir dazu ein Aufklärungsvideo von 2018 geschickt – von Barack Obama. Und da sagt ein Fake-Obama Sachen wie "Trump is a complete dip-shit". Das war eines der ersten qualitativ sehr hochwertigen Videos zu der Zeit, und es hat mich komplett geflashed.

**Ich habe ein Zitat aus dem Silicon Valley gelesen, das besagt, in drei bis fünf Jahren wird alles, was wir in den sozialen Medien sehen, Fake sein. Inwiefern haben Deepfakes schon jetzt Einfluss auf die Gesellschaft?**

Da müsste man sich verschiedene Ebenen anschauen. Auf individueller Ebene können Deepfakes beispielsweise für Betrug genutzt werden oder auch, um den Ruf einer Person zu schädigen. Es ist schon vorgekommen, dass es einen Fake-Anruf gab und man eine Stimme gehört hatte, die wie der Chef klang, und die hat dem Mitarbeiter gesagt, dass er unbedingt Geld transferieren soll – da wurde dann ein wirtschaftlicher und finanzieller Schaden angerichtet. Wenn es Privatpersonen trifft, könnte man das auch Enkeltrick 2.0 nennen.

Und das gleiche gilt auch bei der Rufschädigung. Deepfakes wurden anfänglich genutzt, um pornografische Inhalte, meistens von Frauen, insbesondere Schauspielerinnen, zu erstellen – es wurde quasi das Gesicht dieser Personen auf die Pornodarstellerin geschnitten. Das fand auch bei Privatpersonen, jungen Frauen, statt und das kann zu negativen Konsequenzen wie Mobbing in der Schule führen und schwere psychische Schäden anrichten.

**Es gab zahlreiche Fälle, in denen Fakes für politische Propaganda eingesetzt wurden. Eine belgische Partei hat 2018 ein gefälschtes Video von Donald Trump erstellt, in dem er das Land aufgefordert hat, aus dem Pariser Klimaabkommen auszusteigen. Oder denken wir an die Fotos von ihm, auf denen er den Papst befummelt. Wie sehr wird das unsere Wahrnehmung von Realität und Wahrheit beeinflussen?**

Es werden derzeit sehr viele Projekte initiiert, um Deepfakes zu analysieren. Diese stammen vorwiegend aus der Computerwissenschaft und Informatik, aber auch Rechtswissenschaften. Es wird an Technologien gearbeitet, die Deepfakes anhand bestimmter Merkmale erkennen sollen: etwa anhand der Blinzelrate oder Unterschieden in Kontrast und Licht, Hintergrund und Vordergrund oder etwa Lippen- und Mundbewegungen. Und auch in den Rechtswissenschaften wird geschaut, wie man damit umgehen kann. Das ist aber je nach Land und Gesetzeslage unterschiedlich. Spannend ist auch die Frage: Welche Wirkung haben Deepfakes auf das Publikum? Damit beschäftigen sich jetzt auch die Sozialwissenschaften oder wir in der Kommunikationswissenschaft. In einer repräsentativen Bevölkerungsumfrage von circa 1.500 Personen haben wir gefragt, ob sie wissen, was ein Deepfake ist. Und da hat sich gezeigt, dass drei Viertel der Personen keine oder nur eine geringe Kenntnis darüber haben und nur circa ein Viertel einen mittleren oder hohen Kenntnisstand hat. Die Mehrheit schätzt die eigenen Kompetenzen, Deepfakes zu erkennen, sehr gering ein. Viele haben Bedenken, dass die öffentliche Meinung beeinflusst wird oder Deepfakes der Demokratie schaden.

Noch – so mein Gefühl – halten sich die Risiken tatsächlich aber in Grenzen. Die Qualität ist oft nicht ausreichend, um Journalistinnen und Journalisten oder gar ganze Massen zu täuschen, wie es vorhergesagt wird. Es ist jedoch möglich, dass die öffentliche Meinung beeinflusst und auch der demokratische Diskurs gestört werden kann. Hier stellt sich die Frage, inwiefern wir diesen Herausforderungen gewachsen sind.

### **Wie groß schätzt du die Gefahr in der Zukunft ein?**

Man braucht nicht unbedingt ein Deepfake-Video, um die öffentliche Meinung oder die öffentliche Ruhe zu stören. Es reichen auch Statements, die aus dem Kontext gerissen sind. Es reichen bereits Bilder oder Videoausschnitte, die nicht mal mit Künstlicher Intelligenz bearbeitet sein müssen, um einer Person bestimmte Worte in den Mund zu legen.

Aber was natürlich schon ein Problem sein kann, ist, wenn viel Trubel um so ein Video oder so ein Bild entsteht. Es besteht die Gefahr, dass immer etwas hängen bleibt im Gedächtnis der Menschen, auch wenn es wieder revidiert werden sollte. So könnte zum Beispiel die Einstellung gegenüber einer Partei oder einer Person nachhaltig beeinflusst werden.

### **Wird KI die Lösung sein, um Deepfakes zu identifizieren?**

Es ist ein Kopf-an-Kopf Rennen. Natürlich werden Technologien entwickelt, um immer bessere Deepfakes zu erstellen, und parallel werden Technologien entwickelt, um Deepfakes zu identifizieren. Je besser erstere ist, desto schlechter ist dann wieder die Software zur Identifikation. Man wird wahrscheinlich auch nie allein auf die Ergebnisse von so einer Identifikationssoftware bauen können, weil es da trotzdem oft eine gewisse Fehlerquote gibt. Der Fokus liegt daher

nach wie vor primär auf den journalistischen Fähigkeiten wie der Recherchearbeit und Faktenprüfung, auch wenn diese in Zukunft und im Umgang mit Deepfakes wohl komplexer werden.

### **Welche Möglichkeiten bringt die Technologie mit sich?**

Sie kann im Bereich Online-Shopping und Handel eingesetzt werden, zum Beispiel, um virtuell Kleidung anzuprobieren. Oder auch in der Film- und Entertainmentbranche, wo Schauspieler\*innen für kurze Sequenzen nicht unbedingt noch mal reingeholt werden müssen, sondern Szenen digital nachbearbeitet werden. Auch Personen, die schon längst verstorben sind, können auf der Leinwand wieder zum Leben erweckt werden. Kurze Audiospuren können erstellt und leichter bearbeitet werden, sodass eine Person nicht noch mal ins Tonstudio muss, um etwas einzusprechen. Inhalte können automatisiert und personalisiert werden, ganze Podcasts können in andere Sprachen übersetzt werden. Einzelpersonen wie Influencer können die Technologie zur Erstellung von Inhalten und zur Visualisierung nutzen. Zukünftig ist man wahrscheinlich nicht mehr an ein hohes Budget und große Medienhäuser gebunden, sondern auch kleinere mit einem geringeren Budget aufgestellte Medienhäuser können die Technik einsetzen und Inhalte produzieren. Man spricht hier auch von einer Demokratisierung der Kreativität.

Interview: *Rita Schuhmacher*

### **So wird ein Fake draus**

Um ein Deepfake zu erstellen, wird der KI-Algorithmus mit einem großen Datensatz von Bildern oder Videos der Zielperson trainiert, so dass er lernen kann, deren Bewegungen, Mimik und Stimme nachzuahmen. Anschließend kann der Algorithmus verwendet werden, um ein neues Video oder Bild der Zielperson zu erstellen.

Teilen:

Schlagwörter:

Medien

Weitere Artikel aus der Rubrik »Spezial«

- "Menschliche Filter für das Grausamste, was es gibt"
- Kurzmeldungen Digitalisierung
- Mensch vor Maschine
- Preisrätsel 4\_2023