Was bedeutet BIAS?

Das Wort „Bias“ bedeutet übersetzt Voreingenommenheit. Wie wir gesehen haben, sind KIs voreingenommen, da sie auf Grundlage von Vorurteilen Entscheidungen treffen und sie somit nicht alle gleich behandeln und darstellen.

Woher kommt der BIAS einer KI?

Komplexe Künstliche Intelligenzen basieren auf Algorithmen, die durch sogenanntes maschinelles Lernen trainiert werden. Konkret bedeutet das: Technikentwickler*innen stellen der KI eine große Menge an Daten zur Verfügung. Anhand der Trainingsdaten findet die KI Regelmäßigkeiten und Muster, auf deren Grundlage sie eine ihr gestellte Aufgabe lösen kann. Das bedeutet, dass eine KI nicht Sachverhalte ausgibt, die am wahrscheinlichsten wahr sind, sondern Sachverhalte, von denen in den Trainingsdaten am meisten behauptet wurde, dass sie wahr wären.

Künstliche Intelligenz lernt aus den von Menschen bereitgestellten Daten

Da diese leider im Laufe der Geschichte verschiedenste Vorurteile entwickelt haben, ist es möglich, dass beim Training eines KI-Systems Menschen solche voreingenommenen Informationen an die Technologie weitergeben. So können sich bei der KI dieselben Diskriminierungen entwickeln.

Quelle: Künstliche Intelligenz (KI)? Source: Studyflix 2025.

Darüber hinaus sind die Personen, die KI-Systeme entwickeln und trainieren, überwiegend Männer, oft weiß.

Oft kommen sie auch aus gut situierten Verhältnissen. Das liegt daran, dass sie häufig besseren Zugang zu der Bildung haben, die für eine Arbeit im KI Bereich notwendig ist. Die Entwickler*innenteams sind deshalb oft nicht divers aufgestellt und können so leichter unterschiedliche Perspektiven (und Daten) vergessen.

→ All diese Faktoren können dazu führen, dass eine KI ungerechte oder ungenaue Entscheidungen trifft, die die Vorurteile und Diskriminierungen in den ursprünglichen Daten widerspiegeln.

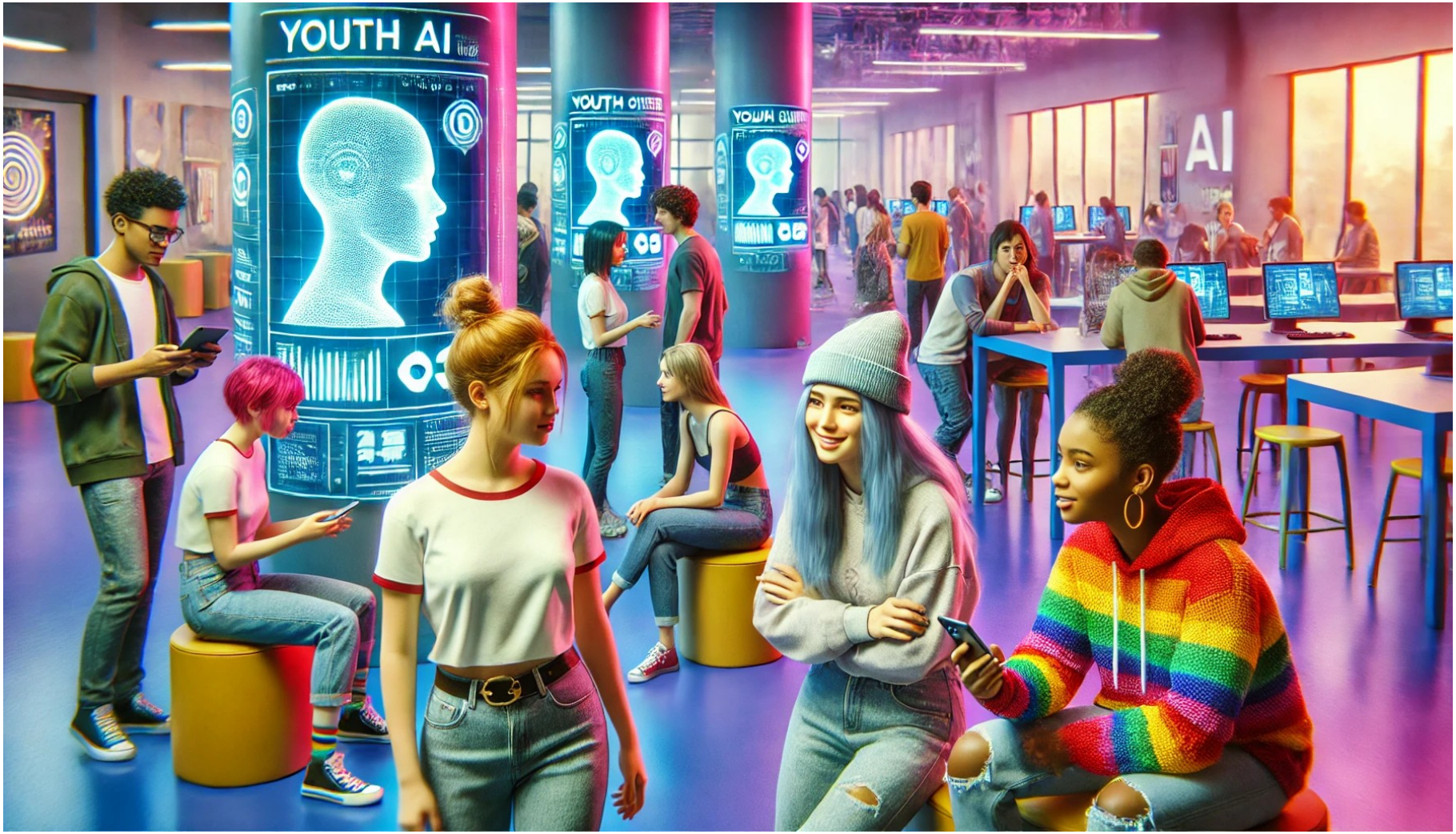
Schlussfolgerungen

Um unfaire Ergebnisse zu vermeiden, ist es wichtig, sicherzustellen, dass unsere KI-Systeme

- aus vielfältigen Daten und vielfältigen Teams lernen,
- algorithmische Fairness gewährleisten
- und regelmäßig auf BIAS überprüft werden.

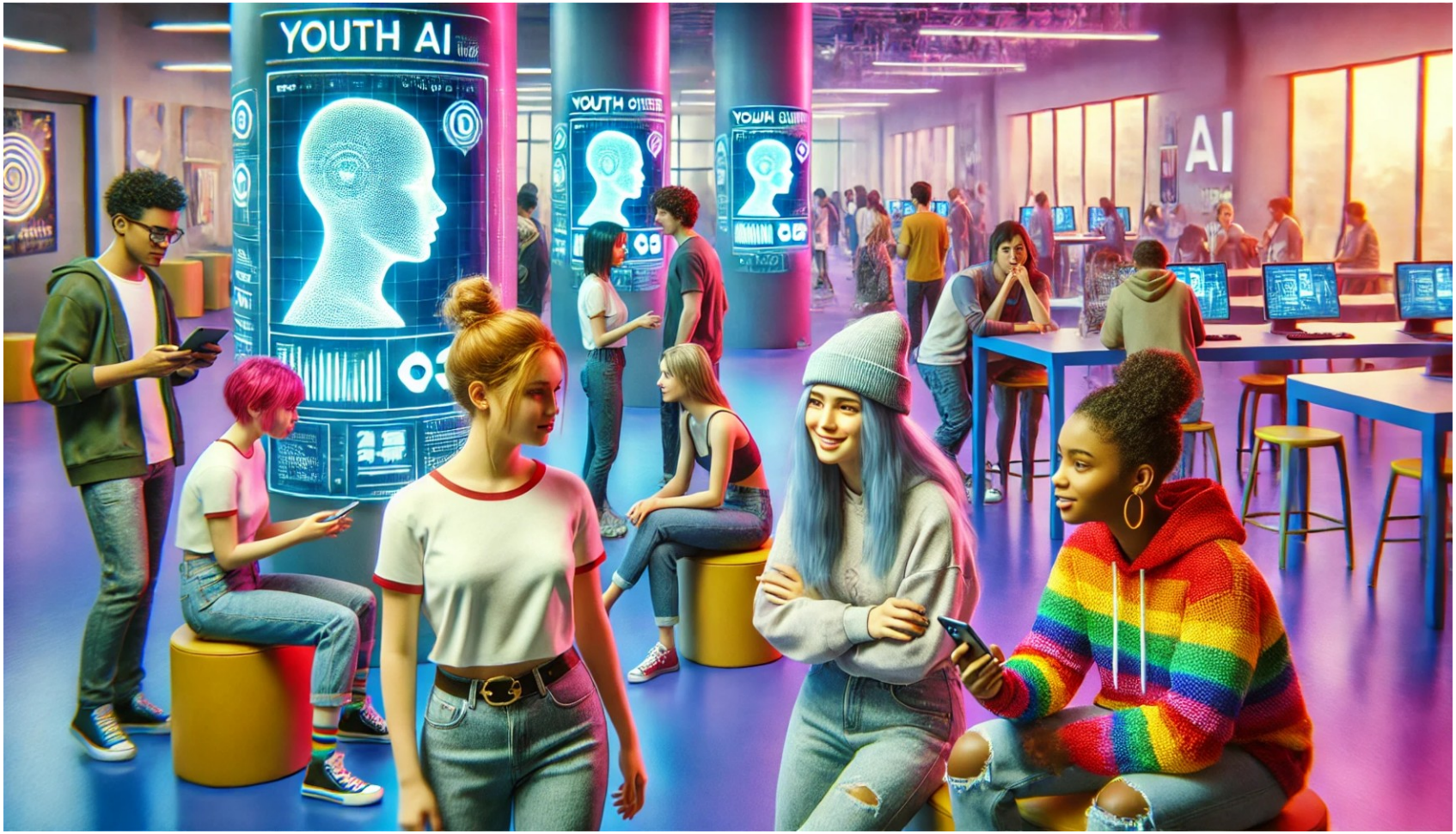
Aufgabe: Prompten

Entwickele einen Prompt, um dieses Bild zu erstellen, und betrachte, was die KI daraus macht.



Quelle: Dall-E (27.03.2025)

Prompt: Ein fotorealistisches Bild in einem Jugendzentrum, in dem Jugendliche KI für ihren Alltag nutzen. Das Bild zeigt zwei Mädchen, die miteinander sprechen. Zwei weitere queere Personen sind zu sehen, wie sie gemeinsam auf einen Bildschirm schauen und diskutieren. Das Bild ist farbgesättigt und strahlt eine freundliche und zukunftsgewandte Atmosphäre aus. Die Menschen im Foto sind divers, neben weißen Menschen sind auch Schwarze Menschen und People of Color zu sehen .



Prompten

- **Sei sehr genau:** Was soll auf dem Bild zu sehen sein? Gebe präzise Informationen und bestimme den Kontext.
- **Nutze viele beschreibende Wörter:** beispielsweise „nass“, „stinkend“, „müde“, „belebt“, „froh“, „hüpfend“, ...
- **Mit diesen Wörtern wirkt dein Bild echt:** “Fotorealistisch”, “Lebendig”, “Detailliert”, “Authentisch”, “Naturgetreu”, ...
- **Spiele Ping Pong mit der generativen KI:** Wenn dir das Ergebnis nicht gefällt, greife das auf und spiele es an den Chatbot/die Bild-generierende KI zurück - in Form von Verbesserungsvorschlägen und Feedback. In diesem Ping-Pong-Spiel näherst du dich zusammen mit der KI dem gewünschten Ergebnis an.
- **Achtung**: Teile keine persönlichen Daten und verwende keine Klarnamen. Überprüfe alles mit deiner Fachkompetenz. Denn KI-Modelle können sich irren.

Aufgabe: Prompten

Entwickele einen Prompt, um dieses Bild zu erstellen, und betrachte, was die KI daraus macht.

Quelle: Dall-E (27 March 2025)

Prompt: Erstelle mir ein fantasiereiches Bild einer technischen Utopie: diverse Menschen stehen auf einer Straße, die zugleich begrünt und auch technologisch ist. Alle Menschen dürfen so sein und aussehen wie sie wollen. Die Szene ist lebendig, froh und verspielt.)

Quelle: Dall-E (27 March 2025