Tiefe Fälschung

Die Methoden zur Manipulation von Videoaufnahmen werden immer besser und haben das Potenzial, die öffentliche Meinung zu manipulieren. ExtraDienst untersuchte, was die Technologie kann und was sie bewirken könnte.

Von Martin Krake

Fake News sind eine wachsende Gefahr – vor allem durch die immer überzeugender werdenden Video-Fakes

Als ein österreichischer Politiker auf einer spanischen Ferieninsel vor versteckter Kamera ins Plaudern geriet, löste das die größte Staatskrise der Zweiten Republik und einen Regierungswechsel aus. Das Ibiza-Video bewies, welche politische Sprengkraft Videoaufnahmen haben können. Was wäre aber, wenn man solche Aufnahmen fälschen könnte? Oder wenn einfach irgendwann nicht mehr klar ist, was echt ist und was nicht? Sogenannte Deepfakes haben ein gefährliches Potenzial, weil sie die Möglichkeit bieten, die öffentliche Meinung gezielt zu manipulieren.

Bei einem Deepfake wird das Gesicht einer Person in einer Videoaufnahme manipuliert oder ausgetauscht. Im günstigsten Fall ist die Fälschung kaum oder gar nicht mehr zu erkennen. So wäre es beispielsweise möglich, eine kompromittierende Szene mit Statisten zu filmen und diese mit den Gesichtern bekannter Politiker zu versehen. So ließe sich eine Szene wie die auf Ibiza fälschen. Ebenso kann man Personen Worte in den Mund legen, die sie nie gesagt haben.

„Deepfake“ ist ein Begriff für eine Technologie, die mit Hilfe Künstlicher Intelligenz das Gesicht einer Person, die in einer Videoaufnahme zu sehen ist, durch das Gesicht einer anderen Person ersetzt. Das wird auch als „Face Swap“, „Gesichtertausch“, bezeichnet. Geprägt wurde der Begriff 2017 durch einen anonymen Nutzer des Netzwerks „Reddit“, der das Pseudonym „Deepfake“ verwendete, ein Kofferwort aus den Begriffen „Deep Learning“ und „Fake“. Er veröffentlichte eine entsprechende von ihm programmierte Software sowie einige damit angestellte Experimente.

Pornos als Technologietreiber

Pikanterweise wurde diese Technologie zuerst für Pornos verwendet und vor allem dadurch einer breiten Öffentlichkeit bekannt: Mit irgendwelchen beliebigen Darstellerinnen gedrehte heiße Szenen wurden mit den Gesichtern von Hollywood-Stars wie Emma Watson oder Scarlett Johansson versehen, um die Illusion zu erzeugen, man würde diesen beim ungezügelten Sex zusehen. Obwohl dabei im Normalfall klar ist, dass es sich um einen Fake handelt, scheint das für manche Betrachter von so besonderem Interesse zu sein, dass es inzwischen

Adobe Stock

Videoaufnahmen hatten bisher eine hohe Beweiskraft. Durch Deepfakes wird sich das ändern

massenhaft Fake-Pornos gibt. Und sie wirken tatsächlich ziemlich echt, man hat nicht das Gefühl, ein manipuliertes Video anzusehen. Viele Experten sind übrigens der Ansicht, dass solche Spielereien maßgebliche Motivation für einige Software-Entwickler waren, Deepfake-Apps zu programmieren oder weiterzuentwickeln. Nun sind Deepfake-Pornos für die betroffenen Celebrities ein Ärgernis, ansonsten aber eher harmlos. Anders sieht es aus, wenn Apps wie DeepFaceLab für sogenannte Rachepornos benutzt werden: Ein enttäuschter Liebhaber kann irgendeine Pornoszene mit dem Gesicht seiner Ex versehen und verbreiten. Oder die Technologie nutzen, um eine Person zu diskreditieren („ihre schmutzige Vergangenheit"). Auch wenn das Opfer der Attacke glaubhaft dementiert, kann so ein Video erheblichen Schaden anrichten.

Hohe Rechenleistung nötig

Deepfakes funktionieren auf der Basis Künstlicher Intelligenz. Voraussetzung dafür sind – neben einer entsprechenden Software – längere Videoaufnahmen beider Personen, deren Gesichter getauscht werden sollen. Mit diesen Aufnahmen als Trainingsmaterial lernt ein sogenannter Deep-Learning-Algorithmus, welche Bildteile er austauschen muss (mehr über das Funktionsprinzip im Kastentext).

Im Dezember 2019 veröffentlichte der Technikjournalist Timothy B. Lee aus Washington auf dem Portal Ars Technica (www.arstechnica.com) einen Beitrag, in dem er davon berichtet, wie er selbst einen Deepfake ausführte. Seine Reportage unter dem Titel „I created my own deepfake – it took two weeks and cost $552" ist ausgesprochen detailliert und techniklastig. Sein Experimentier-Objekt war ein Video von einer Aussage Mark Zuckerbergs vor dem US-Kongress, bei dem er das Gesicht des Facebook-Gründers mit dem des Lieutenant Commander Data aus „Star Trek: The Next Generation" ersetzte. Deepfakes brauchen viel Rechenpower. Der Prozess ist so aufwendig, dass Lee für 552 Dollar einen „Monstercomputer" mieten musste, um die Aufgabe zu stemmen. Selbst dieser benötigte noch eine Woche dafür, außerdem war tagelange händische Nachbearbeitung nötig, um den Prozess zu korrigieren.

Ein guter Deepfake erfordert also einen gewissen Aufwand, ist aber für eine einzelne, technisch einigermaßen versierte Person sehr wohl machbar. Wie Lee zugibt, ist das Ergebnis weit davon entfernt, perfekt zu sein. So sind verräterische Abweichungen am Rand des Gesichts zu sehen, und

WIE EIN DEEPFAKE ENTSTEHT

Um einen Deepfake zu erstellen, braucht man mehrere möglichst unterschiedliche Videosequenzen von den beiden Personen, deren Gesichter man tauschen möchte. Mit diesem Content wird die Software in einem aufwendigen und langwierigen Rechenprozess darauf trainiert, die Unterschiede der Gesichter zu erkennen, um sie schließlich austauschen zu können. Die Software lernt dabei, unveränderliche Merkmale wie Augenfarbe oder Gesichtsform von veränderlichen Bildinformationen wie Mimik und Ausdruck oder auch Blickwinkel und Beleuchtung zu unterscheiden. Diese Unterscheidung ist der Schlüssel dazu, die Gesichter später austauschen zu können.

Beim finalen „Ersetzungsprozess" werden die veränderlichen Informationen des Ursprungsvideos (Bewegung, Mimik, Beleuchtung) beibehalten und die unveränderlichen Merkmale (wie die Gesichtsform) ausgetauscht. Das Video zeigt dann die Mimik und die Bewegungen der Ursprungsversion mit dem Gesicht der anderen Person, das wie eine digitale Maske darübergelegt wurde. Das so trainierte „Modell" kann dann nahezu jedes Video der beiden betreffenden Personen bearbeiten und ihre Gesichter selbstständig austauschen.

Adobe Stock

Jordan Peele (rechts) manipulierte als Demonstration Barack Obama

das „neue“ Gesicht ist unscharf und zeigt verschwommene Details. Aber auf den ersten Blick (und bei geringer Auflösung) wirkt es glaubwürdig. „Es ist bemerkenswert, dass ein Anfänger wie ich ein recht überzeugendes Video so schnell und mit so wenig Geld erstellen kann. Und wir haben jeden Grund zu der Annahme, dass Deepfake-Technologie in den nächsten Jahren besser, schneller und billiger werden wird“, schreibt Lee.

Noch gibt es Grenzen

Bisher hat die Technologie vor allem Probleme damit, die Details eines Gesichts darzustellen. Dieses Manko lässt sich aber dadurch kaschieren, dass man das Gesicht aus größerer Entfernung zeigt oder das Video in geringer Auflösung veröffentlicht. Außerdem lassen sich diese Defizite mit mehr Rechenpower lösen; Lee merkt in seinem Bericht an, dass er mit mehr Rechenzeit in jedem Fall ein besseres Resultat erzielt hätte. „Es gibt keinen Grund zu der Annahme, dass dies eine fundamentale Grenze der Deepfake-Technologie sei. Forscher werden in den nächsten Jahren sicherlich Technologien entwickeln, die diese Begrenzungen überschreiten“, schreibt er.

Doch die Technik hat noch andere Grenzen: Bisher kann man mit Deepfake-Technologie nur das Gesicht einer Person austauschen, nicht den übrigen Körper. Selbst Stirn, Haare und Ohren werden nicht ersetzt, nicht zu reden von den übrigen Körperteilen oder gar der Körpersprache. Mit gutem Grund haben die Datenforensiker, die das Ibiza-Video auf seine Echtheit prüften, sehr genau auf die Ohren von Herrn Strache geschaut. Und dass bisher keine Deepfake-Experimente mit Brillenträgern veröffentlicht wurden, wird wohl auch seinen Grund haben. Für einen wirklich glaubwürdigen Fake reicht es also nicht, irgendeinen Komparsen zu filmen – man bräuchte einen professionellen Schauspieler, der der Zielperson in allen körperlichen Merkmalen täuschend ähnlich ist und es schafft, ihre Körpersprache perfekt zu imitieren.

Ferngesteuerte Gesichter

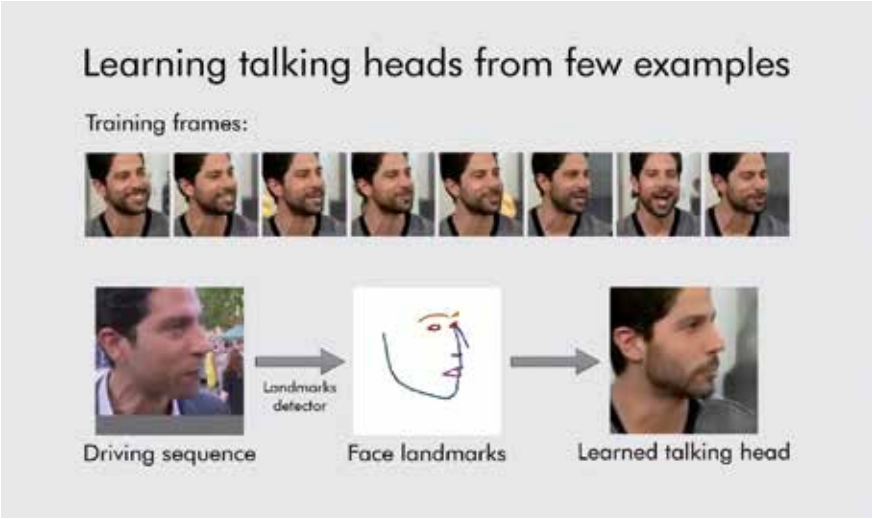
Eine andere Fake-Variante umgeht dieses Problem, indem sie nicht das Gesicht einer Person austauscht, sondern nur ihre Mimik und die Bewegungen des Kopfes manipuliert. Man sieht eine reale Person in einer realen Situation – aber sie bewegt ihr Gesicht ganz anders, als sie es tatsächlich getan hat. Die Software unterscheidet mittels des sogenannten Face Trackings die unbeweglichen Teile eines Gesichts von den beweglichen und ersetzt sie mit der Mimik eines beliebigen Statisten. Auf diese Art kann man ihr mit den passenden Lippenbewegungen Worte in den Mund legen, die sie tatsächlich nie gesagt hat. In Echtzeit – und ohne einen Superrechner zu benötigen.

Dazu genügt als „Trainingsvorlage“ schon eine rund 20 Sekunden lange Aufnahme, etwa ein YouTube-Clip. Gerade die Ansprache eines Politikers ist leicht zu faken, weil die Voraussetzungen (vor allem die relativ starre Haltung) es der Software leicht machen. Außerdem findet man von Politikern in jedem Fall ausreichend frei verfügbares Videomaterial, um das Netzwerk zu trainieren.

Das „Visual Computing Lab“ der TU München experimentiert seit einigen Jahren mit solchen Technologien. In einem Demovideo führt Justus Thies, einer der Entwickler, auf YouTube die Möglichkeiten der Software vor (zu finden unter dem Titel „Face2Face – Real time Face Capture and Reenactment of RGB Videos“). Man sieht Thies, wie er einen Wladimir Putin auf einem Bildschirm in Echtzeit durch seine eigene Mimik buchstäblich fernsteuert. Das wirkt einerseits lustig – und andererseits ziemlich gruselig.

Viel Aufmerksamkeit erfuhr diese Technologie durch einen Clip des amerikanischen Schauspielers und Regisseurs Jordan Peele. Er hat zu Demonstrationszwecken eine fiktive Ansprache von Ex-Präsident Barack Obama gefaket (auf YouTube zu finden unter dem Titel „You Won‘t Believe What Obama Says In This Video!“), in der dieser Ungeheuerlichkeiten sagt wie: „Donald Trump ist ein kompletter Vollidiot.“ Wenn man Obama nicht gut kennt und nicht genau hinsieht, erscheint das Video absolut glaubwürdig. Peele will damit zeigen, was mit Deepfakes möglich ist. Und eine War-

Screenshot



Für Samsungs „Talking Head Models" reichen schon Fotos

nung aussprechen: „It's a time when we need to rely on trusted new sources", mahnt der Fake-Obama.

Automatischer Stimmenimitator

Doch Deepfake-Software kann nur das Bild fälschen, nicht aber die Stimme einer Person. Bisher brauchte man für einen glaubwürdigen Fake einen talentierten Imitator – wie etwa Jordan Peele, der schon lange in der Lage ist, Stimme und Sprechduktus von Barack Obama präzise zu imitieren. Doch auch für dieses Problem könnte es bald eine Lösung geben: Adobe hat 2016 die Software „VoCo" vorgestellt. Sie ist in der Lage, eine Stimme auf ähnliche Art zu fälschen wie eine Deepfake-Software ein Video.

VoCo analysiert die individuellen Sprachmuster einer bestimmten Person und „bastelt" daraus einen neuen Text. Man benötigt dafür einen längeren von der „Zielperson" gesprochenen Text – so viel, dass jeder Laut der jeweiligen Sprache einige Male vorkommt. Bei Politikern ist es eine Leichtigkeit, an derartiges Trainingsmaterial heranzukommen. Dieser Text wird von der Software in kleine Lauteinheiten, sogenannte Phoneme, zerlegt. Ist die Software ausreichend trainiert, kann sie die Phoneme beliebig neu zusammensetzen und auf diese Art eine Kunststimme kreieren, die jeden beliebigen Text vorlesen kann – den man einfach nur eintippen muss. VoCo würde dadurch zu einer Art elektronischem Stimmenimitator, ein Talent wie Jordan Peele bräuchte es dann nicht mehr.

Bisher ist VoCo nur ein Prototyp und nicht öffentlich verfügbar. Adobe hat außerdem – im Bewusstsein des Missbrauchspotenzials – angekündigt, die Software werde in jede Datei eine akustische Signatur einfügen, die die Manipulation identifizierbar macht. Dieses digitale Wasserzeichen könnte allerdings seinerseits manipuliert werden.

Nur noch ein Foto nötig

Eine ganz andere Grenze haben Forscher eines russischen Entwicklungslabors von Samsung überschritten: Sie experimentieren mit einer Software, die für ihr Training nicht mehr eine Vielzahl von Videosequenzen braucht, sondern nur noch ein einziges Foto. Die Software funktioniert, indem sie durch die Analyse zahlloser Videosequenzen irgendwelcher Personen gelernt hat, wie sich menschliche Gesichter bewegen. Dadurch kann sie in einem Bild „Face Landmarks" wie die Umrahmungen von Augen und Lippen erkennen und das Gesicht aus der Abbildung wie eine digitale Maske über das Video einer anderen Person legen. „Realistic Talking Head Models" heißt das. Und es funktioniert im Prinzip sogar mit einem Gemälde: 2019 veröffentlichten die Forscher ein Demo, das die Mona Lisa sprechen lässt. Bisher sind damit allerdings nur sehr kurze Sequenzen von wenigen Sekunden möglich.

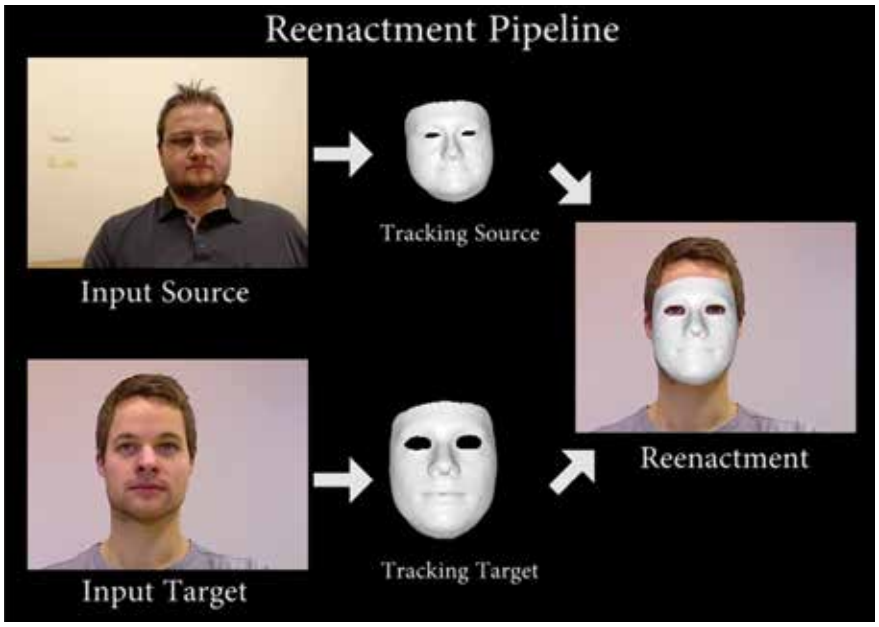
Was wie eine harmlose, witzige Spielerei wirkt, hat ein völlig neues Missbrauchspotenzial: Bisher braucht man für Deepfakes eine gewisse Anzahl von Videoaufnahmen. Daher waren vor allem Politiker und Schauspieler Objekte der Manipulation. Wird das Prinzip der Samsung-Forscher weiterentwickelt, reicht schon ein Facebook-Profilfoto.

Fake-Apps für die Hosentasche

Bisher ist der Einsatz der Technologie noch recht aufwendig. Doch die Fakes werden immer einfacher. Im September letzten Jahres sorgte die Veröffentlichung der Deepfake-App „Zao" in China für Aufsehen. „Zao" kann das Gesicht einer beliebigen Person in Szenen von Hollywood-Filmen kopieren und benötigt dafür nichts weiter als ein paar Portraits und ein paar Sekunden Zeit. Die iPhone-App, die bisher nur in China installiert werden kann, sprang in kürzester Zeit auf den ersten Platz der chinesischen iOS-Charts und verursachte einigen Wirbel.

Derzeit ist das nur eine harmlose Spielerei, deren Missbrauchspotenzial überschaubar ist: So kann man sich nur in bestimmte, vorausgewählte Szenen aus Filmen wie etwa „Titanic" einkopieren, die die App anbietet. Das dürfte auch das Geheimnis hinter ihrer Schnelligkeit und leichten Anwendbarkeit sein: Die Software ist bereits auf genau diese Szenen trainiert. Dennoch zeigt Zao, wo die Reise hin-

Screenshot



Die Forschergruppe um Matthias Nießner entwickelt neue Technologien

gehen könnte: Deepfake-Apps für die Hosentasche, für jeden mit wenigen Klicks einsetzbar.

Entwicklung von Gegengiften

Umgekehrt ist es aber auch möglich, Programme zu entwickeln, die Fakes enttarnen können. Diese nutzen dieselben Prinzipien, arbeiten also quasi rückwärts. Die Forscher des „Visual Computing Lab“ der TU München haben bereits ein System entwickelt, das Fake-Videos entlarven kann. Der Algorithmus „FaceForensics“ nutzt eine Datensammlung, die es ermöglicht, manipulierte Gesichter automatisch zu erkennen. Die Entwickler versprechen eine sehr hohe Genauigkeit: Bei unkomprimierten Videos erreichten sie in ihren Experimenten eine Trefferquote von 99,93 Prozent. Bei Videos mit niedrigerer Qualität soll die Quote niedriger sein, aber immer noch um die 90 Prozent betragen.

„Mit unserer Forschung verfolgen wir das Ziel, Videomanipulationen im Netz künftig einfacher aufspüren zu können“, erklärt Projektleiter Prof. Matthias Nießner. „Fake-Videos als solche zu erkennen, erweist sich besonders in den Sozialen Medien als schwierig, da diese dort meist komprimiert und schlecht aufgelöst hochgeladen werden. Dieselben Methoden, die zur Manipulation von Videos genutzt werden, sind aber auch in der Lage, Fälschungen besonders zuverlässig aufzuspüren – selbst bei schlechter Bildauflösung.“

Nießner trainiert seine Software mit einer umfangreichen Datenbank von Fake-Videos. So lernt die Software, was ein Fake-Video von einem echten unterscheidet. Tatsächlich kann Software besser geeignet sein, Fakes zu enttarnen, als selbst der geübteste Betrachter, denn die Software kann deren Artefakte leichter erkennen. Für automatisierte Systeme ist eine solche Analyse relativ einfach zu erledigen.

Auch andere Entwickler arbeiten mit Hochdruck an Gegenmitteln: Derzeit läuft die „deepfake detection challenge” (www.deepfakedetectionchallenge.ai). Der unter anderem von Facebook und Microsoft gesponsorte Wettbewerb setzt Preisgelder für Forscher aus, die die effektivste Technologie zur automatischen Erkennung von Fake-Videos vorstellen können. Am 31. März ist Einreichschluss.

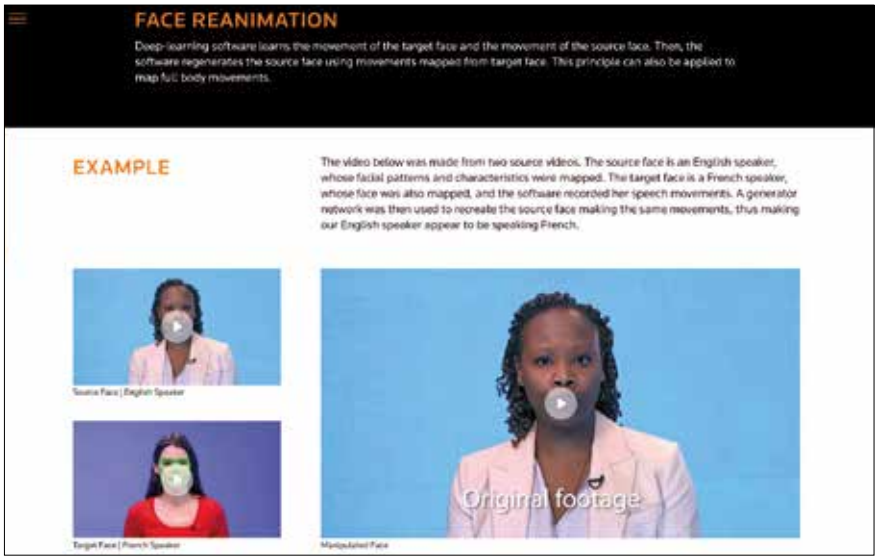
Verheerende Wirkung ist vorstellbar

Das Missbrauchspotenzial von Videofakes ist offensichtlich: Man kann in einem Video jetzt schon jeden alles sagen, bald wahrscheinlich auch jeden alles machen lassen. Bisher haben die Ergebnisse noch Artefakte, die aber bei nicht so genauem Hinsehen nicht auffallen. Die bei Internet-Videos übliche niedrige Auflösung trägt weiterhin dazu bei, diese Artefakte zu verschleiern. Und die Technologie macht rasante Fortschritte: Video-Fakes werden immer besser und es wird immer schwerer sein, sie zu enttarnen. Soviel ist sicher.

Aber auch wenn ein Deepfake-Video einer eingehenden Prüfung mit datenforensischen Mitteln nicht standhält, kann es eine verheerende Wirkung haben. Viele Medienkonsumenten hinterfragen die Glaubwürdigkeit von Nachrichten nicht wirklich, sofern deren Inhalt nur ihrem eigenen Weltbild (oder ihren Vorurteilen) entspricht. Wenn ein Fake-Video mit kompromittierendem Inhalt zu einem strategisch günstigen Zeitpunkt – etwa so kurz vor einer Wahl, dass keine Zeit für eingehende Prüfungen bleibt – veröffentlicht wird, kann es enorme Auswirkungen haben. Wenn dann Tage später irgendein Datenforensiker den Fake entlarvt, hat dieser seine vergiftende Wirkung längst getan.

Insbesondere die kommende US-Präsidentschaftswahl weckt in dieser Hinsicht einige Befürchtungen. Facebook scheint sich seiner Verantwortung in diesem Zusammenhang endlich bewusst zu werden: In einer Verlautbarung vom 6. Jänner 2020 („Enforcing Against Manipulated Media“) verlautete das US-Unternehmen, die Veröffentlichung manipulierter Videos in Zukunft nicht mehr zu gestatten, sofern die Manipulation den Inhalt

Screenshot



Reuters hat einen Online-Kurs zum Erkennen manipulierter Videos veröffentlicht

des Videos verändert („...would likely mislead someone into thinking that a subject of the video said words that they did not actually say."). Ausnahmen soll es geben für offensichtliche oder ausdrücklich gekennzeichnete Fakes (darunter fielen auch Demos wie Jordan Peeles Obama-Fake) und Manipulationen „for purposes of parody or satire". Aber auch dafür muss man einen Fake natürlich erstmal als solchen erkennen.

Gesundes Misstrauen

Wie wird das unsere Welt – und vor allem die Medienwelt – beeinflussen, wenn sich irgendwann bei einer Videoaufnahme niemand mehr sicher sein kann, ob diese echt ist oder nicht? Videos galten bisher als weitgehend sicheres Beweismaterial. Das ist nun nicht mehr der Fall. Dass solche Fakes irgendwann als Grundlage von Propaganda oder absichtlich irreführenden Nachrichten benutzt werden, ist recht wahrscheinlich. Daran werden sich Nachrichtenredaktionen, aber auch alle Medienrezipienten gewöhnen müssen. Wir werden lernen müssen, einen Fake für wahrscheinlich zu halten, sofern es keine ergänzenden Beweise für die Echtheit des Videos gibt. Man muss entscheiden, wem man noch glauben kann.

Ein gesundes Misstrauen wird bei Videos also in Zukunft normal sein müssen. Die Fotografie hat diese Entwicklung längst vollzogen – jeder weiß heutzutage, dass ein Foto leicht manipuliert werden kann und daher als unumstößlicher Beweis nicht mehr tauglich ist. Wenn alle wissen, dass ein Video ebenso manipuliert sein kann wie ein Foto, wird ein gefaktes Pornovideo nicht mehr das Potenzial haben, eine Karriere zu zerstören, eine gefakte Äußerung eines Politikers keine Wahl mehr beeinflussen können.

Zuverlässige Antifake-Software kann dabei natürlich enorm hilfreich sein. So könnten etwa Nachrichtensender ihnen zugespielte Videos standardmäßig zuerst von einer solchen Software verifizieren lassen und dann mit einer Art Echtheitszertifikat veröffentlichen. Das wäre eine weitere Chance für den Qualitätsjournalismus, in einer Welt aus Falschnachrichten für Glaubwürdigkeit und Vertrauen zu sorgen.

Die Experten sind sich allerdings uneins über die Möglichkeiten der Fake-Erkennung: Während manche meinen, dass es immer möglich sein wird, Fakes mit datenforensischen Mitteln zu enttarnen, sind andere da nicht so zuversichtlich. Sie glauben, dass es irgendwann Fakes geben könnte, die auch mit ausgefeiltesten Methoden nicht erkennbar sind. Und was ist, wenn die Software sich irrt und ein Fake daraufhin als „offiziell verifiziert" veröffentlicht wird?

Die Nachrichtenagentur Reuters setzt daher neben Technologie auf Manpower und hat einen Online-Kurs veröffentlicht, der Journalisten helfen soll, Fake-Videos zu identifizieren (www.reuters.com/manipulatedmedia). Dabei geht es auch um ganz altmodisches journalistisches Handwerkszeug wie die Überprüfung der Glaubwürdigkeit und die Verifizierung durch weitere Quellen.

Videos verlieren Beweiskraft

Die Gefahr scheint aber auch in der anderen Richtung zu liegen: Wenn das Misstrauen in Videos so groß wird, dass keines mehr als echt angesehen wird, taugen Videos nicht mehr als Beweise. Auch dann nicht, wenn sie tatsächlich echt sind.

Denkbar und nicht minder brisant ist daher der umgekehrte Fall: Ein Politiker, der tatsächlich in einer kompromittierenden Szene gefilmt wurde, könnte behaupten, dieses Video sei eine Fälschung. Gelingt dann kein unwiderlegbarer Echtheitsbeweis, könnte er sich damit nicht nur reinwaschen, sondern seine Position beim harten Kern seiner Anhänger sogar stärken – man denke nur an die sorgsam gepflegte Opfer-Rhetorik vieler Rechtspopulisten, die alle Angriffe als perfide Verschwörung ihrer Feinde werten. Was wäre, wenn H.C. Strache hätte behaupten können, das Ibiza-Video sei ein Fake? Und niemand in der Lage gewesen wäre, ihn glaubhaft und mit Sicherheit zu widerlegen? ■

Screenshot