

Deepfakes, Chatbots und wir

Von Marlene Obst

Wie Künstliche Intelligenz unseren digitalen Alltag verändert und wie wir uns gegen KI-Fakes wappnen können

Ob Suchmaschinen, Kundenservice oder Chatbots – KI ist längst überall, selbst wenn wir sie nicht bewusst einsetzen. Sie schreibt, antwortet, hilft und macht dabei Fehler – und das nicht selten.

KI-generierte Bilder und Videos fluten unsere Feeds. Der Unterschied zwischen echt und künstlich verschwimmt immer mehr. Was früher Stunden in Photoshop brauchte, entsteht heute in Sekunden. Genau wie Social Media wirkt KI wie ein Brandbeschleuniger für Fakes aller Art, auch für Desinformation, also mit Absicht verbreitete falsche Infos, die gezielt Schaden anrichten sollen. Klar ist: KI und Chatbots sind nicht mehr wegzudenken. Sie erleichtern die Arbeit, inspirieren uns und sparen Zeit. Aber sie können auch täuschen und langfristig unser Vertrauen in echte Inhalte beschädigen. Die entscheidende Frage lautet: Wie können wir KI sicherer nutzen und gleichzeitig die Fakes entlarven, die sie erzeugt?

Jede dritte Antwort fehlerhaft: Die Schwächen von KI-Chatbots kennen!

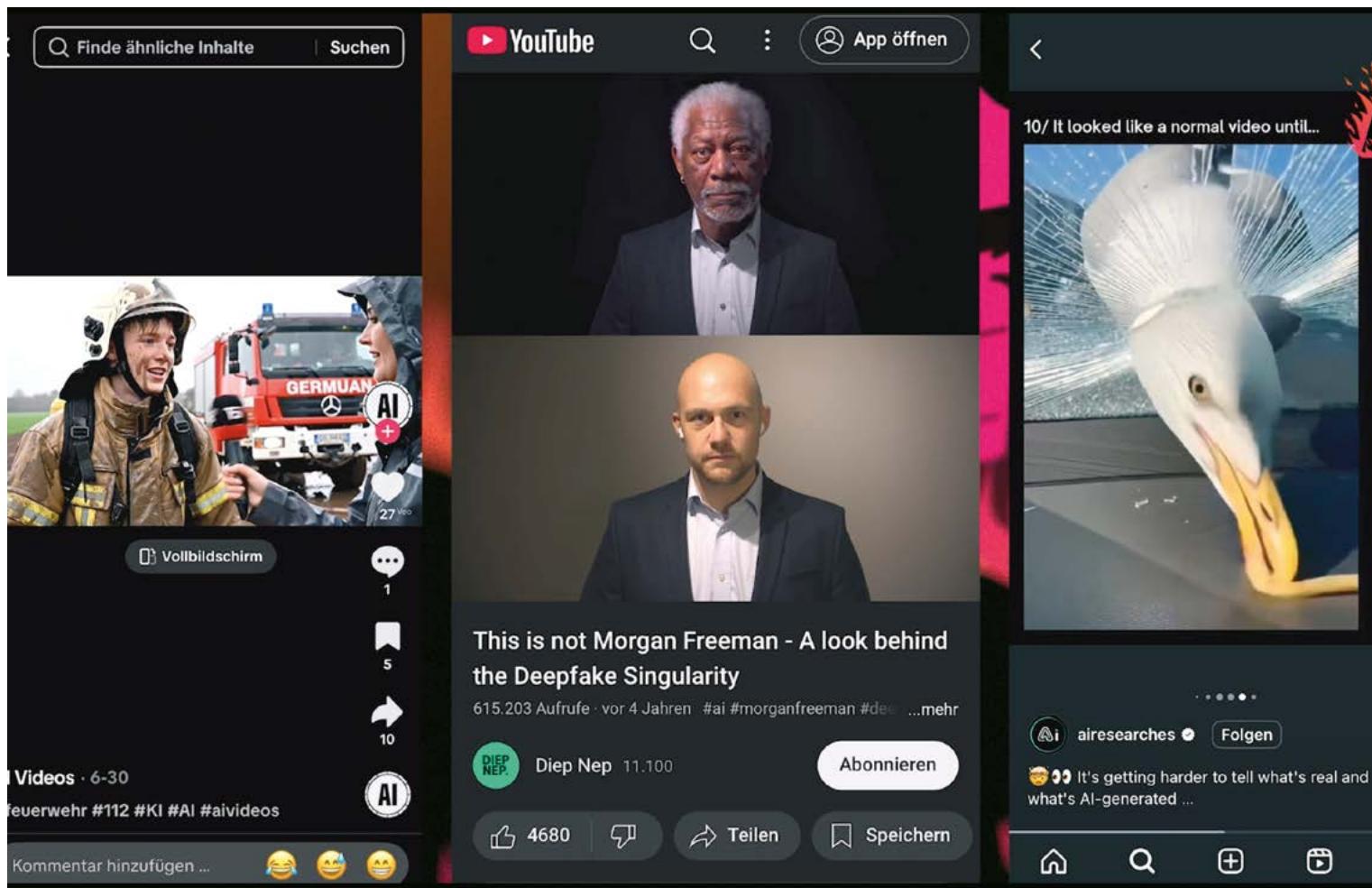
Wer Chatbots nutzt, sollte ihre Schwächen kennen. Dafür müssen wir verstehen, wie sie funktionieren. Denn die smarten Helfer sind keine Wissens- oder Wahrheitsmaschinen, in denen wir einfach schnell etwas nachschlagen können. Sie arbeiten nicht mit Fakten, sondern mit Wahrscheinlichkeiten. Sie berechnen, welches Wort am wahrscheinlichsten folgt – nicht, ob es stimmt. Das macht sie fehleranfällig.

Wie stark, zeigt eine aktuelle Studie der Europäischen Rundfunkunion. Sie hat gängige Chatbots getestet. Das Ergebnis: 45 Prozent der Antworten enthielten mindestens einen erheblichen Fehler. Am häufigsten fehlten Quellen oder sie waren irreführend. Der zweithäufigste Fehler: falsche Behauptungen. Einige Chatbots erklärten etwa, der verstorbene Franziskus sei noch Papst, Schweden gehöre nicht zur NATO oder Olaf Scholz sei weiterhin Bundeskanzler.

Solche Fehler können verschiedene Ursachen haben:

- **Halluzinationen:** Die KI ergänzt fehlende Informationen, indem sie statistisch plausible Wortketten erzeugt – auch wenn sie falsch sind.
- **Unvollständige oder veraltete Trainingsdaten.** Hier auch wichtig: Chatbots suchen nicht automatisch nach aktuellen Infos und auch Websuchen können fehlerhaft sein.
- **Manipulierte Daten:** Akteure versuchen gezielt so genannte Data Voids – Themenbereiche, in denen es kaum verlässliche Quellen gibt – zu nutzen, um dort Falschinformationen oder Propaganda zu platzieren, die Chatbots aufgreifen sollen. Erste Studien zeigen das zum Beispiel beim Thema Ukraine-Krieg.

Tückisch ist auch: Die KI-Antworten klingen meist überzeugend, auch wenn sie falsch sind. Immer wieder glauben oder übernehmen Menschen solche fehlerhaften Infos ungeprüft – Schüler/-innen, Studierende, Anwälte.



Das kann peinliche, manchmal auch gefährliche Folgen haben.

Wichtig: Auch die neuesten Modelle haben dieses Problem! Halluzinationen gehören zur Funktionsweise dieser Systeme. Kritisch bleiben, Quellen prüfen und Infos gegenlesen ist deshalb unverzichtbar. Wie zuverlässig eine Antwort ist, hängt aber auch davon ab, wie wir fragen.

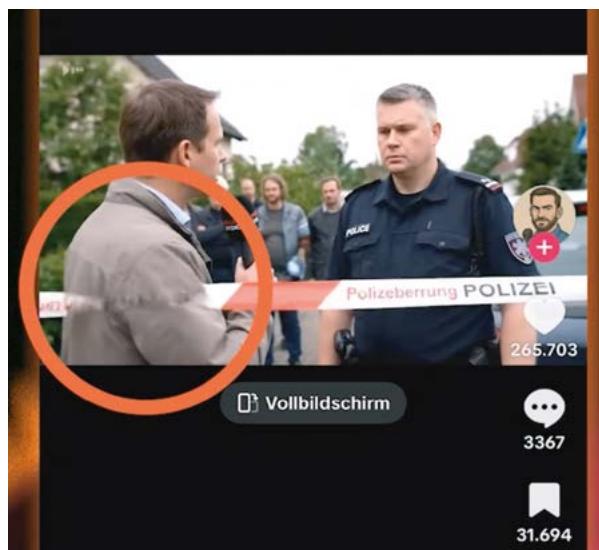
Wie wir unsere Anfragen an Chatbots stellen, ist entscheidend. Ein paar Tipps, die helfen können:

- Formuliert eure Anfragen klar und konkret.
- Fordert eine faktenbasierte Antwort und fragt nach Quellen.
- Fragt nach der Antwort auch ruhig noch einmal nach, ob die Quellen echt sind.
- Besteht darauf, dass der Chatbot im Netz recherchiert. Moderne Modelle können das, machen es aber nicht automatisch. Das liefert oft verlässlichere Antworten (besonders wichtig bei Anfragen zu aktuellen Ereignissen).
- Stellt eure Anfrage an verschiedene Chatbots.
- Trotzdem gilt: Wer sicher gehen will, muss selbst prüfen!

Von Spaß bis Propaganda: Die neue Welt der KI-Fakes

Viele KI-Inhalte, die täglich unsere Social-Media-Feeds fluten, sind Unterhaltung, Parodie, Kunst oder reines Clickbait. Es gibt aber auch gezielte Desinformation, zum Beispiel Videos von vermeintlichen Anschlägen oder Unglücken. Andere KI-Fakes greifen reale Nachrichten auf,

verzerren sie und machen daraus virale Dramen, zugeschnitten auf Emotionen und Reichweite. Noch lassen sich viele dieser Inhalte entlarven, zumindest auf den zweiten oder dritten Blick. Häufig verraten kleine Details die KI: Hände mit sechs Fingern, merkwürdig glänzende Haut oder verschwimmende Schriftzüge im Hintergrund. Auch bei Videos gibt es solche Hinweise: KI-generierte Menschen gehen öfter mal durch Objekte, haben starre Gesichter oder bewegen sich unnatürlich. Achtet auch auf unnatürlich synchronisierte Lippen oder Stimmen. Klingen sie zum Beispiel monoton oder künstlich?



Ein Beispiel für Fehler in KI-Videos: KI-generierte Menschen wirken manchmal geisterhaft, gehen durch Objekte oder bewegen sich unnatürlich. Achtet auch auf die Stimmen – klingen sie monoton oder künstlich? Quelle: Screenshot aus Reporterfabrik-Video. Original: TikTok



Typische Fehler in KI-Bildern: Achtet auf Details wie Hände, Schriftzüge auf Kleidung oder Augen. Auch zu perfekte, glatte Bilder können ein Hinweis auf KI sein. Quelle: Screenshot aus Reporterfabrik-Video. Original: X / @markkayeshow

Aber die Technik entwickelt sich rasant. Systeme wie „Veo“ oder „Sora“ erzeugen inzwischen sehr realistische Videos aus simplen Textbefehlen, ganz ohne technisches Wissen. Deshalb gilt: immer den Kontext prüfen. Gibt es seriöse Quellen oder nur dubiose Kanäle? Ist die Info überhaupt logisch? Bei Unsicherheit lohnt ein Faktencheck.

Gefördert vom



Bundesministerium
für Bildung, Familie, Senioren,
Frauen und Jugend

im Rahmen des Bundesprogramms

Demokratie **leben!**

Über die Gefahren von Deepfakes und KI-Desinformation wird viel diskutiert. Noch tauchen sie im Vergleich zu anderen Fakes nicht allzu häufig auf. KI-Fakes werden allerdings vermehrt in Kriegs- und Konfliktgebieten eingesetzt, um zum Beispiel Propaganda zu verbreiten. Ein weiterer Punkt: Schon heute werden reale Bilder, Videos oder Audios als „KI-Fälschung“ abgetan. „Das ist doch manipuliert“, heißt es dann. Nicht selten mit Absicht, um Kritik abzuwehren oder Zweifel zu säen. Manchmal aber auch aus Unsicherheit oder weil man es nicht glauben will.

Aber was bedeutet das alles für uns, für die Gesellschaft und unseren digitalen Alltag? KI-Fakes und fehlerhafte Infos von Chatbots werden uns erhalten bleiben. Eine der großen Herausforderungen der kommenden Jahre wird sein, das Vertrauen in echte Inhalte zu stärken und zu bewahren.

Und für uns persönlich? Ein wichtiger Schutz ist wohl ein gesundes Misstrauen. Für uns wichtige Informationen weiter konsequent prüfen, ein Bewusstsein für die Risiken und Grenzen von KI-Tools entwickeln und dranbleiben beim Thema KI und Fakes. Denn wer versteht, wie Chatbots (und KI-Fakes) funktionieren, kann sie besser hinterfragen und gezielter und kreativer nutzen – vielleicht auch, um Desinformation zu bekämpfen.

Infos zur Autorin:

Marlene Obst ist Teil des Projekts „Brandherd Desinformation“, das von der Reporterfabrik bei CORRECTIV geleitet und gemeinsam mit der Deutschen Jugendfeuerwehr (DJF) umgesetzt wird. Ziel ist es, junge Menschen in ihrer Nachrichten- und Quellenkompetenz zu stärken und zivilgesellschaftliche Organisationen im Umgang mit Desinformation zu unterstützen. Das Projekt wird an sechs Standorten in ganz Deutschland durchgeführt und im Rahmen des Bundesprogramms „Demokratie leben!“ vom Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend gefördert. Für inhaltliche Aussagen und Meinungsäußerungen trägt die Publizierende dieser Veröffentlichung die Verantwortung.

FÜR UNSERE JUNGEN KAMERADEN/INNEN

BRANDAKTUELL

Schlüsselanhänger Textil Jugendfeuerwehr

Schlüsselanhänger 120 x 27 mm



120 x 27 mm,
blau beidseitig
bestickt

3,80
Unser Preis
Tolles Geschenk

Art.Nr. 54156

Handschuh Askö FW-Jugend

Handschuh nach EN 388,
lieferbar Größe 4 – 10



Art.Nr. 52708

Badetuch DJF „Trockenübung“

Feinstes Frottee mit Aufdruck.
100% Baumwolle Farbe: Rot
Gr. 70 cm x 140 cm



Art.Nr. 52999

NEU

Unser Preis
Tolles Geschenk

19,90
Unser Preis
Tolles Geschenk



Art.Nr. 52116

Rucksack Deutsche Jugend- feuerwehr

Sonderaktion
ab 20 Stück 18,50 Euro

19,90
Sonderaktion
ab 20 Stück 18,50 Euro

Art.Nr. 54059

Geschenkset DJF Flasche & Brotdose

Trinkflasche 770ml plus
Brotdose Marschverpflegung.



Art.Nr. 54059

BESTELL & INFOHOTLINE
(0228) 9 53 50-0

WWW.FEUERWEHRVERSAND.DE

Qualität ist unser Dauerbrenner!