GENERATIVE KI UND DEMOKRATIE[1]

JUDITH SIMON

Einleitung

Generative KI hat die Welt im Sturm erobert. Der jüngste Sommer der KI begann im November 2022 mit der Vorstellung von ChatGPT. Die Zahl der Nutzer:innen explodierte und allein in den ersten beiden Monaten überschritt ChatGPT die Schwelle von 100 Millionen, eine Benchmark, die bedeutende Social Media-Plattformen wie TikTok oder Instagram erst deutlich später erreichten.[2] Seitdem ist der Markt der Generativen KI substanziell gewachsen. In atemberaubender Geschwindigkeit werden neue Produkte und Services eingeführt. Generative KI ist nicht länger auf Text beschränkt, sondern steht jetzt auch für das Generieren von Bildern, Ton- und Videodaten zur Verfügung. Modelle und Tools wie Stable Diffusion, DALL-E und Gemini werden schon vielfach genutzt, während andere, etwa der Videogenerator SORA, zum Zeitpunkt des Verfassens dieses Texts

1 Aus dem Englischen von Sabine Weir

2 Zum Vergleich: TikTok brauchte neun Monate und Instagram zwei Jahre, um die gleiche Schwelle zu erreichen. Siehe: https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.

zwar schon angekündigt waren, aber noch nicht erschienen sind. Diese Tools werden immer häufiger auch in bestehende Services wie Suchmaschinen (ein Beispiel ist Bing) oder Office-Programme (wie Microsoft Copilot) sowie in organisatorische Prozesse in ganz unterschiedlichen Bereiche integriert. Da es nicht gelungen ist, diesen Prozess durch mehr oder weniger glaubwürdige Aufrufe zu Moratorien aufzuhalten oder auch nur zu verlangsamen, wird sich das Tempo der Entwicklung und Verbreitung in nächster Zeit wohl eher erhöhen anstatt zu verlangsamen. In der Folge wird Generative KI zunehmend die Gesellschaft in ihrer Breite beeinflussen und zu Umbrüchen im Journalismus und der Bildung, in der Wissenschaft, der Medizin und der Psychotherapie, in der öffentlichen Verwaltung und im Rechtswesen führen.

Kern Generativer KI ist die Fähigkeit, auf der Grundlage von aus riesigen Datenmengen exzerpierten Mustern, neue verbale oder visuelle Produkte zu erzeugen, deren Qualität immer besser wird. Der Unterschied zwischen Generativer KI und zuvor entwickelten Systemen liegt nicht allein in der verbesserten Performance, sondern auch in der Tatsache, dass diese Tools nicht mehr auf spezifische Bereiche beschränkt sind. Aufgrund der tragenden Rolle, die Sprache und Bilder für die zwischenmenschliche Kommunikation spielen, sollte die Fähigkeit der KI, Texte, Bilder oder Videos – von hoher Plausibilität, aber ohne Wahrheitsbezug – zu jedem denkbaren Thema zu generieren, nicht unterschätzt werden: Während Sprache das zentrale Medium menschlicher Kommunikation ist, sind Bilder und Videos in Zusammenhang mit Beweisführung, Zeug:innenschaft, Erinnerung sowie Emotionen besonders wichtig.

Neben der hohen Qualität der Ergebnisse und den breiten Anwendungsmöglichkeiten spielt noch ein weiterer Aspekt eine wichtige Rolle, der erklärt, warum ChatGPT und andere Tools, die auf Generativer KI basieren, einen noch nie dagewesenen Erfolg haben: die Benutzer:innenfreundlichkeit durch einfache Interfaces und und die freie Verfügbarkeit über das Internet. User:innen brauchen fast keine Vorkenntnisse und auch die technischen Anforderungen sind gering, um Texte, Bilder und Videos von sehr hoher Qualität in wenigen Sekunden zu produzieren und zu verbreiten. Eine einfache Anfrage als Prompt genügt, um in kürzester Zeit und mit geringem Aufwand Text oder Bilder zu erstellen und anpassen zu können.

All das führt zu einer schnellen Verbreitung von ChatGPT, Dall-E und Co – mit all den positiven wie negativen Folgen, die diese KI-Systeme mit sich bringen. Generative KI hat inzwischen viele Millionen regelmäßiger Nutzer:innen, Milliarden von Anfragen und entsprechenden Ergebnissen, die für eine Vielzahl von Zwecken genutzt und auch missbraucht werden können. Daher müssen wir die realen Herausforderungen und Gefahren für die Demokratie abwägen und bekämpfen – und dürfen uns nicht von Scheindebatten ablenken lassen. Zu diesen gehören Diskussionen über die Singularität und das Ende der Menschheit genauso wie die irreführende Debatte darüber, ob ChatGPT Anzeichen einer generellen Künstlichen Intelligenz zeigt und wirklich versteht oder gar ein Bewusstsein entwickelt. Klar ist: Keines der aktuell gebräuchlichen KI-Systeme verfügt über ein echtes Verständnis dessen, was es produziert, geschweige denn über Bewusstsein. ChatGPT erkennt basierend auf der Analyse großer Textmengen Sprachmuster, die Wahrscheinlichkeit von Wortkombinationen und linguistische Strukturen verschiedener Textgenres und produziert neue Texte auf der Grundlage dieser gelernten Muster. Zwar ließe sich argumentieren, dass diese Art des Erkennens linguistischer Muster auch die Grundlage menschlichen Verstehens ist, und dass multimodale KI-Systeme, die Texte und Bilder miteinander verbinden, in der Tat eine Art Vorstufe des „Verstehens" erreichen, doch scheint es weit hergeholt, zu behaupten, dies entspräche in vollem Umfang menschlichem Verstehen. Selbst wenn es also den Anschein macht, als verständen uns ChatGPT und Co, wenn sie auf unsere Prompts antworten, kann nicht oft genug betont werden, dass der generierte Output ausschließlich auf einer statistischen Analyse und der Reproduktion von Text basiert und nicht auf einem echten Verständnis des Inhalts.

Die Probleme der Täuschung

Diese Annahme zeigt allerdings ein zentrales Problem Generativer KI auf: das der Täuschung. Genauer: Generative KI verursacht mindestens vier verschiedene Probleme der Täuschung.[3]

3 Für eine umfassende Analyse der vier Arten der Täuschung durch Generative KI siehe: Simon (2024).

Zunächst einmal besteht die Gefahr, dass Nutzer:innen fälschlicherweise annehmen, sie würden mit einem Menschen und nicht mit einer Maschine interagieren, zum Beispiel bei der Kommunikation mit einem Kundenservice, oder – und das ist wesentlich problematischer – in therapeutischen Kontexten.

Abgesehen von diesem vordergründigsten Problem der Täuschung besteht ein weiteres Problem auch im Hinblick auf Täuschungen hinsichtlich der Fähigkeiten der KI. Auch wenn aktuelle KI-Systeme weder wirklich verstehen können noch ein Bewusstsein haben, kann es Nutzer:innen so erscheinen – und das selbst wenn sie wissen, dass sie mit einer Maschine interagieren. Diese menschliche Neigung Maschinen zu überschätzen zeigte sich nicht erst heute bei ChatGPT, sondern bereits in der Nutzung von Weizenbaums ELIZA (1966), einem Programm, das natürliche Sprache verarbeitet. Es ist mitunter nicht leicht zu unterscheiden, ob Nutzer:innen wirklich glauben, ChatGPT, Lambda und andere KI-basierte Chatbots verstünden sie oder hätten ein Bewusstsein, oder ob zumindest manche von ihnen ganz bewusst den KI-Hype-Cycle befüttern. Ohnehin sagt ein solches Zuschreiben von Fähigkeiten mehr über die menschliche Neigung aus, Technologie zu anthropomorphisieren, als über die Maschine selbst. Das Unvermögen, zwischen Sprech-Performanz und Denk-Kompetenz unterscheiden zu können, findet sich im Diskurs um Künstliche Intelligenz von Anfang an und lässt sich zurückverfolgen bis zum Turing-Test (1950) und dessen Kritik durch Searle (1981).[4]

Indem ich auf diese Verwischung der Differenz zwischen (nichtexistierenden) Denk-Kompetenz von Maschinen und der Art und Weise, wie Menschen von deren Sprech-Performanz getäuscht werden, hinweise,

4 Mit dem so genannten „Turing-Test“ (1950) behauptete Alan Turing, dass es ein Zeichen für maschinelle Intelligenz sei, wenn ein Mensch nicht unterscheiden könne, ob die Antworten auf seine Fragen von einer Maschine oder von einem anderen Menschen kommen. John Searle widersprach dieser Schlussfolgerung mit seinem berühmten Gedankenexperiment „Das chinesische Zimmer“ (1981), in dem er sagte, dass die erfolgreiche Manipulation chinesischer Symbole durch das Befolgen von Sprachregeln vom Verstehen der chinesischen Sprache, also dem Erkennen der Bedeutung dieser Symbole, unterschieden werden müsse.

möchte ich mitnichten diese menschlichen Zuschreibungsfehler belächeln. Im Gegenteil: vielmehr will ich vor der performativen Macht der Simulation warnen: Das Simulieren von Intelligenz, Auffassungsgabe oder sogar Emotionen und Mitgefühl, selbst wenn es eben nur simuliert ist, hat ernstzunehmende Folgen und macht uns als Menschen angreifbar. Wir reagieren auf besondere Weise kognitiv und emotional auf Sprache und Bilder – und das ist es, was diese neuen Technologien gleichermaßen mächtig und gefährlich macht.

Die dritte Form der Täuschung betrifft dann die Täuschungsergebnisse, die diese Systeme hervorbringen. Diese reichen von lustigen Fotos von Papst Franziskus in Alltagsklamotten zu Videos, in denen historische Figuren „auferstehen" oder Familienangehörige „weiterleben", von sogenanntem Racheporno über für Propagandazwecke produzierte Fake News und Deepfakes bis hin zum kriminellen Gebrauch gefakter Stimmen, um Verwandte zu täuschen. Gerade dieses Täuschungsproblem stellt eine ernstzuehmende Bedrohung für die gesellschaftliche Kommunikation und die Stabilität von Demokratien dar. Natürlich sind Täuschung, Propaganda und Manipulation keine neuen Phänomene. Doch die Leichtigkeit und Schnelligkeit, mit der heute Texte, Bilder, Tonspuren und Videos in hoher Qualität produziert und in Echtzeit über Social Media und Messenger-Dienste verbreitet werden können, eröffnet eine ganz neue Dimension für möglichen Missbrauch. Wird der öffentliche Raum mit gefälschten aber plausibel wirkenden Inhalten geflutet, stellt die Generative KI eine echte Bedrohung für unsere Demokratien dar, denn fundamentale Informations- und Kommunikationsprozesse können schnell, einfach und mit potenziell schwerwiegenden und nachhaltigen Auswirkungen gestört werden.

Die vierte und letzte Form der Täuschung betrifft Probleme, welche durch die Integration von Generativer KI in bestehende Services und Produkte entstehen, wie bspw. Suchmaschinen, E-Mail-Programme und Office-Suiten. Bereits bei seiner Einführung wurde ChatGPT als die Zukunft der Suche gefeiert obwohl die zugrundeliegenden Funktionsweisen und Zwecke sich fundamental unterscheiden. Dieses Verwischen der Unterscheidung zwischen Informationserstellung und Informationssuche erschwert die adäquate Bewertung und Nutzung von Informationen.

Zusammengenommen lässt sich sagen, dass diese vier Typen der Täuschung schwere epistemische, ethische und politische Schäden anrichten können. Täuschung führt nicht nur dazu, dass falsche Überzeugungen entstehen können, dass wir etwas glauben, was falsch ist. Die zunehmenden Schwierigkeiten bei der Bewertung des Wahrheitsgehalts von Informationen können auch das Vertrauen in Praktiken und Institutionen der Informationsbewertung insgesamt schwächen. Wenn Menschen nicht mehr davon ausgehen, dass sie Wahrheitsgehalt und Qualität von Informationen und deren Quellen verlässlich bewerten können, dann hat dies schwerwiegende Folgen für die öffentliche Kommunikation und die Demokratie.

Wie geht es jetzt weiter?

Was können wir nun angesichts der hier umrissenen Herausforderungen und speziell der Probleme der Täuschung unternehmen? Ich bin der Meinung, dass ein wirksames Vorgehen mehrere Instrumente miteinander verbinden muss. Regulierung, Technologieentwicklung oder Bildung sind je für sich genommen nicht ausreichend, aber zusammen stellen sie die beste aktuell verfügbare Möglichkeit dar, um den Herausforderungen zu begegnen, die Generative KI in Bezug auf die Demokratie mit sich bringt.

Regulierung kann über verschiedene Formen von Hard und Soft Law erfolgen. Ich bin insgesamt eher skeptisch, was die Selbst-Regulierung in diesem Kontext betrifft, da die Köpfe der Tech-Industrie bisher wenig ethische Sensibilität gezeigt haben und der Wettbewerb und Angst in diesem Wettbewerb zu unterliegen zu stark ist. Wir können uns also nicht darauf verlassen, dass die Stakeholder der Tech-Industrie ihre Produkte freiwillig angemessen kontrollieren, sondern müssen eine vernünftige demokratische Kontrolle dieser Technologien sicherstellen.

Im europäischen Kontext gibt es bereits eine Reihe von Gesetzen oder Gesetzesvorhaben, die sich mit KI befassen, wie die Datenschutz-Grundverordnung, das Gesetz über digitale Dienstleistungen und das Gesetz über digitale Märkte. Das wichtigste Gesetz in diesem Zusammenhang ist aber die europäische Verordnung über künstliche Intelligenz, die 2024 verabschiedet wurde. Die bereits vor der Einführung

der Generativen KI initiierte KI-Grundverordnung befürwortet einen risikobasierten Ansatz, bei dem die Regulierung der KI von dem jeweiligen Kontext und Einsatzgebiet abhängt. Sie schlägt daher spezifische Verpflichtungen für die Anwendung von KI vor, allerdings nur in Bereichen und Kontexten mit hohem Risiko wie bspw. der Medizin oder der Bildung. Zwar hat der Ansatz, nur kritische Anwendungsbereiche und nicht KI generell zu regulieren, auch Vorteile, doch die Generative KI, oder die sogenannte General Purpose-KI, also KI-Systeme mit allgemeinem Anwendungszweck, zeigt auch die Grenzen eines solchen Ansatzes auf, denn es gehört gerade zu ihren Wesenseigenheiten, dass sie über verschiedene Sparten und Bereiche hinweg angewendet wird.

Wie also sollen wir Generative KI regulieren? Lediglich in kritischen Bereichen und so dass die Verantwortung, sie zu regulieren, vor allem bei professionellen Entwickler:innen und Anwender:innen läge? Oder wollen wir die Produzent:innen von Generativer KI in die Verantwortung nehmen? Während des Trilogs zum KI-Gesetz im Dezember 2023 wurde über diese Fragen zwischen dem Europäischen Parlament, dem Europäischen Rat und der Europäischen Kommission hitzig debattiert. Am Ende einigte man sich darauf, den vorherigen Vorschlag für das KI-Gesetz zu ändern und Regeln für einflussreiche General Purpose-KI-Modelle, die künftig systemische Risiken verursachen können, mit aufzunehmen. Da das KI-Gesetz noch nicht in Kraft getreten ist, kann aktuell weder zu dessen Auslegung noch seiner Wirksamkeit etwas gesagt werden, und auch die Diskussion darüber, wie genau Verpflichtungen zur Achtung von Grundrechten und gesellschaftlichen Werten zwischen Hersteller:innen, Anwender:innen und (professionellen) Nutzer:innen solcher Systeme gerecht und effektiv verteilt werden können, muss noch geführt werden.

Für die oben geschilderten Probleme der Täuschung sind insbesondere auch Maßnahmen zur Erhöhung von Transparenz relevant. Die erste, naheliegende Lösung ist die Verpflichtung, KI-Content zu kennzeichnen, eine Maßgabe, die tatsächlich bereits in der KI-Verordnung der EU enthalten ist. Darüber hinaus werden derzeit technische Möglichkeiten entwickelt, mit denen sich Fakes erkennen oder echte Inhalte durch Wasserzeichen kennzeichnen lassen. Die Kennzeichnungspflicht und die technischen Entwicklungen sind wichtig, aber sie allein reichen nicht

aus, um gegen die Probleme der Täuschung vorzugehen. Zum einen kann damit weder krimineller Missbrauch noch informationelle Kriegsführung gänzlich verhindert werden. Zum anderen werden Menschen durch sie auch nicht davon abgehalten, Technologien Eigenschaften zuzuschreiben, die diese nicht haben. Es besteht also auch ein Bedarf, neue Normen und Kompetenzen für den Umgang mit diesen KI-Systemen zu entwickeln welche Möglichkeiten aber auch Grenzen dieser Systeme klar erkennen und benennen.

Mehr Transparenz kann auch auf der Ebene der Modelle selbst erreicht werden, über Open Access. Zwar steht ChatGPT kostenlos zur Verfügung, doch handelt es sich dabei um ein geschlossenes, proprietäres System. Dem gegenüber gibt es Open Source-Alternativen wie bspw. BLOOM oder Stable Diffusion, bei welchen die zugrundeliegende Technologie eingesehen, getestet und sogar modifiziert werden kann. Natürlich gehen mit dieser Offenheit wieder Probleme einher, denn Open Access ermöglicht auch neue Formen des Missbrauchs. Es wird also wichtig sein, sorgfältig zu prüfen, welche Formen der Offenheit die meisten Vorteile und die wenigsten Nachteile mit sich bringen. Ein einfacher und freier Zugang zu einem ansonsten völlig intransparenten und proprietären System wie im Fall von ChatGPT scheint die schlechtestmögliche Kombination zu sein.

Zum Abschluss möchte ich auf das Verhältnis von Generativer KI und Bildung in dreifacher Hinsicht eingehen. Zunächst einmal ist Bildung selbst ein Feld, welches von Generativer KI kontinuierlich vor große Herausforderungen gestellt wird da die Verfügbarkeit von Generativer KI grundlegende Fragen zu Wert und Wesen von Bildung eröffnet. Zweitens gilt der Bildungsbereich selbst in der KI-Grundverordnung als ein Hochrisikobereich für die Anwendung von (Generativer) KI. Und drittens ist Bildung in ihrer Breite und jenseits digitaler Themen zentral, um den Herausforderungen zu begegnen, die KI für die Demokratie mit sich bringt.

Der rasante Aufstieg von ChatGPT konfrontiert Universitäten und Schulen sehr schnell mit der Herausforderung, ihre Prüfungen so betrugssicher wie möglich gestalten zu müssen. Dabei ging es zunächst vor allem um die Frage, wie Fairness gewährleistet werden kann, wenn einige

Schüler:innen ChatGPT für ihre Hausarbeiten nutzen und andere nicht. Noch grundlegender ist jedoch der Umstand, dass ChatGPT die Möglichkeit eröffnet und auch die Notwendigkeit mit sich bringt, dass wir uns mit dem Wesen und dem Wert von Bildung auseinandersetzen. Wenn sogar Student:innen der Literaturwissenschaften ihre Essays von ChatGPT schreiben lassen und scheinbar wissenschaftliche Texte auf der Grundlage gefälschter Quellen verfasst werden, was sagt das über die Ziele von Bildung und die Rahmenbedingungen an Universitäten aus? Welche Fähigkeiten und Fertigkeiten müssen wir angesichts neuer technologischer Möglichkeiten neu erwerben, welche müssen wir uns ergänzend aneignen und welche werden womöglich nicht länger gebraucht? Der Deutsche Ethikrat hat einige Leitlinien zum Beantworten dieser Fragen in seiner Stellungnahme „Menschen und Maschine – Herausforderungen durch Künstliche Intelligenz" (Deutscher Ethikrat, 2023) zur Verfügung gestellt. Die zentrale Frage darin lautet, wie KI so gestaltet und genutzt werden kann, dass die menschliche Handlungsfähigkeit und die Autor:innenschaft der verschiedenen beteiligten Akteur:innen gestärkt anstatt geschwächt werden. Es liegt also nahe, dass Bildung auf allen Ebenen und in all ihren verschiedenen Formen ein umfassendes Verständnis des Wesens, der Voraussetzungen und der Folgen der technologischen Vermittlung unserer Lebenswelten beinhalten muss. Dazu gehört nicht nur wissenschaftliches, technologisches und mathematisches Wissen. Um die Früchte der Generativen KI zu ernten, ohne den Täuschungen durch sie zu erliegen, ist es notwendig, das mit ihr verbundene Wissen zu erweitern: um kritisches Denken, solides Wissen, Expertise und Erkenntnisse aus den Sozial- und Geisteswissenschaften sowie den Künsten. Dieses Wissen muss ein Teil der Ausbildung in Informatik und Data Science werden, um das Verantwortungsbewusstsein beim Gestalten und Entwickeln von KI-Technologie von Anfang an zu fördern. Wir müssen uns wieder ins Bewusstsein rufen, dass es nicht das grundlegende Ziel von Bildung ist, Lernende zu befähigen, verkäufliche Technologieprodukte zu nutzen, sondern die politische Reife und das Verantwortungsbewusstsein von Bürger:innen zu befördern. Letztendlich könnte dies eine der größten Herausforderung sein, wenn es darum geht, eine demokratische und nachhaltige Zukunft zu sichern.

Literaturverzeichnis

Deutscher Ethikrat. (2023, March 20). *Stellungnahme Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf

Searle, J. (1981). Minds, Brains, and Programs, *Behavioral and Brain Sciences*, 3, 417–57. https://doi.org/10.1017/S0140525X00005756

Simon, J. (2004/forthcoming). Generative AI, Quadruple Deception & Trust, *Social Epistemology, Special Issue: The Mind-Technology Problem: Rethinking Minds, Humans and Artefacts in the Age of AI*.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–60. https://doi.org/10.1093/mind/LIX.236.433

Weizenbaum, J. (1966). ELIZA-A Computer Program for the Study of Natural Language Communication Between Men and Machines. *Communications of the ACM*, 9, 36–45.