MEDIEN
KLICKS, LÜGEN UND VIDEO

Womöglich vertrauen wir bald keiner Nachricht mehr. Denn künstliche Intelligenz ermöglicht es inzwischen jedem, täuschend echte Film- und Tonsequenzen zu erzeugen.

Man verwandelt ein Gesicht in ein anderes, indem man beide mit einem gedachten Gitternetz überzieht und korrespondierende Punkte aufeinander abbildet. Mit diesem »motion capturing« kann man insbesondere die Mundbewegungen einer Person einer anderen aufprägen.

SPENCER_WHALEN / GETTY IMAGES / ISTOCK

Brooke Borel ist Journalistin und Autorin des Buchs »The Chicago Guide to Fact-Checking«. Sie ist kürzlich im »Fact-checking« gegen eine künstliche Intelligenz angetreten und hat mit einem Besorgnis erregenden Vorsprung gewonnen.

spektrum.de/artikel/1603758

Im April 2018 erschien ein neues Video von Barack Obama im Internet. Er sieht aus wie bei jeder seiner Reden: blütenweißes Hemd, dunkler Anzug mit einer Flaggennadel am Revers, im Hintergrund die amerikanische und die Präsidentschaftsflagge. Obama blickt in die Kamera und spricht mit ausdrucksvollen Handbewegungen die Worte: »Präsident Trump ist ein totaler Volltrottel.«

Ohne auch nur zu zwinkern, fährt er fort: »Nun, sehen Sie, ich würde so etwas nie sagen. Jedenfalls nicht in einer öffentlichen Rede. Aber jemand anderes würde es tun« – woraufhin der Schauspieler Jordan Peele in der rechten Hälfte des Bildes auftaucht.

Obama hatte nichts dergleichen gesagt; eine Aufzeichnung einer echten Rede war so manipuliert worden, dass seine Mund- und Handbewegungen denen Peeles folgen. Seite an Seite sieht man die beiden sprechen, während Peele als gleichsam digitaler Bauchredner Obama die Worte in den Mund legt (Bild S. 74 unten).

Produziert hat dieses – inzwischen millionenfach angeklickte – Video das Medienunternehmen Buzzfeed News auf eigene Kosten als Demonstration einer neuen Technik aus der künstlichen Intelligenz (KI). Mittlerweile kann die Software für Ton und bewegte Bilder nämlich dasselbe tun wie Photoshop für gewöhnliche Bilder: die Realität täuschend echt verfälschen.

Noch sind die Ergebnisse verbesserungsfähig. Obamas Stimme klingt ein bisschen näselnd, und wer genau hinschaut, sieht seinen Mund für kurze Momente zur Seite verrutschen. Aber das wird sich geben. Die ursprünglich für Kinofilme und Videospiele entwickelte Technik macht große Fortschritte und lässt inzwischen bei Sicherheitsexperten und Medienwissenschaftlern die schwärzesten Fantasien aufkommen. Aller Wahrscheinlichkeit nach wird die nächste Generation dieser Softwarewerkzeuge sich nicht mehr auf die Manipulation vorhandenen Materials beschränken, wie bei der beschriebenen Obama-Ansprache, sondern von Grund auf neue Szenen erschaffen können, die in Wirklichkeit nie stattgefunden haben – nicht einmal annähernd.

Die Folgen für den gesellschaftlichen Diskurs wären verheerend. Man stelle sich vor, dass in einem Kopf-an-Kopf-Rennen zweier Kandidaten kurz vor dem Wahltag ein Video auftaucht, das den Ruf eines von ihnen beschädigt. Oder ein Film, der kurz vor dem Börsengang einer großen Firma deren Vorstandschef in ein schlechtes Licht rückt. Eine Gruppe könnte einen islamistischen Terroranschlag inszenieren, mit gefaktem Bildmaterial Nachrichtenagenturen und Blogger hinters Licht führen und dadurch Vergeltungsaktionen provozieren.

Auch wenn sich ein solches Video später als Fälschung erweist, wird die Öffentlichkeit am Ende trotzdem die Geschichte dahinter für wahr halten? Und vielleicht am beunruhigendsten: Werden wir unter der Masse der Fälschungen gar nicht mehr glauben, was wir sehen und hören – auch die Wahrheit nicht?

Viele Fachleute erkennen an, dass ihre technischen Entwicklungen Gelegenheit zu weit reichendem Missbrauch geben. Aber während sie sich »für Verfahren zur Erkennung und Offenlegung von Fälschungen begeistern, denken sie kaum darüber nach, inwieweit ihre Bemühungen die Überzeugungen der Konsumenten beeinflussen können«, sagt Nathaniel (»Nate«) Persily, Professor für Rechtswissenschaft an der Stanford University. Persily untersucht unter anderem, wie das Internet die Demokratie beeinflusst, und vertritt gemeinsam mit einer wachsenden Gruppe von Forschern die Überzeugung, dass die Eindämmung sich rasch verbreitender Fehlinformationen weit mehr erfordert als technische Lösungen. Vielmehr müsse man Psychologen, Sozialwissenschaftler und Medienexperten zu Rate ziehen, um die Wirkungen der Technik auf die Menschen zu erforschen.

AUF EINEN BLICK
DIE STUNDE DER WAHRHEIT

1 Mit Hilfe der künstlichen Intelligenz lassen sich sehr einfach täuschend echte Videosequenzen erstellen. Das bedroht die Glaubwürdigkeit der Medien und das Vertrauen in demokratische Institutionen allgemein.

2 Gefälschte Videos sind besonders geeignet, Angst zu erregen – die wiederum die Nutzer veranlasst, die Falschmeldung umso bereitwilliger weiterzuverbreiten.

3 Algorithmen zum automatischen Erkennen gefälschter Videos sind in der Entwicklung; aber selbst wenn sie funktionieren, wird die Entlarvung von Fake News ihrer Verbreitung stets hinterherhinken.

»Und das müssen wir jetzt tun«, sagt Persily, »denn im Moment führen die Techniker – zwangsläufig – die Diskussion darüber an«, welche Arten von Videos die künstliche Intelligenz möglich macht. Derweil bröckelt das Vertrauen in demokratische Institutionen wie Regierung und Presse. Dass die Leute mittlerweile ihre Informationen vorrangig über die sozialen Medien beziehen, macht das Geschäft der Fälscher noch einfacher. Und da bislang niemand einer immer raffinierteren Technik eine durchdachte Strategie entgegenzusetzen hat, ist unser ohnehin brüchiges Vertrauen in die klassischen Institutionen erst recht in Gefahr.

Harmlose Anfänge

Die Kunst der digitalen Spezialeffekte entwickelte sich im Kino zunächst mit Sciencefiction-Animationen, die durch kein reales Vorbild gebremst waren. Einen ersten Höhepunkt bildete der 1993 erschienene Film »Jurassic Park«. Ein Jahr später konnte man man bereits altes Material so glaubwürdig manipulieren, dass Forrest Gump John F. Kennedy die Hand schüttelt (Bild S. 74 oben).

Der 2016 erschienene Film »Rogue One« spielt unmittelbar vor der ersten Geschichte der (ersten) »Star Wars«-Trilogie; also mussten die vertrauten Gesichter von Gouverneur Tarkin und Prinzessin Leia auf der Leinwand erscheinen. Da aber der Darsteller Peter Cushing 1994 gestorben war und Carrie Fisher nicht mehr jugendlich genug aussah, ließen sie für deren Rollen andere Schauspieler auftreten und ersetzten mittels »motion capturing« deren Gesichter Bild für Bild durch die altvertrauten.

Anfangs pflegten die Grafiker mit dem Computer dreidimensionale Modelle zu erzeugen und Oberflächenstruktur und andere Details mit der Hand nachzutragen – ein Zeit raubender Prozess, der keine Massenfertigung zulässt. Neue Ideen kamen vor etwa 20 Jahren aus einer anderen Richtung: »Computer vision« (Computersehen) heißt die Kunst, einem Rechner das Erkennen von Gegenständen in Kamerabildern beizubringen. Also muss das Gerät aus den Daten Modelle der Objekte in seiner Umgebung errech-

nen. Da bietet es sich an, diese Modelle gleich weiterzuverwenden, statt sie einzeln von Hand zu fertigen. 1997 entwickelten Wissenschaftler der kurzlebigen, aber äußerst einflussreichen Interval Research Corporation in Palo Alto (Kalifornien) das Programm Video Rewrite, das bestehende Aufnahmen zerlegt und neu zusammenfügt. Die Forscher gestalteten unter anderem einen Video-Clip, in dem John F. Kennedy sagt: »Ich habe Forrest Gump nie getroffen.« Mittlerweile arbeiten sie alle bei Google.

Bald darauf brachten Wissenschaftler aus der Arbeitsgruppe von Heinrich Bülthoff am Max-Planck-Institut für biologische Kybernetik in Tübingen einem Computer bei, Merkmale aus einem Datensatz von 200 dreidimensionalen Scans von menschlichen Gesichtern zu extrahieren und damit neue Gesichter zu gestalten.

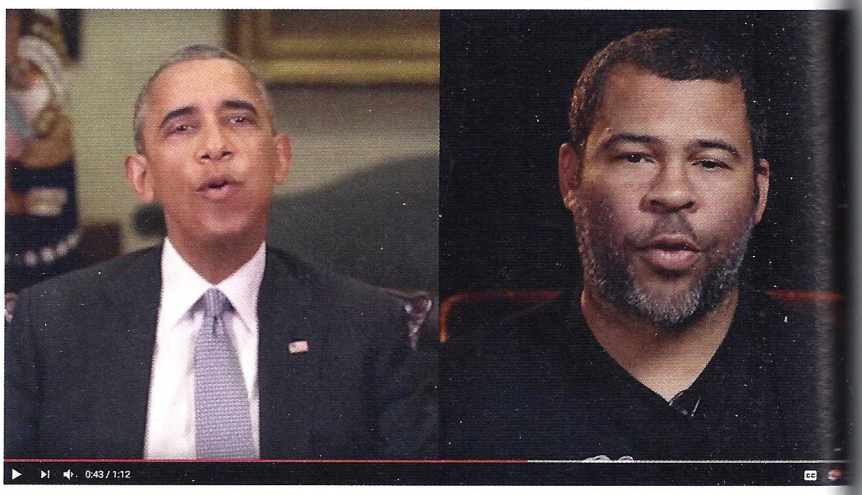
Einen gewaltigen Sprung nach vorn machte das Forschungsgebiet um 2012 mit dem Aufkommen einer KI-Technik namens »Deep Learning« (**Spektrum** Januar 2018, S. 12). Im Gegensatz zu den Anwendungen der späten 1990er Jahre können die »tiefen neuronalen Netze« tatsächlich aus neuen Daten lernen und dadurch immer besser werden. Außerdem erübrigt sich die Notwendigkeit, überhaupt Modelle zu erstellen. »Dies ist der Moment, in dem Ingenieure sagen: Wir modellieren keine Dinge mehr«, sagt Xiaochang Li, Postdoktorandin am Max-Planck-Institut für Wissenschaftsgeschichte in Berlin. »Wir modellieren unser Unwissen über die Dinge und lassen dann das System einfach laufen, um Muster zu finden.«

Man bringt einem tiefen neuronalen Netz bei, zum Beispiel menschliche Gesichter zu erkennen, indem man ihm Hunderte bis Tausende von Fotos vorlegt und jedem Bild die Information mitgibt, ob es sich um ein Gesicht handelt oder nicht. Bei diesem Lernprozess legt das Netz sich die Merkmale zu, die ein menschliches Gesicht ausmachen – allerdings ohne dass diese Merkmale irgendwie beschreibbar wären. Dennoch kann das Netz, sowie es ein neues Bild erblickt, errechnen, in welchem Ausmaß die Merkmale vorliegen, und daraufhin die Entscheidung treffen »Dies ist ein Gesicht« – oder eben nicht.

Der nächste Schritt ist der von der Analyse zur Synthese. Auch ein so genanntes generatives Netz wird mit Tausenden von Bildern trainiert; aber nun nutzt es die so erworbenen Merkmale, um neue, echt wirkende Gesichter zu erzeugen.

Einige Unternehmen verwenden inzwischen den gleichen Ansatz für Ton- an Stelle von Bildaufnahmen. Anfang 2018 stellte Google ein KI-Assistenzprogramm namens Duplex vor, das auf der Software WaveNet basiert. Duplex klingt am Telefon wie ein echter Mensch – inklusive Verlegenheitslauten wie »äh« und »hm«. In Zukunft wäre also Jordan Peele als Sprecher für ein gefälschtes Obama-Video entbehrlich. Im April 2017 veröffentlichte das kanadische Startup-Unternehmen Lyrebird künstlich erzeugte Audiodateien, die echten Aufnahmen von Obama, Trump und Hillary Clinton erschreckend ähnlich klingen.

OBEN: STANDBILD AUS FORREST GUMP, PARAMOUNT PICTURES, 1994 / SCIENTIFIC AMERICAN OKTOBER 2018;
UNTEN: STANDBILD AUS YOU WON'T BELIEVE WHAT OBAMA SAYS IN THIS VIDEO!, MONKEYPAW PRODUCTIONS AND BUZZFEED, 17.04.2018 (WWW.YOUTUBE.COM/WATCH?V=CO54GDM1EL0)



Der fiktive Forrest Gump ist dem echten John F. Kennedy nie begegnet; vielmehr wurde das Gesicht des Schauspielers in historische Filmaufnahmen einmontiert (oben). Heute erfordert es keine besondere Kunstfertigkeit mehr, Gesichtsbewegungen einer Person (Jordan Peele, unten rechts) auf eine andere (Barack Obama, links) zu übertragen.

Generative Netzwerke benötigen große Datenmengen für das Training und entsprechend viel menschliche Arbeit für die Bereitstellung des Trainingsmaterials. Also versucht man, auch diesen Teil der Arbeit an die Maschine zu delegieren: Das Netz trainiert sich selbst.

Im Jahr 2014 entwickelten Ian Goodfellow und seine Kollegen an der Université de Montréal (Kanada) ein Paar aus gegnerischen generativen Netzen (generative adversarial networks, GAN). Das eine Netz, der »Generator«, erzeugt gefälschte Bilder, und das zweite, der »Diskriminator«, lernt, zwischen echt und falsch zu unterscheiden. Mit einem Minimum an menschlichen Eingriffen trainieren sich die Netzwerke gegenseitig, indem sie gegeneinander ausgespielt werden: Der Generator bekommt als Lernziel, dass sein Produkt vom Diskriminator als echt deklariert wird, und der Diskriminator, dass er echte und gefälschte Bilder als

solche erkennt. Jedes Mal, wenn der Diskriminator ein Erzeugnis des Generators als Fälschung enttarnt, verändert der Generator sein Bild, so dass es überzeugender wirkt, und jedes Mal, wenn es dem Generator gelingt, den Diskriminator zu täuschen, bessert dieser seine Kriterien nach.

GANs können alles Mögliche generieren. An der University of California in Berkeley bauten Jun-Yan Zhu und seine Kollegen ein Paar von Netzen, das Pferde auf Fotos in Zebras oder ein impressionistisches Gemälde von Monet in eine scharfe, fotorealistische Szene verwandeln kann.

Im Mai 2018 brachten Forscher des Max-Planck-Instituts für Informatik in Saarbrücken und ihre Kollegen das Programm »Deep Video« heraus, das ebenfalls ein GAN verwendet. Mit seiner Hilfe kann ein Schauspieler einer anderen Person in voraufgezeichnetem Filmmaterial seine Mund-, Augen- und Gesichtsbewegungen aufprägen, so wie Jordan Peele das mit Obama gemacht hat. Deep Video funktioniert bislang nur, wenn das »Opfer« frontal in die Kamera schaut. Bewegt sich der Schauspieler zu sehr, weist das resultierende Video auffällige Artefakte auf; zum Beispiel verschwimmt die Umgebung des Gesichts.

Noch können GANs eine detailreiche Szene nicht aus dem Hut zaubern, ohne dass der Unterschied zu einer echten Aufnahme auffällt. Manchmal produzieren sie Abstrusitäten wie ein Auge, das einem Menschen aus der Stirn wächst. Im Februar 2018 gelang es jedoch Forschern der Firma NVIDIA, mit einem GAN unglaublich hochaufgelöste Gesichter zu produzieren, indem sie das Training mit relativ kleinen Fotos begannen und dann die Auflösung Schritt für Schritt erhöhten. Und Hao Li, Assistenzprofessor für Informatik an der University of Southern California und Chef von Pinscreen, einer Startup-Firma für »angereicherte Wirklichkeit« (augmented reality), hat mit Hilfe von GANs realistische Bilder von notorisch schwierigen Körperpartien erschaffen: Haut, Zähne und Münder.

Keine dieser Technologien ist für Laien einfach zu bedienen. Aber das Experiment von BuzzFeed lässt ahnen, wohin die Entwicklung geht. Das eingangs genannte Video stammt von der kostenlosen Software FakeApp, die ihrerseits Deep Learning (ohne GANs) verwendet. Die so erstellten Videos heißen in der Szene Deepfakes nach einem Nutzer der Website Reddit, der sich das Pseudonym »Deepfakes« (zusammengezogen aus »Deep Learning« und »Fake«) zulegte und Aufsehen erregte, weil er nicht nur als einer der Ersten die Technik beherrschte, sondern auch mit ihrer Hilfe die Gesichter der Darsteller in Pornofilmen durch solche von Prominenten ersetzte.

Seitdem haben Amateure unzählige Videos mit FakeApp produziert. Die meisten von ihnen sind relativ harmlose Streiche. Man sieht den Schauspieler Nicolas Cage in zahlreichen Filmen, in denen er nicht mitgespielt hat, oder Trumps Kopf auf dem Körper von Angela Merkel. Bedrohlicher ist, was man mit der Technik anstellen könnte, nachdem im Prinzip jeder sie nutzen kann.

Experten befürchten seit Langem, dass die neuen Manipulationsmöglichkeiten unsere Wahrnehmung der Realität ruinieren könnten. Bereits im Jahr 2000 warnte ein Artikel im »MIT Technology Review«, dass durch Produkte wie Video Rewrite der Augenschein nicht mehr als Vergewisserung dienen könne (»seeing is no more believing«). Schließlich könne »ein Bild in den Abendnachrichten durchaus eine Fälschung sein – das Produkt einer schnellen neuen Video-Manipulations-Technologie«.

So schlimm ist es 18 Jahre später noch nicht. Allem Anschein nach sind Fake-Videos in der Tagesschau bislang kein Problem. Eine richtig gute Fälschung ist nämlich immer noch schwierig. Die Leute von BuzzFeed haben für ihr Obama-Video 56 Stunden gebraucht, mit professioneller Unterstützung.

Vor den Präsidentschaftswahlen 2016 hat jeder vierte Amerikaner eine Fake-Nachrichtenseite angeklickt

Allerdings hat sich unser Konsumverhalten, was Informationen angeht, erheblich verändert. Nach einer Untersuchung des Pew Research Center schauen heute nur noch etwa die Hälfte der amerikanischen Erwachsenen die Fernsehnachrichten, während zwei Drittel zumindest einen Teil ihrer Informationen aus den sozialen Medien beziehen. Im Internet haben sich zahlreiche verschworene Gemeinschaften (»Echokammern«) etabliert, die nur noch Meldungen aus der eigenen Gruppe zur Kenntnis nehmen, und Websites, die zur Verfolgung ihrer Ziele vor der Erregung von Wut und Angst nicht zurückschrecken – gegen alle journalistischen Standards (**Spektrum** November 2017, S. 58). Je abenteuerlicher die Meldung, desto viraler ist sie, das heißt, desto schneller verbreitet sie sich übers Netz, sagt Persily. Und die Ungenauigkeiten bei gefälschten Videos fallen auf dem kleinen Handydisplay weniger auf als auf dem häuslichen Fernseher.

Was passiert, wenn ein Deepfake-Video mit einer politisch oder sozial einflussreichen Botschaft viral wird? »Die kurze Antwort lautet: Wir wissen es nicht«, sagt Julie Carpenter, die sich in der Ethics + Emerging Sciences Group an der California State Polytechnic University in San Luis Obispo mit der Interaktion von Menschen und Robotern beschäftigt. Vielleicht wissen wir es schon, wenn dieses Heft erscheint; denn Anfang November fanden in den USA wichtige Halbzeitwahlen (midterm elections) statt.

Einmal haben wir das fatale Zusammenspiel von Desinformation und enger Vernetzung bereits erlebt. Im amerikanischen Präsidentschaftswahlkampf 2016 tauchten zahlreiche Falschmeldungen auf, die darauf angelegt waren, sich viral zu verbreiten. Und sie haben ihre Empfänger erreicht: Nach einer gemeinsamen Untersuchung der Princeton University, des Dartmouth College und der University of Exeter in England hat etwa jeder vierte Amerikaner zwischen dem 7. Oktober und dem 1. November 2016 eine Fake-Nachrichtenseite angeklickt; den meisten wurde sie von Facebook angeboten. Im selben Jahr fiel das öffentliche Vertrauen in den Journalismus auf einen Tiefpunkt. In einer Umfrage schenkten nur 51 Prozent unter den Anhängern der Demokraten und 14 Prozent von denen der Republikaner den Nachrichten der Massenmedien tatsächlich Glauben.

Es gibt nicht viele Untersuchungen über gefälschte Nachrichten. Aber einige deuten darauf hin, dass es schon reicht, eine falsche Informationen nur einmal gesehen zu

haben, um sie später für plausibel zu halten, sagt Gordon Pennycook, Dozent für Organsationsverhalten an der University of Regina in Saskatchewan (Kanada). Wenn wir hören, wie Obama Trump mit einem Schimpfwort belegt, und nach einer Woche auf eine weitere Fälschung treffen, in der Obama gegenüber Trump obszön wird, neigen wir dazu, das schon deshalb für echt zu halten, weil uns das Muster vertraut ist.

Laut einer Studie aus dem Massachusetts Institute of Technology (MIT), die zwischen 2006 und 2017 insgesamt 126 000 verschiedene Storys auf Twitter verfolgte, teilen wir falsche Nachrichten bereitwilliger als echte; und politische Falschmeldungen verbreiten sich weiter und rascher als die über Geld, Naturkatastrophen oder Terrorismus. Die Autoren der im März 2018 in »Science« erschienenen Arbeit vermuten, dass die Menschen nach Neuheiten gieren. Allgemein zielen gefälschte Nachrichten auf unsere Emotionen und veranlassen uns dadurch, sie weiterzuleiten, bevor wir die Information wirklich verarbeitet haben und entscheiden können, ob eine Verbreitung sich lohnt. Je mehr ein Inhalt uns überrascht, erschreckt oder empört, desto schneller und häufiger scheinen wir bereit, ihn zu teilen.

Allem Anschein nach ist die Darreichungsform Video besonders geeignet, Angst zu schüren. »Wenn Sie Informationen visuell verarbeiten, glauben Sie, dass die Sache Ihnen räumlich, zeitlich oder sozial nähersteht«, sagt Elinor Amit, Assistenzprofessorin für Kognitionswissenschaft, Linguistik und Psychologie an der Brown University. Ihre Arbeit beschäftigt sich mit den Unterschieden in unseren Reaktionen auf Texte und Bilder. Amit vermutet evolutionäre Ursachen: Unsere Sehfähigkeit ist älter als die Sprache, und in – echten oder vermuteten – Gefahrensituationen geben wir den unmittelbaren Sinneseindrücken den Vorrang.

Gefälschte Videos haben in der Tat bereits in politische Kampagnen eingegriffen. Im Juli 2018 veröffentlichte Allie Beth Stuckey, eine Moderatorin des Medienunternehmens Conservative Review, auf Facebook ein Interview mit Alexandria Ocasio-Cortez, einer demokratischen Kandidatin für den Kongress aus New York City. Das Video war kein Deepfake, sondern ganz traditionell zusammengeschnitten aus einem echten Interview und veränderten Fragen, die nicht zu den Antworten passten, so dass die Befragte wirkte, als sei sie nicht recht bei Verstand. Das mag man je nach der eigenen politischen Orientierung als Verleumdung auffassen oder als Satire, wie Stuckey später zu ihrer Verteidigung behauptete. Jedenfalls brachte es das Video auf 3,4 Millionen Aufrufe innerhalb einer Woche. Etliche der mehr als 5000 Kommentare ließen erkennen, dass die Zuschauer es für echt und die Kandidatin damit für dumm gehalten hatten.

Schlimmer als die Fake-Videos selbst sind deren Folgen für den gesellschaftlichen Diskurs. Jeder Politiker, der bei einer Missetat gefilmt wird, kann Zweifel säen mit der Behauptung, das Video sei gefälscht. Bereits das Wissen, dass glaubwürdige Fälschungen möglich sind, kann das Vertrauen in alle Medien untergraben, sagt Raymond J. Pingree, Assistenzprofessor für Massenkommunikation an

»Wir können das Spiel nicht gewinnen. Wir können es den Bösen nur immer schwerer machen«

der Louisiana State University. Pingree untersucht, inwieweit Menschen sich fähig fühlen, Echtes von Falschem zu unterscheiden, und wie das ihre Bereitschaft beeinflusst, am politischen Leben teilzunehmen. Wer diese Selbstsicherheit verliert, »fällt eher auf Lügner und Kriminelle herein«, sagt er, »und im Extremfall gibt er die Suche nach der Wahrheit auf«.

Katz-und-Maus-Spiel zwischen Fälschern und Entlarvern

Für einen Computerwissenschaftler besteht die Behebung eines Defekts häufig aus noch mehr Computerwissenschaft. Das aktuelle Problem ist zwar alles andere als ein Programmierfehler und vor allem viel komplizierter; dennoch hält sich in der Szene die Vorstellung, man könne Algorithmen entwerfen, die Fälschungen erkennen und mit Warnhinweisen versehen.

»Man kann sicherlich gewisse Techniken gegen das Problem anwenden«, sagt R. David Edelman von der Internet Policy Research Initiative des MIT. Edelman, früher Berater von Präsident Obama, ist beeindruckt von den gefälschten Videos über seinen ehemaligen Chef. »Ich kenne den Mann. Ich habe Reden für ihn geschrieben. Und ich könnte ein echtes Video nicht von einem falschen unterscheiden.« Aber im Gegensatz zu ihm würde ein Algorithmus digitale Merkmale finden, die mit bloßem Auge nicht zu sehen sind.

Bisherige Lösungen sind von zweierlei Art. Man könnte erstens jedem Video einen Echtheitsnachweis mitgeben, vergleichbar den Hologrammen, Wasserzeichen und anderen Mitteln, mit denen Banknoten gegen Fälschungen gesichert sind. Jede Digitalkamera hätte eine eindeutige Signatur, die – theoretisch – schwer zu kopieren wäre.

Zweitens könnte man in der Tat ein Programm schreiben, das ein gefälschtes Video automatisch als solches erkennt. Das bisher wohl bedeutendste Unternehmen in dieser Richtung heißt Media Forensics oder kurz MediFor. Die Forschungsabteilung des amerikanischen Verteidigungsministeriums (Defense Advanced Research Projects Agency, DARPA) hat es 2015 auf den Weg gebracht, nicht lange, nachdem ein russischer Nachrichtensender gefälschte Satellitenaufnahmen ausgestrahlt hatte, auf denen ein ukrainischer Kampfjet auf das – später abgestürzte – Verkehrsflugzeug der Malaysia Airlines (Flug MH 17) feuerte.

Eine Gruppe internationaler Ermittler hat inzwischen festgestellt, dass das Flugzeug von einer russischen Rakete abgeschossen wurde. Die falschen Satellitenbilder wurden nicht mit Hilfe von Deep Learning angefertigt, aber die DARPA sah die revolutionären Fortschritte kommen und wollte rechtzeitig dagegen gewappnet sein, sagt David Doermann, der frühere Chef von MediFor.

Das Programm arbeitet mit drei Verfahren, die sich mit Deep Learning automatisieren lassen. Das erste sucht auf der Ebene der einzelnen Bits nach Anomalien. Das zweite prüft nach, ob die gezeigte Szene den Gesetzen der Physik folgt, zum Beispiel ob die Gegenstände Schatten in die richtige Richtung werfen. Das dritte gleicht mit externen Daten ab: War das Wetter an dem Tag, an dem angeblich gefilmt wurde, wirklich so, wie es aussieht? Aus den Ergebnissen der drei Teilprüfungen soll eine Gesamtpunktzahl errechnet werden, die angibt, mit welcher Wahrscheinlichkeit das vorliegende Video eine Fälschung ist.

Ein weiteres interessantes Konzept stammt von der genannten Saarbrücker Arbeitsgruppe. Aus dem Paar gegnerischer generativer Netze ist der Diskriminator optimal darauf trainiert, das Werk des Generators als Fälschung zu erkennen. Also ist er auch zum Entlarven fremder Machwerke nutzbar.

Nur würde die andere Seite mit geeigneten Maßnahmen dagegenhalten. Die Fälscher würden sich bemühen, digitale Wasserzeichen nachzumachen und ihrerseits mit Hilfe von Deep Learning die Prüfalgorithmen hereinzulegen. Am Ende treiben sich spezielle gegnerische generative Netze, der Generator bei den Kriminellen und der Diskriminator bei den Aufpassern, jeweils zu Höchstleistungen. »Wir können dieses Spiel nicht gewinnen«, sagt Alexei Efros, Professor für Informatik und Elektrotechnik an der University of California in Berkeley und beteiligt an MediFor. »Wir können es den Bösen nur immer schwerer machen.«

Twitter und Facebook fördern durch ihre Konstruktion die Verbreitung von Fake News

Selbst wenn die automatische Entlarvung funktioniert, wird zwischen dem Auftauchen einer Lüge und ihrer Widerlegung immer eine gewisse Zeit vergehen. Deshalb kommt es entscheidend darauf an, dass die Social-Media-Branche der Verbreitung von überzeugenden gefälschten Videos Einhalt gebietet. »Man muss an der Verteilung ebenso ansetzen wie an der Erzeugung«, sagt Edelman. »Wenn ein Deepfake im Wald umfällt, hört das niemand, es sei denn, Twitter und Facebook verstärken es.«

Es ist derzeit weder klar, inwieweit die Social-Media-Unternehmen zu entsprechenden Schritten rechtlich verpflichtet sind, noch, wie eine solche Regulierung praktiziert werden könnte, ohne das verfassungsmäßige Recht auf Freiheit der Rede zu verletzen. Immerhin hat Facebook-Chef Mark Zuckerberg zugegeben, dass seine Plattform bei der Verbreitung von Fake News eine Rolle gespielt hat – wenn auch mit mehr als zehn Monaten Verzögerung nach den Wahlen von 2016. Facebook ist schließlich so konzipiert, dass Nutzer konsumieren und verbreiten, was sie interessiert, womit dem Populären systematisch der Vorrang vor der Wahrheit eingeräumt wird. Mit mehr als zwei Milliarden aktiven Benutzern ist die Plattform ein riesiges Pulverfass, an das jeder eine Lunte legen kann, der eine Wut erregende Lügengeschichte verbreiten will.

Seitdem hat Zuckerberg versprochen zu handeln. Als Erstes schiebt er Arbeit – und damit auch Verantwortung, sagen einige – an die Nutzer ab, indem er sie bittet, die Vertrauenswürdigkeit von Nachrichtenquellen zu bewerten. Außerdem plane er, Desinformationen mit Hilfe von KI zu kennzeichnen. Über Einzelheiten gibt Facebook keine Auskunft. Einige Computerwissenschaftler sind skeptisch, was den Einsatz von KI betrifft, darunter Hany Farid, Professor für Computerwissenschaft am Dartmouth College, der die Versprechungen für »spektakulär naiv« hält. Bisher konnten nur wenige unabhängige Wissenschaftler untersuchen, wie sich gefälschte Nachrichten auf Facebook verbreiten, weil die Firma viele der relevanten Daten unter Verschluss hielt.

Mehr Wissen auf **Spektrum.de**

Unser Online-Dossier zum Thema finden Sie unter **spektrum.de/t/kuenstliche-intelligenz**

ISTOCK / ADVENTTR

Alle Entlarvungsalgorithmen der Welt werden jedoch nicht helfen, wenn die Entwickler der Technologie, die das Fälschen erst ermöglicht, sich nicht damit auseinandersetzen, wie ihre Produkte verwendet oder eben missbraucht werden. »Ich appelliere an die Vertreter der harten Wissenschaft«, sagt Persily, »sich mit den Psychologen, den Politikwissenschaftlern und den Kommunikationsspezialisten zusammenzutun, die seit einiger Zeit an diesen Fragen arbeiten«. Bisher geschah das eher selten.

Im März kündigte das Finnish Center for Artificial Intelligence eine entsprechende Initiative an. Psychologen, Philosophen, Ethiker und andere sollen KI-Forschern helfen, die weit reichenden sozialen Folgen ihrer Arbeit zu erfassen. Und im April hat Persily zusammen mit Gary King, einem Politikwissenschaftler an der Harvard University, die Social Data Initiative ins Leben gerufen. Erstmals sollen Sozialwissenschaftler auf Facebook-Daten zugreifen können, um detailliert zu untersuchen, wie Desinformation sich ausbreitet.

Da an der Spitze ein Verantwortungsvakuum herrscht, liegt es an Journalisten und engagierten Bürgern, die Fake-Videos bloßzustellen. Gegen Ende des Deepfake-Clips von Obama und Peele sagen beide Männer: »In Zukunft müssen wir wachsamer sein in Bezug auf das, was wir aus dem Internet glauben können. Wir leben in einer Zeit, in der wir vertrauenswürdige Nachrichtenquellen benötigen.« Vielleicht hat Obama diesen Satz in Wirklichkeit nie gesagt; wahr ist er trotzdem. ◂

QUELLEN

Lazar, D. M. J. et al.: The Science of Fake News. In: Science 359, S. 1094–1096, 9. März 2018

Marwick, A. E.: Why Do People Share Fake News? A Sociotechnical Model of Media Effects. In: Georgetown Law Technology Review 2, S. 474–512, 2018