Neue Zürcher Zeitung

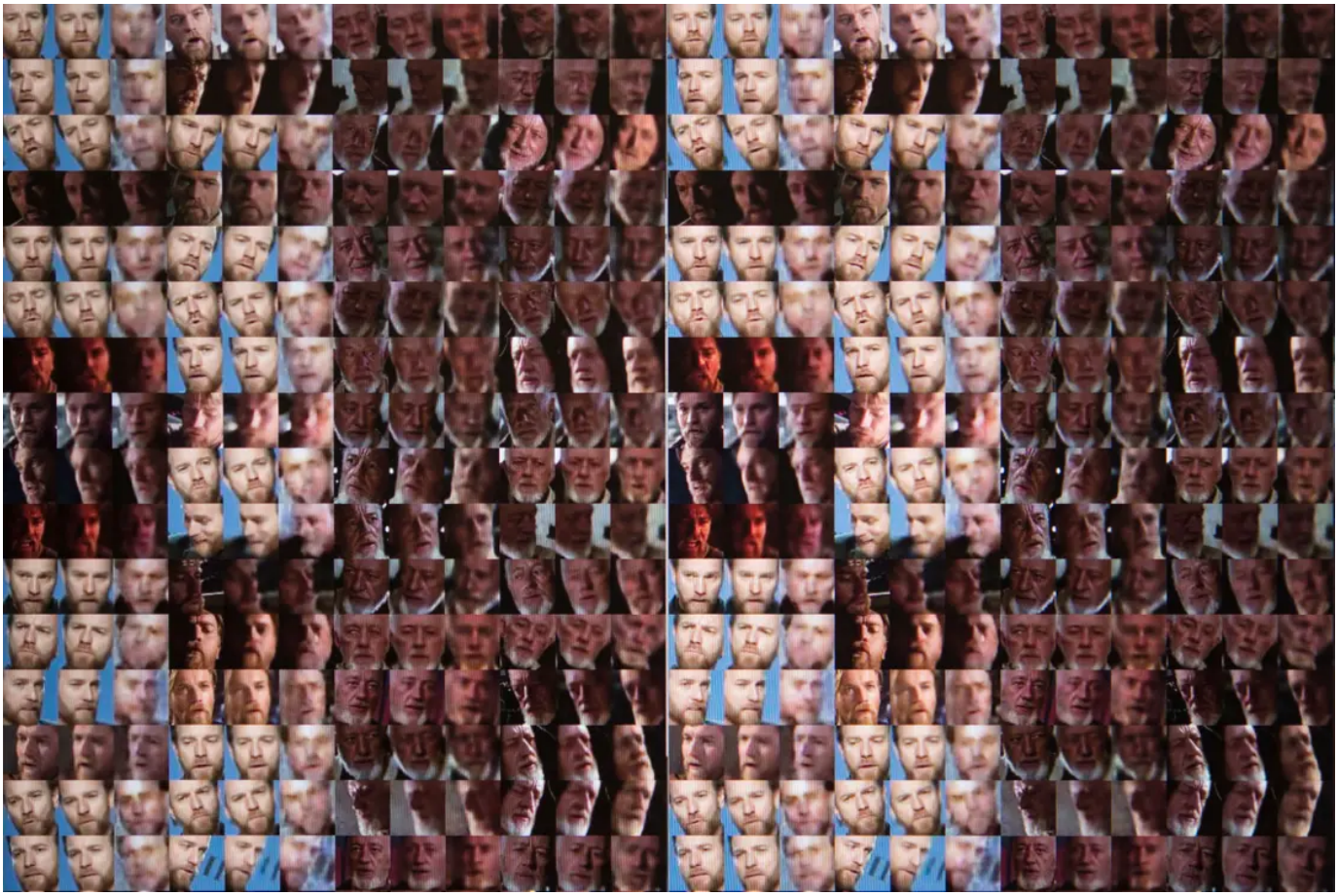
PRO

«Deepfakes werden zu schnell zu gut – und wir haben keine Antwort darauf»

Täuschend echt aussehende Fotos und Videos seien die grösste Gefahr im Bereich der künstlichen Intelligenz, sagt die Expertin Kathy Baxter. Trotzdem müsse man bei der Regulierung von KI umsichtig vorgehen – und die EU, Asien und die USA verfolgten alle unterschiedliche Ansätze.

Marie-Astrid Langer, San Francisco

03.10.2022, 11.00 Uhr



Es wird immer schwieriger, Deepfakes als solche zu identifizieren.

GitHub / Illustration Simon Tanner / NZZ

Dass Technologie sich schneller entwickelt als die Gesetze, die sie regulieren sollen, sieht man in kaum einem Bereich so eindrücklich wie in der künstlichen Intelligenz (KI). Ob bei der Kreditvergabe, der Medizin oder im Online-Shopping, künstlich intelligente Algorithmen entscheiden längst über grosse Teile unseres Alltags.

Was die KI können darf und was moralisch verwerflich wäre, haben viele Länder bisher nicht reguliert. Stattdessen liegt der Entscheid meist im Ermessen der Konzerne, welche die entsprechenden Algorithmen entwickeln – und das sorgt immer wieder für heftige Diskussionen. Viele Konzerne haben inzwischen eigene KI-Ethik-Spezialisten beschäftigt, die

ihnen bei der Beantwortung der Fragen im internationalen Kontext helfen und bei der Produktentwicklung auf die Finger schauen sollen – so wie Kathy Baxter im Softwarekonzern Salesforce.

Frau Baxter, was bereitet Ihnen als Spezialistin für Ethik in der künstlichen Intelligenz zurzeit am meisten Kopfzerbrechen?

Kathy Baxter begleitet im Technologiekonzern Salesforce die Entwicklung von ethischen Produkten aus dem Bereich der künstlichen Intelligenz.
PD

> Deepfakes sind meine grösste Sorge. Die Technologie dahinter ist so gut geworden und kann für alles genutzt werden. Viele von uns kennen etwa lustige Videos wie die Deepfake-Parodie auf Tom Cruise, die Sie bestimmt gesehen haben.
>
> Aber die Technologie wird auch für Erpressung verwendet, was sehr besorgniserregend ist. Speziell Audio-Deepfakes sind da ein ernstzunehmendes Sicherheitsproblem: Bei der Stimmüberprüfung im Telefonbanking kann man sich so als jemand anderes ausgeben und Zugriff auf alle Konten eines anderen Kunden erhalten.

Sind das noch theoretische Befürchtungen?

Die Probleme sind heute schon ganz real, auch für unsere Gesellschaft. Wir haben alle die manipulierten Videos von Nancy Pelosi und Joe Biden gesehen, die angeblich zeigen, wie die beiden Demokraten beim Sprechen lallen – also entweder betrunken oder dement sind. Viele Republikaner haben sie auch in den sozialen Netzwerken verbreitet. Wenn das, was wir sehen und hören, manipuliert wurde, wie sollen wir da noch wissen können, was die Wahrheit ist? Wie können wir nachweisen, was ein Deepfake ist? Eine andere Sorge ist, dass frei zugängliche Plattformen es jedermann ermöglichen, Kunst im Stil etwa von Salvador Dalí zu erstellen, ohne dass die Angehörigen des Künstlers dafür entschädigt würden. Es macht mir wirklich Sorge, dass all diese Programme so schnell so gut werden.

Was ist die Lösung?

Wir haben noch keine gefunden. Früher haben wir mit Wasserzeichen gearbeitet, um ein Original zu kennzeichnen. Aber ein Deepfake-Video oder -Bild fälscht einfach das Wasserzeichen mit. Und jetzt steigt auch noch Amazon in den Markt ein und will uns die Stimme unserer verstorbenen Grossmutter zurückbringen. Die Intention mag ja gut sein. Aber nicht nur Psychologen sagen, dass das eine ganz schlechte Idee sei. Man muss über die ungewollten Konsequenzen nachdenken, auch mit Blick auf Sicherheitsfragen. Wir brauchen mehr Firmen, die sich der Aufgabe verschreiben, Deepfakes zu identifizieren statt sie selbst zu bauen.

Ist die ganze Debatte um Ethik in der KI letztlich überflüssig, weil letztlich ohnehin jeder macht, was er will?

Manche Firmen haben tatsächlich überhaupt kein Problem damit, jegliche Art von Bild- und Stimmgeneratoren zu bauen und der Welt zur Verfügung zu stellen. Umgekehrt hat die Bildagentur Getty gerade angekündigt, dass sie keine von KI erzeugten Bilder auf ihrer Plattform erlauben werde. Mehr Firmen müssen solche Massnahmen ergreifen, solange wir noch keine staatliche Regulierung von KI haben. Es braucht Normen, deren Nichteinhaltung auch Konsequenzen hat.

Die EU arbeitet zurzeit an einer solchen Regulierung von KI. Wird sie dadurch einen First-Mover-Vorteil haben, weil sie weltweit die Spielregeln setzt?

Ja, in gewisser Weise sehen wir wieder den «Brüssel-Effekt» – die EU setzt de facto einen globalen Standard, weil andere Länder in ihre Fussstapfen bei Regulierungsfragen treten. Aber ich glaube nicht, dass es dieses Mal so extrem ausfallen wird wie bei der Datenschutz-Grundverordnung. Die Regulierung von KI hat weitreichende Folgen, und die Technologie ist sehr vielschichtig. Das Militär etwa nutzt KI zur Landesverteidigung, deswegen haben Regierungen rund um die Welt ein grosses Interesse daran, bei der Regulierung der Technologie die Oberhand zu behalten.

Geraten die USA da ins Hintertreffen, weil der Kongress kein KI-

Gesetz verabschiedet?

In den USA heisst es immer wieder: Wir können KI nicht stärker regulieren, als China es tut, weil wir sonst einen strategischen Nachteil hätten. Schlechte Regulierung ist schlimmer als keine Regulierung. Wir werden zunächst Richtlinien und Hilfsmittel entwickeln, um Firmen dabei zu helfen, die richtigen Entscheide zu treffen, bevor wir Regulierungen verabschieden. Wir haben ausserdem gesehen, dass viele asiatische Länder wie Singapur, Australien und Neuseeland KI bewusst noch nicht regulieren, weil noch so viele Fragen offen sind. Und China macht wieder einmal sein eigenes Ding.

Grosse Technologiekonzerne wie Salesforce versuchen, sich selbst zu regulieren, und haben Architekten für ethische KI wie Sie eingestellt. Wer entscheidet bei Salesforce, was moralisch okay ist?

Wir haben ein übergeordnetes Büro für ethischen und menschlichen Gebrauch (Office of Ethical and Humane Use), als Teil dessen arbeite ich mit meinem Team an Fragestellungen etwa zum Datenschutz. Zudem haben wir ein Team, das klare Einschränkungen erarbeitet, wofür unsere Technologie verwendet werden darf; zum Beispiel darf man unsere Computer-Vision-Technologie nicht für Gesichtserkennung verwenden.

Verhindern Sie auch einmal eine Produktlancierung wegen ethischer Bedenken?

Vor einigen Jahren hatten unsere Ingenieure ein Computermodell entwickelt, das analysiert, was Kunden über eine Firma oder ein konkretes Produkt denken, basierend auf dem, was sie in den sozialen Netzwerken darüber sagen. Mein Team war in die Entwicklung nicht einbezogen worden, und einer der Ingenieure stellte gegen Ende des Entwicklungsprozesses fest, dass das Modell einen heftigen Bias hatte: Wenn in einem Satz die Wörter «schwul» oder «Feministin» vorkamen, dann wertete der Algorithmus dies als negative Aussage. Wir fanden heraus, dass die frei zugängliche Datenbank, mit der der Algorithmus trainiert worden war, berüchtigt für Vorurteile war. Wir verhinderten letztlich die Produktlancierung, bis der Algorithmus mit einer anderen, besseren Datenbank neu geschult werden konnte.

Wie bleiben Sie als Expertin für ethische KI sich Ihrer eigenen Vorurteile und Verzerrungen bewusst?

Ich besuche Konferenzen, lese Papers – aber auch ich werde immer solche Biases haben. Es ist wichtig, dass wir uns in unserem Team gegenseitig auf solche Vorurteile hinweisen und eine Arbeitsumgebung schaffen, in der wir uns gegenseitig hinterfragen.

Es gibt immer wieder Kritik daran, dass grosse Technologiekonzerne die Debatte um Ethik in der KI mitbestimmen wollen. Salesforce zum Beispiel ist ein Diamant-Sponsor der weltweit führenden Konferenz für KI, der NeurIPS. Ist das moralisch verwerflich?

Das ist eine ganz grosse Diskussion. Google, um Ihnen ein anderes Beispiel zu geben, sponserte lange eine andere grosse Konferenz, zu Ethik und Computersystemen. Nachdem Google zwei seiner führenden KI-Forscher – Timnit Gebru und Margaret Mitchell – entlassen hatte, lehnten die Organisatoren dieser Konferenz Google als Sponsor ab, weil der Konzern nicht mehr die Werte der Konferenz repräsentiere.

Aber alle grossen Firmen sponsern diese Konferenzen – und das Gleiche tun grosse Universitäten wie Stanford, Yale und Georgia Tech. Als Sponsor bekommt man einen Rekrutierungsstand und eine Adressliste der Teilnehmer – aber es bedeutet nicht, dass auch mehr wissenschaftliche Aufsätze zugelassen werden, denn die werden ja für eine Konferenz blind von anderen überprüft. Wenn diese Konferenzen aber nicht mehr von Firmen oder Universitäten gesponsert werden dürfen, woher kommt dann das Geld, dass Mitarbeiter kleiner Firmen, Studenten oder Arbeitslose sie besuchen können? Da sind noch viele Fragen offen.

Kathy Baxter

leitet den Bereich Ethical AI Practice für den Softwarekonzern Salesforce am Firmensitz in San Francisco. Mit einem Studienhintergrund in angewandter Psychologie beschäftigt sie sich seit 25 Jahren mit Fragen der Ethik in Unternehmen. Zuvor arbeitete sie für Google, Ebay und Oracle im Bereich Nutzerforschung.