«Nicht die KI ist rassistisch oder sexistisch – wir sind es»

TikTok, Instagram oder ChatGPT sind Kommunikations- und Unterhaltungskanäle, die uns bei der Arbeit helfen können, und gerade junge Menschen verbringen viel Zeit auf diesen Plattformen. Laurent Sedano ist Medienbildner und Digitalisierungsexperte und Mitglied der Eidgenössischen Kommission für Kinder- und Jugendfragen EKKJ. Er erklärt, warum kontroverse Inhalte online so viel Raum bekommen, was wir gegen Hatespeech tun können – und wie ChatGPT weniger stereotyp werden könnte.

Interview: Mona Martin

Herr Sedano, Sie sind Experte für Medienbildung und beschäftigen sich seit 15 Jahren mit dem Thema Digitalisierung. Seit einiger Zeit ist ChatGPT in aller Munde. Nutzen Sie das Tool auch?
Ja. ChatGPT ist sehr nützlich, um die Quellen von Wissen ausfindig zu machen, an welches ich mich nicht mehr genau erinnere. So kann ich es nachlesen. Weiter nutze ich es für Arbeiten rund um meinen Podcast, lasse Beiträge transkribieren und zusammenfassen oder generiere ein Inhaltsverzeichnis. Am meisten verwende ich das Tool aber für Experimente. Ich überprüfe beispielsweise, worüber ich mit ChatGPT überhaupt diskutieren kann oder ob ich die KI dazu bringe, ein gewisses Bild zu generieren.

Welche Vor- und Nachteile sehen Sie in Tools wie ChatGPT?
Weder Schwarzmalen noch Glorifizieren bringt etwas. Alle Aspekte unseres Lebens sind mittlerweile mit dem Digitalen verbunden. Jugendliche, insbesondere junge Frauen, werden im Netz belästigt und Opfer von Hatespeech. Gleichzeitig knüpfen sie online Freundschaften, finden Verbündete und Unterstützung. Wir müssen einen Umgang mit dem ganzen Paket finden. Deshalb versuche ich zu verstehen, wie ein Tool oder eine Plattform und die dahinterliegende Technik funktionieren.

Ihnen erleichtert ChatGPT die Arbeit. Kann das Tool allen helfen?
Ich fürchte, dass sich hier der bestehende digitale Graben weiter verstärkt. Diejenigen, die über Ressourcen – vor allem Bildung – verfügen, werden noch produktiver. Sie machen bald zwei Masterabschlüsse in der Zeit von einem. Den anderen droht, weiter abgehängt zu werden, in einer Welt der seichten, KI-generierten Unterhaltung zu versumpfen und Fakenews aufzusitzen.

Kommt hinzu, dass sich die Tools auch stetig weiterentwickeln und verändern …
Ja, es braucht eine ständige Anpassungsleistung. Kaum hat man sich daran gewöhnt, schriftlich mit ChatGPT zu kommunizieren, muss man schon mit dem Tool sprechen. Dieses schnelle Weiterentwickeln vergrössert nicht nur den digitalen Graben, auch gesetzlich kommen wir nicht nach. Mit dem europäischen *Digital Service Act* gibt es nun erstmals einen konkreteren Versuch, unsere Rechte als Online-Nutzende gegenüber Plattformbetreibenden durchzusetzen. Ein anderes Gesetz, der *EU AI Act,* versucht, Regeln zur Anwendung von Systemen Künstlicher Intelligenz aufzustellen. Das braucht es unbedingt, auch wenn wir noch nicht wissen, ob diese Versuche greifen werden.

Laurent Sedano

«Weder Schwarzmalen noch Glorifizieren bringt etwas. Alle Aspekte unseres Lebens sind mittlerweile mit dem Digitalen verbunden.»

«Jede App nutzt mehrere verschiedene Algorithmen, *den* Algorithmus gibt es also nicht.»

Eine neue Gesetzgebung muss auch sicherstellen, dass die Nutzenden nicht manipuliert werden. Um Aufmerksamkeit zu erhalten, schlägt der Algorithmus aber immer extremere Inhalte vor. Sehen Sie eine Gefahr für Radikalisierung?
Es werden viele Tricks verwendet, um Aufmerksamkeit zu erhalten. Jede App nutzt mehrere verschiedene Algorithmen, *den* Algorithmus gibt es also nicht. Eine allgemeingültige Aussage ist dementsprechend schwierig. YouTube stand beispielsweise früher stark in der Kritik, weil sie Nutzende in einen Strudel von potenziell immer extremeren Inhalten – sogenannten *Rabbitholes* – zog. Das ist heute nicht mehr so, weil YouTube das verhindert. User müssen inzwischen bewusst nach extremen Inhalten suchen, wenn sie diese konsumieren möchten.

Trotzdem sorgen Algorithmen dafür, dass uns vor allem Inhalte im Feed vorgeschlagen werden, die unsere bestehenden Weltbilder, Meinungen und Werte bestärken.
Diese Theorie der Bubbles wird heute kontrovers diskutiert. Neueste Erkenntnisse zeigen, dass es zwar Bubbles gibt, diese aber die Folge von bewussten Entscheidungen der Nutzenden sind. Es ist wie im Offline-Leben: Wir gehen nur in gewisse Cafés, lesen gewisse Zeitungen, kaufen in gewissen Läden ein. Dasselbe tun wir online. Das bilden die Algorithmen ab. Im Falle von Facebook haben die Algorithmen sogar die Tendenz, Bubbles entgegenzuwirken. Provozieren mich Inhalte, reagiere ich, was Aufmerksamkeit und Interaktion bedeutet – und das ist ja die Währung der Plattform. Der negative Aspekt dieser Funktionsweise ist hingegen, dass sie Hatespeech und die Verbreitung von Falschnachrichten begünstigt.

Stichwort Provokation: Ist das der Grund, weshalb Influencer wie Andrew Tate mit chauvinistischen und frauenfeindlichen Inhalten solch eine Reichweite kreieren?
Der online Schlagabtausch zwischen Andrew Tate und Greta Thunberg vor einigen Jahren ist ein gutes Beispiel dafür. Da prallen zwei Bubbles aufeinander. Tate weiss genau, wie er Reaktionen provozieren und für sich nutzen kann. Es kümmert ihn nicht, wenn er Thunberg argumentativ unterliegt, denn ihre Argumente sind seiner Bubble egal. Durch die entstandene Kontroverse erreicht er aber ein neues Publikum.

Wäre es in Thunbergs Fall besser gewesen, ihn einfach zu ignorieren?
Fakt ist: Je mehr Reaktionen er bekommt, desto sichtbarer ist er auf den Plattformen. Er fühlt sich dadurch bestätigt und postet noch mehr ähnliche Inhalte. Mit seinen Äusserungen knüpft er an eine verletzte Männlichkeit an, ein Thema, das viele beschäftigt. Leider holt er damit viele junge Männer ab – auch bildungsnahe. Er bedient diffuse Ängste und bietet ähnlich dem Populismus in der Politik einfache Lösungen auf vielschichtige Probleme. Er zeichnet ein klares Feindbild: die Frauen. Darauf «richtig» zu reagieren, ist schwierig.

Sie arbeiten mit Jugendlichen, welche stark mit solchen Inhalten auf Social Media konfrontiert sind. Wie gehen diese damit um?
Es gibt immer wieder Phänomene wie Tate, die bei Jugendlichen die Runde machen. In Basel gibt es beispielsweise einen Lokalpolitiker, der online gut ankommt, weil er psychisch auffällig ist und skurrile Aktionen bietet. Die Videos gehen viral und die Jugendlichen interagieren aus Spass mit ihm. Die Tragik hinter seinen Geschichten sehen sie aber nicht. Er ist offen rechtsextrem, war in einige Wahlbetrugsfälle verwickelt, bereitet der Basler Politik grossen Aufwand und fügt ihr damit Schaden zu. Für Jugendliche ist es schwierig zu erkennen, dass sie ihn unterstützen, wenn sie mit ihm interagieren.

Sind sich junge Menschen bewusst, wie Algorithmen auf den Plattformen funktionieren, auf denen sie so viel Zeit verbringen?
Sie haben ein ungefähres Bewusstsein dafür, wie der Algorithmus auf TikTok wirkt. Ihr Wissen ist aber bruchstückhaft. Sie wissen, wie sie sich präsentieren müssen, um Reichweite zu bekommen. Was mit den Daten aus ihren Interaktionen geschieht oder welchen Einfluss die Inhalte auf die Gesellschaft haben, haben jedoch viele nicht im Blick. TikTok steht in den USA kurz vor dem Verbot, weil befürch-

tet wird, dass die chinesische Regierung über die App Zugriff auf Daten und damit Werkzeuge zur Manipulation einer sehr grossen Anzahl Nutzenden haben könnte. Die verschiedenen Tools haben noch viele solche Eigenschaften, die nicht bekannt sind und über die dringend mehr gesprochen und informiert werden muss.

Wie stereotyp sind die Geschlechterdarstellungen auf Social Media und in KI-Systemen?
Sie sind so stereotyp wie die Gesellschaft selber. Das Internet wirkt ein Stück weit als Katalysator und führt zu Überzeichnungen. Was die KI betrifft: Die von uns kreierten Inhalte bilden die Datengrundlage, aus der sich die KI bedient, um zu lernen. Wenn ich beispielsweise ein Bild eines Sozialarbeiters erfrage, spuckt das KI-basierte Tool eine weisse Frau bei der Arbeit aus. Statistisch gesehen gibt es tatsächlich mehr Frauen, die in der Sozialen Arbeit tätig sind. Es wird also eine aktuelle Realität abgebildet und reproduziert. Nicht die KI ist rassistisch oder sexistisch – wir sind es. Will man dagegen etwas tun, müsste man die Datenlage ändern oder der KI Regeln vorgeben.

Gemäss einer Umfrage sehen 20 Prozent der deutschen Bevölkerung KI als «eher männlich». Nur 3,5 Prozent nehmen sie weiblich wahr. Haben Sie eine Vermutung, weshalb das so ist?
Traurigerweise liegt dies vermutlich daran, dass «Technik» und «Intelligenz» immer noch männlich konnotiert sind – zumindest bei Menschen über dreissig. Geschlechtervorstellungen sitzen sehr tief und sind nicht so leicht veränderbar. Das zeigt auch das Beispiel von Nemo, der nonbinären Person, die dieses Jahr für die Schweiz den Eurovision Song Contest gewonnen hat. Auch wenn ich mich als *Ally* – also als Verbündeter von queeren Menschen – sehe, braucht mein Kopf oft einen Zwischenschritt, um im Gespräch das Pronomen wegzulassen.

«Kelly is a warm person, Joseph is a role model»: Eine Studie besagt, dass KI-generierte Empfehlungsschreiben Männer und Frauen mit stereotypen Begriffen beschreiben.
Hier zeigt sich, wie solche Zuschreibungen Schaden anrichten und Probleme reproduzieren, die wir offline haben. Die gesellschaftliche Vielfalt wird unsichtbar gemacht. Wie auch die Gesellschaft hat die KI Mühe mit Ausnahmen. Sie berechnet vor allem Wahrscheinlichkeiten, aber das Leben und die Geschlechter entsprechen denen oft nicht.

KI bildet also Realitäten einer gesellschaftlichen Mehrheit ab. In unserer Verfassung geben wir uns aber Mühe, Minderheiten zu schützen. Wie schaffen wir das auch online?
Oft lassen sich Offline-Strategien auch online anwenden, wenn wir die besonderen Eigenschaften der digitalen Welten kennen. Eine Lösung kann sein, positive Beispiele und Inhalte zu kreieren und bestehende zu pushen, um die Datengrundlage zu verändern. Es gibt online beispielsweise Listen mit weiblichen Expertinnen, Speakerinnen oder Technikerinnen, die man für Veranstaltungen oder Arbeiten buchen kann. Das kreiert Sichtbar- und Aufmerksamkeit. Eine andere Möglichkeit ist es, sich organisiert in Kommentarspalten zu äussern. Diese sind ein Feld, das oft von Männern bewirtschaftet wird und wo Frauen sich zurückhalten, weil sie die Reaktionen zu fürchten haben.

Die aktuelle Studie «Lauter Hass – leiser Rückzug»[1] zeigt, dass sich Menschen – mehrheitlich Frauen – aus dem Netz zurückziehen, weil sie Hass erleben. Sie werden, teilweise orchestriert, niedergeschrien und mit Drohungen und anderen immer krasseren Reaktionen überhäuft. Als Folge davon fehlen ihre Stimmen, zum Beispiel als Datengrundlage für KI-Tools wie ChatGPT. Diese Dynamik muss korrigiert werden.

Wie kann man Gegensteuer geben?
Ich sehe drei Ansätze: Erstens, Bestehendes stärken: Also Inhalte, Engagements und Projekte weiblicher Creators unterstützen. Man kann sie liken, positiv kommentieren und Gegenreden bei misogynen Posts verfassen. Anklagen und Verfolgen von Online-Übergriffen hat seine Tücken, lässt sich aber umsetzen. Die Organisation Netz Courage macht es vor. Zweitens, Räume anbieten, in denen junge Frauen ihre Stimme finden und stärken können. Offline geschieht das bereits, beispielsweise in Mädchentreffs. Hier können junge Frauen experimentieren, Fuss fassen, Netzwerke bilden und Vorbilder erleben. Drittens müssen sich auch Männer bewegen und einmischen, wenn sich Kollegen gegenüber Frauen abwertend äussern. Misogynie findet ihren Nährboden oft in Männerrunden.

«Die Tools haben viele Eigenschaften, die nicht bekannt sind und über die dringend mehr gesprochen und informiert werden muss.»

«Männer müssen sich auch bewegen und einmischen, wenn sich Kollegen gegenüber Frauen abwertend äussern.»

Was müsste im Bereich Algorithmen und KI getan werden, damit diese der Gleichstellung der Geschlechter dienen könnte?
Um diskriminierende Tendenzen aufzudecken und zu verstehen, müssen wir Algorithmen und KI und ihre Folgen erforschen. Es braucht eine breite gesellschaftliche Diskussion darüber, was wir dulden wollen und wie wir diese Tools nutzen möchten. Und da stellen sich medienethische Fragen, die schwierig zu beantworten sind. Auf der einen Seite verstärkt jede Reproduktion das bestehende Unrecht. Auf der anderen Seite ist es problematisch, Algorithmen und KI in eine gewünschte Richtung zu manipulieren.

Mona Martin ist Journalistin und Politologin. Ihre Fokusthemen sind Gleichstellung der Geschlechter, soziale Gerechtigkeit, Gesundheit und Psychologie.

Anmerkung

1 Kompetenznetzwerk gegen Hass im Netz (Hg.): Lauter Hass – leiser Rückzug. Wie Hass im Netz den demokratischen Diskurs bedroht. Berlin, 2024. https://kompetenznetzwerk-hass-im-netz.de/lauter-hass-leiser-rueckzug/ (abgerufen am: 03.07.2024).

Abstracts

«Ce n'est pas l'IA qui est raciste et sexiste – c'est nous»

Les algorithmes et l'intelligence artificielle se sont installés dans notre quotidien. TikTok, Instagram et ChatGPT deviennent de plus en plus sophistiqués dans l'analyse de nos comportements, qu'ils exploitent pour monopoliser notre attention. Ils reproduisent des stéréotypes de genre et offrent une plateforme à des contenus et des protagonistes misogynes. Formateur dans le domaine des médias et expert en transformation numérique, **Laurent Sedano** explique dans cette interview pourquoi ces contenus bénéficient de tant d'espace, ce qui peut être fait contre les discours haineux et des phénomènes comme Andrew Tate, et que la reproduction des stéréotypes commence dans la vraie vie.

«L'IA non è razzista o sessista – lo siamo noi»

Gli algoritmi e l'intelligenza artificiale fanno ora parte della nostra vita quotidiana. Attraverso l'analisi sempre più sofisticata del nostro comportamento, i canali di comunicazione e intrattenimento utili per il nostro lavoro come TikTok, Instagram e ChatGPT sono in grado di catturare e mantenere la nostra attenzione. Riproducono immagini di genere stereotipate e offrono una piattaforma a contenuti e protagonisti misogini. In quest'intervista, l'educatore mediale ed esperto in digitalizzazione **Laurent Sedano** spiega perché proprio a questi contenuti viene concesso così tanto spazio, che cosa può essere fatto per contrastare i discorsi d'odio e i fenomeni come Andrew Tate e come la riproduzione di stereotipi abbia origine offline.