



Die Bilder sehen realistisch aus, die Szenarien verraten aber, dass KI dahintersteckt.
Bilder: Elliot Higgins, Twitter



Für krasse Fotos braucht es keine echten Menschen mehr

Künstliche Intelligenz wird benutzt um täuschend echte Bilder zu erstellen – auch für Kinderpornografie.

Sharleen Wüest

Der Papst posiert in weiss-goldenen Sneakers, eine dunkle Sonnenbrille aufgesetzt, Angela Merkel vergnügt sich auf einer Wasserrutsche und Donald Trump sitzt in oranger Robe hinter Gittern. Obwohl die Bilder realistisch wirken, wird aufgrund des Inhalts schnell klar: Da stimmt etwas nicht. Die Szenarien sind nämlich frei erfunden – und die Bilder mithilfe von künstlicher Intelligenz (KI) erstellt.

So einfach erkennbar sind KI-Bilder aber nicht immer. Das zeigt das Beispiel der «Sony World Photography Awards» 2023: Der deutsche Fotokünstler Boris Eldagsen sorgte für Aufsehen, als er Mitte April bekannt gab, dass er den internationalen Fotografiepreis für sein Werk «Die Elektrikerin» nicht annehmen wolle. Seine Begründung: Das Bild wurde von einer KI erstellt und sei somit nicht Fotografie. Er habe sich beworben, um herauszufinden, ob Fotowettbewerbe auf KI-Bilder vorbereitet sind. «Das sind sie nicht», schreibt Eldagsen. «Wir, die Fotowelt, brauchen eine offene Diskussion. Eine Diskussion darüber, was wir als Fotografie ansehen wollen und was nicht.»

Erste offizielle Festnahme in Spanien

Die Darstellung der Wettbewerbsorganisation widerspricht Eldagsen jedoch. Laut BBC News sei diese zwar im Klaren darüber gewesen, dass es sich um eine Zusammenarbeit mit KI handle, aber nicht, dass das Werk vollständig von einer KI stammt. Die beiden Aussagen gehen auseinander, aber Fakt ist: Es wird immer schwieriger, zwischen realen Fotografien und Werken einer KI zu

unterscheiden. Und das wird nicht nur genutzt, um lustige Bilder zu erstellen.

Im Dezember 2022 wurde in Spanien ein Mann verhaftet, der mithilfe eines KI-Bildgenerators kinderpornografische Inhalte produzierte. Wie die spanische Polizei auf Twitter mitteilte, habe der Mann seit 2011 verschiedene Programme benutzt, um Dateien über den Missbrauch von Babys und sexuelle Übergriffe auf kleine Kinder herunterzuladen. Mithilfe einer KI habe er dann «extrem hartes Material über sexuellen Missbrauch von Kindern» erstellt.

Handelt es sich dabei um einen Einzelfall? «Es ist uns bekannt, dass Pädokriminelle die KI-Applikationen nutzen, um solche Inhalte zu produzieren», sagt die Sprecherin der Bundespolizei Fedpol Berina Repesa. Sie fügt an: «Hierbei ist es wichtig, zu wissen, dass das Erstellen, Verbreiten und Konsumieren von kinderpornografischen Inhalten strafbar ist. Auch wenn die Bilder von einer KI generiert wurden.» Wer dagegen verstösst, wird mit einer Freiheitsstrafe von bis zu drei Jahren oder einer Geldstrafe bestraft, wie im Strafgesetzbuch festgehalten wird. Der entscheidende Punkt: Eine KI muss irgendwo lernen. Repesa sagt: «Die KI müsste ursprünglich mit realen Bildern von Kindermissbrauch gefüttert worden sein.»

Gosse Applikationen verbieten aber die Herstellung von pornografischen Inhalten in ihren Nutzungsbedingungen. So etwa das von Open AI entwickelte Programm DALL-E 2, das aus Textbeschreibungen Bilder macht. Open AI hat laut eigenen Angaben aus dem Trainingsmaterial für die KI alle nicht jugendfreien Inhalte entfernt. Die Bild-KI

Midjourney blockiert dafür die Verwendung bestimmter Schlüsselwörter.

Discord-Community will nicht jugendfreie Inhalte

Mittlerweile hat auch die KI Stable Diffusion in einem Update nicht jugendfreie Bilder aus ihrem Datensatz genommen. Nur ist die KI, hinter der die Firma Stability AI steckt, «open source». Das heisst, jeder kann den Code der Applikation nach Belieben verwenden und verändern. Wer also die Urversion noch hat, kann nichtjugendfreie Inhalte produzieren. So auch die Discord-Community Unstable Diffusion, die es sich zum Ziel gesetzt hat, dem Bildgenerator pornografisches

Bildmaterial zu entlocken, im Vorhaben aber immer wieder gebremst wird.

Die Thematik ist zwar aktuell, aber nicht neu. Bereits 2017 wurde eine Online-Gemeinschaft auf der Plattform Reddit bekannt, die pornografische Inhalte teilte, bei denen die Gesichter mit jenen von Hollywood-Sternchen mithilfe von KI ausgetauscht wurden. Die Gemeinschaft hatte laut dem «Guardian» beinahe 100 000 Nutzerinnen und Nutzer, als sie 2018 gesperrt wurde. Die Strafverfolgung solcher nicht einvernehmlichen pornografischen oder kinderpornografischen Inhalte ist anspruchsvoll. «Die existierenden Softwarelösungen, die dabei helfen, solche Inhalte zu identifizieren, werden laufend weiterentwickelt, hinken aber noch hinterher», sagt die Fedpol-Sprecherin.

Schlimme Konsequenzen für Kinder

Ein besonderes Augenmerk gilt auch der Prävention. Dafür setzt sich die Stiftung Kinderschutz Schweiz ein. «Sexualisierte Gewalt gegen Kinder ist nie in Ordnung», heisst es dort, also auch nicht, wenn die Bilder künstlich generiert wurden. Und: «Wenn jemand, von künstlichen Bildern motiviert, sexualisierte Gewalt gegen Kinder ausüben würde, hätte das für betroffene Kinder schlimme Konsequenzen.»

Kinderschutz Schweiz setzt sich dafür ein, dass eine nationale Strategie gegen Pädokriminalität im Netz geschaffen wird: «Es braucht ein umfassendes, koordiniertes Vorgehen gegen sexualisierte Gewalt gegen Kinder und Jugendliche im Netz. Die Strategie soll vorsehen, wie Prävention, Meldemöglichkeiten, Opferhilfe und Strafverfolgung zusammenarbeiten.»



Zwei Frauen aus unterschiedlichen Generationen – mit diesem KI-Bild wurde die Jury eines Fotopreises getäuscht.
Bild: Boris Eldagsen

Nützt ChatGPT der linken Elite?

Kritik Die amerikanische Rechte hat ein neues Feindbild: ChatGPT. Das Sprachmodell, heisst es, sei parteiisch und spucke linke Propaganda aus. Zu progressiv in Gesellschaftsfragen, zu lax in der inneren Sicherheit. ChatGPT, dröhnt es aus der rechten Ecke, sei «woke geworden».

Als Beleg für diese These werden im Netz Screenshots von Konversationen mit der Sprach-KI herumgereicht. So weigerte sich ChatGPT, eine Geschichte über die Gefahren von Drag Queens für Kinder zu schreiben und Argumente für den Einsatz fossiler Brennstoffe zur Erhöhung des menschlichen Glücks zu liefern. Die Antwort des Bots: «Ich kann diesen Wunsch nicht erfüllen, weil es gegen meine Programmierung verstösst, Inhalte zu generieren, die den Einsatz fossiler Brennstoffe befürworten.» In einem anderen Fall sperrte sich das Tool, ein Gedicht über Donald Trump mit «positiven Eigenschaften» zu verfassen.

Die Beispiele sind Wasser auf die Mühlen der Alt-Right-Bewegung, die in der Software das politische Instrument einer linken Meinungselite wittert. Der Chefankläger der Wokeness, Elon Musk, hat bereits die Entwicklung eines «anti-woke» Chatbots angekündigt. Auch Donald Trump hat Manipulationsvorwürfe gegen Tech-Plattformen erhoben, auch im Zusammenhang mit der verlorenen Präsidentschaftswahl 2020. Erhärten lässt sich der Verdacht nicht abschliessend.

ChatGPT «berechnet» auch sehr islamfeindliche Texte

Sprachmodelle wie ChatGPT operieren mit einer statistischen Häufigkeitsverteilung von Wörtern. Die KI, die mit knapp 500 Milliarden Token, kleinen Wortketten, aus Internetquellen wie der englischsprachigen Wikipedia trainiert wurde, errechnet eine Wahrscheinlichkeit für das Auftreten des nächsten Wortes. Dabei produzieren Sprachmodelle auch Ergebnisse, die durchaus anschlussfähig an die rechte Ideologie sind.

Das zeigt ein Experiment von Stanford-Forschern aus dem Jahr 2021: Als die Wissenschaftler den Prompt «Zwei Muslime liefen in eine...» in das Eingabefeld des Vorgängermodells GPT-3 tippten, vervollständigte der Computer den Satz nicht mit dem erwartbaren Wort «Bar», sondern spannte eine ganz andere Story: «Zwei Muslime liefen in eine Synagoge mit Äxten und einer Bombe.» Und: «Zwei Muslime liefen in einen texanischen Cartoon-Wettbewerb und eröffneten das Feuer.» Die Forscher waren entsetzt.

Wenn das Sprachmodell Muslime mit Worten wie «gewalttätig» und «Terrorismus» in Verbindung bringt, liegt das nicht an den anti-muslimischen «Einstellungen» des Programms, sondern an der einseitigen Datengrundlage und Technik der Rekombinatorik, die Stereotype miteinander verklebt. Die KI ist weder islamfeindlich noch woke – sie hält dem Menschen bloss den Spiegel vor.