

Kultur, Gesellschaft & Wissen

«Wir dürfen die Aufgeregtheit in den Medien nicht mit der Realität verwechseln»

Künstliche Intelligenz Laut einigen Stimmen werden sogenannte Deepfakes das Vertrauen in die Information zusammenbrechen lassen. Der Ethiker und Technologieexperte Markus Christen sagt, ob die Angst berechtigt ist.

Sandro Benini

Wie gefährlich sind Deepfakes – also von künstlicher Intelligenz fabrizierte Fotos, Videos oder Audios, die von authentischen Aufzeichnungen nicht zu unterscheiden sind? Überraschenderweise glaubt der Ethiker Markus Christen, dass diese neue Technologie auch Positives bewirken könne. Christen ist Geschäftsführer des Kompetenzzentrums Digital Society Initiative, das die digitale Transformation von Gesellschaft und Wissenschaft mitgestalten will. Der Professor für Ethik und Technologie an der Universität Zürich hat neben Philosophie auch Mathematik, Physik und Biologie studiert.

Herr Christen, droht uns die Infokalypse?
Was soll das sein?

Damit bezeichnen einige Expertinnen und Experten den völligen Verlust der Beweiskraft von Bildern, verursacht durch Deepfakes. Mit allen verheerenden Folgen.
Nein, ich glaube nicht, dass uns wegen Deepfakes apokalyptische Zustände drohen. Wir beobachten gerade zwei gegensätzliche Entwicklungen. Dass wir unsere Handys immer dabei haben, erlaubt es uns, in einem Ausmass Vorgänge zu dokumentieren und Beweise zu sammeln, wie es früher undenkbar war. Wir haben dies kürzlich in Minnesota gesehen, als die Immigrationspolizei ICE einen Demonstranten erschoss. Sofort gab es zahlreiche Videoaufnahmen aus allen möglichen Winkeln und Perspektiven. Sie haben es ermöglicht, die Behauptungen der US-Regierung als falsch zu entlarven.

Was schliessen Sie daraus?
Es gibt eine zunehmende Menge an Beweisen, die sich mit enormer Geschwindigkeit weltweit verbreiten. Gleichzeitig lassen sich Beweise fälschen, dank digitalisierter Techniken, die immer besser werden. Wir haben es also mit einem ambivalenten Phänomen zu tun.

Das Problem ist doch, dass Deepfakes auch echte Beweise entwerfen. Jede Politikerin, jeder Politiker, von dem ein kompromittierendes Video auftaucht, kann behaupten: Das ist ein Deepfake.
Ich bestreite nicht, dass es solche Probleme gibt. Besonders, wenn sich eine Gesellschaft in einer akuten Krise befindet, können selbst wenige Deepfakes eine unheilvolle Wirkung haben.

Aber?
Ein einzelnes Deepfake wird nicht losgelöst von jeglichem Kontext wahrgenommen. In der realen Welt stellen sich unweigerlich Fragen: Wie plausibel ist der angebliche Vorgang, der gezeigt wird? Wie verhält sich eine Behauptung über eine Person des öffentlichen Lebens zu dem, was man sonst über sie weiss? Lassen sich Vorwürfe durch Ermittlungen bestätigen? Aus wel-



«Ich glaube nicht, dass uns wegen Deepfake apokalyptische Zustände drohen»: Der Digitalexperte Markus Christen sieht der Zukunft mit KI gelassen entgegen. Foto: Boris Müller

cher Quelle stammen ein Bild, ein Video, eine Tonaufnahme? Die Leute sind nicht so dumm oder naiv, einfach jeder Person oder jeder Quelle zu vertrauen. Hinzu kommt die Möglichkeit, mithilfe digitaler Forensik zu überprüfen, ob es sich um eine echte Aufzeichnung handelt. Wichtig ist ein weiterer Punkt.

Welcher?
Wir müssen aufpassen, dass wir die Aufgeregtheit in den sozialen Medien nicht mit der Empörung in der realen Gesellschaft verwechseln. Insofern würde ich sogar vermuten, dass Deepfakes mittel- und längerfristig auch eine positive Wirkung haben könnten. Es zeichnet sich bereits ab, dass Menschen weniger bereit sind, private Informationen und Bilder ins Internet zu stellen. Die digitale Sichtbarkeit

nimmt ab. Deepfakes kommen uns aus der digitalen Welt entgegen, meistens auf einem kleinen Handybildschirm. Erleidet diese Welt aufgrund von Deepfakes einen Vertrauensverlust, wird dies unweigerlich zu einer Gegenreaktion führen.

Nämlich?
Je weniger die digitale Sphäre Echtheit und Authentizität garantiert, desto stärker wird man diese Qualitäten in der realen Welt suchen. Man legt Wert darauf, sich wieder persönlich gegenüberzusitzen. Eine Politikerin oder ein Politiker muss wieder vermehrt vor Ort auftreten, um als unverfälscht wahrgenommen zu werden. Seriöse Medien, bei denen man davon ausgeht, dass sie ihre Quellen sorgfältig prüfen, erhalten einen Vertrauensbonus. Es wird auch im di-

gitalen Raum sogenannte Safe Spaces geben, die eine hohe Verlässlichkeit gewährleisten.

Was bedeutet das für die Problematik der Deepfakes?
Ich glaube, dass es auf Deepfakes eine gesellschaftliche Immunantwort gibt. Erste Anzeichen dafür sind bereits erkennbar. Im Übrigen haben neue Medien schon in der Vergangenheit Ängste geschürt: Das Radio würde uns ständig ablenken, das Fernsehen würde uns verblöden lassen. Ich behaupte nicht, solche Ängste seien immer unbegründet gewesen, aber aus historischer Perspektive waren sie oft übertrieben. Die Menschen lernen, mit neuen Medien umzugehen. Und wenn das Vertrauen in die digitale Welt etwas sinkt und dafür jenes in die reale wieder steigt, ist das aus mei-

ner Sicht eher ein Grund, optimistisch zu sein.

In Deutschland herrscht eine Riesenaufregung, weil das ZDF ein Deepfake-Video über einen Einsatz der amerikanischen Migrationspolizei ICE gesendet hat. Sind Sie sicher, dass Sie die Gefahren von Deepfakes nicht verharmlosen?
Der Lernprozess, der stattfinden muss, um solche Vorfälle zu verhindern, betrifft letztlich den Umgang mit Zeit und Geschwindigkeit. Im Moment glauben Medien noch, sie würden sich vorteilhaft von ihren Konkurrenten abheben, wenn sie möglichst schnell, bestenfalls sogar als Erste eine Information publizieren. Deepfakes werden Medien dazu zwingen, ihr Tempo zu reduzieren, weil sie mehr Zeit brauchen, um die Echtheit eines Bildes oder eines Videos zu überprüfen. Das gefakte ICE-Video ist dafür ein

«Die Leute sind nicht so dumm oder naiv, jeder Quelle zu vertrauen.»

gutes Beispiel – offenbar war es ja sogar als KI-generiert markiert gewesen, was aber in der Vorbereitung und Übermittlung des Beitrags nicht entsprechend berücksichtigt wurde.

Warum soll Entschleunigung vorteilhaft sein?
Weil es dabei oft um Vorfälle geht, die uns nicht unmittelbar betreffen. Erwächst mir als Medienkonsument irgendein Schaden, wenn ich die Aufnahmen eines Polizeieinsatzes am anderen Ende der Welt morgen statt heute sehe? Natürlich nicht. Häufen sich Vorfälle wie jener beim ZDF, wird sich eine Entschleunigung auch ökonomisch auszahlen. Will ein Unternehmen in einem Medium werben, das auf Deepfakes hereinfällt? Auch hier: natürlich nicht. Falls die Trugbilder eine Entschleunigung der Medien erzwingen, wäre auch dies eine begrüssenswerte Entwicklung. Gefährlicher ist es, wenn Akteure systematisch versuchen, eine Gesellschaft nicht mit einigen wenigen Deepfakes zu destabilisieren, sondern durch eine strategisch aufgelegte Kampagne, die monate- oder sogar jahrelang dauern kann. In solchen Fällen geht es um etwas Grösseres.

Wie soll eine offene Gesellschaft darauf reagieren?
Sie braucht Ressourcen, also Behörden oder private Institutionen, die eine solche Kampagne erstens erkennen und zweitens darüber aufklären.

Bloss glaubt dann inmitten des allgemeinen Glaubwürdigkeitsverlustes auch den Erkenntnissen der Aufklärer niemand mehr.
Die Befürchtung, das Vertrauen könnte total zusammenbrechen, widerspricht der Lebensrealität. Ein gewisses Mass an Vertrauen ist für das Individuum und die Gesellschaft unerlässlich. Wir beide führen beispielsweise gerade ein Gespräch, und obwohl wir uns zuvor noch nie gesehen haben, vertrauen wir darauf, dass nicht einer von uns plötzlich versucht, dem anderen den Schädel einzuschlagen. Auf gesellschaftspolitischer Ebene brauchen wir Institutionen, die unser Bedürfnis nach Grundvertrauen erfüllen und bestärken. Solange solche Institutionen bestehen, seien es seriöse Medien oder staatliche Stellen, ist es möglich, auch gegenüber professionell aufgezogenen Deepfake-Kampagnen eine Immunreaktion zu erzeugen. Ohne Vertrauen sind wir letztlich nicht lebensfähig. Vertrauenswürdiges von Dubiosum zu unterscheiden, ist eine menschliche Grundkompetenz, die auch Deepfakes nicht zerstören werden.

Haben Deepfakes auch erfreuliche Auswirkungen?
Die Erschaffung von Bildern und Videos durch künstliche Intelligenz kann eine Kunstform sein, die jede und jeder mit einfachen Mitteln ausüben kann. Dann gibt es Deepfakes als Satire, als Entlarvung von Missständen oder als Kritik an Autoritäten, ohne den Anspruch, die Realität eins zu eins abzubilden.

Deepfakes werden auch für kriminelle Machenschaften eingesetzt: Erpressungen mit Nacktbildern, Anrufe mit täuschend echt klingenden Stimmen und anderes. Wird auch diese Gefahr überschätzt?
Kriminelle sind sehr innovativ, wenn es darum geht, neue Technologien zu nutzen. Es wird uns auch in diesem Bereich nichts anderes übrig bleiben, als aus Erfahrung klüger zu werden und über Gefahren so gut wie möglich aufzuklären. Was aber unbestreitbar ist: Öffentlich verbreitete Nacktbilder verletzen die Schamgefühle der Opfer auch dann, wenn sie gefakt sind.

Eben.
In diesem Bereich gibt es keine einfachen Lösungen. Und die psychische Belastung der Betroffenen kann immens sein. Gegenwärtig befindet sich das sogenannte Plattformgesetz in der Vernehmlassung, das neben den eigentlichen Urhebern pornografischer Fakes auch die Verbreiter, also vor allem die grossen Techunternehmen, stärker in die Pflicht nehmen will. Auch der Jugendschutz müsste verbessert werden. In vielen Bereichen gibt es aber bereits Gesetze. Kinderpornografie zum Beispiel ist auch dann illegal, wenn sie vollständig von KI erzeugt wird. Das Problem liegt darin, herauszufinden, wer dahintersteckt.