Fotomontage, gemeinfrei.

„Putin reißt Neugeborenem das **Herz aus der Brust**“

So oder so ähnlich könnte die Befehlseingabe lauten, um das per künstlicher Intelligenz gesteuerte Bilderstellungsprogramm DALL·E 2 [1] ein fotorealistisches Bild erstellen zu lassen. Die weltweite Empörungswelle wäre gesichert und der Kriegseintritt der NATO-Staaten wäre – wenn schon nicht völkerrechtlich – dann aber moralisch gerechtfertigt.

Autor: Björn Gschwendtner

Freischaffender Künstler, Aktivist im Internet und im echten Leben. Der gelernte Biologielaborant hat diverse Internetprojekte mit Augenmerk auf der Förderung von Demokratie, Freiheit und Bildung gestartet. U.a. auch **gegendarstellung.org**

Echt oder nicht echt, das ist hier die Frage

Seit einigen Jahren entstehen mittels sog. Neuronaler Netzwerke (zu „künstlichen Gehirnen“ verbundene Netzwerke, allgemein unter dem Begriff „künstliche Intelligenz“ bekannt) künstlich erstellte Videos, Fotos und Audioaufnahmen, die von der Realität nicht mehr zu unterscheiden sind. Die künstlich erschaffenen Menschenabbilder sind so realitätsgetreu, dass sie das sog. Uncanny Valley („unheimliches Tal“; Akzeptanzlücke [2]) überwunden haben und eine hohe Akzeptanz bei der Bevölkerung erreichen. Das schafft, neben dem großen Unterhaltungswert, den diese Technologien bieten können, ein großes Problem in die Glaubwürdigkeit von Beweismitteln und Nachrichten.

Die in der Einleitung erwähnte Software DALL·E 2 ist die bekannteste ihrer Art, aber es gibt auch einige alternative Anbieter. Mittels einer Textzeile kann der Nutzer dem Programm das gewünschte Motiv beschreiben und das Programm spuckt nach einer kurzen Rechenzeit verschiedene Bilder mit dem beschriebenen Inhalt aus. Gibt man dem Programm vor, es solle das Bild als Manga-Comic erstellen, kommt auch ein Manga-Comic-Bild heraus. Gibt man dem Programm vor, es solle das Bild fotorealistisch erstellen, kommt am Ende ein fotorealistisches Bild heraus.

Zwar verbieten die Allgemeinen Geschäftsbedingungen der Software dem Nutzer Befehlseingaben, wie sie in der Überschrift des Artikels verwendet wurden, aber die Technologie ist da und Ge-

schäftsbedingungen können geändert oder auch umgangen werden.

Diese Person existiert nicht

Ruft man die Domain https://thispersondoesnotexist.com („Diese Person existiert nicht“) auf, sieht man das Porträtfoto einer unbekannten Person. Rufen Sie die Seite erneut auf, erscheint ein weiteres Porträtfoto – jung oder alt, männlich oder weiblich. Ist dies die Webseite eines Porträtfotografen, der alle seine Modelle in einer Präsentationsschleife zeigt?

Nein, die Bilder, die Sie sehen, sind durch keine Fotokamera der Welt aufgenommen worden. Die Bilder werden bei jedem neuen Aufruf der Seite von einer künstlichen Intelligenz neu erstellt. Kein Bild gleicht dem Bild davor und das Foto, das Sie sehen, wird kein Anderer sehen können (es sei denn, Sie speichern es auf Ihrem Computer und teilen dann diese Datei).

Vor ein paar Jahren konnte man häufiger noch Fehler oder Artefakte in den Bildern erkennen, was die Identifizierung als künstliches Produkt leichter machte. So sah man zum Beispiel Brillen, deren Gestell nur einen Bügel hatte, oder Ohrringe, die seltsam am Ohr der Trägerin hingen. Haare wurden früher häufig auch mit offensichtlichen Problemen dargestellt. Mittlerweile generiert die Webseite aber immer seltener diese Fehler, sodass selbst geübte Augen schwer zwischen Foto und Fake unterscheiden können.

Eine Studie stellte sogar fest, dass die künstlich erschaffenen Gesichter für Testpersonen nicht nur ununterscheidbar waren, sondern dass den gefakten Fotos statistisch mehr Glaubwürdigkeit zugesprochen wurde [3].

Praktisch hierbei: Die nicht unter Urheberrecht stehenden und demnach frei verfügbaren Bilder lassen sich für Produktbewertungen und sog. „Testimonials“ verwenden. Produktbewertungen erzeugen unterbewusst in uns eine höhere Glaubwürdigkeit, wenn uns dazu ein Gesicht mit Namen präsentiert wird, als wenn wir nur einen Text zu lesen bekommen (der häufig auch einfach nur frei erfunden ist). Um nicht lange über glaubwürdige Namen und Berufsbezeichnungen nachdenken zu müssen, bedient man sich einfach einer Webseite wie https://www.fakenamegenerator.com und erhält, je nach gewünschtem Land, echt wirkende Personendaten inklusive Fake-Telefonnummer, Fake-Geburtsdatum, und Fake-Lieblingsfarbe.

Bild links – Texteingabe bei DALL-E 2: „Cover eines Comics aus den 1960er Jahren, mit Bruce Lee, der durch ein fehlgeschlagenes wissenschaftliches Experiment zum Hamster geworden ist.“
Bild rechts – Texteingabe bei DALL-E 2: Texteingabe: „Foto eines verwirrten Grizzlys im Mathematikunterricht“

Deep Fake – Tiefe Vertrauenskrise in Medieninhalte

Im Sommer 2022 erfuhren die Bürger über die Nachrichtenkanäle, dass die Bürgermeisterin von Berlin, Franziska Giffey, bei einer Videokonferenz mit Kievs Bürgermeister Vitali Klitschko mittels eines sog. Deep Fakes hereingelegt worden sein soll [4]. Nach 30 Minuten wurde die Verbindung dann abgebrochen, da Giffey das Gespräch zunehmend als skurril empfand.

Recherchen ergaben in diesem Fall, dass es sich hierbei nicht um ein Deep Fake gehandelt haben soll, sondern um ein sog. Shallow Fake/Cheap Fake („oberflächliche Fälschung/ billige Fälschung“) [5]. Dabei werden bereits existierende, echte Videoaufnahmen einer Person im Vorhinein so zusammengeschnitten, dass ein neuer Kontext entsteht. Dies war auch der Grund, weshalb das Gespräch nach einer gewissen Zeit des Vertrauensvorsprungs als seltsam empfunden wurde: Es gab keine bereits existierenden Gesprächsinhalte auf Giffeys Fragen, die sinnvoll hätten im Voraus zusammengeschnitten werden können.

Mit diesem Vorfall jedoch wurden größere Teile der Bevölkerung mit dem Thema „Deep Fake“ vertraut gemacht, wenn sie es nicht vorher schon waren. Bei Deep Fakes wird das Gesicht eines Körperdoubles per Computer mit einer anderen – meist prominenten – Person ausgetauscht (Face swapping = „Gesichtstausch“). Sämtliche Bewegungen und Lichtverhältnisse im Originalvideo berechnet der Computer für das einzubindende Gesicht so genau, dass das Endergebnis absolut glaubwürdig ist. Im Idealfall kann das Körperdouble die Stimme der gewünschten Person bereits imitieren oder man bedient sich dafür einer weiteren Software: Der Audio-Editor „Voco“ (Voice Conversion) kann z.B. nach 20-minütigem Anlernen beliebige Texte in der Zielstimme sprechen lassen [6].

Im medialen Zeitalter gibt es von jeder öffentlich agierenden Person genügend Material, das zum Anlernen entsprechender Deep-Fake-Software ausreicht, um realistische Fälschungen zu erstellen.

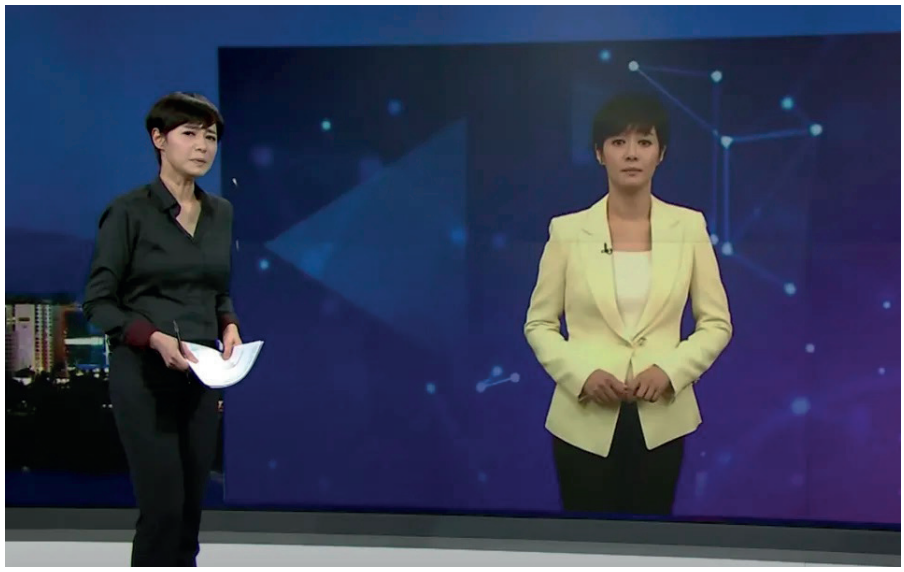
Ein TikTok-Kanal ist z.B. darauf spezialisiert, Deep-Fake-Videos von Tom Cruise zu veröffentlichen, in denen der Schauspieler in für ihn untypischen Alltagssituationen parodiert wird [7]. Von Cruise gibt es unzähliges Originalmaterial aus seinen Filmen, aus Interviews und anderen Formaten, mit dem die neuronalen Netzwer-

Publiziert auf Free21 am 28.12.2022, 6:00 Uhr

ke gefüttert werden können, um einen falschen Cruise zu erzeugen.

Ebenso wurde die südkoreanische Nachrichtensprecherin Kim Joo-ha für eine große Zuschauerschar mittels Deep Fake nachgeahmt. Die echte Moderatorin unterhielt sich sogar in der Sendung mit ihrem künstlichen Pendant, das auf deren Fragen antworten konnte [8]. Stellen Sie sich vor, dies wäre bei Franziska Giffey eingesetzt worden. Wäre der Fake dann jemals aufgeflogen?

Mit dem Programm auf der Webseite https://synthesia.io können Präsentationsvideos anhand von Texteingaben erstellt werden [9]. Eine frei wählbare Person – ob schwarz oder weiß, weiblich oder männlich, alt oder jung – fungiert als Erzähler(in) – in über 60 frei wählbaren Sprachen.



Die Nachrichtensprecherin Kim Joo-ha (links) unterhält sich mit ihrem synthetischen Abbild. Bild: Youtube - https://www.youtube.com/watch?v=k8X_Em-NQno&t

Die zwei Seiten der echten falschen Bilder

Es gibt zwei Seiten des Deep Fake: Einerseits ist es unterhaltsam und spart Ressourcen, andererseits zeigt es uns, welche Gefahren für die Glaubwürdigkeit von Bildern in Deep Fake stecken.

Beliebte Schauspieler könnten vollständig per Computer im Film nachgespielt werden. Ein geringer bezahlter unbekannter Schauspieler leiht der Figur seinen Körper und auf das Gesicht wird das eines beliebten, aber teureren - oder verstorbenen – Schauspielers gesetzt. Der beliebte Schauspieler würde, allein nur dafür, dass er die Reputation seines Gesichts zur Verfügung stellt, eine entsprechende Bezahlung erhalten, ohne, dass er jemals am Filmset auftauchen muss. Studio und Arbeitskosten könnten reduziert werden. Bruce Willis hat zwar seinen Rückzug aus dem Filmgeschäft aus gesundheitlichen Gründen bekannt gegeben, aber verleiht sein Gesicht gegen Lizenzgebühr für zukünftige Filme [10].

Was aber durch die verblüffenden Ergebnisse dieser Technologie spätestens jetzt Fakt geworden ist: Man darf keinen Bildmaterialien mehr – ob Bewegtbild, Tonaufnahme oder Fotografie – blindlings vertrauen. Es wird einem nichts anderes übrig bleiben, als alles, was einem vorgesetzt wird in Frage zu stellen, denn die Nutzer können es nicht auseinanderhalten – nein, sie trauen den künstlichen Kreationen sogar mehr. Wenn wir grundsätzlich allen Videos, Fotos und Stimmen misstrauen müssen, haben wir ein großes Problem mit Nachrichten. Und vor allem mit denen, die die Medienmacht inne haben.

Es ist auch davon auszugehen, dass im Moment mehr gefälschtes Material im Umlauf ist, als uns bewusst ist. Zudem muss unterstellt werden, dass entsprechend technologisch und finanziell unterstützte Stellen derzeit bereits über weit fortschrittlichere Methoden der Fälschung verfügen, als es Otto Normalverbraucher zur Verfügung steht – GPS und Internet waren zunächst militärische Entwicklungen, die erst später Zugang zum Massenmarkt fanden. Militär und Geheimdienste auf der ganzen Welt haben ein Interesse daran, die als Gegner auserkorene Seite in jeglicher Hinsicht zu kompromittieren, zu täuschen oder auch die eigene Bevölkerung in die gewünschte Richtung zu beeinflussen, in die man sie haben möchte (s. Überschrift dieses Artikels – in Anlehnung an die Brutkastenlüge, die die moralische Rechtfertigung für einen Kriegseintritt der USA in den Irak-/Kuwaitkrieg lieferte [11]).

Etliche Firmen spezialisieren sich im Moment darauf, Deep Fakes zu entlarven. Diese Technologien könnten dann so

Quellen:

[1] DALL·E 2 Webseite <*https://openai.com/dall-e-2/*>

[2] Das „Uncanny Valley“, oder Akzeptanzlücke: „Während man zunächst annehmen würde, dass Zuschauer oder Computerspieler ihnen dargebotene Avatare umso mehr akzeptieren, je fotorealistischer die Figur gestaltet ist, zeigt sich in der Praxis, dass dies nicht stimmt. Menschen finden hochabstrakte, völlig künstliche Figuren mitunter sympathischer und akzeptabler als Figuren, die besonders menschenähnlich bzw. natürlich gestaltet sind. Die Akzeptanz fällt der Theorie zufolge ab einem bestimmten Niveau des Anthropomorphismus schlagartig ab und steigt erst ab einem bestimmten, sehr hohen Grad wieder an. Die Akzeptanz wäre dann am höchsten, wenn sich die Imitationen überhaupt nicht mehr von echten Menschen unterscheiden ließen.“ Wikipedia: <*https://de.wikipedia.org/wiki/Uncanny_Valley*>

[3] PNAS - Proceedings of the National Academy of Sciences of the United States of America, Studie „AI-synthesized faces are indistinguishable from real faces and more trustworthy“ von Sophie J. Nightingale und Hany Farid am 14.02.2022 <*https://www.pnas.org/doi/10.1073/pnas.2120481119*>

[4] ZDF, „Giffey fällt auf "Fake-Klitschko" herein“ von Julia Klaus am 24.06.2022 <*https://www.zdf.de/nachrichten/panorama/giffey-deepfake-falscher-klitschko-ukraine-krieg-100.html*>

[5] Tagesschau, „Was gegen ein Deepfake spricht“ von Daniel Laufer am 28.06.2022 <*https://www.tagesschau.de/investigativ/rbb/deep-fake-klitschko-101.html*>

[6] Wikipedia, „Adobe Voco“ https://de.wikipedia.org/wiki/Adobe_Voco›

[7] CNN Business, „How a deepfake Tom Cruise on TikTok turned into a very real AI company“ von Rachel Metz am 06.08.2021 <*https://edition.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company/index.html*>

Quellen:

[8] Yahoo, „So realistisch sieht die erste KI-Moderatorin aus“ von Antonia Wallner am 19.11.2020 *<https://de.style.yahoo.com/so-realistisch-sieht-die-erste-ki-moderatorin-aus-114021224.html>*

[9] Synthesia Präsentationsvideos auf AI-Basis *<https://www.synthesia.io/>*

[10] CBR.com, „Bruce Willis Sells Deepfake Likeness Rights So His 'Twin' Can Star in Future Movies“ von Narayan Liu um 29.09.2022 *<https://www.cbr.com/bruce-willis-sells-deepfake-likeness/>*

[11] Wikipedia, „Brutkastenlüge“ *<https://de.wikipedia.org/wiki/Brutkastenl%C3%Bcge>*

eingesetzt werden wie Anti-Virus-Filter, die uns bei erfolgreicher Detektion sofort darauf hinweisen, dass dieses Video oder jenes Foto zu soundsoviel Prozent ein möglicher Fake ist. Da die Entwicklung der Deep Fakes aber immer weiter voranschreitet, werden auch genau jene Hinweise, die momentan noch auf einen Fake hinweisen, vom Computer in Zukunft auch bedacht werden. Es zeichnet sich also ein Katz-und-Maus-Spiel zwischen Fälschern und Aufdeckern ab.

Wird es uns dann noch Spaß bereiten, Medieninhalte zu nutzen, wenn überall der Verdacht lauert, es könne sich um eine Fälschung handeln? Wenn uns ständig ein Pop-Up darauf hinweist, dass auf der besuchten Seite ein Fake-Profil oder ein Fake-Video abgebildet ist? Die jetzt schon immer weiter um sich greifenden Hinweise und Warnmeldungen (z.B. über sog „medizinische Fehlinformationen“ im Zusammenhang mit der Corona-Krise und „Impfung“ genannte, experimentelle Gentherapien) würden um weitere Warnmeldungen ergänzt werden.

Es wird auch viel leichter sein, diese Technologien zu nutzen um Finanzbetrug, Datenschutzverletzungen oder Phishing-Betrug zu begehen. Identitäten können digital gestohlen und Verbrechen im Namen unbescholtener Bürger durchgeführt werden. Schreiten die Bemühungen etlicher Staaten voran, Bargeld zu verbieten und werden Bürger zur Nutzung digitaler Währungen gezwungen, wird der Taschendiebstahl zwar aussterben, aber durch Identitätsdiebstahl ersetzt werden.

Die Versprechen, die uns seitens der Politik gemacht werden, dass mehr (intelligente) Technologien eine gerechtere Welt für uns erschafft und Kriminalität weiter zurückgedrängt würden, sind nichts weiter als Deep Fake.

Publiziert auf Free21 am 28.12.2022, 6:00 Uhr