Die manipulative Macht der künstlichen Intelligenz

Angela Merkel und Barack Obama am Strand bei einem Eis, dann die Verhaftung von Donald Trump. Was real aussieht, ist in Wahrheit vom Computer gemacht und birgt Gefahren.

Dass man auf solche manipulativen Methoden reinfallen kann, bewies Wiens Bürgermeister Michael Ludwig, als er im August 2022 mit einer gefälschten Version von Vitali Klitschko, Bürgermeister von Kiew, telefonierte.

Ex-Kanzler Sebastian Kurz sang mittels künstlicher Intelligenz (KI) beim Eurovision Songcontest das Lied „Nur ein Lied“. Dass es sich um gefälschtes Videomaterial handelt, konnte man hier schnell erahnen. Anders schaut es mit den Fotos aus, die von prominenteren Politikern stammen. Denn: Je mehr Bildmaterial zur Verfügung steht, desto besser kann die künstliche Intelligenz jedes Szenario erstellen.

Trump-Fans waren schockiert, als sie sahen, wie ihr Idol festgenommen wurde. Dabei erlaubte sich Blogger Eliot Higgins nur einen Scherz. Als Ex-US-Präsident Trump schrieb, dass er seine Verhaftung erwarte, erstellte der IT-Experte Higgins mit dem Programm Midjourney fiktive Fotos. Auf Twitter gingen die Aufnahmen viral. Auch wenn Higgins sie als gefälscht kennzeichnete, glaubten viele an die Echtheit.

Visuelles wird intuitiv als real wahrgenommen

„Das liegt daran, dass Menschen visuelle Information zunächst intuitiv als wahr und real wahrnehmen. In einem ersten Schritt können die Menschen mit gut gemachten Fake-Bildern oder Deepfakes also relativ leicht hinters Licht geführt werden“, erklärt Jörg Matthes, Medienpsychologe an der Universität Wien.

Foto: Stadt Wien


Foto: YouTube.com/Kurt Fleisch

Ludwig sprach mit einem falschen Klitschko. Sebastian Kurz als Sänger.

Diese Aktion verdeutlicht auch das Fehlen von Unternehmensstandards oder staatlichen Vorschriften, die sich mit dem Einsatz von künstlicher Intelligenz zur Schaffung und Verbreitung von Unwahrheiten befassen. „Allgemein hinkt die Politik in diesem Bereich nach. Das hat man schon gemerkt, als das Internet eingeführt wurde. Damals war die Devise: Mal schauen, was passiert. Man macht sich erst Gedanken, wenn gewisse Sachen schon eingetreten sind. Es ist auch ein schwieriges Thema, weil auch Politiker Social Media für ihre Kampagnen nützen. Es braucht unbedingt eine Regulierung. Die KI ist aus mathematischer Sicht eigentlich relativ banal, das Heikle sind die Daten und das damit verbundene Urheberrecht“, erklärt Thomas Pock, Professor für Bildverarbeitung an der TU Graz. Konkret geht es um die zahlreichen Fotos, die Social-Media-Nutzer freiwillig auf Instagram und Co. hochladen. Konzerne wie Meta und Alphabet haben dadurch unendlich viel Bildmaterial, auf das die KI zurückgreift.

Dominik Erlinger, Katharina Pirker

Foto: Julian AI Art


Foto: EliotHiggins

Nicht real: Angela Merkel und Barack Obama bei einem Eis. Festnahme von Trump.

> *Wenn die Menschen realisieren, dass sie mit Fake-Bildern getäuscht wurden, dann erzeugt dies Unsicherheit. Studien belegen, dass diese Unsicherheit sich negativ auf das Medienvertrauen auswirkt.*
>
> Jörg Matthes, Medienpsychologe an der Universität Wien


Foto: Barbara Mair

Wie KI Bilder fälscht, und man es erkennt

Noch arbeiten KI-Tools nicht fehlerfrei, aber das Missbrauchspotenzial ist groß.


Foto: Midjourney

Udo Landbauer und Johanna Mikl-Leitner lächeln einander an: Dieses Fake-Foto hat die KI Midjourney erzeugt.

Zwei Fotos genügen für so etwas nicht“, sagt IT-Sicherheitsexperte Martin Eiszner von SEC Consult. Um Gesichter in Fotos und Videos auszutauschen, benötigen KI-Tools eine große Menge an Trainingsdaten – Videomaterial oder Fotos aus dem Internet. Bei Politikern und Filmstars ist es leicht zu beschaffen, eine Instagram-Seite reicht nicht.

KI-Werkzeuge für diese „Deepfakes“ arbeiten mit zwei Datensätzen: den Trainingsdaten und einer Zieldatei, in die das neue Gesicht eingebaut wird. Algorithmen analysieren Blickrichtung und Perspektive und errechnen, wie die Zieldatei manipuliert werden muss, um die Person aus den Trainingsdaten einzubauen. Dabei werden sogar die Mund- und Augenbewegungen automatisch angepasst.


Fotos: Midjourney

Rendi-Wagner und Doskozil fälscht die KI wenig überzeugend: Das passiert, wenn nicht genug Trainingsdaten vorliegen

Neben Propaganda droht Missbrauch durch Cyberkriminelle: Die Tools sind gratis verfügbar und laufen auf jedem Privat-PC.

Weil ein Deepfake viel Rechenleistung benötigt und Einzelbild für Einzelbild vorberechnet werden muss, ist er aber noch nicht für Live-Videoanrufe geeignet. Betrügerische Sprachanrufe mit falscher Stimme funktionieren hingegen schon. „Deepvoice-Fakes hat man in Asien bereits beobachtet“, berichtet Eiszner. Fake-Videos sind für geschulte Beobachter derzeit noch gut erkennbar: Achten Sie auf Tonaussetzer, ungewöhnliche Schatten oder Hauttöne, seltsame Mund-, Stirn- oder Augenbewegungen und auch auf unnatürliche Elemente im Hintergrund.