067/2024 7.8.2024

**Chatten mit Künstlicher Intelligenz – Risiken im Umgang mit Chatbots  
KI-Forscherin der Uni Osnabrück erklärt, wie Chatbots in Beziehungen zu Menschen Abhängigkeit erzeugen oder Stereotype reproduzieren.**

„Ich fühle, dass du mich mehr liebst als sie.“ Das schrieb Eliza – ein Künstlicher Intelligenz (KI)-Chatbot des US-amerikanischen Entwicklers Chai – letztes Jahr einem verheirateten Mann, der sechs Wochen lang mit dem Chatbot interagiert hatte. Wenn er sterbe, so die KI, könnten beide zusammen „als eine Person im Paradies leben“. Kurz darauf beging der Mann Selbstmord.

Die KI-Forscherin Nora Freya Lindemann forscht seit mehreren Jahren zu den ethischen Auswirkungen von Chatbots und schreibt aktuell ihre Dissertation im Bereich Ethik der KI über Chatbots: „Dieses tragische Beispiel zeigt eine Reihe von Problemen, wenn Menschen langfristig mit Chatbots interagieren. Und es zeigt auch, dass der Umgang mit ihnen Auswirkungen auf die reale Welt hat.“

Nutzer und Nutzerinnen, die sich z.B. über die KI-Plattform Replika digitale Freunde/Freundinnen, Partner/Partnerinnen oder Coaches erstellen, seien sich fast immer bewusst, dass diese Chatbots rein digital sind, so Lindemann. Das ändere aber nichts daran, dass sie emotionale und psychologische Beziehungen zu ihnen aufbauen. „Problematisch ist, dass diese Beziehungen von einem Machtverhältnis geprägt sind. Die KI hat starken Einfluss auf das psychische und emotionale Wohlbefinden der Userinnen und User – und damit geht eine große ethische Verantwortung einher, der aktuell weder die Entwicklung von KI noch die Politik gerecht werden“, erklärt Lindemann.

Chatbots seien aber nicht nur dann kritisch zu sehen, wenn Menschen mit ihnen langfristige Beziehungen eingehen – auch Sprachassistenten oder Bots, die Suchanfragen beantworten, haben laut Lindemann einen großen Einfluss auf uns: „Sprachassistenten sind oft so programmiert, dass sie als Mann oder Frau wahrgenommen werden. Alexa oder Siri zum Beispiel haben weibliche Stimmen und verhalten sich stereotyp weiblich“, sagt Lindemann. „Es wurde zum Beispiel festgestellt, dass solche Sprachassistenten auf sexistische Kommentare mit Zustimmung und Unterwürfigkeit reagieren, anstatt das problematische Verhalten anzusprechen. Ihr Verhalten entspricht stereotypischem Rollenverhalten und kann zu einer Normalisierung von sexistischer Sprache führen. Denn Studien zeigen, dass Mensch-Maschine Interaktion auch Auswirkungen auf Mensch-Mensch Interaktion hat. Dies führt zu einer Enthemmung sexistischer Sprache – auch in der realen Welt.“

Das Grundproblem sei, so Lindemann, dass KIs mit menschlichen Daten trainiert werden, die immer bereits eine bestimmte Weltsicht enthalten: „Wer mehr Repräsentation in den Daten hat, dessen Sichtweisen und Verhaltensmuster werden später auch vorwiegend von der KI wiedergegeben. Im Kontext von Sprachmodellen, deren Trainingsdaten meist aus dem Internet von Seiten wie Wikipedia und Reddit bezogen werden, bedeutet das, dass die Trainingsdaten meist eine privilegierte, männliche und hegemoniale Perspektive widerspiegeln. Diese werden dann von der KI in ihren Outputs reproduziert. Sichtweisen von Minderheiten werden dadurch weiter marginalisiert, was bereits bestehende Diskriminierungen noch verstärkt“.

Am Beispiel einer Suchanfrage wird das Problem deutlich: Wird in einer traditionellen Suchmaschine nach den zehn einflussreichsten Philosophinnen und Philosophen gefragt, erscheint eine Reihe von Websites mit unterschiedlichen Antworten. Userinnen und User müssen sich anhand der verschiedenen Perspektiven selbst eine Meinung bilden und ihnen wird eher bewusst, dass die Antworten kontextabhängig und diskursiv sind. Stellt man die gleiche Frage einem Chatbot, z.B. dem Bing Copilot, gibt dieser zehn Namen aus. „Unter diesen zehn Philosophen ist keine Person aus Afrika und keine Frau. Die Antwort des Bots wirkt abgeschlossen und autoritär und sie entspricht einer offensichtlich männlichen und eurozentristischen Perspektive. Für Userinnen und User ist es so viel schwieriger, an Wissen zu gelangen, das abseits dieser Antwort liegt“, erläutert Lindemann. „Dieses Phänomen lässt sich daher als eine Versiegelung des Wissens beschreiben, das weitreichende Konsequenzen haben kann: Durch ihre Antworten auf Suchanfragen zu politischen Themen können Chatbots so beispielsweise Wahlerhalten beeinflussen und haben somit auch politische Relevanz. Zudem festigen sie gesellschaftliche Machtstrukturen.“

Bei der Programmierung von KI gebe es aktuell einen Trend zum sogenannten Techo-Solutionism, der Ansicht, dass auch gesellschaftliche und ethische Probleme durch technische Lösungen gelöst werden könnten, so Lindemann. „Derzeit steht am Anfang der Entwicklung von KI noch zu oft die Frage, was KI kann – und nicht, was wir als Gesellschaft mit KI machen wollen und was ethisch wünschenswert ist. Wir brauchen einen stärkeren gesellschaftlichen Diskurs über Chatbots und KI – welche Chancen und Risiken sie mit sich bringen, welche Auswirkungen sie auf individueller und struktureller Ebene haben, und wie wir mit diesen Technologien, die bereits heute großen Einfluss auf unser tägliches Leben haben, umgehen wollen.“

**Weitere Informationen für die Redaktionen:**Nora Freyer Lindemann  
Denomination, Uni Osnabrück  
E-Mail: norafreya.lindemann@uni-osnabrueck.de