

# Nichts, was sich nicht fälschen liesse

## Das wirkliche Leben könnte eine Renaissance erfahren, weil die Virtualität immer mehr zum Betrug verkommt

*Marco Wehr*

Die Systeme, die unsere Mimik und Sprache imitieren, werden in wenigen Jahren perfekt sein. Bettmann/Getty

Nach einer euphorischen Gründerphase, in der vor allen Dingen Transparenz und der freie Austausch von Daten im Zentrum der Aufmerksamkeit standen, sind mittlerweile Lug und Trug im Internet so verbreitet, dass man nicht weiss, wo man beginnen soll: Beim Cybergrooming? Dem Heranpirschen pädophiler Männer, die sich mit Fake-Accounts und falschen Bildern auf Foren herumtreiben, in denen sich junge Mädchen ihre Reiterlebnisse erzählen. Oder bei Hackern, die es auf die Sabotage von Kraftwerken oder Krankenhäusern abgesehen haben? Die Liste liesse sich beliebig verlängern, und dennoch haben wir uns bis dato nicht aus den digitalen Informationsparadiesen vertreiben lassen. Jetzt tauchen allerdings neue dunkle Wolken auf: In naher Zukunft wird es für den normalen Internetnutzer nicht mehr möglich sein, Filme, in denen uns vertraute Menschen zu sehen und mit ihrer Stimme zu hören sind, von raffinierten Fälschungen zu unterscheiden. Damit transformiert sich der digitale Raum zum potenziellen Lügenland.

Bilder und Töne lassen sich besonders gut digitalisieren. Das Internet spricht nicht die Gesamtheit unserer Sinne an, sondern bevorzugt Augen und Ohren. Für andere Sinnesqualitäten ist das nur in eingeschränkter Masse oder gar nicht möglich. Man kann sich eigentlich jeden Computerbenutzer als ein riesiges Auge vorstellen, an dem zwei grosse Ohren kleben. Dieser sensorische Homunculus besitzt weder eine Nase noch einen Mund, er hat keine Haut und auch keine komplexe Innenwahrnehmung.

Seit den Pionierarbeiten von Ivan Sutherland in den siebziger Jahren des vorigen Jahrhunderts wird zwar viel vom sogenannten «ultimate display» gesprochen – eine Art omnipotente Computer-Mensch-Schnittstelle. Sie wäre in der Lage, die gesammelte Sinnesfülle unserer Alltagserfahrungen zu reproduzieren. Aber über vibrierende Joysticks mit variablem Widerstand und ein paar andere Gimmicks ist man noch nicht hinausgekommen.

### **In wenigen Jahren perfekt**

Augen und Ohren lassen sich jedoch exzellent durch hochauflösende Kameras und empfindliche Mikrofone ersetzen. Lautsprecher und Bildschirme wiederum verwandeln den Input in einen natürlich erscheinenden Output. Damit wächst auch die Möglichkeit des Missbrauchs. Als Fotos noch aufwendig in der Dunkelkammer entwickelt wurden, war viel handwerkliche Finesse notwendig, um ein Bild glaubwürdig zu manipulieren, auch wenn etwa Hitlers Leibfotograf Heinrich Hoffmann beim Retuschieren schon beachtliches Geschick besass. Das Gleiche galt für die in dieser Zeit gemachten Filme. Aus diesem Grund konnte man als Betrachter lange wenigstens einigermassen sicher sein, dass das, was auf dem Bild zu sehen war, auch «in Wirklichkeit» so war. Doch schon 1994 konstatierte der Medienwissenschaftler William J. Mitchell, dass die Sicherheit, dass ein Foto die Wirklichkeit dokumentiere, unwiederbringlich vorbei war.

Schon mit damaliger, noch rudimentärer digitaler Technik gelang es, Fotos, auf denen Politiker zu sehen waren, so zu arrangieren, dass der Kontext der Beziehung zwischen den Personen verändert wurde. Aber das war nichts im Vergleich zu gegenwärtigen Möglichkeiten. Im optisch-akustischen Bereich gibt es eigentlich nichts mehr, was sich nicht überzeugend fälschen liesse.

Beginnen wir mit der Stimme. Ihr individueller Klang macht jeden Menschen unverwechselbar. Bis vor kurzem jedenfalls. Heute reichen zwanzig Minuten Sprachmaterial, das viele Menschen nichtsahnend im Netz verfügbar machen, damit ein Computer in der Lage ist, eine Stimme ziemlich echt zu imitieren. Mithilfe eines Programms wie VoCo von Adobe kann jeder Satz, den man mit der Tastatur eingibt, mit der gesampelten Stimme wiedergegeben werden. Damit kann man jemanden Dinge sagen lassen, die er nie sagen würde.

Damit sind die Möglichkeiten des Betrugs nicht ausgeschöpft. Es ist der Stand der Technik, auf der Grundlage von Filmaufnahmen genauestens zu analysieren, in welcher Weise bestimmte Menschen beim Sprechen ihren Mund bewegen und mit welcher Mimik sie das tun. Damit kann man eine Zielperson in einer computergenerierten Filmaufnahme jeden denkbaren Satz artikulieren lassen: mit ihrer persönlichen Stimme und ihrer eigenen Mimik. Die Systeme werden in wenigen Jahren perfekt sein.

Doch das Internet wird noch aus anderen Gründen ein ungemütlicher Raum. Betrachten wir eine alltägliche Gesprächssituation: Der Inhalt der Worte ist nur ein kleiner Teil der Information, die zwischen den Sprechenden ausgetauscht wird. Viele andere Dinge schwingen mit: die Körperhaltung, der räumliche Abstand, das Mienenspiel, Stimmlage und Betonung. Obwohl wir diese Informationen meist unbewusst wahrnehmen, sind sie essenziell, um das Gesagte richtig zu bewerten.

Interessanterweise lassen sich Deep-Learning-Algorithmen so trainieren, dass sie in solchen Kontexten genauer hinschauen können als Menschen und ihnen deshalb im fein gewebten Spiel der Emotionen nichts entgeht. Natürlich verstehen die Computer die Gefühle nicht. Sie lesen aber deren Zeichen. Sie können mittels subtiler Informationen auf die seelische Verfasstheit der observierten Menschen schließen. So lässt sich etwa das Mienenspiel minutiös analysieren. Und Körpersprache und Gangbild zeigen, ob sich jemand in irgendeiner Weise auffällig macht. Auch die Stimme gibt Geheimnisse preis. Sie verrät dem Computer zum Beispiel, ob der Observierte depressiv ist oder Gefahr läuft, an Parkinson zu erkranken.

### **«Find Face»**

Es gibt bereits Algorithmen, die Probanden umfassend charakterisieren und sauberlich in Schubladen einordnen. Die deutsche Firma Precire will die gesamte Persönlichkeit auf der Grundlage von Sprachdaten entschlüsseln. Sie ordnet Probanden in Kategorien ein wie «selbstorganisiert», «ausgeglichen» oder «autonomiebedürftig» und verteilt Punkte. Es gibt bereits Unternehmen, welche die Software verwenden, um Bewerber zu bewerten. Wann werden Krankenversicherungen solche Mittel für ihre Zwecke nutzen? Laut der Informatikerin Julia Hirschberg von der Columbia State University ist die Bewertung der Persönlichkeit auf der Basis von Sprachdaten seriöse Wissenschaft. Sie selbst hat ein Programm entwickelt, das Lügner besser enttarnt als jede andere Methode.

Unterm Strich sind dystopische Szenarien denkbar. Einen Vorgeschmack gab es 2016, als in Russland die App «Find Face» auf den Markt kam. Ein Foto eines Menschen lässt sich mit dieser App einem Profilbild in den sozialen Netzwerken zuordnen. Bis vor kurzem war die App auf das russische Facebook-Pendant VK.com beschränkt. Das wird wohl nicht so bleiben. Dann ist es denkbar, auf der Strasse von einer Person schnell ein heimliches Bild zu machen, um sie dann in den sozialen Netzwerken aufzuspüren. Es dauerte nicht lange, bis «Find Face» in einem anderen Kontext zur Anwendung kam. Man enttarnte Pornodarstellerinnen, um sie dann zu erpressen. Doch damit sind mögliche Szenarien nur angerissen.

Vor diesem Hintergrund sollte uns die Einschätzung von Artem Kucharenko, dem Chefentwickler von «Find Face», in den Ohren klingen. In einem Interview mit «jetzt.de» sagt er, dass durch Software wie «Find Face» unsere Privatsphäre in grossem Masse zerstört werde.

Wer diese – zumindest in Massen – erhalten will, muss sich deshalb genau überlegen, was er von sich selbst und seinen Nächsten ins Netz stellt. Vermutlich wäre es noch besser, konsequent zu handeln: Um Manipulationen auszuschliessen, trifft man sich mit vertrauten Menschen aus Fleisch und Blut und sorgt dafür, dass Handys und Computer abgestellt sind. In diesem Sinne war der kürzlich verstorbene Karl Lagerfeld, der sowohl Handy als auch Uhr ablehnte, nicht ein aus der Zeit gefallener Hinterwäldler, sondern eher ein Visionär, der seiner Zeit voraus war.

**Marco Wehr** ist Physiker und Philosoph. Er leitet in Tübingen das Philosophische Labor ([www.das-philosophische-labor.de](http://www.das-philosophische-labor.de)).