

Wochenkommentar 23/2023 von Matthias Zehnder

# Führt KI zu mehr Fake News?



Bild: KEYSTONE/DPA/Thomas Trutschel

*ChatGPT hat die Möglichkeiten der generativen KI einem breiten Publikum bewusst gemacht und auch gezeigt, wie einfach sich die Möglichkeiten missbrauchen lassen.*

**Letzte Woche habe ich mich an dieser Stelle mit Fake News und dem Problem der bröselnden Wahrheit beschäftigt. Geendet habe ich mit einer Frage: Was ändert die breite Verfügbarkeit von künstlicher Intelligenz? Sorgt die KI für mehr Fake News? Diese Fragen schauen wir uns heute gemeinsam an. Wissenschaftliche Studien dazu gibt es noch nicht. Die generative KI ist erst seit wenigen Wochen breit verfügbar. Einige Entwicklungen lassen sich dennoch bereits beobachten. Und wir können gemeinsam darüber nachdenken, welche Folgen die breite Verfügbarkeit von KI-Werkzeugen wie ChatGPT haben wird. Welche Konsequenzen es hat, wenn die AfD in Deutschland oder Republikaner in den USA manipulierte Bilder einsetzen. Was das für den Umgang mit der KI bedeutet. In meinem Wochenkommentar geht es deshalb diese Woche um die Frage, ob generative KI zu Fake-News-Schleudern werden.**

Bevor wir einsteigen, müssen wir kurz klären, worüber wir sprechen. «Fake News» sind bewusst verbreitete Falschmeldungen. Im harmloseren Fall sollen solche Meldungen vor allem Aufmerksamkeit erregen. Oft handelt es sich aber um gezielte politische Desinformation oder Propaganda. Solche Falschmeldungen spielen mit den Ängsten der Menschen,

indem sie zum Beispiel behaupten, die Bewohner eines Altersheims seien zum Auszug gezwungen worden, weil der Staat dort Geflüchtete unterbringen wolle.

Von künstlicher Intelligenz sprechen wir dann, wenn Computersysteme selbstständig lernen können. Die Programmierer geben also nicht vor, welche Antworten der Computer in diesem oder jenem Fall geben soll. KI-Systeme können Antworten selbst finden und Probleme bis zu einem gewissen Grad eigenständig lösen. Dazu werten sie sehr grosse Datenmengen aus und suchen darin nach Mustern. ChatGPT und ähnliche KI-Systeme gehen noch einen Schritt weiter: Sie sind in der Lage, Sprache, Programmcode oder Bilder zu generieren. Um solche Systeme geht es, wenn von KI-generierten Fake News die Rede ist. Die grosse Angst ist, dass solche generativen KI-Systeme im grossen Stil Fake News produzieren und die Welt mit Falschmeldungen fluten.

### **Keine Computer-Frankensteine**

Ganz wörtlich genommen ist diese Angst unbegründet. Dahinter steckt die Vorstellung, die KI sei eine autonome Maschine, die sich ins Internet einklinkt und dann nach eigenem Gutdünken Nachrichten verbreitet. Diese Vorstellung ist falsch. ChatGPT, Google Bard und ähnliche Systeme sind Werkzeuge. Sie haben keinen eigenen Willen und handeln nicht aus eigenem Antrieb. Es sind also keine Computer-Frankensteine, die unkontrolliert im Internet ihr Unwesen treiben. Das heisst aber nicht, dass die Systeme nicht gefährlich werden können. Wenn man sie zu bedienen weiss, sind generative KI-Systeme extrem leistungsfähige Werkzeuge.

In der harmloseren Variante erstellen Menschen mit Hilfe von KI Fake News und verbreiten sie dann zum Beispiel über Twitter, Facebook oder andere soziale Medien. Im Zentrum steht also eine Senderin oder ein Sender. Diese Person will eine bestimmte Botschaft aussenden. Dazu bedient sie sich der KI: Mit der Hilfe der neuen Systeme kann die Person schneller als bisher eine Nachricht oder ein Bild fälschen. Dabei tut die KI nichts anderes, als die Person selbst mit herkömmlichen Mitteln am Computer tun könnte: Sie schreibt einen Text oder bearbeitet ein Bild. Neu daran ist, dass die KI unfassbar viel schneller arbeitet als ein Mensch und sich um einiges einfacher bedienen lässt. So kann die KI sehr schnell Texte in einem bestimmten Stil generieren, übrigens auch in Fremdsprachen, und Bilder lassen sich mit Hilfe der KI viel einfacher bearbeiten und fälschen.

### **Nicht jeder Trottel kann KI**

Wichtig ist dabei das Wort «einfacher»: Mit KI kann der Photoshop beispielsweise schneller als bisher einen Hintergrund ersetzen oder verändern. Trotz aller KI bleibt Bildbearbeitung aber immer noch etwas für Profis: Wer nicht weiss, wie Ebenen, Masken und Gradationskurven funktionieren, kommt auch mit der KI im Photoshop nicht so weit wie der Fachmann. Es gibt zwar Onlinedienste, die mit KI Bilder generieren, aber auch diese Dienste setzen voraus, dass jemand die richtigen Anweisungen gibt. Da geht es weniger um bearbeitungstechnische Befehle als

vielmehr um konkrete Anweisungen und Vorstellungen, wie ein Bild aussehen soll. Nur weil die KI Befehle unfassbar schnell umsetzt, heisst das nicht, dass jeder Trottel gute Fälschungen produzieren kann. Man muss schon wissen, was man will, und etwas Erfahrung haben im Umgang mit den KI-Tools.

Einer, der damit offenbar viel Erfahrung hat, ist Norbert Kleinwächter, AfD-Bundestagsabgeordneter aus Augsburg und stellvertretender Vorsitzender der AfD-Bundestagsfraktion. Auf Twitter und Instagram setzt Kleinwächter immer wieder KI-generierte Bilder ein, um seine politischen Botschaften zu illustrieren. Im harmloseren Fall ist eine Zombie-Teufelversion von Wirtschaftsminister Robert Habeck zu sehen und dazu die Schlagzeile «Schützt unsere Häuser vor der Regierung!» Über dem Slogan «Verbrennermotor vor Brüssel retten!» prangt das Bild eines Monsters mit blauen Flügeln und Europasternen, das sich mit seinen Klauen auf ein Auto stürzt. Gesundheitsminister Karl Lauterbach wird gleich selbst als Monster dargestellt. Die Parole «Nein zu noch mehr Flüchtlingen!» wird durch einen aggressiven Mob junger Männer illustriert. Dieses letzte Bild hat Norbert Kleinwächter und seiner Partei viel Kritik eingetragen. Der Vorwurf: Er setze KI zur Hetze ein. Kleinwächter antwortete darauf, natürlich mit einem Posting. Schlagzeile: «Können Fakten Hetze sein?»

### **Der «Stürmer» lässt grüssen**

Die Art und Weise, wie Kleinwächter und die AfD die generative KI einsetzen, erinnert stark an den Einsatz von politischen Karikaturen in den Kampfblättern der Parteien. So verteidigt sich auch die AfD. Das Argument: Man sieht es den Bildern ja an, dass sie künstlich sind. Das stimmt. Aber auch den Karikaturen im «Stürmer» sah man es an, dass sie gezeichnet waren. Trotzdem konnten sie ihre antisemitische Wirkung entfalten. Es sind Bilder, die sich in den Köpfen festsetzen. Im Zeitalter der sozialen Medien kommt noch dazu: Ein starkes Bild bekommt viel mehr Aufmerksamkeit und es wird viel mehr geteilt als ein blosser Text. Das heisst auch: Das überzeichnete Bild wirkt ungleich stärker als die möglicherweise korrekten Fakten in der Bildzeile.

Problematisch an der bildergenerierenden KI ist, dass sie auch fotorealistic Bilder erzeugen kann. Ein berühmtes Beispiel ist das Bild von Papst Franziskus in einer weissen Balenciaga-Jacke. Das Bild wurde im Netz breit geteilt und war im Nu weit verbreitet. Viele Menschen haben das Bild wahrscheinlich für echt gehalten. Aber es hat sich auch bei jenen Menschen ins Gedächtnis eingebrannt, die wussten, dass es sich um ein Fake-Bild handelt. Und genau das ist das Problem mit gefälschten Bildern. Schon jetzt spielen solche Bildfälschungen im amerikanischen Wahlkampf eine grosse Rolle. So zeigte Ron de Santis auf einer Wahlplattform ein Bild, auf dem Donald Trump den Immunologen Anthony Fauci umarmt und küsst. Die Bilder sind eingebettet in einen Videoclip mit realen Videoaufnahmen von Donald Trump. Sie sind nur kurz zu sehen. Vermischt mit den realen Bildern haben sie eine suggestive Wirkung und sind gefährlich, weil sie die Botschaft des Wahlclips untermauern.

### **KI als superschneller, supervergebbarer Assistent**

Wie gross ist der Beitrag, den die KI dabei leistet? Ich vermute, dass Norbert Kleinwächter ohne KI nicht in der Lage wäre, Robert Habeck oder Heiner Lauterbach in Monster zu verwandeln. Dazu bräuchte er eine Assistentin, die ihm die Bildbearbeitung abnimmt. Mit anderen Worten: Die KI erschliesst traditionelle Bildbearbeitung kostengünstig auch Laien und erhöht dabei das Tempo der Bearbeitung drastisch. Die KI übernimmt die Rolle eines Assistenten, der rund um die Uhr zur Verfügung steht. Gefährlich ist dabei nicht die KI-Leistung als solche, sondern die breite und kostengünstige Verfügbarkeit dieser Leistung. Wer also politisch manipulieren will und entsprechende Bilder im Kopf hat, kommt dank KI schneller und einfacher zu einem Ergebnis.

Wie sieht es mit Textmeldungen aus? Führt die KI auch zu mehr Fake News in Textform? Diese Frage ist, Stand heute, schwer zu beantworten, weil es dazu noch keine Studien gibt. Die Antwort lautet wohl: Im Prinzip ja ... Die KI ist eine extrem leistungsfähige Schreibmaschine, die zudem eine Vielzahl unterschiedlicher Sprachen, Schreibformen und Ausdrucksweisen beherrscht. Die KI kann problemlos eine Polizeimeldung, eine politische Rede oder irgendeine andere Textform imitieren. Dabei generiert die KI nicht nur die entsprechenden Formulierungen und Floskeln, sondern auch den Aufbau, die Ausgestaltung und die Unterschrift. Je präziser die KI dabei instruiert wird, desto besser ist das Ergebnis. Man kann die KI also mit einer Sekretärin (oder einem Sekretär) vergleichen. Statt wie in den 50er-Jahren Frau Müller zum Diktat zu bitten, wird die KI angewiesen, dies oder jenes zu schreiben.

### **Die KI weiss nicht, was sie schreibt**

Aber im Gegensatz zu Frau Müller damals hat die KI keine Ahnung, was sie schreibt. KI versteht kein Wort, sie kennt nur Wortstatistiken. Die KI schreibt nicht wirklich, sie imitiert den Schreibprozess und das in jedem beliebigen Stil, absolut furchtlos, ohne Ermüdungserscheinungen und in irrwitzigem Tempo. Wie leistungsfähig die KI dabei ist, hängt wie bei der Bildbearbeitung vom Menschen ab, der sie steuert. Wer genau weiss, was er will und das in präzise, KI-taugliche Befehle umsetzen kann, für den ist die KI extrem leistungsfähig. Man kann aber nicht an den Computer sitzen und der KI «nu mach mal» zuraunzen. Das funktioniert nicht. Auch das im Gegensatz zu den Frau Müllers in den 50er-Jahren, die durchaus in der Lage waren, das Gurren ihrer Chefs in einen wohlformulierten Brief zu übersetzen.

Das ist auch der zentrale Befund bis hier: Die KI kann ein extrem leistungsfähiges Werkzeug sein, wenn sie von einem Menschen bedient wird, der genau weiss, was er will. Dazu muss der Mensch nicht nur präzise Befehle geben, er muss vor allem wissen, welche Botschaft er vermitteln will. Das heisst: wie er die Verführungskraft der KI einsetzen will. Darüber hinaus muss er in der Lage sein, aus vielen mehr oder weniger gelungenen Anläufen und Versuchen das beste Resultat auszuwählen. Auch das erfordert Fingerspitzengefühl, Fachwissen und Erfahrung.

Wo liegt die Gefahr dabei? Bringt die KI die grosse Zeitenwende? Ich glaube nicht. Die hat nämlich schon stattgefunden. Aber nicht auf der Seite der Produzenten solcher Nachrichten und Bilder, sondern auf der Seite der Konsumenten, der Rezipienten. Es ist wie bei einer Zaubershow: Natürlich braucht eine solche Show einen guten Zauberer. Vor allem aber braucht eine Zaubershow ein Publikum, das bereit ist, Zauber zu sehen, wo schnöde Tricks am Werk sind. Mit anderen Worten: Das Problem ist nicht so sehr die durch die KI gestiegene Leistungsfähigkeit der Sender. Das Problem ist, dass das Publikum nicht mehr zwischen Fake und Wahrheit unterscheiden will. Mit Betonung auf «will».

### **KI malt uns die Welt, wie sie uns gefällt**

Viele Fake News sind einigermaßen gut zu erkennen. Das Publikum will sie aber gar nicht als Tricks entlarven, weil es lieber an die Magie des Zauberers glaubt, ganz egal, ob der nun Donald Trump oder Norbert Kleinwächter heisst. Eigentlich weiss man, dass der Zauberer nur trickst, aber man will sich den Spass daran nicht verderben lassen. Oder teilt das sichtlich von einer KI generierte Bild mit dem Kommentar «was für ein geiler Fake!». Das ist eine Reaktion, wie wir sie auch von unseren Kindern kennen, die genau wissen, dass dieser Videoclip von Redbull reine Werbung ist, den Clip aber trotzdem so cool finden, dass sie ihn nicht nur weiterverbreiten, sondern selbst auch Teil der Redbull-Welt sein möchten und deshalb die Koffeinbrause kaufen. Und, schlimmer noch, auch trinken.

Die grösste Gefahr sehe ich dort, wo Fake News nicht mehr als faule Tricks entlarvt werden, sei es, weil sie so gut gemacht sind, sei es, weil jemand genau das glauben will, was die Fake News aussagen, weil die Fake News präzise den Schwachpunkt der Menschen treffen. Und genau da lässt sich die KI eben auch einsetzen: im Aufspüren dieser Schwachstellen. Da reden wir aber nicht von einer generativen KI wie ChatGPT, sondern von intelligenten Steuerungsmechanismen, die aufgrund von Nutzungsdaten wissen, was die Nutzer auf Facebook, Twitter und Co. anklicken. Diese Art der künstlichen Intelligenz ist keine Zukunftsvision, sondern längst Alltag: Wir alle werden im Internet von den grossen Unternehmen ständig überwacht. Die Inhalte der sozialen Netzwerke und die Angebote von grossen E-Shops sind längst präzise individualisiert und auf unsere Schwachstellen zugeschnitten. Ihre Wirkung wird nun potenziert durch eine generative KI, die in der Lage ist, genau jene Inhaltselemente zu produzieren, die diese Schwachstellen ausnutzen.

Ist die KI also böse? Sind ChatGPT und Google Bard verderbliche Frankenstein-Programme? Wir machen es uns zu einfach, wenn wir die KI verteufeln. Wir können die Verantwortung nicht einfach auf das Werkzeug abschieben. Die Schuld liegt bei jenen Menschen, die KI skrupellos zur Fälschung von Nachrichten einsetzen – und bei denen, die sich von Fake News gerne verführen lassen. Das Problem der KI-generierten Fake News ist also kein Problem der Technik, sondern ein Problem der Ethik

und der Regeln, die wir uns als Gesellschaft im Umgang mit der KI auferlegen. Die Frage ist also: Welche Regeln brauchen wir im Umgang mit der KI? Wie müssen wir rechtlich und ethisch mit KI umgehen? Ich glaube, die Antwort auf diese Frage ist gar nicht so kompliziert. Dazu mehr nächste Woche.

Basel, 9. Juni 2023, Matthias Zehnder [mz@matthiaszehnder.ch](mailto:mz@matthiaszehnder.ch)

PS: Nicht vergessen – [Wochenkommentar abonnieren](#). Dann erhalten Sie jeden Freitag meinen Newsletter mit dem Hinweis auf den neuen Kommentar, einen Sachbuchtip, einen Tipp für einen guten Roman, das aktuelle Fragebogeninterview und mein persönlicher Tipp für ein gutes, digitales Leben. Einfach [hier klicken](#). Und wenn Sie den Wochenkommentar unterstützen möchten, finden Sie [hier](#) ein Formular, über das Sie spenden können.

## Quellen

aam (2023): *AfD Hetzt Mit Künstlich Generierten Bildern Gegen Flüchtlinge*. In: DER STANDARD. [<https://www.derstandard.de/story/2000144906234/afd-hetzt-mit-kuenstlich-generierten-bildern-gegen-fluechtlinge>; 9.6.2023].

Carnevale, Robert (2023): *Did Trump Really Kiss Fauci? DeSantis Ad Seemingly Uses AI Imagery*. In: TheWrap. [<https://www.thewrap.com/trump-kissing-fauci-fake-desantis-ad/>; 9.6.2023].

Caspar, Johannes (2023): *Wir Datensklaven. Wege aus der digitalen Ausbeutung*. Econ, 352 Seiten, 35.90 Franken; ISBN 978-3-430-21081-2; vgl. [<https://www.matthiaszehnder.ch/buchtip/wir-datensklaven/>]

Growcoot, Matt (2023): *Ron DeSantis Campaign Makes Trump Attack Ad With AI Imagery*. In: PetaPixel. [<https://petapixel.com/2023/06/08/ron-desantis-campaign-makes-trump-attack-ad-with-ai-imagery/>; 9.6.2023].

Stritzel, Sophia und Deuschle, Christoph (2023): *KI Und Medien: Wie Künstliche Intelligenz Uns Alle Täuschen Kann*. In: NDR, ZAPP. [<https://www.ndr.de/fernsehen/sendungen/zapp/KI-und-Medien-Wie-Kuenstliche-Intelligenz-uns-alle-taechen-kann,zapp14164.html>; 9.6.2023].

Weiser, Benjamin und Schweber, Nate (2023): *The ChatGPT Lawyer Explains Himself*. In: The New York Times. [<https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html>; 9.6.2023].

Zehnder, Matthias (2019): *Die digitale Kränkung. Über die Ersetzbarkeit des Menschen*. NZZ Libro Verlag, 128 Seiten, 24 Franken; ISBN 978-3-03810-409-4

**Spenden für den Wochenkommentar ist jetzt so einfach wie bezahlen im Hofladen**



**Bequem mit TWINT bezahlen**

Scannen Sie den QR-Code mit Ihrer TWINT App.

Geben Sie den Totalbetrag ein und bestätigen Sie Ihre Zahlung.

**Sie möchten lieber über einen Einzahlungsschein spenden? Kein Problem:**

[https://www.matthiaszehnder.ch/wp-content/uploads/2023/04/Spende\\_EZ\\_Wochenkommentar.pdf](https://www.matthiaszehnder.ch/wp-content/uploads/2023/04/Spende_EZ_Wochenkommentar.pdf)