Wie kann KI-Bias reduziert werden?

Bild: Erstellt mit Dall-E 3

Arbeitsauftrag (Einzelarbeit, 60 Minuten)

1. **Lesen Sie den Infotext oder schauen Sie das YouTube-Video.**

2. **Notieren Sie stichpunktartig, welche Lösungsvorschläge im Video bzw. im Infotext genannt werden, um KI-Bias zu reduzieren.**

3. **Verfassen Sie einen Brief an den Vorsitzenden/die Vorsitzende des Ethikrats mit einer Empfehlung und Lösungsvorschlägen im Umgang mit KI-Bias.**

Hinweise:

- Vermeiden Sie, die Lösungsvorschläge aus dem Video bzw. Infotext einfach zu übernehmen, sondern lassen Sie ebenso eigene Ideen einfließen.
- Gehen Sie dazu auch auf Ihre Handlungsempfehlungen ein, die Sie im Rahmen Ihres Leitfadens (Arbeitsblatt 1) formuliert haben.
- Formulieren Sie adressatengerecht.

How AI Image Generators Make Bias Worse *(London Interdisciplinary School)*
https://youtu.be/L2sQRrf1Cd8?si=oA_dlu4zG8M1YyZx

Hinweis: Das Video ist in englischer Sprache. Aktivieren Sie bei Bedarf die deutschsprachigen Untertitel. Schauen Sie das gesamte Video oder springen Sie direkt zu den Lösungsvorschlägen ab Minute 4.

Lösungsvorschläge zur Reduktion von KI-Bias

Die rasante Entwicklung künstlicher Intelligenz (KI) und deren Fähigkeit, Bilder zu generieren, hat beeindruckende Möglichkeiten eröffnet. Programme wie Midjourney können auf Basis einfacher Texteingaben einzigartige Bilder erstellen. Diese Technologie könnte bald bis zu 90 Prozent der Bilder im Internet generieren. Doch mit dieser Innovation entstehen auch Herausforderungen, insbesondere die Verstärkung und Verbreitung von Vorurteilen und Stereotypen, die in den Trainingsdaten der KI-Systeme verankert sind.

Eine Untersuchung von Leonardo Nicoletti und Dina Bass für Bloomberg Technology, die über 5.000 von der KI Stable Diffusion generierte Bilder analysierten, offenbarte signifikante Verzerrungen entlang der Linien von Geschlecht, Hautfarbe und sozioökonomischem Status. Bilder von Berufen mit höherem Einkommen, wie CEOs, Anwälte und Politiker, wurden überwiegend mit heller Hautfarbe dargestellt, während Berufe mit niedrigerem Einkommen, wie Geschirrspüler, Hausmeister und Fast-Food-Mitarbeiter, häufiger Personen mit dunklerer Hautfarbe zeigten. Ähnliche Verzerrungen zeigten sich bei der geschlechtsspezifischen Darstellung, wobei höher bezahlte Berufe vorwiegend von Männern und niedriger bezahlte Berufe überwiegend von Frauen ausgeübt wurden. Diese KI-generierten Vorurteile spiegeln nicht nur die Realität verzerrt wider, sondern können sie durch die Schaffung und Verstärkung von Stereotypen noch verschlimmern.

Die Wurzel dieses Problems liegt in den Trainingsdaten der KI. Jeder Datensatz ist ein Produkt seiner Zeit und spiegelt die politischen, ökonomischen und sozialen Bedingungen wider, unter denen er erstellt wurde. Die Wissenschaftlerin Melissa Terras betont, dass es keine neutralen Datensätze gibt und dass die Reflexion über die Herkunft der Daten entscheidend ist, um Bias in der Datenwissenschaft zu vermeiden. Doch selbst wenn KI-Systeme lediglich die Vorurteile in ihren Trainingsdaten widerspiegeln, können sie diese durch Feedbackschleifen verstärken und verbreiten, was zur Zunahme und Intensivierung bestehender Vorurteile führt.

Die Lösung dieses Problems erfordert mehr als nur technische Anpassungen; sie erfordert eine Auseinandersetzung mit tiefgreifenden philosophischen Fragen nach der Definition von Fairness und Bias. Die Frage, wie beispielsweise das Geschlechterverhältnis bei der Darstellung von CEOs aussehen sollte, verdeutlicht die Komplexität des Problems. Sollte die Darstellung die reale Verteilung (neun männliche CEOs zu einer weiblichen CEO bei den Fortune-500-Unternehmen) widerspiegeln oder ein idealisiertes Verhältnis anstreben, um Ungleichheiten entgegenzuwirken? Die Antwort auf diese Frage ist nicht einfach und wirft weitere Fragen nach der fairen Darstellung anderer sozialer Gruppen und Identitäten auf.

Regierungen und Aufsichtsgremien könnten Gesetze und Vorschriften erlassen, um der Verbreitung von Vorurteilen durch KI entgegenzuwirken. So könnten Unternehmen mehr in die Verantwortung und im Zweifel zur Rechenschaft gezogen werden. Dies könnte auch die Festlegung von Standards für Trainingsdatensätze, die Förderung von Transparenz und Diversität sowie die Einrichtung von Beschwerdestellen umfassen.

Die Frage, ob die Büchse der Pandora bereits geöffnet wurde oder ob wir den Weg der KI noch zum Wohle der Gesellschaft lenken können, bleibt offen. Die Digitalaktivistin Joy Buolamwini betont, dass es an uns liegt, ob KI unsere Bestrebungen unterstützt oder ungerechte Ungleichheiten verstärkt. Die Lösung komplexer Probleme erfordert interdisziplinäre Ansätze, die Expertenwissen aus Kunst, Wissenschaft und Geisteswissenschaften vereinen, wie es die London Interdisciplinary School (LIS) vorschlägt. Nur durch ein breites Verständnis und Zusammenarbeit können wir hoffen, die Herausforderungen der KI-Bias zu bewältigen und eine gerechtere Gesellschaft zu fördern.

Verwendete Quellen

LIS - The London Interdisciplinary School (11. August 2023): How AI image generators make bias Worse [Video]. YouTube. Abgerufen am 1. März 2024, https://www.youtube.com/watch?v=L2sQRrf1Cd8

Nicoletti, L. & Bass, D. (2023): Humans are biased. Generative AI is even worse. Bloomberg Technology. Abgerufen am 1. März 2024, von https://www.bloomberg.com/graphics/2023-generative-ai-bias/?embedded-checkout=true