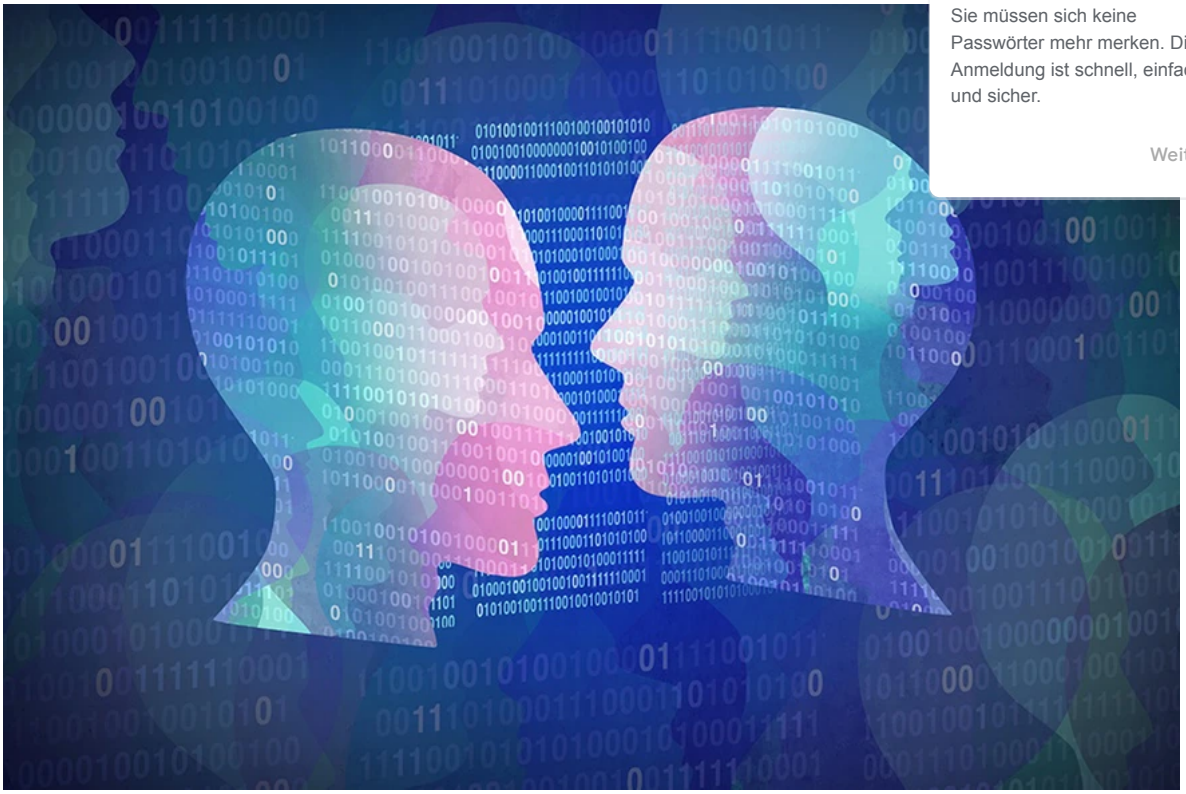
Wissenschaft

Wenn Deepfakes die Realität überholen – und ve



Die Manipulation von Bildern ist mindestens so alt wie die Fotografie. Abraham Lincoln nutzte die Fälschung, um besser auszusehen und seine präsidiale Aura zu verstärken. Stalin und Mao Zedong nutzten sie, um ihre politischen Gegner aus den Geschichtsbüchern zu tilgen.

14. August 2021 - 09:00

Sara Ibrahim

Ich befasse mich mit Fragen im Zusammenhang mit den Auswirkungen der neuen Technologien auf unsere Gesellschaft. Sind wir uns der gegenwärtigen Revolution und ihrer Folgen bewusst? Liebstes Hobby: freies Denken. Mein Tick: Ich stelle mir selber zu viele Fragen.

Doch während früher nur Experten in der Lage waren, das menschliche Auge meisterhaft zu täuschen, ist dies heute ein Kinderspiel geworden. Praktisch jeder kann dies tun. Eine aus dem Internet heruntergeladene Software und ein paar Bilder aus Internet-Suchmaschinen oder sozialen Medien reichen aus, um gefälschte Videos herzustellen, die sich im Internet wie ein Lauffeuer verbreiten.

"Eine Fotografie reicht aus, um ein gutes Deepfake zu erstellen", sagt Touradj Ebrahimi, Leiter des Labors für Multimedia-Signalverarbeitung an der Eidgenössischen Technischen Hochschule Lausanne (EPFL).

Was bedeutet "Deepfake"?

Der Begriff "Deepfake" wurde 2017 geprägt und setzt sich zusammen aus "Deep Learning" (das "tiefgehende" maschinelle Lernen, das auf künstlicher Intelligenz beruht) und "Fake" (falsch im Englischen).

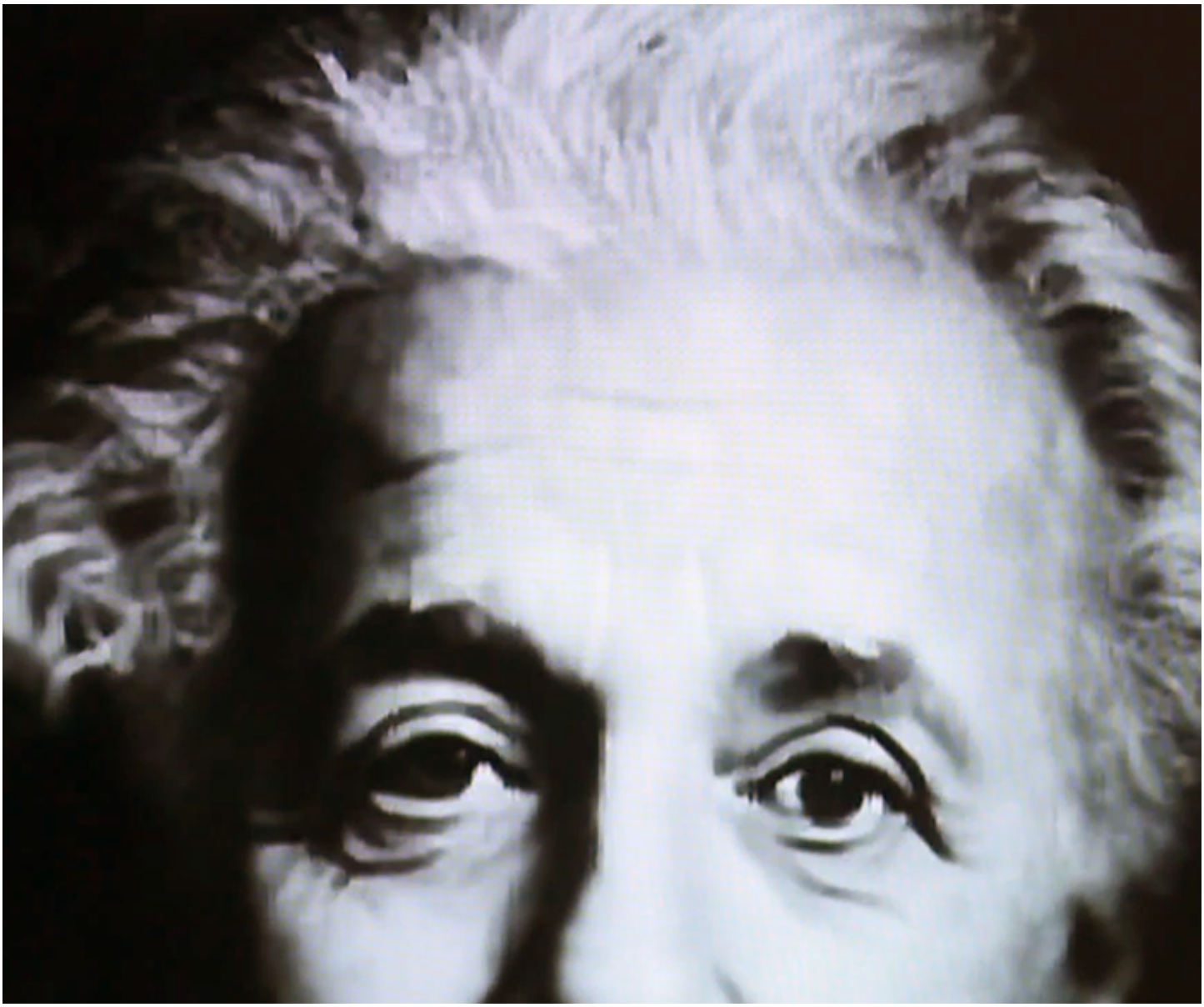
Deepfake nutzt künstliche Intelligenz (KI), um synthetische Bilder zu erzeugen. Sie wirken so echt, dass sie nicht nur das menschliche Auge täuschen, sondern auch die Algorithmen, die sie erkennen sollen. Seit einigen Jahren beschäftigt sich Ebrahimis Team mit Deepfakes. Das Ziel der Forschungen ist die Entwicklung hochmoderner Systeme, die zur Überprüfung und Erkennung der Echtheit von Fotos, Videos, Bildern und Audiodaten dienen.

Doch dieser Forschungs- und Entwicklungsprozess ist zu einem Wettlauf gegen die Zeit und Technologie geworden, denn die Manipulation von Informationen und die Herstellung von mit KI veränderten Videos ist förmlich explodiert.

In vielen Teil der Welt sind Deepfakes sogar zu einem Problem der nationalen Sicherheit geworden – dies mit dem Aufkommen der sozialen Medien. Millionen von Menschen, aber auch Unternehmen und Regierungen, können Inhalte erstellen und frei zugänglich machen, aber eben auch manipulieren.

Laut Ebrahimi gelten Länder wie Russland, China, Iran und Nordkorea – also allesamt autoritäre Staaten oder Diktaturen – als sehr aktiv bei der Verbreitung von Fake News, auch in Form von Deepfakes. Und das sowohl innerhalb als auch ausserhalb ihrer Landesgrenzen.

Ein Spinoff von Forschenden der ETH Lausanne setzt KI und Deep Learning ein, um manipulierte Videos erkennen zu können. Video SRF/swissinfo.ch:



Das Auge will seinen Teil der Wahrheit

Menschen zweifeln kaum daran, was sie mit den eigenen Augen sehen. Eine Studie des Massachusetts Institute of Technology (MIT), hat gezeigt, dass sich gefälschte Nachrichten auf Twitter bis zu sechsmal schneller verbreiten als wahre Nachrichten.

Dies macht das Deepfake-Phänomen laut Ebrahimi besonders besorgniserregend. "Deepfakes sind ein sehr wirkungsvolles Mittel der Fehlinformation, weil die Menschen immer noch dazu neigen, das zu glauben, was sie sehen."

Auch die Qualität der Videofilme nimmt weiter zu, so dass es immer schwieriger wird, echte von gefälschten Videos zu unterscheiden. "Ein Staat, der über unbegrenzte oder fast unbegrenzte Ressourcen verfügt, kann heutzutage gefälschte Videos erstellen, die so echt erscheinen, dass diese Fälschungen selbst von Experten nicht erkannt werden können", sagt Ebrahimi.

Ausgefeilte Software vermag Manipulationen zwar heute noch erkennen, aber der Professor schätzt, dass selbst Maschinen in zwei bis fünf Jahren nicht mehr in der Lage sein werden, echte von gefälschten Inhalten zu unterscheiden.

Auf der Jagd nach Deepfakes

Das Forschungslabor von Touradj Ebrahimi beschäftigt sich seit 20 Jahren mit Problemen medialer Sicherheit in Bezug auf Bilder, Videos, Tonträger und Sprache sowie mit der Überprüfung ihrer Echtheit.

Ursprünglich waren die Manipulationen hauptsächlich eine Frage des Urheberrechts. Später verlagerte sich das Thema auf den Schutz der Privatsphäre und die Videoüberwachung bis zum Aufkommen der sozialen Medien, die zu einer massiven Verbreitung manipulierter Inhalte beitrugen.

Deepfakes sind in der Lage, die zur Erkennung von Fälschungen verwendeten Detektoren zu umgehen. Aus diesem Grund verwendet Ebrahimis Labor ein "Paradigma", die so genannte "Provenance-Technologie", um anonym festzustellen, wie ein Inhalt erstellt wurde und welche Manipulationen vorgenommen wurden.

"Damit die Provenance-Technologie funktioniert, muss sie von einer Vielzahl von Akteuren im Web genutzt werden: von Google über Mozilla, Abode, Microsoft bis hin zu allen sozialen Medien, um nur einige zu nennen", so der Experte. Ziel ist es, sich auf einen JPEG-Standard (für Bild- und Videodateien) zu einigen, der weltweit angewendet werden soll", fügt er hinzu.

"Immer mehr Manipulationen"

Zunächst wurden gefälschte Videos vor allem dazu verwendet, komische Clips von Schauspielern, Politikerinnen und anderen bekannten Personen zu erstellen. Oder sie wurden für Videospiele verwendet. Doch schon bald wurden sie zu einem wirksamen Instrument der Verunglimpfung oder Hate Speech, insbesondere gegen Frauen. Oder zu einem Mittel, um Geld zu erpressen und die öffentliche Meinung zu manipulieren.

Deepfakes haben bewiesen, dass sie in der Lage sind, die Gesichter von zwei verschiedenen Personen zu überlagern, um beispielsweise ein falsches Profil oder sogar eine falsche Identität zu erstellen. Cyberkriminelle haben dies ausgenutzt, um Unternehmen dazu zu bringen, ihnen Geld zu übermitteln, indem sie sich als Administrator ausgaben und eine dringende Anfrage für eine Geldüberweisung vortäuschten.

"Im Moment gibt es nur wenige solcher Manipulationen, aber wenn die Technologie erst einmal ausgereift ist, werden wir immer mehr davon sehen", prognostiziert Sébastien Marcel, Wissenschaftler am schweizerischen Forschungsinstitut für Künstliche Intelligenz Idiap. Marcel erklärt, dass mit der derzeitigen Deepfake-Technologie nur visuelle Inhalte manipuliert werden können, nicht aber Audiodateien.

Die Stimmen werden, wenn sie nicht aus anderen Videos stammen, von einem Profi nachgeahmt. "Audiofälschungen sind immer noch eine Herausforderung, aber in Zukunft werden wir ultrarealistische Fälschungen sehen, welche Bild und Stimme einer Person in Echtzeit reproduzieren können." Dann sind Manipulationen, etwa das Vortäuschen eines Skandals über einen Marktrivalen oder Konkurrenten, leicht möglich.

Biometrische Daten in der Schweiz

Sébastien Marcel leitet die Gruppe für biometrische Sicherheit und Datenschutz am Schweizer Forschungsinstitut Idiap.

Es handelt sich um eine der wenigen Forschungseinrichtungen in der Schweiz, die sich auf biometrische Forschung spezialisiert hat, um die Echtheit und Verlässlichkeit von Fingerabdruck- und Gesichtserkennungssystemen zu bewerten und zu verbessern.

"Die Forschung im Bereich der Gesichtserkennung und der Biometrie im Allgemeinen ist in der Schweiz noch recht unterentwickelt", sagt Marcel.

Die Realität verleugnen

Ebrahim weist seinerseits darauf hin, dass gefälschte Videos auch positive Anwendungen finden können. "Deepfakes wurden bereits in der Psychotherapie eingesetzt, um das Leid derjenigen zu lindern, die einen geliebten Menschen verloren haben", sagt der Professor. Es gab beispielsweise einen Fall in den Niederlanden, wo ein trauerndes Elternteil ein Deepfake seiner zu früh verstorbenen Tochter erstellte, um sich von ihr verabschieden zu können.

Die Genealogie-Website MyHeritage kann Ähnliches: Mit ihrem Tool DeepNostalgia kann sie verstorbene Verwandte "wiederauferstehen" lassen, indem sie deren auf Fotos festgehaltenen Gesichter filmisch belebt.

Da jedoch das Bewusstsein für Deepfakes zunimmt, können sogar echte Videos mit manipulierten Inhalten verwechselt werden. In Gabun (Zentralakfrika) wurde ein Neujahrsvideo von Präsident Ali Bongo, der sich zuvor wochenlang wegen Krankheit nicht in der Öffentlichkeit gezeigt hatte und sich im Ausland aufhielt, von vielen für eine Fälschung gehalten, was zu einem Aufstand einer Handvoll Militärputschisten führte.

Die Ungewissheit darüber, was echt ist und was nicht, kann einen unbeabsichtigten Effekt haben und eine Kultur der "plausiblen Bestreitbarkeit" schaffen. Das würde heissen: Niemand ist mehr bereit, Verantwortung zu übernehmen, weil alles gefälscht sein könnte. So argumentiert zumindest die Autorin Nina Schick in ihrem Buch "Deepfakes: The Coming Infocalypse" (Deepfakes: Die kommende Infokalypse).

"Deepfakes könnten jedem die Möglichkeit geben, alles zu fälschen, und wenn alles gefälscht werden kann, kann jeder eine plausible Bestreitbarkeit behaupten", argumentiert Schick. Sie ist der Meinung, dass dies eine der grössten sozialen Gefahren ist, die von Deepfakes ausgehen. Auch andere Autoren halten Deepfakes für eine Bedrohung für Demokratie und Gesellschaft.

Wie man die Fake News-Kultur bekämpft

Die gute Nachricht ist, dass die Europäische Union das Problem nicht auf die leichte Schulter nimmt. Förderprojekte wie Horizon Europe unterstützen die Forschung zu gefälschten Videos. "Wir gehen davon aus, dass es in den kommenden Jahren mehr EU-Ausschreibungen zu Deepfakes-Forschungsvorhaben geben wird", sagt Sébastien Marcel von Idiap.

Auf technischer Ebene bedeutet die Bekämpfung von Deepfakes, dass man proaktiv vorgehen und sich auf Schwachstellen in Systemen konzentrieren muss. "Aber das ist nicht immer so einfach", argumentiert der Wissenschaftler. "Die akademischen Mühlen, um Fördermittel zu erhalten, mahlen langsam. Unterdessen entwickeln sich die Technologien hinter den Deepfakes immer schneller."

Ebrahimi und Marcel sind sich einig, dass es zur Bekämpfung von Fake News unerlässlich ist, die Bevölkerung zu sensibilisieren und zu erziehen, damit sie ein kritisches Bewusstsein und einen tiefen Sinn für staatsbürgerliche Verantwortung entwickelt. "Wir müssen unseren Kindern beibringen zu hinterfragen, was sie im Internet sehen", sagt Ebrahimi, "und nicht wahllos irgendwelche Inhalte verbreiten".

(Übertragung aus dem Italienischen: Gerhard Lob)

jti CERTIFIED In Übereinstimmung mit den JTI-Standards

Mehr: JTI-Zertifizierung von SWI swissinfo.ch