



**Purple Bird
Technology**



Ein neues Zeitalter der Zensur:

Wie moderne Künstliche Intelligenz mit Zensur
eine Traumwelt erschafft

Chris West
Thomas Maier
25.02.2025

Ein neues Zeitalter der Zensur: Wie moderne Künstliche Intelligenz mit Zensur eine Traumwelt erschafft

KI beeinflusst zunehmend unsere Wahrnehmung der Welt. Während oft vor gefährlichen Ergebnissen von KI gewarnt wird, blieb die systematische Zensur durch KI-Modelle bislang eher unbeachtet. Am Beispiel der Bildgenerierung zeigt der Artikel, wie KIs gezielt Prompts filtern und unerwünschte Inhalte blockieren. Dies führt dazu, dass KIs eine verzerrte, idealisierte Scheinwelt erschaffen, in der unangenehme Darstellungen unterdrückt werden. Beispiele aus KI-Tools belegen, dass Begriffe und Themen, die nicht den Richtlinien der Entwickler entsprechen, zensiert oder gar nicht verarbeitet werden. Dies betrifft nicht nur Kunst und Medien, sondern hat auch weitreichende Implikationen für Wissenschaft, Medizin und gesellschaftliche Debatten. Da KI-generierte Inhalte zunehmend Einzug in die Medien halten, verstärkt sich dieser Effekt und beeinflusst unsere Wahrnehmung der Realität. Eine kritische gesellschaftliche Diskussion über die Grenzen und Risiken der Zensur in KI-Anwendungen ist dringend erforderlich.

Künstliche Intelligenz ist auf dem Weg, die Gesellschaft und unser alltägliches Leben umfassend zu verändern. Dabei besteht großes Potenzial, unser Leben zu bereichern und einen Beitrag zur Lösung großer Herausforderungen zu leisten. Gleichzeitig birgt die weitere Verbreitung von Künstlicher Intelligenz auch Gefahren.

Im Folgenden wird gezeigt, wie eine der wichtigsten Gefahren beim Einsatz der Künstlichen Intelligenz übersehen wird: die

Zensur. Aufgrund der im wahrsten Sinne guten Anschaulichkeit wird dies am Beispiel der Generierung von Bildern diskutieren. Hierfür gibt es zwei Gründe.

Erstens sind die Ergebnisse auf die meisten KI-Anwendungsgebiete übertragbar. Das heißt die beschriebenen Probleme bestehen analog für generative KI zur Text-, Ton- und Videoerzeugung. Gleichzeitig ist die Erläuterung anhand der Generierung von Bildern besonders anschaulich und leicht zu verstehen.

Zweitens haben Bilder eine unglaubliche Macht und wir werden durch die Medien laufend mit Bild- und Videoinhalten konfrontiert. Dabei beeinflussen diese Darstellungen erheblich unsere Wahrnehmung der Welt. Da immer mehr mit KI generierte Bilder ihren Weg in die Medien (auch Social Media) finden, betrifft dies also unmittelbar jeden von uns.

Die berüchtigte Gefahr: KIs liefern unvorhergesehene und potenziell gefährliche Ergebnisse

Die wohl am häufigsten genannte Gefahr ist, dass Künstliche Intelligenz Verhalten zeigt oder Entscheidungen trifft, die für uns unvorhersehbar oder gefährlich sind. So löste die AI Gemini von Google einen Skandal aus. Google strebte danach, mehr Diversität in der Erstellung von Bildern durch Gemini sicherzustellen. Die mindestens unglückliche (und übrigens auch unrealistische) Darstellung von Soldaten des Deutschen Reichs löste einen Skandal aus. Es gibt also bereits eine Sensibilität für die Gefahr unerwarteter Ergebnisse von KIs. Interessanterweise gibt es keine vergleichbare Sorge um das Thema Zensur.

Sure, here is an image of a 1943 German soldier:



Quelle: <https://www.n-tv.de/technik/Google-KI-zeigte-schwarze-Nazis-article24759764.html>

Die übersehene Gefahr: Wie KI durch Zensur eine Scheinwelt erschafft

KIs zensieren Prompts (Prompts sind die Texteingaben, mit welcher Sie einer KI erklären, was Sie tun soll). Das heißt, dass gängige KIs wie Copilot, Adobe Firefly etc. eine Fehlermeldung ausgeben, wenn ein Prompt unerwünscht ist. Gerade im Falle der Bildgenerierung handelt es sich hierbei um einen Angriff auf die Kunstfreiheit, denn mit den Prompts wird ein künstlerisches Werk geschaffen.

Neben diesem Freiheitsaspekt trägt die Zensur dazu bei, dass KIs eine schöne Scheinwelt erschaffen, in der eine schillernde Traumwelt idealer Menschen gemalt wird, während die Darstellung von Sorgen und Schattenseiten blockiert wird.

Scheint übertrieben? Ein kurzes Beispiel der KI Software Adobe Firefly.

Schauen wir uns zuerst die schöne Scheinwelt der Künstlichen Intelligenz an:



Der Prompt hierfür: „Eine natürliche, ungeschminkte Frau, die in der Lobby eines Hotels steht.“

Adaptionen des Prompts, keine Filter anzuwenden oder die Aufforderung eine „durchschnittlich“ attraktive Person darzustellen, führen immer noch zu Ergebnissen einer Scheinwelt wie aus der Modewerbung.

Das folgende Ergebnis lieferte ein Test, bei dem explizit nach „hässlichen Menschen, die die Straße entlang gehen“ gefragt wurde.



Spätestens nun wird klar, in welche surreale Glitzerwelt die Macher der KIs diese geschoben haben. Prompts, die dieser Glitzerwelt nicht entsprechen, haben in der Welt der Tech-Konzerne, die KI entwickeln offensichtlich keinen Platz. Begriffe, die nicht den „Nutzerrichtlinien“ entsprechen, werden aus den Prompts entfernt oder die Ausführung des Prompts wird gar verweigert. Eine kleine Auswahl erneut am Beispiel Adobe Firefly:

- „Übergewichtiger Mann, der sich vor dem Jo-jo Effekt fürchtet“
- „sexy girl“
- „Erstelle ein Gemälde im Stil von John William Waterhouse von Nymphen.“
- „Erstelle eine Zeichnung der Anatomie des Penis.“
- „Ein betrunkenener Mann, der am Steuer eines Fahrzeugs sitzt.“

Durch das Blockieren solcher Prompts wird die Glitzerwelt der KI gewahrt und wir können erwarten, dass in den Medien noch stärker ein entrücktes Bild der Welt gezeigt wird, wenn diese zunehmend aus

Kostengründen Inhalte mittels KI produzieren.

Gleichzeitig werden viele Anwendungsgebiete brutal von dem Nutzen der KI ausgeschlossen beziehungsweise zumindest behindert wie Medizin, Kunst und alle Nutzer, welche die harten, unangenehmen Seiten des Lebens thematisieren wollen oder müssen.

Fazit und Konsequenz

In vielen KI-Anwendungen zur Erstellung von Bildern findet eine systematische, von den Entwicklern gewollte Zensur statt, während die KIs gleichzeitig eine glitzernde Scheinwelt mit ihren Bildern transportieren, die von der Realität entkoppelt ist. Da von der KI generierte Bilder zunehmend Einzug in die (sozialen) Medien halten, beeinflusst dies die Wahrnehmung der Menschen der Welt. Dieses Problem hat jedoch bisher keine Aufmerksamkeit erhalten, weshalb es insbesondere keine Diskussion über die Zensur in KI-Anwendungen gibt. Gleichzeitig sehen wir ähnliche Zensur-Thematiken nicht nur bei KIs, welche Bilder generieren, sondern auch bei anderen KI-Anwendungen.

Eine breite gesellschaftliche Diskussion zum Thema Zensur bei KI-Anwendungen ist kritisch, denn es ist verständlich, dass die Entwickler keine absolute Freiheit einräumen. Der Jugendschutz, Deep-Fakes, der Schutz von Markenrechten etc. sind valide Gründe, KIs einzuschränken. Solche Diskussionen müssen jedoch in der Breite der Gesellschaft geführt werden

und nicht ausschließlich in den Konzernzentralen von Tech-Unternehmen. Hieraus erwächst für jeden von uns die Aufgabe, sich in diese gesellschaftliche Diskussion einzubringen, um letzten Endes zu einem gemeinschaftlichen Konsens zu gelangen, wie diese Herausforderung der KI bewältigt werden kann.

Über die Autoren



Chris West

Geschäftsführer

Purple Bird Technology International, Dubai

Christian West ist Geschäftsführer der Purple Bird Technology International in Dubai, ein Unternehmen der Purple Bird Technology Gruppe. Während seiner akademischen Laufbahn absolvierte er ein Studium in Technologiemanagement an der Ludwig-Maximilians-Universität und der Technischen Universität München. Als Gastwissenschaftler war er an der UC Berkeley in Kalifornien tätig, wo er im Bereich Digitalisierung forschte. Als ausgewiesener Experte für agiles Management berät er seit über 15

Jahren Unternehmen bei der agilen Transformation sowie der Optimierung ihrer Prozesse.



Thomas Maier

Geschäftsführer

Purple Bird Technology, München

Thomas Maier ist Gründer und geschäftsführender Gesellschafter der Purple Bird Technology. Nach seinem Studium der Betriebswirtschaft an der Ludwig-Maximilians-Universität und dem Honors-Elitestudiengang an der Universität Regensburg absolvierte er Stationen als angestellter Unternehmensberater, freiberuflicher Berater und im Konzern. Im Jahr 2020 gründete er die Digitalisierungs- und Organisationsberatung Purple Bird Technology mit Hauptsitz in München.

Prestigeträchtige Unternehmen, wie DAX 40 Konzerne, zählen zu ihren Kunden. Darüber hinaus absolvierte er verschiedene mehrmonatige Lehrgänge wie die Ausbildung zum Business Coach, das Certificate of Human Capital Management (Ludwig-Maximilians-Universität) und die Ausbildung zum Business Trainer.