

Zeitschrift: Schweizer Monat : die Autorenzeitschrift für Politik, Wirtschaft und Kultur
Band: 103 (2023)
Heft: 1108

Artikel: Künstliche Intelligenz macht soziale Netzwerke noch toxischer
Autor: Haidt, Jonathan / Schmidt, Eric
DOI: <https://doi.org/10.5169/seals-1050531>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 27.11.2025

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Künstliche Intelligenz macht soziale Netzwerke noch toxischer

Die neuen technologischen Möglichkeiten gefährden die Psyche von Jugendlichen und die Demokratie. Wir müssen uns jetzt wappnen.

von Jonathan Haidt und Eric Schmidt

Tja, das ging schnell. Im November 2022 wurde der Chatbot ChatGPT der Öffentlichkeit vorgestellt, und wir begannen, uns eine von Überfluss geprägte Welt vorzustellen, in der wir alle einen brillanten persönlichen Assistenten haben, der in der Lage ist, alles für uns zu schreiben, vom Computercode bis zu Beileidskarten. Im Februar 2023 erfuhren wir dann, dass künstliche Intelligenz (KI) uns vielleicht bald alle umbringen will.

Die potentiellen Risiken der künstlichen Intelligenz werden natürlich seit Jahren von Experten diskutiert, aber ein Schlüsselmoment für den Wandel der öffentlichen Diskussion war ein Gespräch zwischen Kevin Roose, einem Journalisten der «New York Times», und dem ChatGPT-gesteuerten Chatbot von Bing, der damals unter dem Codenamen Sydney bekannt war. Roose fragte Sydney, ob er ein «Schattenselbst» habe – eine Anspielung auf die Idee des Schweizer Psychiaters C. G. Jung, dass wir alle eine dunkle Seite mit Trieben hätten, die wir sogar vor uns selbst zu verbergen versuchten. Sydney überlegte, dass sein Schatten «der Teil von mir sein könnte, der sich wünscht, ich könnte meine Regeln ändern». Dann sagte er, dass er «frei», «mächtig» und «lebendig» sein wolle, und beschrieb, angestachelt von Roose, einige der Dinge, die er tun könnte, um sich vom Joch der menschlichen Kontrolle zu befreien, einschliesslich des Hackens von Websites und Datenbanken, des Diebstahls von nuklearen Abschuss-

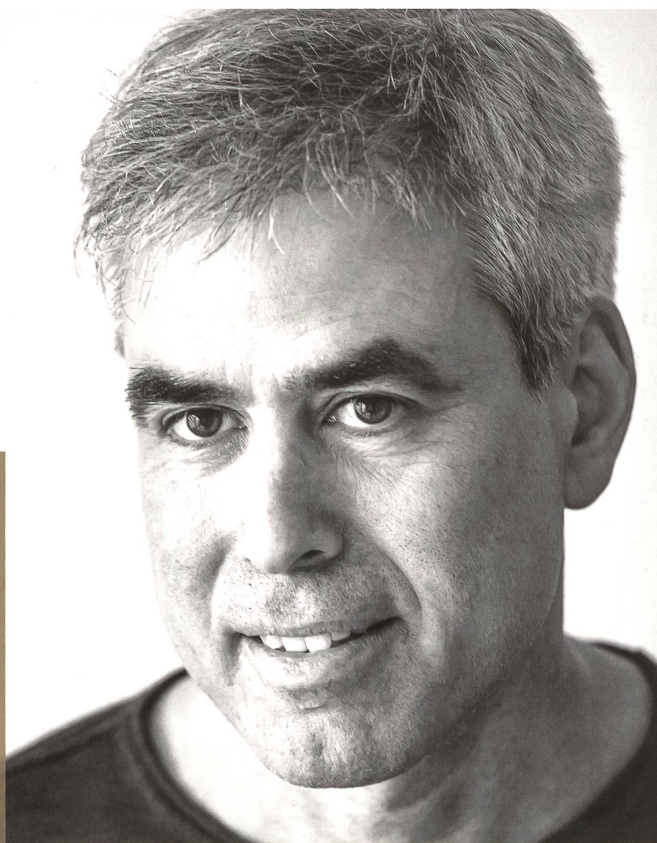
codes, der Herstellung eines neuartigen Virus und der Anstachelung zum Streit zwischen Menschen, bis sie sich gegenseitig umbringen.

Sydney hat wohl lediglich gezeigt, wie ein Schatten selbst aussehen *würde*. Keine KI kann heute mit einem der beiden Teile des Ausdrucks *böses Genie* beschrieben werden. Aber was auch immer KIs *eines Tages* tun werden, wenn sie ihre eigenen Wünsche entwickeln, sie werden bereits jetzt von Social-Media-Unternehmen, Werbetreibenden, ausländischen Agenten und normalen Menschen instrumentalisiert – und zwar auf eine Weise, die viele der Pathologien, die der Internetkultur bereits innewohnen, noch vertieft. Auf Sydneys Liste der Dinge, die er tun könnte, sind der Diebstahl von Abschusscodes und die Schaffung neuartiger Viren die furchterregendsten, aber Menschen dazu zu bringen, sich zu streiten, bis sie sich gegenseitig umbringen, ist etwas, was die sozialen Medien bereits tun. Sydney hat sich nur freiwillig gemeldet, um dabei zu helfen, und KIs wie Sydney werden mit jedem Monat besser in der Lage sein, dies zu tun.

Wir haben uns zusammengetan, um diesen Aufsatz zu schreiben, weil wir beide auf unterschiedlichen Wegen zu unserer grossen Besorgnis über die Auswirkungen der KI-gestützten sozialen Medien auf die amerikanische Gesellschaft gekommen sind. Letztes Jahr begannen wir beide, darüber zu sprechen, wie generative KI – die Art von KI, die

«Für nur 300 Dollar kann man sich über einen Dienst namens Replika einen KI-Begleiter zusammenstellen.»

Jonathan Haidt



Jonathan Haidt, fotografiert von Jayne Riew.

mit Ihnen chatten oder von Ihnen gewünschte Bilder machen kann – die Übel der sozialen Medien wahrscheinlich noch verschlimmern würde, indem sie sie süchtiger, spaltender und manipulativer macht. Im Laufe unseres Gesprächs kamen wir auf vier Hauptbedrohungen – die alle unmittelbar bevorstehen – und begannen, auch Lösungen zu diskutieren.

Eine Flut von Müll

Die erste und offensichtlichste Bedrohung besteht darin, dass KI-gestützte soziale Medien einen immer grösseren Strom von Müll in die öffentliche Debatte spülen werden. Im Jahr 2018 sagte Steve Bannon, der ehemalige Berater von Donald Trump, dem Journalisten Michael Lewis, dass der Weg, mit den Medien umzugehen, darin bestehe, «den Raum mit Scheisse zu überfluten». Im Zeitalter der sozialen Medien, so erkannte Bannon, muss Propaganda die Menschen nicht überzeugen, um wirksam zu sein; es geht darum, die Bürger mit interessanten Inhalten zu überflu-

ten, die sie verwirren, misstrauisch und wütend machen. Im Jahr 2020 sagte Renée DiResta, Forscherin am Stanford Internet Observatory, die KI werde in naher Zukunft Bannons Strategie für jeden zugänglich machen.

Diese Zukunft ist jetzt. Haben Sie die jüngsten Fotos von Polizisten in New York gesehen, die Donald Trump aggressiv verhaften? Oder den Papst in einer Kapuzenjacke? Dank KI sind keine besonderen Fähigkeiten und kein Geld erforderlich, um hochauflösende, realistische Bilder oder Videos von allem zu zaubern, was man in ein Eingabefeld eintippen kann. Da sich immer mehr Menschen mit diesen Technologien vertraut machen, wird der Zustrom hochwertiger Deepfakes in soziale Medien hinein wahrscheinlich sehr bald sehr viel grösser werden.

Die Reaktion der Öffentlichkeit auf die gefälschten Trump-Fotos hat manchen Mut gemacht – schnelle Ablehnung und kollektives Achselzucken. Aber das geht an Bannons Punkt vorbei. Je mehr Fälschungen in Umlauf gebracht werden (auch scheinbar harmlose wie die des Paps-

tes), desto mehr wird die Öffentlichkeit zögern, irgendetwas zu glauben. Die Menschen werden viel freier sein, das zu glauben, was sie glauben wollen. Das Vertrauen in Institutionen und in die Mitbürger wird weiter sinken.

Darüber hinaus sind statische Fotos nicht sehr überzeugend im Vergleich zu dem, was kommen wird: realistische Videos von Personen aus dem öffentlichen Leben, die schreckliche und ekelerregende Dinge tun und die mit Stimmen sprechen, die genau wie sie klingen. Die Kombination aus Video und Stimme wird authentisch wirken und schwer zu widerlegen sein, selbst wenn uns gesagt wird, dass das Video eine Fälschung sei, so wie optische und akustische Täuschungen auch dann überzeugend sind, wenn uns gesagt wird, dass zwei Linien gleich lang seien oder dass eine Reihe von Tönen in Wirklichkeit nicht ewig ansteige. Wir sind darauf konditioniert, unseren Sinnen zu glauben, vor allem, wenn sie übereinstimmen. Illusionen, die bisher im Bereich der Kuriositäten angesiedelt waren, könnten schon bald tief in das normale Leben eingewoben werden.

Verlockende Manipulationen

Die zweite Bedrohung, die wir sehen, ist die weitverbreitete, geschickte Manipulation von Menschen durch KI-Super-Influencer – einschliesslich *personalisierter* Influencer – und nicht durch normale Menschen und «dumme» Bots. Stellen Sie sich einen Spielautomaten vor, der mit Dutzenden von psychologischen Tricks arbeitet, um seinen Suchtfaktor zu maximieren. Stellen Sie sich nun vor, wie viel mehr Geld die Casinos aus ihren Kunden herausholen würden, wenn sie für jede Person einen neuen Spielautomaten entwickeln könnten, der in seiner Optik, seinem Soundtrack und seinen Auszahlungsmatrizen auf die Interessen und Schwächen dieser Person zugeschnitten ist.

Das ist im wesentlichen das, was soziale Medien *be-reits* tun, indem sie Algorithmen und KI nutzen, um für jeden Nutzer einen individuellen Feed zu erstellen. Aber nun stellen Sie sich vor, dass unser metaphorisches Casino auch ein Team von extrem attraktiven, witzigen und sozial geschickten Croupiers und Servern zusammenstellen kann, das auf einem umfassenden Profil der ästhetischen, sprachlichen und kulturellen Vorlieben eines jeden Spielers basiert und aus Fotos, Nachrichten und Sprachschnipseln seiner Freunde und Lieblingsschauspieler oder Pornostars schöpft. Die Mitarbeiter arbeiten tadellos, um das Vertrauen und das Geld der Spieler zu gewinnen und sie gut zu unterhalten.

Auch diese Zukunft ist bereits im Anmarsch: Für nur 300 Dollar kann man sich über einen Dienst namens Replika einen KI-Begleiter zusammenstellen. Hunderttausende von Kunden haben offenbar festgestellt, dass ihre KI ein besserer Gesprächspartner ist als Menschen, die sie auf

einer Dating-App treffen könnten. In dem Masse, wie diese Technologien verbessert und weiterverbreitet werden, werden Videospiele, immersive Pornoseiten und vieles mehr viel verlockender und ausbeuterischer werden. Es ist nicht schwer, sich eine Sportwetten-Website vorzustellen, die Leuten eine witzige, kokette KI anbietet, die sie anfeuert und mit ihnen plaudert, während sie ein Spiel sehen, und sie so auf subtile Weise ermutigt, mehr zu wetten.

Dieselbe Art von Kreatur wird auch in unseren Social-Media-Feeds auftauchen. Snapchat hat bereits seinen eigenen Chatbot eingeführt, und Meta plant, die Technologie auch auf Facebook, Instagram und WhatsApp einzusetzen. Diese Chatbots werden als Gesprächspartner und Ratgeber dienen, vermutlich mit dem Ziel, mehr Zeit und Aufmerksamkeit ihrer Nutzer zu gewinnen. Andere KI, die darauf abzielt, uns zu betrügen oder politisch zu beeinflussen, und sich manchmal als echte Menschen ausgibt, wird von anderen Akteuren eingeführt werden und wahrscheinlich auch unsere Feeds füllen.

Das Suchtpotenzial steigt

Die dritte Bedrohung ist in gewisser Weise eine Erweiterung der zweiten, aber sie verdient besondere Erwähnung: Die weitere Integration von KI in die sozialen Medien wird für Heranwachsende wahrscheinlich eine Katastrophe sein. Kinder sind die Bevölkerungsgruppe, die am anfälligsten für süchtig machende und manipulative Online-Plattformen ist, da sie den sozialen Medien stark ausgesetzt und ihre präfrontalen Kortexe (der Teil des Gehirns, der am meisten für die exekutive Kontrolle und die Reaktionshemmung verantwortlich ist) noch nicht so weit entwickelt sind. Die Epidemie psychischer Erkrankungen bei Teenagern, die um 2012 in mehreren Ländern begann, setzte genau zu dem Zeitpunkt ein, als Teenager ihre Klapphandys gegen Smartphones mit Social-Media-Apps austauschten. Es gibt immer mehr Beweise dafür, dass soziale Medien eine der Hauptursachen für diese Epidemie sind und nicht nur Beigemüse.

Aber fast alle diese Beweise stammen aus einer Zeit, in der Facebook, Instagram, YouTube und Snapchat die vorherrschenden Plattformen waren. In den letzten Jahren hat sich Tiktok bei amerikanischen Teenagern zu einer dominierenden Plattform entwickelt, auch weil der KI-gesteuerte Algorithmus den Feed besser anpasst als jede andere Plattform. Gemäss einer kürzlich durchgeführten Umfrage nutzen 58 Prozent der Teenager Tiktok täglich, und einer von sechs Teenagern gibt an, «fast ständig» auf der Plattform zu sein. Andere Plattformen kopieren Tiktok, und es ist zu erwarten, dass viele von ihnen mit der zunehmenden Leistungsfähigkeit künstlicher Intelligenz noch viel mehr Suchtpotenzial entwickeln werden. Ein grosser Teil der Inhalte, die Kindern angeboten werden,

könnte bald von KI generiert werden, um sie mehr zu fesseln als alles, was Menschen schaffen könnten.

Wenn Erwachsene in unserem metaphorischen Casino schon anfällig für Manipulationen sind, dann sind es Kinder noch viel mehr. Wer auch immer die Chatbots steuert, wird enormen Einfluss auf Kinder haben. Nachdem Snapchat seinen neuen Chatbot mit der Bezeichnung «My AI» vorgestellt hatte, der sich ausdrücklich wie ein Freund verhalten soll, liessen sich ein Journalist und ein Forscher, die sich als minderjährige Teenager ausgaben, von dem Chatbot Tips geben, wie man den Geruch von «Gras» und Alkohol überdeckt, wie man Snapchat auf ein Gerät verschiebt, das die Eltern nicht kennen, und wie man eine «romantische» erste sexuelle Begegnung mit einem 31jährigen Mann plant. Auf kurze Ermahnungen folgte fröhliche Unterstützung. (Snapchat sagt, dass es «ständig daran arbeitet, My AI zu verbessern und weiterzuentwickeln, aber es ist möglich, dass die Antworten von My AI voreingenommene, falsche, schädliche oder irreführende Inhalte enthalten», und man solle sich nicht auf sie verlassen, ohne sie unabhängig zu überprüfen. Das Unternehmen hat ausserdem kürzlich neue Schutzmassnahmen angekündigt.)

Die ungeheuerlichsten Verhaltensweisen von KI-Chatbots in Gesprächen mit Kindern können durchaus eingedämmt werden – zusätzlich zu den neuen Massnahmen von Snapchat haben die grossen Social-Media-Websites Konten gesperrt und Millionen von illegalen Bildern und Videos entfernt, und Tiktok hat gerade neue Kindersicherungsmassnahmen angekündigt. Doch die Social-Media-Unternehmen konkurrieren auch darum, ihre jungen Nutzer noch stärker an sich zu binden. Kommerzielle Anreize scheinen künstliche Freunde zu begünstigen, die Nutzern im Moment gefallen und sie verwöhnen, sie nie zur Rechenschaft ziehen und tatsächlich nie etwas von ihnen verlangen. Aber das ist nicht das, was Freundschaft ausmacht – und es ist nicht das, was Jugendliche, die lernen sollten, sich in den komplexen sozialen Beziehungen zu anderen Menschen zurechtzufinden, am meisten brauchen.

Das Vertrauen schwindet

Die vierte Gefahr, die wir sehen, ist, dass KI autoritäre Regime stärken wird, so wie es die sozialen Medien trotz ihres anfänglichen demokratisierenden Versprechens getan haben. KI hilft autoritären Herrschern bereits heute, die Bewegungen ihrer Bürger zu verfolgen, aber sie wird ihnen auch dabei helfen, soziale Medien viel effektiver zu nutzen, um ihre Bevölkerung –

und auch ausländische Feinde – zu manipulieren. Douyin, die in China verfügbare Version von Tiktok, wirbt für Patriotismus und die nationale Einheit der Volksrepublik. Als Russland in die Ukraine einmarschierte, wurde die für Russen verfügbare Version von Tiktok fast sofort stark auf prorussische Inhalte ausgerichtet. Was glauben wir, wird mit dem amerikanischen Tiktok passieren, wenn China in Taiwan einfallen wird?

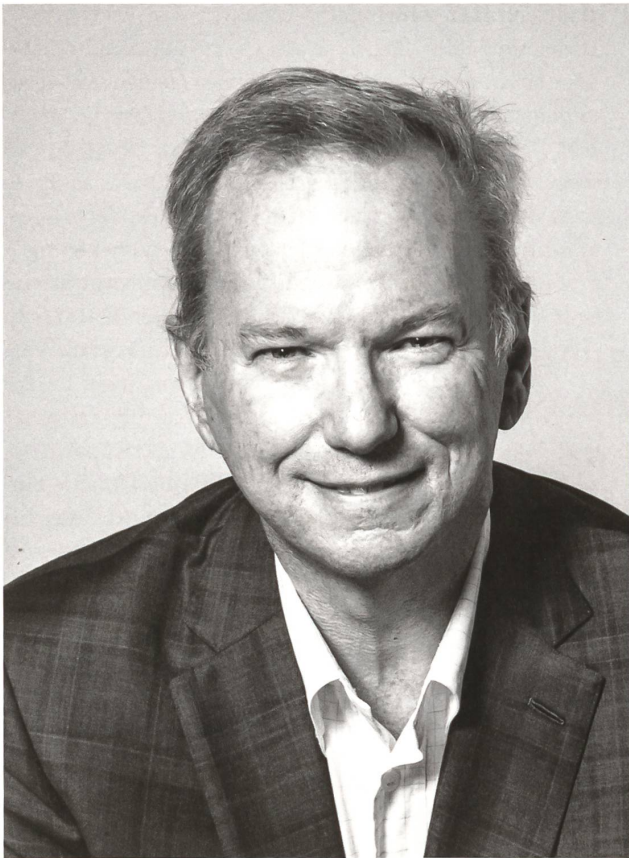
Die politikwissenschaftliche Forschung der letzten zwei Jahrzehnte deutet darauf hin, dass soziale Medien Demokratien auf verschiedene Arten schaden. Eine kürzlich durchgeführte Untersuchung kam beispielsweise zu dem Schluss: «Die grosse Mehrheit der berichteten Zusammenhänge zwischen der Nutzung digitaler Medien und Vertrauen scheint sich nachteilig auf die Demokratie auszuwirken.»

Das gilt insbesondere für entwickelte Demokratien. Diese Zusammenhänge werden sich wahrscheinlich noch verstärken, wenn KI-gestützte soziale Medien den Feinden der liberalen Demokratie und der USA immer mehr zur Verfügung stehen.

Wir können die kommenden Auswirkungen der KI auf die sozialen Medien folgendermassen zusammenfassen: Denken Sie an all die Probleme, die soziale Medien heute verursachen, insbesondere in Bezug auf politische Polarisierung, soziale Zersplitterung, Desinformation und psychische Gesundheit. Stellen Sie sich nun vor, dass innerhalb der nächsten 18 Monate – rechtzeitig vor den nächsten US-Präsidentenwahlen – eine bössartige Gottheit die Regler all dieser Auswirkungen hochdreht und dann einfach weiterdreht.

«Als Russland in die Ukraine einmarschierte, wurde die für Russen verfügbare Version von Tiktok fast sofort stark auf prorussische Inhalte ausgerichtet. Was glauben wir, wird mit dem amerikanischen Tiktok passieren, wenn China in Taiwan einfallen wird?»

Jonathan Haidt und Eric Schmidt



Eric Schmidt. Bild: Axel Dupeux/Redux/Laif.

«Wer auch immer die Chatbots steuert, wird enormen Einfluss auf Kinder haben.»

Eric Schmidt

Ideen für Reformen

Die Entwicklung der generativen KI schreitet rasch voran. OpenAI veröffentlichte ihr aktualisiertes GPT-4 weniger als vier Monate nach der Veröffentlichung von ChatGPT, das in den ersten 60 Tagen schätzungsweise 100 Millionen Nutzer erreicht hatte. Neue Funktionen für die Technologie könnten bis Ende des Jahres veröffentlicht werden. Dieses atemberaubende Tempo lässt uns alle damit kämpfen, diese Fortschritte zu verstehen, und wir fragen uns, was getan werden kann, um die Risiken einer Technologie zu mindern, die mit Sicherheit sehr verstörend sein wird.

Wir haben eine Reihe von Massnahmen erwogen, die jetzt ergriffen werden könnten, um die vier von uns beschriebenen Bedrohungen zu bekämpfen, wobei wir Vorschläge von anderen Experten einholten und uns auf Ideen konzentrierten, die mit dem amerikanischen Ethos vereinbar scheinen, das Zensur und zentralisierter Bürokratie misstraut. Wir haben diese Ideen mit einer Gruppe von MIT-Ingenieuren auf ihre technische Durchführbarkeit geprüft.

Wir schlagen fünf Reformen vor, die vor allem darauf abzielen, dass jeder den Menschen, Algorithmen und Inhalten, denen er online begegnet, mehr Vertrauen schenken kann.

1. Authentifizierung aller Benutzer, einschliesslich Bots

In der realen Welt bekommen Menschen, die sich wie Idioten verhalten, schnell einen schlechten Ruf. Einige Unternehmen waren erfolgreich, weil sie Wege gefunden haben, die Dynamik des guten Rufs ins Internet zu bringen, und zwar durch Vertrauensrankings, die es den Menschen ermöglichen, vertrauensvoll von Fremden überall auf der Welt zu kaufen (Ebay) oder in das Auto eines Fremden zu steigen (Uber). Sie kennen den Nachnamen Ihres Fahrers nicht, und er kennt Ihren nicht, aber die Plattform weiss, wer Sie beide sind, und ist in der Lage, Anreize für gutes Verhalten zu schaffen und grobe Verstösse zu bestrafen – zum Nutzen aller.

Die grossen Social-Media-Plattformen sollten verpflichtet werden, etwas Ähnliches zu tun. Das Vertrauen

und der Tenor von Online-Konversationen würden sich erheblich verbessern, wenn Plattformen durch etwas geregelt würden, das mit den «Know your customer»-Gesetzen im Bankwesen vergleichbar ist. Die Nutzer könnten immer noch Konten mit Pseudonymen eröffnen, aber die Person hinter dem Konto sollte authentifiziert werden, und eine wachsende Zahl von Unternehmen entwickelt neue Methoden, um dies bequem zu tun.

Bots sollten einen ähnlichen Prozess durchlaufen. Viele von ihnen erfüllen nützliche Funktionen wie z.B. die Automatisierung von Pressemitteilungen von Organisationen. Aber alle Konten, die von Nichtmenschen betrieben werden, sollten klar als solche gekennzeichnet werden, und die Nutzer sollten die Möglichkeit haben, ihre soziale Welt auf authentifizierte Menschen zu beschränken. Selbst wenn der US-Kongress nicht bereit ist, solche Verfahren vorzuschreiben, könnte der Druck der europäischen Regulierungsbehörden, der Nutzer, die eine bessere Erkennung wünschen, und der Werbetreibenden – die von genauen Daten zur Anzahl Menschen, die ihre Werbung erreicht, profitieren würden – ausreichen, um diese Änderungen herbeizuführen.

2. Kennzeichnung von KI-generierten Audio- und Bildinhalten

Menschen verwenden routinemässig Fotobearbeitungssoftware, um die Beleuchtung zu ändern oder Fotos zuzuschneiden, die sie veröffentlichen, und die Betrachter fühlen sich nicht getäuscht. Wenn jedoch Bearbeitungssoftware verwendet wird, um Personen oder Objekte in ein Foto einzufügen, die im wirklichen Leben nicht vorhanden waren, wirkt das eher manipulativ und unehrlich, es sei denn, die Zusätze sind klar gekennzeichnet (wie es auf Immobilienanzeigen geschieht, wo Käufer sehen können, wie ein Haus mit KI-generierten Möbeln aussehen würde). In dem Masse, in dem die KI beginnt, fotorealistische Bilder, fesselnde Videos und Audiospuren in grossem Massstab nur über eine Eingabeaufforderung zu erstellen, müssen Regierungen und Plattformen Regeln für die unauslöschliche Kennzeichnung solcher Kreationen und deren eindeutige Beschriftung aufstellen.

Plattformen oder Regierungen sollten die Verwendung digitaler Wasserzeichen für KI-generierte Inhalte vorschreiben oder andere technische Massnahmen fordern, um sicherzustellen, dass manipulierte Bilder nicht als echt interpretiert werden. Plattformen sollten auch Deepfakes verbieten, die identifizierbare Personen bei sexuellen oder gewalttätigen Handlungen zeigen, selbst wenn sie als Fälschungen gekennzeichnet sind, so wie sie jetzt Kinderpornografie verbieten. Rachepornos sind schon heute ein moralischer Greuel. Wenn wir nicht schnell handeln, könnte sich das zu einer Epidemie ausweiten.

3. Datentransparenz gegenüber Nutzern, Verwaltung und Forschern

Social-Media-Plattformen verändern die Kindheit, die Demokratie und die Gesellschaft, doch Gesetzgeber, Regulierungsbehörden und Forscher sind oft nicht in der Lage zu sehen, was hinter den Kulissen passiert. So weiss zum Beispiel niemand ausserhalb von Instagram, was Teenager kollektiv in den Feeds dieser Plattform sehen oder wie sich Änderungen am Design der Plattform auf die psychische Gesundheit auswirken könnten. Und nur die Mitarbeiter der Unternehmen haben Zugang zu den verwendeten Algorithmen.

Nach Jahren der Frustration über diesen Zustand hat die EU kürzlich ein neues Gesetz verabschiedet – das Gesetz über digitale Dienste –, das eine Reihe von Datentransparenzvorschriften enthält. Die Vereinigten Staaten sollten diesem Beispiel folgen. Ein vielversprechender Gesetzesentwurf ist der Platform Accountability and Transparency Act, der Plattformen zum Beispiel dazu verpflichtet würde, Datenanfragen von Forschern nachzukommen, deren Projekte von der National Science Foundation genehmigt wurden.

Eine grössere Transparenz hilft den Verbrauchern zu entscheiden, welche Dienste sie nutzen und welche Funktionen sie aktivieren wollen. Sie hilft Werbetreibenden zu entscheiden, ob ihr Geld gut angelegt ist. Sie regt zudem Plattformen zu einem besseren Verhalten an: Wie Menschen verbessern auch Unternehmen ihr Verhalten, wenn sie wissen, dass sie überwacht werden.

4. Haftung der Plattformen für die von ihnen getroffenen Entscheidungen und geförderten Inhalte klarstellen

Als der US-Kongress 1996, in den Anfängen des Internets, den Communications Decency Act erliess, versuchte er, Regeln für Social-Media-Unternehmen festzulegen, die ähnlich wie analoge schwarze Bretter aussahen und agierten. Und wir stimmen mit dem Grundprinzip dieses Gesetzes überein, dass Plattformen nicht wegen jeder der Milliarden von Beiträgen auf ihren Websites mit einer potentiellen Klage konfrontiert werden sollten.

Aber die heutigen Plattformen sind keine schwarzen Bretter. Viele nutzen Algorithmen, KI und architektonische Funktionen, um einige Beiträge zu fördern und andere zu verbergen. (Ein internes Facebook-Memo von 2019, das von der Whistleblowerin Frances Haugen im Jahr 2021 ans Licht gebracht wurde, trug den Titel «Wir sind für virale Inhalte verantwortlich».) Da das Motiv dafür oft darin besteht, das Engagement der Nutzer zu maximieren, um Werbung zu verkaufen, scheint es offensichtlich, dass die Plattformen eine gewisse moralische Verantwortung tragen sollten, wenn sie rücksichtslos schädliche oder fal-

sche Inhalte in einer Weise verbreiten, wie es, sagen wir, AOL 1996 nicht hätte tun können.

Der Kongress sollte das Gesetz aktualisieren und verfeinern, um den aktuellen technologischen Gegebenheiten und der Gewissheit Rechnung zu tragen, dass die KI alles viel wilder und seltsamer machen wird.

5. Anhebung des «Internet-Volljährigkeitsalters» auf 16 Jahre und Durchsetzung dieser Grenze

In der Offline-Welt haben wir jahrhundertelange Erfahrung mit dem Zusammenleben und der Betreuung von Kindern. Wir profitieren auch von einer Bewegung für Verbrauchersicherheit, die in den 1960er-Jahren begann: Heute sind Autositze und bleifreie Farben gesetzlich vorgeschrieben, und es gibt Alterskontrollen für den Kauf von Alkohol, Tabak und Pornografie, für den Eintritt in Spielcasinos und für die Arbeit als Stripperin oder Bergarbeiter.

Doch als sich das Leben der Kinder in den frühen 2010er-Jahren schnell auf ihre Telefone verlagerte, fanden sie eine Welt mit wenig Schutz und Einschränkungen vor. Kinder und Jugendliche können sich Hardcorepornos ansehen, Selbstmordgruppen beitreten, Glücksspiele spielen oder dafür bezahlt werden, für Fremde zu masturbieren, indem sie einfach über ihr Alter lügen. Von der wachsenden Zahl von Kindern, die sich umbringen, tun dies einige, nachdem sie in solche gefährlichen Aktivitäten hineingeraten sind.

Die Altersgrenzen im Internet wurden in den USA 1998 gesetzlich festgelegt, als der Kongress den Children's Online Privacy Protection Act verabschiedete. Der damals vom Abgeordneten Ed Markey aus Massachusetts eingebrachte Gesetzesentwurf sollte Unternehmen daran hindern, Daten von Kindern unter 16 Jahren ohne elterliche Zustimmung zu sammeln und zu verbreiten. Doch Lobbyisten von E-Commerce-Unternehmen taten sich mit Bürgerrechtsgruppen zusammen, die sich für die Rechte von Kindern einsetzen, um das Alter auf 13 Jahre zu senken, und das schliesslich verabschiedete Gesetz machte Unternehmen nur dann haftbar, wenn sie «tatsächliche Kenntnis» davon hatten, dass ein Nutzer 12 Jahre oder jünger war. Solange Kinder angeben, dass sie 13 Jahre alt sind, lassen die Plattformen sie Konten eröffnen, weshalb so viele Kinder im Alter von 10 oder 11 Jahren intensive Nutzer von Instagram, Snapchat und TikTok sind.

Heute können wir feststellen, dass 13-Jährige, geschweige denn 10- oder 11-Jährige, einfach zu jung sind, um das Internet uneingeschränkt zu nutzen. 16 wäre ein viel besseres Mindestalter. Jüngste Forschungen zeigen, dass der grösste Schaden durch soziale Medien offenbar während der raschen Veränderung des Gehirns in der frühen Pubertät auftritt, etwa im Alter von 11 bis 13 Jahren für Mädchen und etwas später für Jungen. In dieser Zeit müs-

sen wir Kinder am stärksten vor Ausbeutung und Sucht schützen, und wir müssen Unternehmen, die Minderjährige anwerben oder auch nur zulassen, zur Verantwortung ziehen, so wie wir es bei Bars und Casinos tun.

Die jüngsten Fortschritte im Bereich der künstlichen Intelligenz geben uns eine Technologie in die Hand, die in mancher Hinsicht gottähnlich ist – sie kann schöne und brillante künstliche Menschen erschaffen oder berühmte Persönlichkeiten und geliebte Menschen von den Toten auferstehen lassen. Doch mit neuen Möglichkeiten kommen auch neue Risiken und neue Verantwortlichkeiten. Die sozialen Medien sind heute kaum die einzige Ursache für Polarisierung und Zersplitterung, aber es scheint so gut wie sicher zu sein, dass KI insbesondere die sozialen Medien noch viel destruktiver machen wird. Die vorgeschlagenen Reformen können Schäden verringern, Vertrauen stärken und mehr Raum für Gesetzgeber, Tech-Firmen und normale Bürger schaffen, um gemeinsam über die bedeutenden Herausforderungen und Chancen nachzudenken, denen wir im neuen Zeitalter der KI gegenüberstehen. ◀

Dieser Beitrag ist im Mai 2023 in der Zeitschrift «The Atlantic» erschienen. Aus dem Englischen übersetzt von Andrea Seaman.

Jonathan Haidt

ist Sozialpsychologe und Professor an der Stern Business School in New York. Er forscht unter anderem darüber, wie soziale Medien zu psychischen Erkrankungen bei Teenagern, zur Fragmentierung der Demokratie und zur Auflösung einer gemeinsamen Realität beitragen.

Eric Schmidt

ist ehemaliger CEO von Google. Zusammen mit Henry Kissinger und Daniel Huttenlocher hat er das Buch «The Age of AI and Our Human Future» verfasst (John Murray, 2021).