Deepfakes

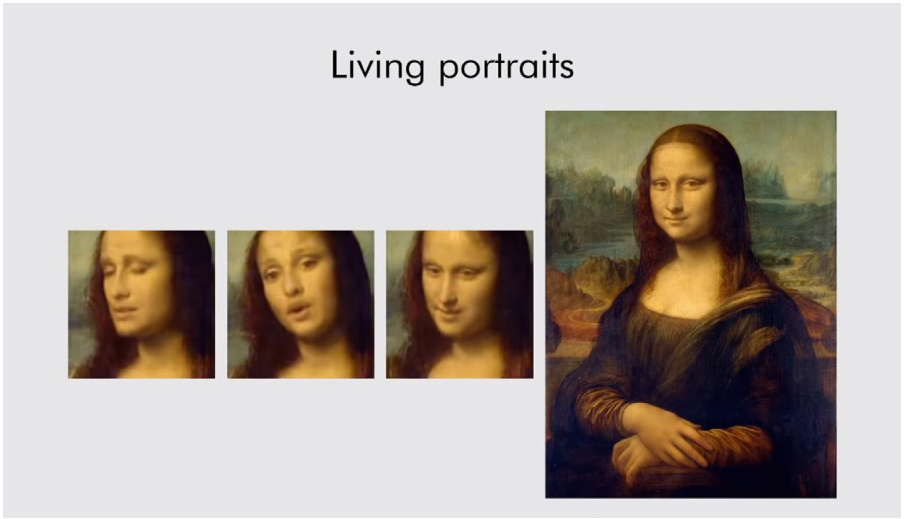
WENN DIE KI DAS MENSCHLICHE AUGE BETRÜGT

Deepfakes dominieren derzeit die Diskussion um die Entwicklung künstlicher Intelligenzen. Wir sprechen deshalb im Detail über Chancen und Gefahren der Technologie. Von Sara Petzold

Wir schreiben das Jahr 2029. Skynet, eine intelligente Maschine, hat nach einem Atomkrieg die Kontrolle über die Erde übernommen und die Menschheit unterjocht. Das Szenario von James Camerons »Terminator« von 1984 ging in die Filmgeschichte ein, steht es doch sinnbildlich für die Angst der Menschen, dass von uns entwickelte Roboter sich irgendwann gegen ihre Schöpfer auflehnen und uns versklaven könnten. Heutzutage sind wir allerdings noch weit von derartiger Science-Fiction entfernt – künstliche Intelligenzen besitzen stark eingeschränkte Fähigkeiten, und die ihnen zugrunde liegenden Algorithmen begrenzen sich auf vordefinierte einzelne Aufgaben. Trotzdem hat die KI-Entwicklung während der vergangenen Jahre derartig große Fortschritte gemacht, dass sich Experten immer wieder dazu veranlasst sehen, vor dem Potenzial bereits existierender neuronaler Netzwerke und Algorithmen zu warnen. Wir werfen deshalb einen Blick auf das interessante Thema der künstlichen Intelligenz. Wir erklären, wozu KI bereits jetzt in der Lage ist, welche potenziellen Gefahren von neuronalen Netzwerken wie Deepfakes ausgehen, und welche Chancen sich für die Zukunft der Spieleentwicklung durch künstliche Intelligenzen ergeben könnten.

Was KI derzeit kann – und was nicht

Als Gari Kasparow am 11. Mai 1997 seinen Läufer von c1 nach f4 zog, beging er einen der größten Fehler seiner Karriere als Profi-Schachspieler. Die Partie endete kurze Zeit später mit dem Sieg des Gegners über den damaligen Schachweltmeister und markierte einen Meilenstein in der Entwicklungsgeschichte künstlicher Intelligenzen. Kasparows Gegner: der IBM-Computer Deep Blue, der als erste Maschine überhaupt einen Schachweltmeister in dem komplexen Brettspiel schlagen konnte. Obwohl Deep Blue strenggenommen keine echte künstliche Intelligenz darstellte – schließlich handelte es sich bei dem Computer nicht um ein lernfähiges System im Sinne des Machine Learning –, gilt er bis heute als Meilenstein der KI-Entwicklung. Nachdem mehr als 20 Jahre seit dem Wettkampf zwischen Deep Blue und Kasparow vergangen sind, hat sich das Machine Learning aber eine zentrale Rolle beim Entwicklungsprozess künstlicher Intelligenzen erkämpft. Computer sind mittlerweile in der Lage, mithilfe entsprechender Lernprozesse menschliche Profis in Spielen wie League of Legends oder Starcraft 2 zu schlagen. Derartige Algorithmen bringen uns zum Staunen, erscheinen mitunter beängstigend – und trotzdem verfügen sie nur über sehr eingeschränkte Fähigkeiten, sodass sie



Samsungs Deepfakes-KI kann auch alternative Mona-Lisa-Mimiken erschaffen.

die Bezeichnung künstliche Intelligenz strenggenommen gar nicht verdienen.

Der Grund dafür, dass künstliche Intelligenzen mittlerweile für außenstehende Beobachter teils beeindruckende Dinge vollbringen können, gleichzeitig aber einer echten Intelligenz im eigentlichen Wortsinn längst noch nicht nahekommen, hängt mit der Technik des Machine Learning zusammen. Machine Learning funktioniert, indem wir einem Algorithmus jede Menge Daten eines bestimmten Typs zuführen – zum Beispiel Fotos von Pizza. Die KI analysiert dann diese Daten und bringt sich selbst bei, welche Bedingungen ein bestimmter Datensatz erfüllen muss, damit er den Regeln der zuvor eingespeisten Daten entspricht.

Füttern wir einen Algorithmus also mit Fotos von Pizza, lernt er, welche Bestandteile zu einer Pizza gehören und kann dann eigene Fotos von Pizzen generieren beziehungsweise Pizza-Bilder von solchen ohne Pizza unterscheiden. Der Vorteil dabei: Weil der Algorithmus sich die notwendigen Regeln selbst beibringt, müssen wir sie nicht alle einzeln programmieren.

KI mit Vorurteilen

Allerdings hat die Sache auch einen Haken: Weil die KI ihr Wissen nur auf den Daten aufbaut, die wir ihr zuführen, besteht die Gefahr, dass sie Vorurteile entwickelt oder sich nur für sehr eingeschränkte Tätigkeiten eignet. Deshalb sind derartige auf Machine Learning basierende KIs auch noch weit entfernt von den Fähigkeiten, die etwa Skynet in »Terminator« besitzt. Trotzdem übertreffen trainierte Algorithmen die menschlichen Denkfähigkeiten bereits heute schon deutlich, auch wenn sie nur auf einzelne Aufgaben beschränkt bleiben.

Experten diskutieren bereits seit Längerem über die Gefahren und Möglichkeiten, die sich aus lernfähigen Algorithmen und künstlichen Intelligenzen ergeben. Die Frage, ob wir bis zum Jahr 2100 eine Super-KI entwickelt haben werden, die das menschliche Denkvermögen deutlich übertrifft, können sie zwar nicht beantworten. Fest steht aber, dass künstliche Intelligenz als solche immer stärker Einzug in unseren Alltag hält und über die kommenden Jahre zu einem selbstverständlichen Bestandteil unseres Lebens avancieren dürfte – egal, ob es sich dabei um Schach- und Gaming-KIs, um Gesichtserkennung und Smart Homes oder um selbstfahrende Autos handelt. Deshalb fordern Wissenschaftler bereits jetzt, dass wir mit KIs und ihrer Entwicklung verantwortungsvoll umgehen und uns der potenziellen Gefahren bewusst werden, die lernfähige Algorithmen schon heutzutage mit sich bringen. Viele KI-Forscher und Programmierer bemühen sich bereits darum, mit eigenen Entwicklungen und Projekten über die Chancen und Probleme von künstlichen Intelligenzen aufzuklären. Die sogenannten Deepfakes-KIs liegen dabei im Fokus ihrer Aufmerksamkeit.

Zehn Prozent der regulären Arbeitszeit investieren einige Mitarbeiter von TNG Technology Consulting in das Programmieren einer Deepfake-KI. TNG macht sonst eigentlich in Firmen-IT.

Die Deepfakes-2.0-KI funktioniert in Echtzeit – allerdings sind die Ergebnisse bei Weitem noch nicht perfekt. (Bildquelle: Jonas Mayer/TNG Consulting)

Errechnete Regeln

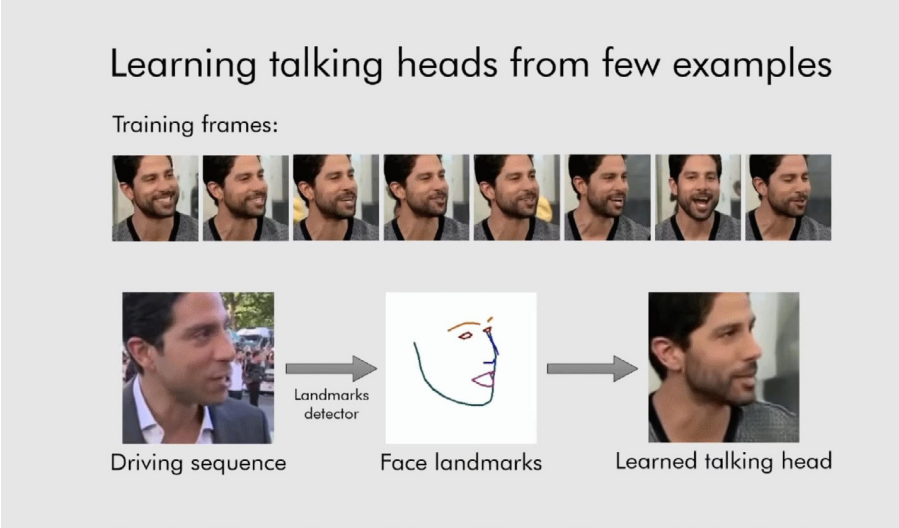
Deepfakes machen sich das maschinelle Lernen als Technik des sogenannten Deep Learning zunutze. Dabei greift der autoadaptive, also sich selbst ohne äußere Einwirkung anpassende Algorithmus auf eine Hierarchie von Konzepten zurück, die er in einem künstlichen neuronalen Netz verarbeitet. Dieses Netz besteht aus mehreren Schichten, wobei die sogenannte Eingabeschicht die von dem menschlichen Programmierer eingegebenen Daten sammelt und an zusätzliche, versteckte Schichten weitergibt. Die Schichten abstrahieren dann nach und nach die Daten, bis das neuronale Netz die Daten bestimmten Merkmalen zugeordnet und sie kategorisiert an die letzte, wiederum sichtbare Schicht ausgegeben hat. Der Algorithmus berechnet also anhand der eingegebenen Daten die Regeln, die wir ohne Deep Learning selbst mühsam programmieren müssten – eine enorme Erleichterung.

Die Geißel des Internetzeitalters?

Die sogenannten Deepfakes machen sich genau dieses Prinzip zunutze und erzeugen mithilfe selbstlernender Algorithmen gefälschte Bilder und Videos, die sich mit bloßem Auge oft gar nicht oder nur schwer von echten Aufnahmen unterscheiden lassen. Der Begriff Deepfakes kursiert seit 2017 und geht auf einen reddit-Nutzer namens deepfakes zurück, der damals die Gesichter bekannter Persönlichkeiten wie Emma Watson und Kate Perry in pornografische Videos verpflanzte. Das sogenannte »Synthesizing Obama«-Programm erschien ebenfalls 2017 und zeigte modifizierte Aufnahmen des ehemaligen US-Präsidenten Barack Obama, für die Forscher ein neuronales Netzwerk verwendeten, um die Form und die Bewegungen von Obamas Mund nachzumodellieren – mit beeindruckendem Ergebnis. Allerdings finden viele Experten die Deepfakes-Technologie nicht nur lustig. Denn das Potenzial einer derartigen KI lässt sich leicht missbrauchen, wie das Programm FakeApp beweist. Nutzer können mithilfe dieser Software verhältnismäßig einfach gefälschte Videos erstellen – entsprechend große Mengen an Daten in Form von Bildmaterial der betroffenen Person vorausgesetzt. Die zunehmende Beliebtheit der App, die sich das Google-Tool TensorFlow zunutze macht, resultierte

2018 in einer Welle aus Deepfakes-Pornos mit Prominenten und auch Privatpersonen – bei Letzteren speziell als Rachepornos.

Die Gefahren von Deepfakes liegen dementsprechend auf der Hand: Wenn wir mit bloßem Auge nicht (mehr) erkennen können, ob das Video einer prominenten Person oder eines Politikers echt oder gefälscht ist, sind Fake-News Tür und Tor geöffnet.



Deepfakes-KIs wie die von Samsung lernen anhand Hunderter Stunden an Videoaufnahmen, wie sie die Gesichtsbewegungen einer Person imitieren können.

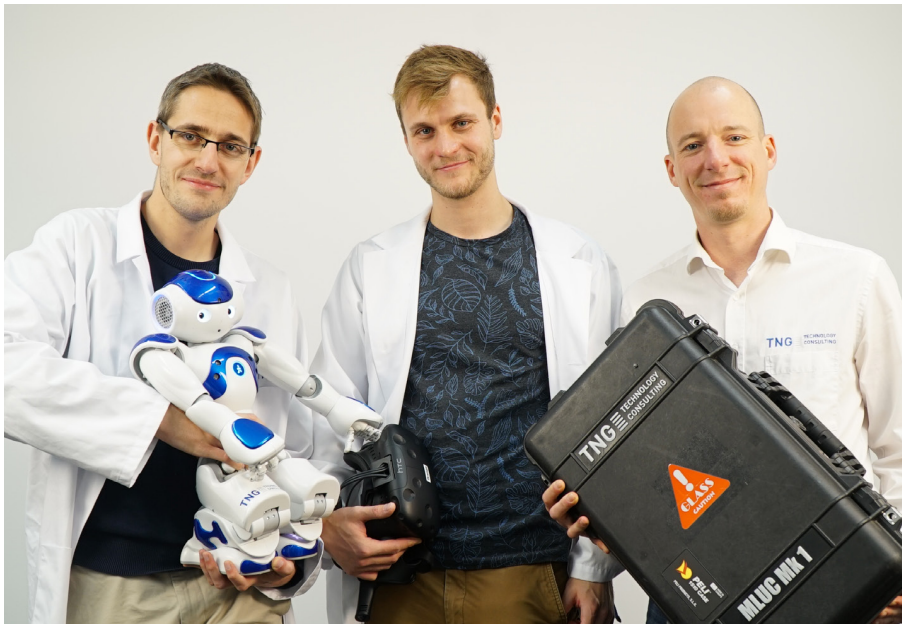
Deepfakes-KI aus Deutschland

Um auf die Ambivalenz der Chancen und Probleme von Deepfakes aufmerksam zu machen, bemühen sich Forscher und Experten zunehmend um Aufklärung und die Entwicklung von Programmen, die Deepfakes entlarven können. Das »Innovation Hacking Team« des deutschen Unternehmens TNG Technology Consulting hat sich unter anderem dieser Aufklärung verschrieben. GameStar hat das Team besucht und das Ergebnis ihrer bisherigen Arbeit angeschaut. Das Team, das aus den beiden Gründern Thomas Endres und Martin Förtsch, dem Software Consultant und Hacker Jonas Mayer sowie wechselnden Mitarbeitern von TNG besteht, nutzt zehn Prozent der regulären Arbeitszeit für eigene Projekte, darunter auch Deepfakes 2.0. Bei Deepfakes 2.0 handelt es sich um eine Deepfakes-KI, die in Echtzeit die Gesichter von Personen durch Deepfakes ersetzen kann. Der aktuelle Prototyp des Projekts nimmt Personen live über eine integrierte Kamera auf und berechnet direkt die entsprechenden Deepfakes. Der Bildschirm des Prototyps gibt anschließend in Echtzeit die Aufnahmen mit den angepassten Köpfen aus – mit beeindruckendem Ergebnis. Denn Deepfakes 2.0 beschränkt sich nicht nur auf das Gesicht der gefilmten Person, sondern ersetzt den gesamten Kopf durch den einer anderen Person. Dabei erfolgt die Gesichtserkennung über ein neuronales Netzwerk, das auf Open CV basiert. Nach der Kopf-Segmentierung entfernt die KI dann potenzielle Makel, indem sie das ursprüngliche Gesicht vor dem vorhandenen Hintergrund entfernt und dadurch Überlappungen mit dem neuen Gesicht vermeidet.

Die Deepfakes-2.0-KI erkennt in Echtzeit das Gesicht einer Person und ersetzt es direkt durch das Gesicht einer anderen. (Bildquelle: Jonas Mayer/TNG Consulting)

Deepfakes 2.0 ist nur der Anfang

Das Ergebnis ist allerdings noch nicht ganz perfekt – hier und da hat die KI noch vereinzelt Probleme mit ungewöhnlichen Kopfformen oder fehlender Kopfbehaarung. Das Entwicklerteam hat GameStar gegenüber dazu erklärt, dass die von der Deepfakes-2.0-KI erzeugten Deepfakes durch »eine niedrige Detailtiefe der Gesichtshaut oder aber auch durch Artefakte um den Kopf als solche noch schnell identifizierbar« seien. Trotzdem können sich die Aufnahmen durchaus sehen lassen und zeigen, was jetzt schon mit Echtzeit-Deepfakes möglich ist beziehungsweise sein wird – insbesondere, weil sich die noch bestehenden Probleme mithilfe »bestimmter Faktoren wie hochauflösendem Trainingsmaterial und einer längeren Trainingsphase mit mehr Rechenleistung wesentlich verbessern« lassen. Besonders beeindruckend: Das Innovation Hacking Team von TNG hat die Deepfakes-2.0-KI in ein paar Monaten quasi nebenbei entwickelt. Umso bemerkenswerter fällt die Qualität der Deepfakes aus, die jetzt schon erschreckend real wirken. Ebenfalls erstaunlich ist die Tatsache, dass das Team den ak-



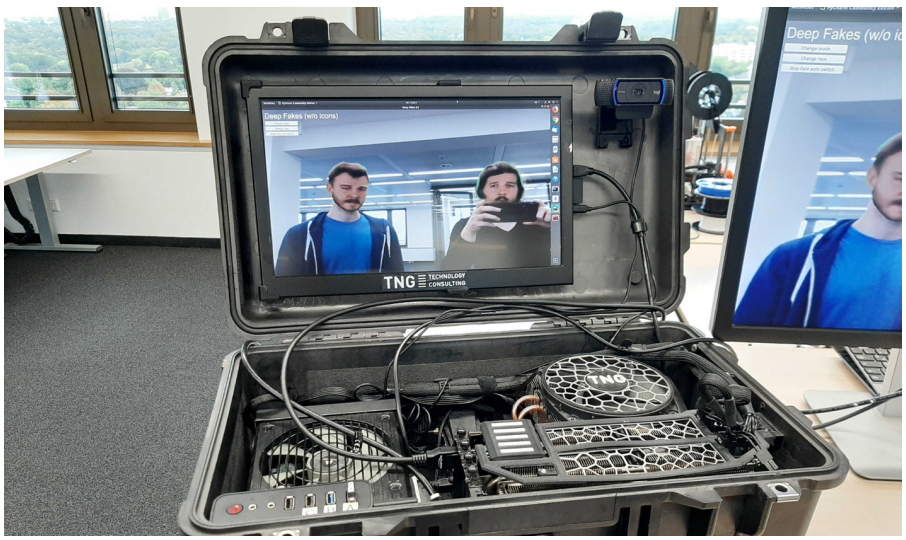
Das Innovation Hacking Team hat eine KI entwickelt, die Gesichter in Echtzeit ersetzt.

tuellen Prototypen der Deepfakes-2.0-KI mit Gaming-Hardware steuert: Das System besteht aus zwei Geforce-1080-Ti-Grafikkarten, einem Intel-Core-i7-8700K-Prozessor auf einem Micro-ATX-Board sowie einem 1000-Watt-Netzteil.

Gefahren aufzeigen

Aber das Team rund um Endres, Förtsch und Mayer hat Deepfakes 2.0 nicht entwickelt, damit sie Deepfakes um der Deepfakes willen erzeugen und einem aktuellen Hacking-Trend hinterherlaufen können. Stattdessen haben sich die Entwickler das Ziel gesetzt, mit ihrer Arbeit auf das Potenzial und die Gefahren von Deepfakes aufmerksam zu machen und aufzuklären. Dabei warnt das Team unter anderem vor Fake News und gefälschter Pornografie: »Deepfakes können in verschiedenen Bereichen missbräuchlich verwendet werden. [...] Man kann Menschen Aussagen in den Mund legen, die sie nie gesagt haben, und Schlimmeres. Um solche gefälschten Inhalte herum könnten gar ganze Verschwörungstheorien ihren Anfang nehmen oder eine diplomatische Krise ausgelöst werden. Darüber hinaus werden Deepfakes bereits heute – wenn auch hauptsächlich auf humoristische Art und Weise – verwendet, um Schauspielern ein anderes Gesicht aufzusetzen. Denkt man diese Szenarien weiter, so könnten in der Zukunft Schauspieler auch vollständig digitalisiert und nur noch als ›Gesichtsmodell‹ für Filme verwendet werden. [...] Eine bereits heute sehr reale Gefahr durch Deepfakes ergibt sich in deren Verwendung zur Fälschung von Pornografie. Unbescholtene Menschen könnten in Verruf gebracht werden, wenn ihr Gesicht plötzlich in einem Pornofilm zu sehen ist. Diese Filme entstehen heute bereits, wobei der Fokus derzeit noch auf Berühmtheiten liegt.«

Das Innovation Hacking Team von TNG verweist in diesem Kontext aber auch zu Recht darauf, dass sich im Fall von Deepfakes das Feuer am besten mit Feuer bekämpfen lässt: Denn während Menschen mit bloßem Auge die Deepfakes teilweise heute schon nicht mehr von echten Aufnahmen unterscheiden können, finden Deepfakes-Algorithmen die Spuren anderer Algorithmen, die diese in Deepfakes-Aufnahmen hinterlassen – und entlarven dadurch die Fälschungen.



Erstaunlich: In diesem Köfferchen verbirgt sich die Technik von Deepfakes 2.0.

Deepfakes und die Zukunft des Gaming

So gefährlich Deepfakes auf den ersten Blick wirken mögen, so viel Potenzial besitzen sie, insbesondere auch für unser Lieblingshobby Videospiele. Auch wenn wir noch ein ganzes Stück davon entfernt sein dürften, dass die Deepfakes-Technologie bei großen Spieleentwicklern zum Standardrepertoire gehört, fällt es nicht schwer, sich Szenarien vorzustellen, in denen Deepfakes für Videospiele und die Gaming-Community eine Rolle spielen könnten.

Nehmen wir etwa das Deepfakes-2.0-Projekt von TNG, das in Echtzeit manipuliert: Eine verbesserte Version dieses Prototyps könnte etwa in Live-Streams auf Twitch zum Einsatz kommen. Streamer hätten die Möglichkeit, ihr eigenes Gesicht live zu verändern und durch eine bekannte Persönlichkeit oder eine Figur aus dem Videospiel zu ersetzen, das sie gerade spielen. Zugegeben, das wäre wohl kaum mehr als eine nette Spielerei. Aber Deepfakes bergen für die Gaming-Branche noch deutlich mehr Potenzial. Stellen wir uns einmal vor, wir könnten mithilfe von Deepfakes unser eigenes Gesicht auf den Charakter eines Spiels übertragen. Statt mehrere Stunden mit einem komplexen Charaktereditor zu verbringen, um dann doch nicht das Ergebnis zu bekommen, das wir uns wünschen, bekämen wir mit einem einfachen Knopfdruck ein echtes virtuelles Alter Ego. Zwar enthalten einzelne Titel wie NBA 2K20 dieses Feature bereits, die Ergebnisse sind jedoch ausbaufähig. Derweil bietet Star Citizen mit »Face over IP« bereits die Möglichkeit, die live von der Webcam aufgenommenen Gesichtszüge des Spielers auf seinen eigenen Avatar zu übertragen. Wenn ihr während des Spiels lächelt, grinst auch euer Alter Ego.

Der Grad an Immersion, den derartige Spielelemente mit sich brächten, wäre immens – denn welcher Gaming-Fan hat nicht schon immer davon geträumt, einen Charakter durch Fallout oder Die Sims zu steuern, der exakt so aussieht wie er selbst? Oder in Sportspielen wie FIFA einen Avatar zum Sieg zu verhelfen, der einem selbst wie aus dem Gesicht geschnitten ist – im wahrsten Sinne des Wortes? Insbesondere Rollenspiele, die vergleichsweise stark personalisierte Erfahrungen ermöglichen, würden sehr von dieser Technik profitieren.

Ob Deepfakes irgendwann Einzug in die Welt der Videospiele halten werden, können und wollen wir an dieser Stelle nicht beantworten. Auch die Klärung der Frage, ob wir uns langfristig der Gefahren erwehren können, die von Deepfakes an sich ausgehen, steht uns nicht zu. Fest steht aber, dass ein verantwortungsvoller Umgang mit künstlichen Intelligenzen letztendlich darüber bestimmen wird, wie das Deepfakes-Szenario der Zukunft tatsächlich aussieht. ★

Die Website wurde nach rechtlichen Schritten von Peterson deaktiviert.