

Macht und Bilder: KI und politische Meinungsbildung

5. Juni 2024, 17:30 Uhr

Online

Begrüßung.....	2
Judith Simon · Deutscher Ethikrat.....	2
Mark Schweda · Deutscher Ethikrat.....	2
Informatik – Computer Vision & Learning.....	2
Björn Ommer · Ludwig-Maximilians-Universität München	2
Thematische Einführung.....	5
Judith Simon · Deutscher Ethikrat.....	5
Politische Theorie und Philosophie	7
Lisa Herzog · Universität Groningen (NL)	7
Wissenschaftskommunikation und Medien	10
Juliane A. Lischka · Universität Hamburg	10
Diskussion.....	13
Moderation: Judith Simon · Deutscher Ethikrat.....	13
Publikumsanwalt: Mark Schweda · Deutscher Ethikrat.....	17
Schlusswort.....	23
Judith Simon · Deutscher Ethikrat.....	23

Hinweis: Bei dem folgenden Text handelt es sich nicht um eine wörtliche Transkription. Der Text wurde lektoriert, um eine gute Lesbarkeit herzustellen.

Begrüßung

Judith Simon · Deutscher Ethikrat

Sehr geehrte Damen und Herren, liebes Online-Publikum, ich begrüße Sie herzlich zu unserem Webinar „Macht und Bilder. Künstliche Intelligenz und politische Meinungsbildung“. Ganz besonders herzlich möchte ich zunächst unsere drei Referentinnen begrüßen. Dies sind Frau Professor Lisa Herzog von der Universität Groningen, Frau Professor Juliane Lischka von der Universität Hamburg und Herr Professor Björn Ommer von der LMU in München.

[Aufgrund von technischen Problemen wurde die weitere Begrüßung und Moderation von Mark Schweda übernommen.]

Mark Schweda · Deutscher Ethikrat

Meine sehr geehrten Damen und Herren, ich begrüße Sie nochmal ganz herzlich zu dieser Veranstaltung des Deutschen Ethikrates „Macht und Bilder. KI und politische Meinungsbildung“. Wir wollen gemeinsam in den folgenden zwei Stunden miteinander diskutieren, wie sich neue Formen von KI-Anwendungen im Bereich der Bildgenerierung auf öffentliche Debatten auswirken können, insbesondere auf politische Auseinandersetzungen in der Öffentlichkeit, wie wir sie jetzt im Vorfeld der vielen Wahlen im Jahr 2024 erleben. Dazu haben wir drei Expertinnen und Experten eingeladen. Judith Simon, Mitglied des Deutschen Ethikrates, hat eine Einführung in die Thematik vorbereitet, die sie uns gleich vortragen wird. Um aber keine weitere Zeit zu verlieren, möchte ich zunächst Herrn Ommer bitten, seinen Vortrag zu starten.

Informatik – Computer Vision & Learning

Björn Ommer · Ludwig-Maximilians-Universität München

Vielen Dank. Es freut mich, heute zu Ihnen zu sprechen.

Generative KI hat in kürzester Zeit begonnen, diverse Anwendungsfelder zu transformieren. Erstens stellt sich die Frage, was eigentlich der Kern dieser Transformation ist. Was sind zweitens die Konsequenzen für unsere Gesellschaft und insbesondere für die Meinungsbildung, über die wir heute sprechen möchten? Und wie können wir drittens verantwortungsvoll mit dieser Technologie umgehen?

Sie sehen hier ein Beispiel, warum sich generative KI so schnell durchsetzen konnte. Ich denke, generative KI ist so schnell in unserer Gesellschaft angekommen, weil sie zum Nachbrenner für Fertigkeiten und Fähigkeiten und Tätigkeiten geworden ist, die wir bereits am Personal Computer gemacht haben. Das heißt, der Computer ist dadurch für gewöhnliche Benutzer zu einem leistungsfähigeren und potenteren Werkzeug geworden. Ich kann jetzt mit natürlicher Sprache dem Computer Anweisungen geben und der setzt sie um, ohne mich spezifisch in eine Applikation hineinfuchsen zu müssen oder gar programmieren zu können.

Ich kann eine Szene beschreiben, und ein System – Stable Diffusion, das wir entwickelt haben – setzt das beispielsweise in ein Bild um. Die dabei gelernte Repräsentation kann dann für andere Aufgaben verwendet werden, um in Bildern etwas suchen zu können, um Klassifikationsaufgaben oder Ähnliches zu bearbeiten.

Gleichzeitig kann ich auch Oberflächen von Programmen skizzieren. Ich kann aber auch Bilder,

die ich gerne hätte, vorzeichnen und den Computer das dann in Programmcode oder in ein richtig ausgestaltetes Bild verwandeln lassen.

Das Neue an generativer KI ist dabei, dass ich ein neuronales Netz nicht mehr mit einer spezifischen Aufgabe trainiere, beispielsweise ein neuronales Netz auf Krebsklassifikation trainiere, sondern dass ich die Trainingsdaten, Bilder beispielsweise, nehme und das Netz bitte, sie zu vervollständigen, oder dass ich Text nehme und dieses neuronale Netz darauf trainiere, diesen Text zu vervollständigen.

Das Interessante ist, dass dabei nicht einfach nur ein Papagei herauskommt, der nachplappert, was ich im Training vorne reingesteckt habe, sondern dass sich Emergenzeffekte gezeigt haben. Es hat sich gezeigt, dass das System die Datenpunkte, die ich ihm während des Trainings gezeigt habe, verbindet und dazwischen Neuland erschließt, interpolieren kann und neue Dinge erschließen kann, die es im Training nicht gesehen hat. Eine Extrapolation, etwas komplett Neues zu finden, ist eine deutlich schwierigere Aufgabenstellung.

Man kann zusammenfassen, dass durch generative KI der Personal Computer zugänglicher geworden ist und ein leistungsfähigeres Werkzeug geworden ist für gewöhnliche Nutzer, die nicht programmieren können, die damit anfangen können, gewissermaßen die Magie zu bewerkstelligen, die man vorher nur durch Programmierung durchführen konnte, und dass der leistungsfähigere PC dadurch demokratisiert worden ist. Das Spannende ist, dass diese Form von generativer KI ein großes Potenzial hat, aber wie jede leistungsfähige Technologie gebraucht, aber auch missbraucht werden kann.

Interessant finde ich hierbei, dass wir Bilder, Videos und anderen multimedialen Content immer *über* den Text gestellt haben und gesagt haben, dass Text natürlich erfunden werden kann. Aber

Bilder haben für uns eine besondere Rolle gespielt, die sie eigentlich nicht als Alleinstellungsmerkmal haben sollten.

Wir sehen hier, dass die Fotografie, solange sie existiert, bereits gefälscht worden ist. Schon bevor es neue generative KI gegeben hat, gab es Shallowfakes, hier aus dem vormaligen amerikanischen Präsidentschaftswahlkampf, wo ein Video von Nancy Pelosi langsamer abgespielt wurde, um sie als alt und debil darzustellen. Es gab ganze Youtube-Channel, die Prominente entsprechend in anderen Kontexten dargestellt haben.

Das Fazit, das man daraus ziehen kann, ist, dass sich die Implikationen der neuen generativen KI-Technologien oder die Technologie der generativen KI sich als ein Vergrößerungsglas zeigt für Probleme, die wir vorher schon hatten, für Hausaufgaben, die wir vorher nicht genügend bewerkstelligt haben.

Schauen wir uns das am Beispiel von Falschnachrichten an. Wir haben vorher Quellen vertraut, die eigentlich nicht vertrauenswürdig waren. Wir haben Individuen, die einen enorm großen Einfluss haben und sehr viele Follower haben, potenziell mehr als die Abendnachrichten, aber nicht den Pressekodex unterlegen, die diese Nachrichten unterlegen. Und wir haben eine Aufmerksamkeitsökonomie, die betrieben wird und algorithmisch unterstützt wird, sodass es dabei auch zu algorithmischen Selbstverstärkungsprozessen kommt, bei denen Meinung gemacht wird und gewissermaßen aus dem Nichts Polarisation betrieben werden kann.

Was ist die Lösung? Wenn wir begreifen, dass die Probleme eigentlich schon vorher da waren, dass sie unter dieser Technologie nur wie unter einem Brennglas verstärkt werden, was können dann die Lösungen sein?

Lassen Sie mich als Erstes das Bild herausgreifen. Das einzelne Bild kann sich nicht aus sich selbst rechtfertigen. Wenn ich in einem Kriegsgebiet bin und ein briefmarkengroßes Bild bekomme, das verwaschen, verwischt ist, dann kann ich nicht herausfinden, ob dieses selbst denn korrekt ist. Das funktioniert nicht. Das haben wir in der Vergangenheit auch schon gesehen.

Wir haben, wie Sie hier am Beispiel des Abschlusses der MH 17 sehen, nur eine Möglichkeit, indem wir nicht nur ein Bild, sondern sehr viele Bilder ansehen, aber indem wir darüber hinaus andere Zeugen, Augenzeugen befragen und uns selbst ein Bild der Lage machen. Die Wahrheit aus dem einen Bild herauszufinden und dementsprechend aufgrund eines einzigen Bildes einer Fälschung zu unterliegen, da sollten wir sehr vorsichtig sein. Also investigativer Journalismus, wenn man so will.

Der zweite Punkt ist der Einzelne. Hier zeigt sich, dass ein großes Bewusstsein in der Bevölkerung da ist: 85 Prozent der Bevölkerung sehen in Desinformation im Internet ein großes Problem, wie diese Bertelsmann-Studie gezeigt hat¹. 70 Prozent sehen die Gefahr, dass sich andere dadurch beeinflussen lassen. Nur 16 Prozent denken jedoch, dass sie selber gefährdet sind, und wenn ein einzelner Klick hier den Unterschied macht in der Verbreitung von Falschnachrichten, dann können Sie sich die Probleme dabei dann vorstellen.

Lassen Sie mich einen dritten Aspekt herausgreifen, und das ist unsere Bevölkerung als Ganzes. Was Sie hier, in dieser Allegorie sehen, ist ein Kohlenstoffatom. Wenn Sie dieses Kohlenstoffatom mit seinen direkten Nachbarn verbinden, dann entsteht ein Gitter, ein Netz, was das zusammenhält und zum stärksten Element, Material auf unserem Planeten macht: einem Diamanten.

Wenn Sie diese direkten Nachbarschaftsbeziehungen durch beliebige, Freunde in diesem Netz austauschen, dann wird aus diesem härtesten Material nur noch Asche, reiner Kohlenstaub.

Was ich damit sagen will, ist, dass wir Menschen soziale Wesen sind, die von diesen direkten Abhängigkeiten leben, von dem direkten Feedback, das diese Abhängigkeiten zueinander mitbringt, mit Leuten, mit denen wir Lebenswirklichkeit teilen. Wenn wir das durch beliebige, weit entfernte Kontakte in einem sozialen Netzwerk ersetzen, dann bekommen wir ein Problem, denn Vertrauen skaliert nicht mit der Konnektivität in einem solchen Netzwerk beliebig. Dann zeigt sich gerade das Gegenteil, dass das Vertrauen abnimmt.

Lassen Sie mich mit einer positiven Nachricht schließen. Unsere Gesellschaft ist sehr heterogen, und wie wir im politischen Bereich sehen, gibt es einen hohen Grad an Polarisierung und unterschiedliche Pole, bei denen unterschiedliche Meinungen propagiert werden. Die Datenanalyse, Data Analytics, hat über Jahre versucht herauszufinden, wie man möglichst viele Menschen direkt ansprechen kann. Dabei hat sich gezeigt, dass es sehr schwierig ist, Punkte zu finden, mit denen man viele Menschen zusammenbringt, dass es aber viel einfacher ist herauszufinden, was Menschen voneinander trennt. Diese Triggerpunkte algorithmisch zu identifizieren – das ist das, was wir in den letzten Jahren und Jahrzehnten gesehen haben – funktioniert sehr einfach und ermöglicht es, eine One-size-fits-all-Kommunikation durchzuführen, eine Kommunikation, die alle über einen Kamm schert. Die Kommunikation wird dadurch immer lauter, immer einfacher und immer polarisierender und grenzt Teile von anderen Teilen der Bevölkerung ab.

¹ Bernhard et al., „Verunsicherte Öffentlichkeit“, Bertelsmann Stiftung, 2024.

Wenn wir jetzt aber der Bevölkerung zugestehen, dass sie nicht eindimensional ist, sondern dass wir, egal wie das Lager ist, auf jeder Seite sehr hohe Grade an Heterogenität sehen, wenn wir uns andere Dimensionen ansehen, dann wird daraus ein Punkt, nämlich: Generative KI ermöglicht es uns, die Kommunikation mehr zu personalisieren, um auf diese Art und Weise Individuen, kleinere Teile der Bevölkerung leiser und sehr direkt anzusprechen und auf diese Art und Weise komplexere, abgewogenere Nachrichten rüberzubringen. Ich bin mir bewusst, dass zuerst die Gefahr gesehen wird, dass man mit generativer KI auch polarisieren kann. Aber wie ich eben gezeigt habe, brauche ich für diese Triggerpunkte keine komplexe Technologie mehr.

Lassen Sie mich schließen. Wir sehen großes Potenzial darin, den Computer zugänglicher und leistungsfähiger für die allgemeine Bevölkerung zu machen und ihn dadurch zu demokratisieren. Natürlich hat eine leistungsfähige Technologie das Potenzial, missbraucht zu werden. Ein großer Fehler lag aber darin, dass wir Bilder als viel vertrauenswürdiger gesehen haben als andere Nachrichtenformen, und das sind sie nicht. Viele von den Problemen, die wir gehabt haben, waren vorher schon da (Shallowfakes und ähnliche Dinge), und diese neuen Technologien werden ein Vergrößerungsglas für Probleme, die wir schon vorher hatten. Der Einzelne kann aber einen Unterschied machen, hatte dafür aber das Bewusstsein nicht direkt, und in den sozialen Netzwerken besteht das Problem, dass die Konnektivität und das Vertrauen leider nicht miteinander einhergehen, sondern genau das Gegenteil der Fall ist.

Damit bedanke ich mich für Ihre Aufmerksamkeit.

Thematische Einführung

Judith Simon · Deutscher Ethikrat

Herzlichen Dank für den tollen Vortrag. Bei mir ist leider für ein paar Minuten das Internet ausgefallen. Jetzt bin ich wieder da und würde das wiederholen, was ich vorher gern hätte sagen würden, und danach gebe ich direkt weiter an den nächsten Vortrag von Lisa Herzog.

Zunächst der Hinweis, dass dieses Webinar jetzt das zweite ist. Wir haben im März 2023 die Stellungnahme zu künstlicher Intelligenz veröffentlicht; in dem Kontext haben wir uns vor allen Dingen mit KI zur Entscheidungsunterstützung beschäftigt und noch nicht mit generativer KI, die in den folgenden Monaten viel Aufmerksamkeit bekommen hat. Wir haben dann im Dezember das erste Webinar „KI im Klassenzimmer“ gemacht, vor allem zu ChatGPT. Heute beschäftigen wir uns insbesondere mit Bildern, und da am Sonntag auch Europawahlen sind, ist es sicherlich ein guter Zeitpunkt, sich einmal Gedanken zu machen über die politische Wirkung von solchen Bildern.

Ein paar Worte zur Stellungnahme: Sie ist mit 403 Seiten sehr umfangreich und hat drei Teile. Im ersten Teil legen wir die technologischen, aber auch die ethischen und philosophischen Grundlagen dar. Dann schauen wir uns den Einsatz von künstlicher Intelligenz in vier Anwendungsbereichen an: in der Medizin, in der schulischen Bildung, in der öffentlichen Kommunikation und Meinungsbildung, also dem Bereich, dem wir uns heute auch widmen (damals noch nicht mit dem Fokus auf generative KI, sondern allgemeiner auf Content Moderation, auf das Generieren oder auf die Auswahl von Technologien), und in der öffentlichen Verwaltung. Wir haben zehn Querschnittsthemen identifiziert, auf die ich später noch eingehen werde.

Der Fokus unserer Stellungnahme war die Mensch-Technik-Relation. Wir haben uns angeschaut die Wechselwirkungen zwischen Mensch und Maschine oder genauer: Was passiert, wenn wir Tätigkeiten, die zuvor Menschen durchgeführt haben, delegieren an Maschinen, menschliche Tätigkeiten durch Maschinen ersetzen? In welcher Art und Weise erweitert oder vermindert dies die Handlungsmöglichkeiten, die Handlungsoptionen (wir haben es auch Autorschaft genannt) der unterschiedlichen Beteiligten und Betroffenen?

Wir haben wie gesagt bei dem Blick auf den Einsatz von KI in diesen vier Sektoren unterschiedliche Querschnittsthemen identifiziert, die immer wieder, wenn auch nicht unbedingt auf die gleiche Art und Weise hervorgetreten sind. Ich habe die unterschiedlichen Themen farblich gruppiert:

Das sind einerseits viele Aspekte, die oft diskutiert werden. Aufgrund der Tatsache, dass diese KI-Systeme datenbasierte Systeme sind, entstehen bestimmte Herausforderungen in Bezug auf den Schutz der Privatsphäre und Fragen der Überwachung, aber auch von systematischen Verzerrungen und Diskriminierungen: Behandeln diese datenbasierten Systeme Personengruppen gleich oder ungleich? Natürlich gibt es Diskussionen rund um Intransparenz und mangelnde Nachvollziehbarkeit von insbesondere Deep Learning.

Ein großer Teil, den wir uns angeschaut haben mit dem Fokus der Stellungnahme, ist, in welcher Art und Weise sich KI auf unsere Handlungsmöglichkeiten auswirkt, diese erweitert oder vermindert. Die Auswirkung auf menschliche Kompetenzen, aber auch auf Fragen der Autonomie sind hier blau markiert.

Gelb markiert ist ein Punkt, der auch heute wichtig sein dürfte, wenn wir generative KI diskutieren, da viele dieser digitalen Technologien mittlerweile kritische Infrastrukturen sind und wir uns

hier auch Fragen stellen müssen in Bezug auf Abhängigkeiten oder Missbrauch von diesen Technologien.

Der letzte Punkt, hier grün markiert, ist die Frage, was es mit uns macht, wenn wir in zunehmend vielen Bereichen unseres gesellschaftlichen und alltäglichen Lebens über Statistik und Wahrscheinlichkeitstheorie Voraussagen treffen über Menschen und dementsprechend auf einer Wahrscheinlichkeitstheoretischen Art und Weise über Menschen, Individuen und Gesellschaft nachdenken und welche Folgen das hat.

Das Fazit unserer Stellungnahme war: Wenn wir KI einsetzen, sollte das immer das Ziel haben, menschliche Handlungsmöglichkeiten und Autorschaft zu erweitern und Verminderungen zu verhindern. Die Auswirkungen sind jedoch sehr unterschiedlich für verschiedene Betroffene. Deswegen muss man den Fokus insbesondere auf jene legen, die bereits vulnerabel sind. Aber hier – ganz wichtig – steckt der Teufel im Detail und deswegen brauchen wir einen genauen Blick auf Technologien, aber auch auf institutionelle und organisationale Rahmenbedingungen.

Ich schlage die Brücke zur Frage: Was ist denn neu an generativer KI? Welche neuen Herausforderungen gibt es da? Das schlägt auch ein bisschen die Brücke zu dem, was Herr Ommer schon ausgeführt hat.

Einer der Gründe für diese extrem schnelle Durchdringung und die disruptiven Effekte, die sich durch diese unterschiedlichsten Technologien der generativen KI zeigen, ist die Breite der Anwendbarkeit. Bild und Sprache haben für Information, Kommunikation und Emotionen eine zentrale Bedeutung. Dementsprechend sind diese Technologien der generativen KI so breit anwendbar.

Darüber hinaus haben wir es mit Technologien zu tun, die geringe Hürden aufweisen sowohl in Bezug auf die technischen Voraussetzungen, um sie zu nutzen, als auch in Bezug auf die Kompetenzen der Nutzerinnen und Nutzer. Es gibt einen freien Zugang zu Basistechnologien oder Basisfunktionen für viele Tools online, und es gibt sehr einfache Interfaces, mit denen nahezu jeder über simple Prompts mit diesen Technologien interagieren kann.

Das Dritte ist sicher die hohe Geschwindigkeit, mit der diese Materialien hergestellt werden, und die Geschwindigkeit, mit der dieses Material über soziale Medien weiterverbreitet werden kann. In dem Fall ist da schon eine Veränderung zu beobachten im Vergleich zu vorherigen Technologien.

Dadurch ergeben sich bestimmte Probleme und Herausforderungen. Das eine ist, was ich das doppelte Täuschungsproblem nennen würde: teilweise eine Täuschung darüber, in welcher Art und Weise diese Bilder, Materialien oder Texte generiert werden (nämlich auf eine rein statistische Art und Weise), und dass man sich klarmachen muss, dass diese Bilder, aber auch Texte eine hohe Qualität haben, aber nicht notwendigerweise einen Wahrheitsbezug. Davon abhängig geht es auch um gezielte Täuschung, gezielte Manipulation für Deepfakes und Propaganda. Auch das ist nichts Neues, aber die Geschwindigkeit und die Einfachheit hat sich verändert.

Daraus ergibt sich die Frage in Bezug darauf, was ich ein doppeltes Vertrauensproblem nennen würde. Das eine ist die Gefahr, dass ich Fehlinformationen vertraue und dementsprechend in die Irre geführt werde oder mir etwas Falsches aneigne, und auf der anderen Seite vielleicht das noch viel größere Problem, dass, wenn ich den Eindruck habe, nicht mehr zwischen wahren und Falschmeldungen unterscheiden zu können oder

zunehmend schwieriger unterscheiden zu können, dass vielleicht auch mein Vertrauen in vertrauenswürdige Informationen und Informationsanbieter insgesamt sinkt.

Da ist wieder der Bogen zurück zur Stellungnahme: Alle herkömmlichen Herausforderungen für KI, die wir vorher beschrieben haben, rund um Bias und Diskriminierung, mangelnde Transparenz und Nachvollziehbarkeit, aber auch Fragen zum Schutz der Privatsphäre und des geistigen Eigentums, aber auch so etwas wie prekäre Arbeitsbedingungen in der Herstellung oder der Energieverbrauch, spielen auch bei generativer KI eine große Rolle.

Damit beendete ich meinen Kurzinput zur Stellungnahme und zur generativen KI und gebe das Wort ab an Lisa Herzog.

Politische Theorie und Philosophie

Lisa Herzog · Universität Groningen (NL)

Herzlichen Dank. Ich knüpfe da an, wo Judith Simon aufgehört hat, nämlich bei der Frage nach Vertrauen. Demokratie benötigt den öffentlichen Diskurs, das ist klar, und dieser Diskurs benötigt Vertrauen und Vertrauenswürdigkeit.

Das Phänomen Vertrauen ist asymmetrisch: Es ist sehr leicht zu zerstören und schwer wieder aufzubauen. Das kann man sich aus zwischenmenschlichen Beziehungen sehr gut erklären, gilt aber auch gegenüber Institutionen oder Organisationen. Und wie Judith Simon ausgeführt hat, reicht schon die Möglichkeit, dass bestimmte Inhalte KI-generiert und falsch sind, obwohl sie noch so echt aussehen, um auch das Vertrauen in andere Inhalte zu zerstören oder zumindest brüchig werden zu lassen, denn man weiß nicht mehr, welche der Bilder echt sind und welche nicht, und wie Herr Ommer ausgeführt hat: Wir haben vielleicht

zu lange zu sehr den Bildern vertraut, wurden immer schon manipuliert. Aber durch den Einsatz dieser neuen Technologien ist es noch viel einfacher geworden.

Es gibt inzwischen nicht nur Bilder, sondern auch Videos, die gefälscht werden, auf eine sehr authentisch wirkende Art und Weise. Für den öffentlichen Diskurs ein besonders schwieriges Beispiel sind Fälle, wo Videos von Menschen, die sprechen, mit einem anderen Text unterlegt werden und dann die Mundbewegungen so angepasst werden, dass es so aussieht, als würde diese Person bestimmte Aussagen machen, obwohl das überhaupt nicht der Fall ist. Diese Art von Deepfakes, wie die heißen, ist besonders gefährlich, weil wir es einfach gewohnt sind, uns auf Videos, wo Leute etwas sagen, zu verlassen, und das nicht mehr unbedingt der Fall ist.

Generell gibt es, auch das hat Judith Simon schon angesprochen, viel Regulierungsbedarf beim Gebrauch von KI-generierten Inhalten und algorithmischen Inhalten allgemein. Zum Beispiel die Kuratierung, das bedeutet: Was wird mir in sozialen Medien eigentlich angezeigt? Das wird algorithmisch gesteuert, aber diese Algorithmen gehören den Firmen, die die Plattformen bereitstellen, und aus Sicht der Verbraucherinnen und Verbraucher ist nicht ganz klar, was da eigentlich passiert. Es ist klar, es geht irgendwie darum, die Aufmerksamkeit zu bündeln. Personalisierte Werbung ist ein weiterer Bereich. Man erhält Werbung, auch Wahlwerbung angezeigt und weiß aber nicht: Wird die jetzt breit gezeigt oder nur einem selber gezeigt?

Bei KI-generierten Inhalten gibt es noch mal spezifische Punkte, wo besonders viel und dringender Regulierungsbedarf herrscht. Hier geht es oft um die Persönlichkeitsrechte Betroffener. Ein Beispiel, wo das besonders offensichtlich ist, sind

Sexvideos, wo die Köpfe von Personen auf irgendwelche Pornokörper montiert werden, um Leute damit zu diskreditieren, und das zirkuliert dann online. Es ist klar, dass das mit Persönlichkeitsrechten nur schwer vereinbar ist.

Eine zweite Frage betrifft Offenlegungspflichten, also: Müssen Inhalte, die KI-generiert sind, als solche gekennzeichnet werden? Und wie kann dadurch die Vertrauenswürdigkeit des öffentlichen Diskurses insgesamt bewahrt werden?

Ich sehe durchaus die Möglichkeit eines Verbots von Deepfakes in spezifischen Kontexten, zum Beispiel im Wahlkampf. Das ist ein starker Eingriff, aber man könnte sich vorstellen, dass so etwas vor bestimmten wichtigen Wahlen durchaus gerechtfertigt wäre.

Bei den Regulierungsansätzen scheint mir eine wichtige Frage zu sein, für wen man eigentlich reguliert und welches Menschenbild dabei die Gesetzgebung anleitet. Es ist natürlich immer sehr schmeichelhaft, von einem Bild der mündigen Bürgerinnen und Bürger auszugehen, aber alle von uns neigen dazu, sich zu überschätzen. Das hat die Studie, die Herr Ommer zitiert hat, schön gezeigt. Man sieht die Fehler immer bei den anderen, nicht bei sich selbst.

Es gibt viel psychologische Forschung dazu, die auch zeigt, in welchen geistigen Zuständen Menschen besonders anfällig für Manipulation sind. Es gibt das Stichwort Scarcity Mindset, also der Geisteszustand, wenn man Knappheit erfährt. Das ist für die Nutzung von Online-Medien besonders interessant, und zwar wenn man sich in einer Situation befindet, in der man irgendetwas als knapp empfindet (wenn man zum Beispiel Zeitnot hat, Geldnot erlebt oder das Gefühl hat, man muss jetzt irgendeine Aufgabe schnell lösen, und fühlt sich unter Druck gesetzt), dass sich dann das menschliche Gehirn auf dieses Problem fokussiert und alles andere ausblendet oder zumindest

an den Rand des Tunnels drängt und man nur noch nach vorne gerichtet schaut, wie man jetzt das anstehende Problem lösen kann. Die gleichen Personengruppen, das hat man in Experimenten festgestellt, sind unterschiedlich gut im Lösen kognitiver Aufgaben, je nachdem, ob sie gerade in so einem Knappheitszustand sind oder nicht.

Natürlich ist jeder und jede von uns immer wieder mal in solchen Situationen. Man ist übermüdet, man hat gerade Zeitnot, es kommen drei Anrufe und fünf Apps gleichzeitig rein und dann scrollt man noch über eine Webseite oder muss dringend irgendwohin oder dringend irgendwas erledigen. Das sind Situationen, in denen Menschen nicht gut in der Lage sind, sich auf das zu fokussieren, was ihnen online vorgesetzt wird, und wo sie dann zum Beispiel auch eine kleine Fußnote irgendwo in der Ecke eines Bildes (übrigens auch KI-generiert) sehr leicht übersehen würden und man sich die Frage stellen muss: Bräuchten wir nicht eine viel deutlichere, psychologisch informierte Art und Weise, wie solche Markierungen erfolgen?

Das ist das eine, dass jeder und jede von uns in solchen Situationen sein kann. Das andere ist aber, dass hier auch eine Gerechtigkeitsfrage dahintersteht, und zwar insofern, als es benachteiligte Gruppen gibt, die sich strukturell viel öfter als andere in derartigen Situationen befinden. Zum Beispiel ist gut erforscht, über viele westliche Gesellschaften hinweg, dass alleinerziehende Eltern kleiner Kinder sehr zeitarm sind, und wenn die sich online irgendwelche Sachen anschauen, übermüdet sind, das Kind hat die Nacht hindurch geschrien, kann man sich sehr gut vorstellen, dass, ohne dass man sie dafür moralisch zur Verantwortung ziehen wollte, ihre Aufnahmefähigkeit einfach anders aussieht. Wenn noch Geldnot dazukommt, weil sie sozioökonomisch benachteiligt sind, dann wird es noch schlimmer. Solche Gruppen haben es tendenziell schwerer, und als

demokratische Gesellschaft müssen wir uns die Frage stellen, wie wir auch diese Gruppen schützen können, sodass sie trotzdem weiterhin auch am öffentlichen Online-Diskurs teilnehmen können.

Es gibt in der Philosophie das Stichwort epistemische Gerechtigkeit. Epistemisch heißt, vom Altgriechischen herkommend, dass es um Wissen geht. Epistemologie ist die Lehre vom Wissen. Epistemische Gerechtigkeit meint Gerechtigkeit für Individuen als Trägerinnen und Träger von Wissen. Das wurde von Miranda Fricker und anderen in der Philosophie in den letzten Jahren stark diskutiert: Wer hat mehr Glaubwürdigkeit? Wer hat die Ressourcen, um die eigene Welt sich darzulegen? Da sieht man, dass es viele subtile und weniger subtile Unterschiede gibt und dass bestimmte Gruppen in der Gesellschaft oft ernster genommen werden mit ihren Wissensinhalten, die sie vortragen und in die Öffentlichkeit bringen, als andere.

Auch als Empfängerinnen und Empfänger von Informationen gibt es so eine Frage von Gerechtigkeit: Wer hat Zugang zu vertrauenswürdiger Kommunikation? Denn in Online-Medien sind der Empfang und die Reaktion immer sehr eng miteinander verbunden. Das ist ja gerade der Reiz, dass man es teilen, kommentieren kann usw. Wenn wir aber wollen, dass diese Diskurse wirklich epistemisch gerecht funktionieren, dann muss der Maßstab sein, dass es demokratische Teilhabemöglichkeiten für alle Teilnehmer der Gesellschaft gibt.

Vielen Dank.

Judith Simon

Herzlichen Dank, Lisa Herzog, für diesen großartigen Vortrag. Ich gebe jetzt das Wort weiter an Juliane Lischka für den dritten Beitrag.

Wissenschaftskommunikation und Medien

Juliane A. Lischka · Universität Hamburg

Vielen Dank. Ich möchte die Debatte um Deepfakes aus der Perspektive des Journalismus beleuchten. Ich bin Professorin für digitalen Journalismus und Kommunikationswissenschaftlerin und werde die Deepfakes-Debatte im Journalismus anhand von drei Hauptaussagen beleuchten und dann versuchen, diese drei Hauptaussagen, die man durchaus als angstgeprägt betiteln kann, mit kommunikationswissenschaftlicher Forschung zu belegen oder zu entkräften. Wir werden sehen, wo wir landen.

[1.]Bei der Deepfake-Debatte innerhalb des Journalismus, die Journalist:innen miteinander führen, ist das erste wichtige Thema, um das geht, eine dystopische Zukunftsvision, die auf der Grundlage von Deepfakes aufgemacht wird, nämlich: Böse Mächte, die Deepfakes entwickeln, werden unsere demokratischen Entscheidungsprozesse untergraben, weil (wie wir schon in den vorhergehenden Vorträgen gehört haben) wir nicht in der Lage sein werden, Wahrheit von Unwahrheit zu unterscheiden. Das ist also ein sehr negativer Ausblick.

Dazu habe ich mal eine Studie herausgesucht, die mit Fact-Checker:innen gesprochen hat, vom International Fact-Checking Network². Das ist ein großes Netzwerk von fast 80 Nachrichtenorganisationen weltweit. Den Fact-Checker:innen, deren Fokus es ist, Informationen zu prüfen und zu verifizieren, wurde gesagt: Wir haben beide Seiten der Debatte, auf der einen Seite dieses Dystopische, diese dystopische Vision: [auf der Folie weiß hinterlegt] Deepfakes kommen, wir können nicht länger an den Wahrheitsgehalt von dem

glauben, was wir sehen, und wir müssen uns darauf vorbereiten; und auf der anderen Seite diese Position [grün hinterlegt] innerhalb des Journalismus zu sagen: Okay, es gibt diese Deepfakes, aber die sind, was die Propaganda oder die politische Kommunikation betrifft, noch kein reales Problem. Dafür werden sie eher wenig genutzt.

Die Fact-Checker:innen stimmen eigentlich eher dieser grünen Position zu: Deepfakes sind zumindest noch kein Problem, und man erwartet zumindest im Moment auch nicht, dass sie für Wahlen ein Problem werden. Vielmehr sagen die Fact-Checker:innen: Womit sie Probleme haben, sind echte Videos, die dekontextualisiert sind, die also aus ihrem Zusammenhang gerissen werden und bei denen Fact-Checker:innen prüfen müssen, inwieweit die Videos auch das sagen, was behauptet wird.

Ein gutes Beispiel für solche dekontextualisierten Videos sind diese Film-your-hospital-Videos, die wir im Rahmen der Coronapandemie gesehen haben, wo die Leute in leere Abschnitte von Krankenhäusern gegangen sind, diese gefilmt haben und eben leere Betten, leere Zimmer gezeigt haben mit der Frage: „Ja, wo ist denn das Virus? Hier sind ganz viele Betten. Warum sind wir im Lockdown?“ Also dekontextualisierte echte Videos. Das sind keine Deepfakes. Also hier wieder der Punkt: Deepfakes fassen Probleme zusammen, vergrößern Probleme, wie Herr Ommer gesagt hat, die vorher schon existiert haben.

Dieser erste Punkt der Dystopie ist noch nicht so stark, ist noch nicht allumfassend. Es ist noch kein dominantes Problem unserer digitalen Öffentlichkeiten, wenn wir an politische Entscheidungsprozesse denken.

² Weikmann, T., & Lecheler, S. (2023). Cutting through the Hype: Understanding the Implications of Deepfakes

for the Fact-Checking Actor-Network. *Digital Journalism*, 1–18.

[2.] Eine Grundannahme für diese Dystopie ist diese Annahme, wir können Deepfakes nicht erkennen; Bürger:innen haben wenig Medienkompetenz, sind nicht in der Lage, diese Deepfakes zu erkennen. In der Vergangenheit konnte man vielleicht noch daran orientieren, ob der Papst jetzt sechs Finger hat oder fünf und ob diese Deepfakes blinzeln oder nicht. Aber irgendwann werden wir wirklich nicht mehr in der Lage sein, das gut zu erkennen, und da hilft vielleicht auch die höchste Medienkompetenz nicht.

Es gibt eine Studie, die die Medienkompetenz der deutschen Bevölkerung gemessen hat.³ Sie werden gleich schockiert sein. An diesem Test (Sie können auch gern selber mal versuchen, das zu machen) kann man bis zu 30 Punkte erreichen. Meine Studierenden, die im Mediensystemseminar bei mir sind, erreichen ein sehr hohes Level, über 25 Punkte. Die deutsche Bevölkerung schneidet im Mittel ungefähr bei 13 Punkten ab. Die normale Medienkompetenz ist also gering bis mittel. Es ist nicht gut um die Medienkompetenz bestellt. Allerdings glaube ich auch, es kann nur eine Lösung sein, hier anzusetzen, weil die einzelne Person letztlich überfordert ist, Deepfakes erkennen zu können.

Es gibt erste kommunikationswissenschaftliche Experimentalstudien, die versuchen zu testen, was passiert, wenn ich Deepfakes ausgesetzt bin. Eine Studie zeigt verschiedenen Gruppen verschiedene Arten von Deepfakes: nur Audio-Deepfakes, audiovisuelle Deepfakes, und auch 360-Grad-Videos als Deepfake. Bei allen drei Gruppen kann sie zeigen: Wenn man den Leuten dann sagt: „Übrigens, das war ein Deepfake, den du gerade gesehen hast“, steigt die allgemeine

Unsicherheit über den Wahrheitsgehalt von Medieninhalten. Die Frage ist dann wirklich: Sobald ich einmal mit Deepfakes in Kontakt gekommen bin, habe ich selber nicht mehr das Vertrauen in mich selbst, dass ich Deepfakes erkennen kann, und auf der anderen Seite fange ich an, generell sehr skeptisch gegenüber Nachrichteninhalten zu sein. Das kann auf der einen Seite gut sein, aber wenn ich zu skeptisch bin und alles anzweifle, ist das nicht mehr gut für unsere Öffentlichkeiten.

Eine weitere spannende Studie hat den Leuten echte Videos gezeigt und dann ein Face Warp zu bestimmten Themen. Hier konnte man zeigen, dass Menschen Deepfakes mit diesem Face Warp glaubwürdig finden, wenn der Inhalt mit der eigenen Meinung übereinstimmt. Sie finden es nur nicht glaubwürdig, wenn der Inhalt *nicht* mit der eigenen Meinung übereinstimmt. Das ist eine Tendenz, die sehr problematisch werden kann.

Was wir bisher auch schon in einer Debatte zu Desinformation gesehen haben: Menschen teilen eher Informationen, die mit ihrer eigenen Meinung übereinstimmen, auch wenn es Desinformation ist. Das ist also auch nichts Neues, aber auch problematisch.

[3.] Der dritte Punkt ist eher lösungsorientiert. In dieser angstgeprägten Debatte geht es um Gegenmaßnahmen. Was können wir machen, um Herr über diese Deepfakes zu werden?

Es gibt vier Punkte, die im Journalismus adressiert werden. Auf zwei davon möchte ich später näher eingehen.

Erstens: Wir können KI-Analysesoftwares nutzen, um Deepfakes zu identifizieren. Dann sagen aber Journalist:innen richtigerweise: „Was hilft es mir, wenn mir die Maschine sagt: 67 Prozent

³ Meßmer, A.-K., Sänglerlaub, A. & Schulz, L (2021). „Quelle Internet“? Digitale Nachrichten- und Informationskompetenz der deutschen Bevölkerung im Test.

Deepfake? Ich brauche entweder die Gewissheit: 100 Prozent ja oder 100 Prozent nein.“ Die Anwendungsmöglichkeiten werden hier als nicht besonders gut beurteilt.

Das Zweite ist die große Debatte um die Regulierung sozialer Plattformen. Die sozialen Plattformen sagen: Reguliert uns gerne. Aber wer sind wir, zu entscheiden, was die Wahrheit ist, was Deepfakes sind und was nicht? Das können wir auch gesellschaftlich nicht den sozialen Plattformen überlassen.

Der dritte Punkt ist dieser Punkt: Lasst uns doch die digitale Medienkompetenz der Bevölkerung steigern, dann wird alles gut. Ich bin da eher pessimistisch, denn wir können das Problem nicht auf die Fähigkeiten des Einzelnen abwälzen. Das wird kein Problem sein, was wir mit Medienkompetenz lösen können, weil irgendwann die Deepfakes wirklich nicht mehr erkennbar sind.

Viertens bleiben den Journalist:innen nur noch manuelle Verifikationspraktiken.

Ich zeige Ihnen ein Slide zu einem Test, den das Reuters Institute for the Study of Journalism an der Oxford-Universität mit öffentlich zugänglichen Analysesoftware gemacht hat, und dann zeige ich Ihnen eine Liste von manuellen Verifikationspraktiken, wo Sie sehen werden: Das können wir als normale Bürger:innen nicht anwenden.

Schauen wir uns zuerst diese Analysesoftware an. Hier hat das Reuters Institute ein Foto genommen aus dem Gaza-Krieg mit Israel und hat dann mit Dolly diese Bereiche erweitert. Das Foto war echt, aber die Bereiche sind jetzt KI-generiert. Daran knüpft sich die Frage an: Ab wann ist etwas ein Deepfake? Das kann durchaus teilweise KI-generiert sein wie hier. Dann haben die das in Tests durch KI-Analysesoftware gejagt, und zum Beispiel diese Analysesoftware sagt: Dieses

Foto ist zu einer hohen Wahrscheinlichkeit von einem Menschen geschossen. Analysesoftware können uns auch nicht mit Sicherheit sagen, woher dieses Video kommt. Das ist ein Beispiel, wie schwierig es sein wird, sich auf Analysesoftware zu verlassen.

Was dann noch bleibt, sind diese manuellen Verifikationspraktiken. Die stammen zu einem großen Teil aus dem Berkeley-Protokoll für Verstöße gegen Menschenrechte und wurden entwickelt, um Menschenrechtsverletzungen zum Beispiel in Kriegsgebieten zu verifizieren und festzustellen, hier ist wirklich eine Menschenrechtsverletzung passiert oder hier ist ein Video, das eine Menschenrechtsverletzung darstellt, das nicht echt ist.

Daran orientiert sich auch der Journalismus. Da gibt es im Prinzip zwei Gruppen von Praktiken, die man machen kann: erstens eine Bild- und Videoanalyse. Normalerweise steigen Fact-Checker:innen hier mit einer umgekehrten Bildersuche ein und testen, ob das Bild nicht aus einem anderen Kontext kommt und schon mal veröffentlicht worden ist. Dann gibt es, was sehr zeitaufwendig ist, eine Frame-for-Frame-Analyse. Hier gucken die Fact-Checker:innen, ob es Glitches gibt, die die KI vielleicht gemacht hat, wo Sachen verschwimmen oder sich verändern und eigenartig aussehen. Sie schauen sich Videos bis zu 100 Mal an, um diese detaillierten Analysen zu machen. Das ist etwas, was wir einfach nicht machen konnten, auch weil wir in dieser Zeitnot sind, wie das Frau Herzog dargestellt hat.

Zweitens gibt es den großen Bereich der Kreuzverifikation der Metadaten mit digitaler Open-Source-Information. Die Fact-Checker:innen oder die forensischen Analytiker:innen gucken sich dann im Detail an: Sind da irgendwelche Schilder in den Videos? Oder Uniformen von offiziellen Personen, Polizist:innen, die vielleicht auch Badges haben? Verändern die sich innerhalb

des Videos? Das wäre ein Indikator für Deepfakes. Kann ich anhand der Vegetation ablesen, ob das wirklich an diesem Ort stattgefunden hat? Sehe ich die Schatten, kann ich die Schattenlänge messen? Das kann ich zur Zeitverifikation nutzen. Der Schatten hat nur an zwei Tagen im Jahr eine ganz bestimmte Länge, und daran kann man versuchen zu testen oder herauszufinden, wann das Video gedreht worden ist. Oder man kann historische Wetterinformationen oder die Geolocation abgleichen.

So etwas kann gemacht werden. Das ist aber alles sehr zeitintensiv. Wir können das nicht auf die einzelnen Bürger:innen abwälzen, aber auch nicht auf die einzelnen Journalist:innen innerhalb von Nachrichtenorganisationen, sondern wir brauchen ein Netzwerk von Verifikationsspezialist:innen, die sich mit dem Identifizieren von Deepfakes auseinandersetzen. Wir brauchen eine Netzwerkarchitektur, die im Idealfall global angesiedelt ist.

Damit möchte ich enden, vielen Dank.

Diskussion

Moderation: Judith Simon · Deutscher Ethikrat

Herzlichen Dank, Juliane Lischka, für den dritten großartigen Vortrag. Wir steigen jetzt direkt in die Diskussion ein. Ich möchte mit einer etwas offenen Frage starten, weil ich mir nicht sicher war, ob wir die gleiche Einschätzung haben in Bezug auf die Frage, ob jetzt tatsächlich eine neue Gefahr da ist oder ob alles so ist wie vorher.

Daher wäre meine erste Frage in die Runde: Wie groß sehen Sie die Gefahren von Manipulation und gezielter Propaganda durch generative KI in Kombination mit sozialen Medien oder Ähnlichem? Haben Sie Sorgen in Bezug auf Mis- und Desinformation in Bezug auf Wahlkämpfe?

Björn Ommer

Ja, in der Tat. Ich habe Sorge. Ich sehe das als eine sehr potente Technologie, und für jede potente Technologie ist nicht nur der Gebrauch, sondern auch der Missbrauch dort kritisch. Das war der Grund, warum mein Lehrstuhl, als wir damals Stable Diffusion entwickelt haben, das trainierte Modell nicht herausgegeben haben. Waren dann Firmen, die dort ein bisschen schnell weitergegangen sind; bei anderen ist ja immer noch Geld mit im Spiel, aber weil das etwas ist, was mich durchaus umgetrieben hat.

Gleichzeitig fände ich es, wie ich versucht habe kundzutun, es etwas verkürzt, zu sagen, dass sich das Problem nur durch diese neue Technologie stellt. Denn wenn dort (Frau Herzog hat ja drastische Beispiele erwähnt) mein Kopf auf irgendwas anderes draufmontiert wird, ob das jetzt mit Photoshop oder irgendeiner neuen Technologie gemacht wird, ist für mich als Einzelner erst mal sekundär im Videobereich; da gibt es natürlich Dinge, die vorher schwieriger gingen. Aber wir sollten uns klar darüber sein, dass es viele Akteure außerhalb unseres Landes gibt, die das in der Vergangenheit im amerikanischen Präsidentschaftswahlkampf oder ähnlichen Dingen gemacht haben und keine Kosten, keine Mühen gescheut haben, dass sich also – egal ob mit neuer oder alter Technologie – dieses Problem schon gezeigt hat. Das macht mir Sorgen, ja.

Lisa Herzog

Da würde ich mich anschließen. Es multiplizieren sich die Möglichkeiten des Einsatzes. Es gibt diesen Begriff, „ein Problem auf Steroiden“, also noch mal hochgejazzt durch neue technologische Möglichkeiten, und so sehe ich es hier auch.

Für die Vertrauenssituation ist es natürlich viel schädlicher, wenn ich weiß, das kann im Grunde jeder Teenager daheim auf dem Handy machen,

als wenn ich davon ausgehe, okay, vielleicht passiert es manchmal, wenn es gut finanzierte Kräfte irgendwo machen. Denn die Ubiquität, also die ständige mögliche Anwesenheit solcher Inhalte sieht dadurch noch mal anders aus als vorher.

Judith Simon

Ich gebe die Frage weiter an Juliane Lischka, weil auch die Empirie unterschiedliche Antworten gegeben hat. Wie ist deine Einschätzung, wie würdest du das einordnen?

Juliane A. Lischka

Ich würde es auch auf jeden Fall eher dystopisch einordnen, obwohl ich weiß, dass auch der Journalismus im Moment sagt, es ist noch kein wesentliches Problem. Aber allein das Potenzial, dass es zu einem Problem werden kann, das doppelte Vertrauensproblem, ist sehr problematisch für unsere Öffentlichkeit. Wenn wir an nichts mehr glauben, was wir sehen, funktioniert nichts mehr, funktioniert keine Meinungsbildung mehr. Ich sehe tatsächlich die Gefahr, dass wir mehr und mehr auf so einen Zustand zusteuern, dass wir eine generelle Skepsis gegenüber allen Quellen und allen Informationen entwickeln und es umso schwieriger wird, unsere politische Meinungsbildung so zu steuern, dass wir gut informiert sind und uns auch gut informiert fühlen.

Judith Simon

Ich versuche das zusammenzufassen: Wir sehen alle eine Gefahr und ein Problem, die allerdings nicht originär nur aus der Technik kommt, sondern auch aus anderen Quellen, aber wo die Technik eine Art Verstärker ist und gerade durch die Geschwindigkeit, durch die Schnelligkeit auch einen qualitativen Unterschied macht, der nicht nur quantitativer Natur sein könnte.

Wenn wir sagen, ein Teil der Probleme besteht in Polarisierung und Manipulation im Allgemeinen,

die verstärkt wird, dann stellt sich die Frage zunächst an Lisa Herzog, wie dem begegnet werden kann, auch aus einer Perspektive der politischen Theorie.

Lisa Herzog

Ein Punkt, wie man sich dieser Frage nähern kann, ist: Wie entsteht vertrauenswürdige Kommunikation zwischen Menschen? Ich denke, das hat zwei Ebenen: Das eine ist das Vertrauen der Bürgerinnen und Bürger untereinander, und das andere ist das Vertrauen in Institutionen, wo ich auch Medien dazurechnen würde.

Bei Vertrauen in Institutionen und Medien sind wir in einer Phase, wo es immer noch die Medien gibt, die traditionellen journalistischen Standards entsprechen, aber zunehmend hinter Paywalls. Das heißt, das ist auch eine Frage des Geldbeutels, wer sich das leisten kann. Wir haben zum Glück noch die öffentlichen Medien. Es gibt Forschung, die Länder mit und ohne öffentlichen Rundfunk oder Medien vergleicht, und zeigt, dass das für die allgemeine politische Bildung einen tendenziell positiven Effekt hat.

Die zweite Frage, die sich stellt, ist: Was passiert eigentlich zwischen den Menschen untereinander? Ich suche immer noch nach empirischer Forschung dazu, aber ich glaube fest daran, dass der persönliche, echte 3D-Kontakt (also nicht was wir hier machen, sondern wirklich gemeinsam in einem Raum) weiterhin einen großen Stellenwert hat. Aber wir haben da das Problem, dass durch die immer stärkere sozioökonomische Ungleichheit, auch bei den Entwicklungen auf dem Arbeitsmarkt, sich zunehmend Leute begegnen, die von ihren Hintergründen her ähnlich sind, ähnlich denken. Es gibt diesen Begriff der Bubble: meine Bubble, das sind die Leute, die wie ich denken.

Für uns als Gesellschaft ist eine große Frage: Wo begegnen sich eigentlich Leute, die unterschiedliche Meinungen haben, aber trotzdem auf gut informierten Diskussionen oder da landen können, wo sie sich gegenseitig auch ernst nehmen und respektieren? Lange war ja die Hoffnung der demokratischen Theorie: Das passiert in den sozialen Medien. Toll, da begegnen sich Leute, die sich sonst nie begegnet wären, das war der Enthusiasmus in der Anfangsphase des Internets und der sozialen Medien. Das müsste doch ganz viel neues demokratisches Potenzial bieten. Es hat sich dann aber schnell gezeigt: So einfach ist es nicht, weil sich auch da Gleichgesinnte oft stärker treffen. Und wenn das jetzt noch durch künstlich generierte Fake-Nachrichten und Fake-Profile zum Beispiel unterminiert wird, dann ist eine große Frage, wie dieses horizontale Vertrauen der Bürgerinnen und Bürger untereinander eigentlich stabilisiert werden kann.

Dann muss man meines Erachtens auch fragen: Wie organisieren wir eigentlich unsere Arbeitswelt, unsere Wirtschaftswelt? Wie sind unsere Städte organisiert? Dann geht es sozusagen in die Hardware und nicht nur in die Software des sozialen Zusammenlebens.

Judith Simon

Ich würde da einen Punkt rausgreifen, weil du von Vergleichen zwischen Ländern und öffentlichem Rundfunk davon gesprochen hast. Juliane, bei deinen Analysen habe ich mich auch gefragt: Wird da auch zu allen Ländern geforscht? Und zu welchen Ländern haben wir Kenntnisse darüber, wie wirksam Fake News und Propaganda sind?

Es gibt Länder, die noch ein funktionierendes Mediensystem jenseits sozialer Medien haben, und andere, wo soziale Medien für die öffentliche Kommunikation viel zentraler sind, wo dann vielleicht Manipulation noch stärker durchgreift, weil es eben ein zentraleres Medium ist.

Vor ein paar Tagen gab es eine erste Studie von OpenAI, dass die Manipulation so effektiv gar nicht sei. Ich frage mich: Wohin guckt man da? In welche Länder schaut man da? Das ist für mich eine relevante Frage, ob man da auch blinde Flecken hat, wohin man da schaut.

Juliane A. Lischka

Ja, das ist auf jeden Fall wichtig, dass man das Mediensystem mit betrachtet, wenn man Studien durchführt. Auf jeden Fall sind die USA ein polarisiertes Mediensystem, wo der Öffentlich-Rechtliche im Prinzip inexistent ist und wir wirklich ein polarisiertes System haben, wo man ungefähr weiß, wie die politische Positionierung eines Medien-Outlets ist und wo gleichzeitig Social Media und News Snacking eine große Rolle spielen, vor allen Dingen für das jüngere Publikum.

Das ist in Europa noch ein bisschen anders. Hier spielt der öffentlich-rechtliche Rundfunk noch eine größere Rolle, und hier haben wir ein Nord-Süd-Gefälle. Die nordeuropäischen Länder haben noch ein größeres Medienvertrauen als die mittel- und südeuropäischen Länder. Wir liegen in Deutschland so im mittleren Bereich. Wir haben noch ein einigermaßen gut, aber kein hoch ausgeprägtes Vertrauen mehr in Nachrichten oder in den Journalismus. Gleichzeitig haben wir aber nicht das Problem wie in den USA, dass wir sehr polarisiert sind, also sehr weit links oder rechts außen mit unseren Nachrichtenorganisationen.

Dennoch zeigen sich in diesen Studien, die dieses doppelte Vertrauensproblem testen – da gibt es eine Studie, die versucht, das in den Niederlanden zu testen im Vergleich zu den USA, und eine Studie, die das in Österreich testet. Man kann in den verschiedenen Ländern, zumindest in den westlichen Demokratien nachweisen, dass der Vertrauensverlust, wenn ich das Gefühl oder den Eindruck habe, es gibt immer mehr Deepfakes oder immer mehr Desinformation (und auch einfach

die Warnung, dass es Desinformation gibt), dazu führt, dass ich weniger Vertrauen in Medien habe. Das Vertrauen in Medien ist ein wichtiger Indikator dafür, dass ich politischen Prozessen vertraue, dass ich mich gut informiere und nicht zu alternativen Medien gehe. Wenn der Vertrauensverlust einmal da ist, werde ich vielleicht eher zu nicht traditionellen Nachrichtenorganisationen ziehen und diese Nachrichten nutzen. Dann ist es sehr schwer, diese Leute wieder zurückzuholen und das Vertrauen wiederherzustellen, wenn das einmal verloren ist.

Judith Simon

Was können wir denn tun jeweils in unseren unterschiedlichen Rollen, um problematischen Entwicklungen entgegenzutreten und einen sinnvollen Umgang mit generativer KI im öffentlichen Raum zu fördern? Ich würde Herrn Ommer die Frage stellen, insbesondere mit dem Blick darauf, was Informatikerinnen und Informatiker tun können. Das frage ich auch deswegen, weil ich an einem Fachbereich für Informatik bin. Wie sehen Sie die Rolle? Was ist die Aufgabe von Informatik, da wirksam zu sein?

Björn Ommer

Das ist eine sehr wichtige Frage. Wenn Bilder generiert werden, stellt sich natürlich die Frage, ob man nicht mit technologischen Mitteln die Fakes automatisch selektieren kann. In den Vorträgen vorher haben wir schon gesehen, dass es an vielem scheitert. Ich hatte versucht, es auszuführen, dass in vielen Bereichen die Bilder einfach so klein, alleingestellt sind und Ähnliches, dass sich das Bild nicht aus sich selbst verifizieren lässt, oder dass das Bild sogar richtig ist, aber dekontextualisiert ist und auf diese Art und Weise eine andere Meinung entfaltet. Deswegen wäre ich sehr vorsichtig, in diesem Fall auf eine rein technologische Lösung zu vertrauen. Es werden dann

Möglichkeiten wie Wasserzeichen und Ähnliches angeführt.

Frau Herzog, Sie hatten gesagt, dass wir bei der Regulierung im Auge behalten sollten, für wen wir regulieren, dann schützenswerte Teile der Bevölkerung erwähnt und als Gegenmaßnahmen mehr Teile dieser Regulierung genannt, die typischerweise aufgebracht werden. Mir ist wichtig, dass wir auch kundtun, wen wir dort regulieren, nicht nur für wen, sondern wen wir regulieren. Bei einigen dieser Positionen macht es mir ein bisschen Sorge, weil wir bei dieser Regulierung auf Kooperation des Regulierten setzen oder die erfordern. Das funktioniert sehr gut mit großen Firmen, die bei uns Geld machen wollen. Wir haben es im AI Act gesehen, dass wir potenziell sogar für Firmen, die nicht hier sind, eine Regulierung hoffen durchsetzen zu können, weil wir eine relativ große Marktmacht haben und die eine Technologie nicht zweimal entwickeln wollen und dann hoffen, dass sie, wenn sie die woanders entwickeln, weil sie sie bei uns ausspielen wollen, sich dann auch an unsere Regeln halten.

Das gilt aber nicht für Leute, die Deepfakes bewusst in die Welt setzen und, ich will es mal direkt nennen, die kriminell sind. Das kann hier oder da noch der einzelne Jugendliche sein, der sich nichts dabei gedacht hat. Aber in den Vorträgen von Ihnen sind ja viele Beispiele gebracht worden von Dingen, die man als handfest kriminell bezeichnen würde. In dem Fall können wir über Regulierung reden, aber ich glaube, da sollten wir dann eher über Strafen reden, denn wir haben ja schon beschlossen oder festgehalten, dass das etwas ist, was wir nicht haben wollen. Das ist also da. Jetzt ist nur die Frage: Wie kommen wir dem Ganzen bei? Wir wissen, je nachdem, wie kriminell das Ganze ist, dass es dann sehr schwierig wird; wenn es Akteure aus dem Ausland sind,

noch mal umso schwieriger. Das macht den Prozess kritisch.

Wir haben natürlich Medienhäuser und Ähnliches. Wir können umgekehrt denken, dass wir Wasserzeichen einführen für originäre Dinge, wo sie herkommen. Da gibt es natürlich die Plattformen, zu denen man direkt gehen kann, zu den Medienhäusern, und sich dann mit Informationen versorgen kann.

Der letzte Punkt: Ich möchte ein bisschen von der Dystopie weg, in der wir gerade gefangen sind. Frau Lischka, Sie hatten zu Beginn gesagt, wenn diese Welt entsteht, in der wir nichts mehr trauen können – natürlich ist das etwas, was mich auch im Bildbereich sorgt. Aber ich möchte den Blick auf den Text lenken. Wir wissen seit jeher, dass jeder alles Mögliche schreiben kann. Wir sind alle Wissenschaftlerinnen und Wissenschaftler. Jeder könnte ein Paper produzieren, das komplett sinnvoll aussieht. Gleichzeitig haben wir es im Textbereich geschafft, Authentifizierungssysteme zu schaffen, Plattformen zu schaffen. Das ist das Glück der großen Verlagshäuser. Jeder könnte ein Paper heutzutage mit LaTeX auch ohne die drucken. Aber die bürgen für Authentizität. Das haben wir im Text geschafft, während im Internet jeder alles Mögliche behaupten kann, und es ist nicht so, dass wir jetzt auf einmal Worten nicht mehr glauben können.

Dementsprechend glaube ich, dass der direkte Bezug zu Verlagshäusern, Medienkonzernen und Ähnlichem, die Vertrauen aufgebaut haben, zu Akteuren, die vertrauenswürdig sind, in diesem Fall wichtig ist, und nicht ein Bild, das irgendwo im Internet herfloatet. Das dürfte dementsprechend auch keine Bedeutung für die Wahrheitsfindung haben.

Judith Simon

Vielen Dank. Ich würde jetzt unsere Diskussion öffnen für die Zuhörer:innen und Zuseher:innen im Livestream und das Wort abgeben an Mark Schweda als unseren Publikumsanwalt.

Publikumsanwalt: Mark Schweda · Deutscher Ethikrat

Hier im Slido türmen sich die Fragen und Rückmeldungen. Erst mal gibt es Dank für die interessanten Vorträge. Es gibt aber auch eine Reihe von Fragen, die sich auf die Einschätzung, Gewichtung und Dimensionierung des Problems beziehen, weil das in den Vorträgen durchaus ein bisschen hin und her zu gehen schien, also: Widmen wir dem Thema Deepfakes zu viel Aufmerksamkeit? Müssten wir mehr auf Dekontextualisierung achten?

Juliane A. Lischka

Aus Sicht des Journalismus auf jeden Fall. Der Journalismus oder auch die Forschung zu Desinformation sagt: Das größte Problem sind eigentlich Sachen, die nicht richtig falsch sind, aber die aus dem Kontext gezogen sind. Diesen Aspekt sehen wir auch in der politischen Kommunikation. Es werden Informationen aus dem Kontext genommen, um die eigene Position zu stärken und um bestimmte Meinungen zu fundieren. Das ist eine Argumentationsweise, die mit Deepfakes nicht neu ist und die auch im Bereich der Desinformation nicht neu ist. Aber wir sollten unsere Antennen diesbezüglich stärken, um uns zu fragen: Inwieweit ist diese Information jetzt vielleicht aus dem Zusammenhang gerissen? Gehe ich da gerade etwas auf den Leim, was eigentlich ganz anders gemeint war? Das ist schon eine alte Frage. Das ist nicht neu.

Mark Schweda

Es gibt auch die Frage, die ein bisschen auf den Titel unserer Veranstaltung zielt: Was macht die

besondere Macht der Bilder aus? Haben Bilder gegenüber anderen Informationsformen eine besondere politische Macht, einen besonderen politischen Einfluss, den es zu reflektieren, einzuhegen, zu beachten gilt?

Juliane A. Lischka

Man weiß, dass wir Bilder einfacher verarbeiten können. Unser kognitiver Aufwand, um Bilder zu verarbeiten, ist geringer, als wenn wir Text verarbeiten müssen, als wenn wir Sachen lesen müssen. Von daher sind unsere Filter, vielleicht auch unsere Wachsamkeit gegenüber Bildern nicht so aktiviert, weil wir das in so einer Leanback-Reflexionssituation machen können. Wenn wir Text lesen, müssen wir dafür viel stärker unsere kognitive Leistung erhöhen.

Studien zeigen, dass Bilder zum Beispiel im Zusammenhang mit Verschwörungsnarrativen etwas besser funktionieren, etwas überzeugender sind. Nicht riesig, aber man sieht man schon einen signifikanten Unterschied, der groß genug, also messbar genug ist im Vergleich zu Text. Wenn ich Verschwörungsnarrative nur in Textform rezipiere, finde ich das nicht so glaubwürdig, wie wenn ich auch ein Bild oder ein Video dazu habe. Von daher ist es durchaus gerechtfertigt zu sagen, dass Bilder mächtig sind.

Was man zusätzlich sagen kann: Zum Beispiel an diesen Fotografien zur Mondlandung würden jetzt Verschwörungsanhänger:innen argumentieren, dass die Beweis dafür sind, dass es *nicht* stattgefunden hat. Die eigene Meinung beeinflusst also, wie wir Bilder rezipieren, während diejenigen, die sagen: „Hey, die Mondlandung hat wirklich stattgefunden“, die Bilder genau als diesen Beweis ansehen. Unsere Voreinstellung prägt also, wie wir Bilder und jegliche Informationen wahrnehmen und wofür die Beweis sind. Das ist sehr spannend und problematisch.

Mark Schweda

Wenn ich mich kurz mal aus der Rolle des Publikumsanwalts herausbegeben darf, würde mich auch die Perspektive der politischen Theorie, des politischen Denkens auf die Frage der Macht der Bilder interessieren.

Lisa Herzog

Ich würde zwei Punkte ergänzen zu dem, was Frau Lischka gesagt hat. Das Erste ist, dass Bilder sehr stark Emotionen wachrufen können, denen man sich teilweise kaum entziehen kann. Es gab vor einigen Jahren dieses Beispiel eines ertrunkenen Flüchtlingsjungen an der türkischen Küste, das durch die Welt ging. Heutzutage wird sowas dann ein Meme, aber das spricht einen auf einer anderen existenziellen Ebene an als Texte. Das Interessante ist, dass Texte, die besonders wirkmächtig sein wollen, oft auch bildhafte Vergleiche oder Metaphern oder so verwenden, um genau diesen Effekt herbeizurufen.

Das Zweite ist viel banaler, aber vielleicht trotzdem wichtig. Bei Texten sind wir es schon länger gewohnt, wie einfach die gefälscht werden oder einfach verkehrt oder manipulativ sein können. Bei Bildern gibt es auch die historischen Beispiele, aber wir erleben gerade jetzt diese Phase, dass die Möglichkeiten der Manipulation technisch gepimpt werden und wir deswegen neu darüber nachdenken können, wie wir damit umgehen.

Björn Ommer

Ich glaube, der wesentliche Unterschied ist, dass wir bei Text immer wissen, dass es eine Interpretation ist, die jemand geschrieben hat. Der Text ist nicht die Realität. Aber das Bild haben wir begriffen als eine Abbildung der Wirklichkeit, die sozusagen die rohe Wirklichkeit darstellt, die wir uns ansehen können. Festzustellen, dass allein schon der Blick, wie wir den lenken, eine Interpretation

der Wirklichkeit darstellt, ist ein Bewusstsein, das erst kommen muss.

Das war das, was wir bisher gehabt haben: Wo hat der Fotograf hingeguckt und was er uns nicht gezeigt? Und dann kommt die zweite Dimension dazu, dass jetzt auf einmal das, was dargestellt wird, noch mal zusätzlich verändert, verarbeitet wird und auf diese Art und Weise definitiv nicht mehr die rohe Wirklichkeit darstellt, sondern eine Interpretation derselben ist oder noch viel mehr, eine Fälschung der Realität ist. Das war etwas, was beim Text immer klar war, und dieses Bewusstsein muss sich erst langsam schärfen, aber es wird Zeit dafür. Denn jeder, der bisher im Internet unterwegs war, sollte sich eigentlich dessen schon bewusst gewesen sein, dass das, was er dort sieht, nicht unbedingt die Realität ist.

Mark Schweda

Herr Ommer, damit sprechen Sie einen weiteren Komplex von Fragen und Kommentaren im Slido an, nämlich Fragen, die darauf zielen, wie wir mit der Entwicklung von Deepfakes umgehen, mit der Entwicklung von bildgenerierender KI im öffentlichen Diskurs. Die erste Frage, die ich rausgreifen will, zielt auf Bildung: Brauchen wir mehr Medienkompetenz? Und wie können wir an den Schulen, im Schulunterricht Medienkompetenz schärfen, entwickeln, um wahr von falsch unterscheiden zu können?

Juliane A. Lischka

Ja, das brauchen wir auf jeden Fall. Wir brauchen mehr Medienkompetenz, die sehr früh in der Bildung ansetzt. Aber ich bin nicht optimistisch, dass das eine Lösung des Problems sein wird. Als ich diese Medienkompetenz-Studie vorgestellt habe, bei der man 30 Punkte erreichen kann und meine Studierenden diejenigen sind, die 25 plus erreichen – es kann ja nicht die Lösung sein, dass man das erst studieren muss. Auf der einen Seite ist es

sehr wichtig, dass wir da ansetzen, dass wir bereits in den Schulen Medienkompetenz als Fach integrieren und das auch weiterhin über den Lebenslauf eine Rolle spielt. Man sieht tatsächlich ein Altersgefälle: Die Jüngeren sind medienkompetenter als ältere Teile der Bevölkerung.

Aber wie gesagt, wir können diese Verantwortung nicht auf die Einzelnen abwälzen. Das fände ich ungerecht und nicht zielführend. Das wird auch nicht die Lösung des Problems sein, sondern nur ein Baustein dieser Lösung.

Björn Ommer

Ich denke, das würde es auch ein bisschen verkürzen, weil es sonst nur den Blick darauf lenkt, dass wir Rezipienten sind von Nachrichten, die von außen irgendwie über uns kommen. Aber wir haben solche Dinge wie den Confirmation Bias, dass sich Leute bewusst das herauspicken, was ihre eigene Meinung bestärkt. Gleichzeitig sind sie Sender von Informationen, entweder indem sie nur mit dem Link dem Ganzen mit den Füßen sozusagen noch beistimmen oder es noch explizit nach außen senden, weil sie ihre Meinung kundtun wollen. Das heißt, Medien werden auch von den Einzelnen gebraucht. Es ist nicht so, dass wir ihnen nur Medienkompetenz beibringen müssen, um ihnen mitzuteilen, dass sie von außen manipuliert werden sollen, sondern wissentlich oder unwissentlich tun sie das auch mit anderen.

Da weiß ich nicht, ob das nur Kompetenz ist oder ob wir wieder bei banalen ethischen Fragen sind: Was möchte ich in dem Fall tun? Möchte ich selbst von anderen und möchte ich das mit anderen dann tun? Ist es das wert (das haben wir ja im politischen Diskurs), nur um Meinungen durchzupushen, dass ich das mache, indem ich dann vielleicht hier oder da die Wahrheit etwas verzerre oder Ähnliches? Geht es nur darum, den Punkt zu landen? Oder sollten wir vielleicht im Diskurs alle ein bisschen weiter abrüsten? Aber es wird sehr

schwer sein, das durch irgendwelche Regulierung durchzubekommen.

Lisa Herzog

Vielleicht noch als kleine Ergänzung. Studien aus der Psychologie zeigen: Eine Art und Weise, wie man das Bewusstsein für mögliche Manipulation sehr gut stärken kann, ist, wenn man das die jungen Leuten selber machen lässt, also wenn man sie „impft“, so wird das genannt, indem sie selber mal probieren, wie leicht man so einen Fake bauen kann. Das prägt sich besser ein, als wenn man es nur theoretisch vorstellt. Und das hat auch diese aktive Rolle, die gerade betont wurde.

Mark Schweda

Super, das sind ganz konkrete didaktische Ansätze. Medienkompetenz, Schulunterricht allein genügt nicht, sagen Sie.

Im Slido wird auch gefragt nach der Rolle von Fact-Checkern und Fact-Checkerinnen und vor allem der Gewichtung. Wie viel kann man den Fact-Checkern und Fact-Checkerinnen zumuten und wie viel muss die KI selbst und müssen technische Lösungen hinzugezogen werden?

Björn Ommer

Das erfolgt ja schon. Die schiere Menge an Daten, insbesondere an Bilddaten, die in sozialen Netzwerken unterwegs ist, könnte gar nicht als solche verarbeitet werden, wenn nicht KI sozusagen gegen KI entsprechend verwendet wird.

Der zweite Teil ist, dass die Hoffnung besteht, dass Sie in Zukunft nicht mehr Menschen haben, die sich den ganzen Müll ansehen müssen, der dort da ist. Da ist vieles, wo sich die Leute danach in Behandlung begeben müssen, wenn sie das den ganzen Tag machen müssen. Zumindest einen Teil davon herauszufiltern, da ist natürlich die Hoffnung, dass wir algorithmisch weiterkommen. Aber, wie wir eben schon diskutiert haben: Da

wird es noch sehr viele Falsch-Positive geben, die sehr schwer weiterzuverarbeiten sind.

Aber noch mal: Es wäre schon etwas gewonnen, wenn jedes einzelne von diesen Artefakten, was sich ein Mensch eigentlich nicht antun möchte und was sich ein Fact-Checker nicht antun müsste, durch die KI rausgefiltert würde.

Mark Schweda

Eine weitere Frage zu dem Zusammenhang aus dem Slido betrifft die Idee des Verifikationsnetzwerkes, von dem Sie, Frau Lischka, gesprochen haben. Wie hat man sich das vorzustellen? Wie könnte so ein Verifikationsnetzwerk aussehen? Gibt es Ansätze oder Beispiele, auf die man da verweisen kann?

Juliane A. Lischka

Da fallen mir konkret zwei Beispiele ein, wo da Netzwerke entstehen. Das eine ist, was ich angesprochen hatte, das International Fact-Checking Network, wo in Deutschland zum Beispiel die dpa [Deutsche Presse-Agentur] und auch Correctiv Teil ist. Sie erarbeiten bestimmte Prozesse und Prozeduren, wie Fact-Checking ablaufen sollte, und entwickeln Guidelines, woran man sich orientieren sollte, wie das funktionieren kann.

Dann gibt es in Zusammenarbeit mit Amnesty International, aber auch mit anderen Nachrichtenorganisationen und Forschungsinstitutionen in Bezug auf Journalismus ein Media Lab, das heißt Witness. Dieses Witness Media Lab versucht, Prozeduren zu entwickeln, um Deepfakes zu identifizieren oder Videos zu identifizieren, von denen behauptet wird, es seien Deepfakes, die aber vielleicht wahr sind. Sie sind dabei, einen Deepfake Rapid Response Force aufzustellen. Der soll im ersten Quartal des nächsten Jahres aktiv sein oder vielleicht schon in diesem Jahr aktiv werden, wo dann Fact-Checker:innen und Forensiker:innen

miteinander arbeiten und versuchen, Tools zu entwickeln und anzuwenden, die möglichst noch nicht öffentlich sind und möglichst nicht öffentlich werden, sondern dem Journalismus oder diesem Netzwerk vorbehalten sind, um automatisiert oder auf Grundlage bestimmter Schritte zu identifizieren, ob etwas ein Deepfake ist oder nicht. Das wird im Moment gemacht.

Mark Schweda

Vielen Dank. Das macht es sehr anschaulich.

Ein dritter Komplex von Fragen und Rückmeldungen bezieht sich auf die politische Ebene, politische Lösungsansätze. Welche Maßnahmen sollten Regierungen konkret ergreifen, um den Missbrauch von KI-Bildern in der Kommunikation zu verhindern? Ist vielleicht das Strafrecht in dem Zusammenhang ein geeignetes Instrument?

Lisa Herzog

Ich würde zwei Ebenen unterscheiden. Das eine ist: Welche gesetzlichen Regelungen braucht man? Gibt es die schon? Liegen die im Strafrecht oder liegen die in anderen Bereichen? Ich hatte ja ein paar Punkte genannt. Ich könnte mir gut vorstellen, dass man sagt: Sechs Wochen vor Wahlen dürfen keine Deepfakes gezeigt werden oder so was, oder nicht von Politikern. Für gerechtfertigt halte ich auch eine Markierungspflicht, die mit psychologischen Erkenntnissen darüber arbeitet, wie Menschen Medien nutzen und wahrnehmen. Das kann man alles diskutieren.

Die zweite Ebene ist aber: Wie wird es durchgesetzt? Da liegt im Moment nach dem, was ich so beobachte, vieles im Argen. Es dauert oft sehr lange. Es ist für einzelne Nutzerinnen und Nutzer oft unklar, was für Kosten auf sie zukommen. Sie trauen sich vielleicht auch nicht immer, Dinge zu melden, und wenn es internationale Netzwerke sind, stellt sich auch die Frage nach den Durchsetzungsmöglichkeiten. Es gibt die Diskussion

darüber, inwieweit hier mehr investiert werden müsste in Online-Staatsanwaltschaften, in Meldepunkte, einfach in die Ressourcen, um Dinge, die gesetzlich schon geregelt sind und die in Zukunft noch geregelt werden müssen, auch durchzusetzen.

Ich weiß nicht, inwieweit es Möglichkeiten gäbe, dafür wiederum KI einzusetzen und die Rechtsprechung vielleicht in manchen Bereichen zu erleichtern. Das ist natürlich ein anderes Feld von KI-Anwendung, das im Moment viel diskutiert wird. Aber wenn wir nur Regelungen und Gesetze erlassen, ohne uns die Frage zu stellen, wie das durchgesetzt werden soll, dann kommen wir da nicht weiter.

Björn Ommer

Um da anzuknüpfen: Ich finde es schwierig. Sie hatten die Wahlen erwähnt. Natürlich kann man verbieten, dass davor irgendwelche Deepfakes von Politikern durchgeführt werden oder auch sonst. Ich glaube, in vielen Fällen möchte man die sonst auch gar nicht erst haben. Nur wie der amerikanische Präsidentschaftswahlkampf damals gezeigt hat, wenn die Sachen dann aus Russland reinkommen – da können Sie viel verbieten. Wir haben Russland auch versucht zu verbieten, in die Ukraine einzumarschieren. Hat sie auch nicht viel gekümmert, ja? Und dann kann man sich vorstellen, wie viel sie darum geben, wenn wir sagen, dass wir ein paar Pixel von ihnen nicht haben wollen. Da ist der Arm unseres Gesetzes nicht besonders lang.

Ansonsten reden wir, das hatten wir schon diskutiert, dann eher auf der Ebene des Strafrechts, was wir denn tun, wenn wir beschlossen haben, dass es nicht in Ordnung ist. Ich glaube, da liegt häufiger noch was im Argen. Ich hatte neulich einen Fall, wo mich eine Journalistin aus den USA kontaktiert hatte; das war noch vor der Zeit der neuen

generativen KI. Da hatte jemand eine Bildmanipulation mit einer Klassenkameradin durchgeführt, und das Recht dort sah noch so aus: Na ja, das ist ja nicht die Realität gewesen. Das sind ja nur Bilder. Und es war sehr schwer, ihrer dann habhaft zu werden. Das ist irgendwann in späteren Fällen gelungen, aber das muss man sich mal vorstellen. Da reden wir erst mal darüber, dass wir definieren können, was überhaupt geduldet werden kann und was nicht geduldet werden kann. Und dann kommt der zweite Teil, der kritischer ist: Wer entscheidet auf diesen Netzwerken? Wie setzen wir das durch? Und mit welcher Geschwindigkeit setzen wir das durch?

Da gibt es relativ viele Studien dazu, dass die Latenz in dem Fall nicht gerade die beste ist, um es mal vorsichtig zu sagen. Und was nutzt es, wenn wir definiert haben, dass etwas nicht in Ordnung ist, dass es identifiziert und sogar gemeldet worden ist, wenn es dann aber nicht durchgesetzt wird oder an anderer Stelle schneller wieder neu hinzukommt, als es beseitigt wird? Das wäre etwas, bei dem nachgearbeitet werden muss. Aber ich finde, dann sollte man eher an der Quelle ansetzen. Wenn ich so ein mächtiges Werkzeug habe wie die Presse, dass ich etwas instantan ausspielen kann an Millionen von Leuten, dann sollte man eher dort gucken, was man dort zulässt und wo man dort Grenzen einzieht.

Mark Schweda

Vielen Dank. Für das Ende der Diskussion habe ich mir eine schöne Frage aufgehoben, die auch zu diesem Rahmen als Veranstaltung des Deutschen Ethikrates passt. Herr Ommer, Sie hatten es angesprochen: Vielleicht brauchen wir nicht nur Medienkompetenz, sondern auch so etwas wie Tugenderziehung, Tugenderziehung zur Medienutzung. Kann es ethische Deepfakes geben, also

Deepfakes, die für gute Zwecke eingesetzt werden? Oder ist das Konzept das Deepfakes per se verwerflich und problematisch? Wer möchte?

Lisa Herzog

Das kommt ein bisschen darauf an, welche ethische Position Sie einnehmen. Wenn Sie utilitaristisch argumentieren, dann zählt hier das Endergebnis, und man kann sich gut vorstellen, dass zum Beispiel Hilfsorganisationen versuchen würden, durch besonders emotional anrührende Videos ihr Spendenaufkommen zu erhöhen. Und wenn man der Meinung ist, der Zweck heiligt die Mittel, dann würde man sagen, das ist ein ethisches Deepfake.

Wenn man eher der Meinung ist, dass es auch Pflichten gibt, die man nicht verletzen sollte, und dazu gehört eine Wahrhaftigkeitspflicht, dann sieht das anders aus. Ich glaube, dass Hilfsorganisationen nicht diejenigen sind, die das machen würden, weil für die das Reputationsrisiko so groß ist, dass es unwahrscheinlich ist, dass sie es täten.

Insofern tue ich mich ein bisschen schwer damit, ein gutes Beispiel einer ethisch zulässigen Deepfake-Manipulation zu finden. Vielleicht müsste man da suchen, wo man auch sagt, dass bestimmte Formen von Lügen manchmal gerechtfertigt sein können. Denn es hat ja etwas von einer Lüge, wenn man durch so einen Deepfake eine andere Wirklichkeit nahelegt als die, die eigentlich existiert. Da müsste man suchen. Ich bin nicht überzeugt, dass man was finden würde.

Björn Ommer

Ich glaube, mit dem Fake selbst setzen wir einen Tick zu spät an. Aber die Technologie, die den Fake generiert, kann sehr hilfreich sein. Das hatte ich versucht, am Ende noch mal kurz als positiven Output zu geben. Wenn ich auf diese Art und

Weise Kommunikation sehr spezifisch auf das Individuum oder einen Teil der Gruppe ausrichten kann, dann kann ich versuchen, auch kompliziertere Botschaften zu kommunizieren. Stable Diffusion ist zum Beispiel im Bereich des Klimawandels von, ich glaube, der WEF war es, oder wem auch immer verwendet worden, um zu visualisieren. Denn das macht es greifbarer für uns, abstrakte Dinge zu erfassen, um zu visualisieren, was das für uns bedeuten würde. Das sind Bereiche, bei denen generative KI helfen kann.

Mark Schweda

Da würde mich jetzt auch die Perspektive einer journalistischen Berufsethik interessieren. Frau Lischka, ethische Deepfakes?

Juliane A. Lischka

Ich denke, dass das möglich ist. Nicht aus einer Perspektive der Sorgfaltspflicht, aber durchaus, wenn wir an so etwas wie politische Satire oder an Meinungsfreiheit denken. Wir können den Leuten nicht verbieten, Deepfakes zu produzieren. Das ist ein Teil unserer Meinungsfreiheit – der Pressefreiheit eher nicht, weil da die Sorgfaltspflicht dahintersteht. Aber solange das markiert ist und als Deepfake ausgewiesen ist, dürfte das in unseren öffentlichen Diskursen nicht verboten werden. Sonst sind wir da schnell an einer Zensurdiskussion. Von daher kann ich mir vorstellen, dass Deepfakes auch unterhaltsam sein können und eben wie politische Satire ein zweischneidiges Schwert sein können, also nicht nur schlecht sind.

Mark Schweda

Vielen Dank. Damit beschließe ich die Inputs aus dem Slido und gebe zurück an Judith Simon.

Schlusswort

Judith Simon · Deutscher Ethikrat

Herzlichen Dank, Mark. Das war ein schönes Schlusswort, weil wir am Ende noch mal über Tugenden geredet haben und über etwas Positives. Das ist eine schöne Wendung zum Ende.

Ich möchte mich an dieser Stelle ganz herzlich bei allen Beteiligten bedanken. Das sind zunächst die drei Referent:innen Lisa Herzog, Juliane Lischka und Björn Ommer. Dank auch an unseren Publikumsanwalt Mark Schweda, aber auch an die Technik, an die Gebärdendolmetscherinnen, an die Mitglieder der Vorbereitungs-AG ein besonderer Dank und natürlich auch an die Mitglieder der Geschäftsstelle, insbesondere Lilian Marx-Stölting und Steffen Hering, die viel in der Vorbereitung des Webinars gearbeitet haben.

Danke natürlich an Sie alle, die Sie im Livestream dabei gewesen sind. Ich habe gehört, es waren um die 700 Zuschauerinnen und Zuschauer im Livestream. Das ist wunderbar.

Damit bedanke ich mich ganz herzlich bei allen und wünsche noch einen schönen Abend.