



Neue, leistungsfähige KI-Systeme wie ChatGPT werden gerade heiß diskutiert: Welchen Einfluss werden sie auf Beruf und Alltag haben? Dr. Anne Lauscher ist Professorin für Data Science an der Universität Hamburg. Sie forscht zu künstlicher Intelligenz (KI) mit einem Schwerpunkt auf effektiver, fairer, inklusiver und nachhaltiger Kommunikation. Im Interview erklärt sie, was die KI-Revolution für die Wissenschaft bedeutet und vor welchen Herausforderungen wir stehen.

PAUL SUTTER IM GESPRÄCH MIT ANNE LAUSCHER

# „KI wird unsere Gesellschaft transformieren!“

*Frau Lauscher, der Chatbot ChatGPT des US-Unternehmens OpenAI ist seit Ende letzten Jahres in aller Munde. Was macht das Programm besonders?*

Aus meiner Sicht als Forscherin im Bereich „Natural Language Processing“ ist das Programm erst mal nicht besonders. ChatGPT ist das Resultat einer jahrelangen Entwicklung, die Stück für Stück und Schritt für Schritt erfolgt ist. Die KI basiert auf Large Language Models, großen künstlichen neuronalen Netzwerken, die auf großen Mengen von Textdaten, wie z.B. der gesamten englischen Wikipedia, trainiert wurden. Diese wurden zum ersten Mal Ende 2018 präsentiert und haben damals schon einen großen Entwicklungsschritt hinsichtlich der Performance von KI gebracht. Das Neue an

ChatGPT ist zum einen, dass die Qualität nach und nach so gestiegen ist, dass es tatsächlich sehr flüssige Antworten gibt. Und zum anderen, dass das Grundprinzip des Trainings dieser Sprachmodelle mit einer weiteren Art des Trainings basierend auf menschlichem Feedback kombiniert wurde. Somit werden bessere Antworten bevorzugt durch das Modell generiert. Das führt dazu, dass die Antworten sprachlich sehr gut sind und gleichzeitig sehr gut an menschliche Präferenzen angepasst sind. Auch die Kompetenzen sind sehr weitreichend. Ich glaube aber, das Besondere, was nun zu diesem Hype geführt hat, war, dass dieses System der breiten Masse zugänglich gemacht wurde.

# „KI ist, global betrachtet, eines der größten Forschungsfelder mit sehr viel Industriebeteiligung.“

*Sehen Sie Einsatzbereiche von ChatGPT und ähnlichen Systemen im akademischen Bereich?*

Ja, absolut. Erst mal muss man sagen: Im akademischen Bereich kann ChatGPT somit vor allem erste Hilfestellungen leisten: Die KI kann mir Textbausteine erzeugen, mich inspirieren und mir verschiedene Ideen liefern. Sie kann auch Texte auf Basis von Stichworten formulieren und somit helfen, Schreibblockaden zu lösen oder erste Ideen zu finden. Schlussendlich unterstützt mich die KI bei der Lösung von Problemen, sodass ich nicht bei null anfangen muss.

Aber die KI hat auch sehr viele Schwächen: Wir brauchen einen kritischen Blick, um genau zu überprüfen, ob der Output stimmt und angemessen ist. Ich als Wissenschaftlerin bin ja bereits Expertin in meinem Feld und kann kritisch beurteilen, was die KI mir ausgibt. ChatGPT kann ich jedoch auf keinen Fall zur Literaturrecherche einsetzen, weil die Daten, auf denen es trainiert wurde, nur bis Ende 2021 reichen und somit veraltet sind (und sich die KI-Welt so schnell bewegt). Auch Student\*innen können sich z.B. für eine initiale Recherche zu einem Thema, für das sie ja üblicherweise noch keine Expert\*innen sind, Ideen suchen oder Nachfragen stellen. Ganze Hausarbeiten schreiben zu lassen, ist aber noch schwierig – vor allem, wenn man noch nicht bewerten kann, ob der Output angemessen ist.

*Müssen Nutzer\*innen von ChatGPT oder ähnlichen Systemen die Wirkweisen des Systems verstehen? Also das, was hinter dieser schönen Fassade passiert?*

Das ist das Charmante und Gefährliche bei den Systemen: Dadurch, dass ChatGPT mit natürlichsprachigem Input funktioniert und mir auf eine ganz normale Frage eine Antwort gibt, habe ich vielleicht den Eindruck, dass ich das nicht verstehen muss, weil ich es ja auch so benutzen kann. Dennoch brauchen wir eine gewisse

KI-Literacy. Ich muss schon zumindest den grundlegenden Aufbau der KI nachvollziehen, um die Schwächen zu verstehen und den Output kritisch bewerten zu können. Dafür brauchen wir eben KI-Kompetenz an verschiedenen Stellen. Ich finde, das sollte schon in der Schule anfangen. Aber auch die Hochschullehre muss nach und nach entsprechend angepasst werden. Das ist der zentrale Punkt: verstehen, wie die Systeme funktionieren – zumindest grob. Die Kompetenz zu den Themen, zu denen ich die Systeme als Werkzeuge einsetzen möchte, brauche ich natürlich trotzdem.

Aber ich möchte auch umgekehrt sagen, dass es eine bi- oder multidirektionale Interdisziplinarität geben muss. Wir brauchen z.B. Ethik in jedem Computer Science Curriculum, um die Student\*innen früh für solche gesellschaftlich relevanten Fragestellungen zu sensibilisieren. Und wir brauchen KI-Kurse in anderen Fächern außerhalb der Computer Science, damit auch dort eine Reflexion stattfindet.

*Sie haben zu Stereotypen und Vorurteilen in Pre-Trained Language Models geforscht. Mit welchen ethischen Problemen müssen wir heute bei der Verwendung von KI-Systemen rechnen?*

Das ist meiner Meinung nach ein ganz wichtiges Thema. Wir leben in einer Welt, in der es Stereotype, Vorurteile und Diskriminierung gibt. All das ist auch reflektiert in den Daten, die uns zum Training von KI-Systemen zur Verfügung stehen. Nur ein gewisser Teil der Weltbevölkerung hat überhaupt Zugriff auf das Internet und erzeugt eben diese Daten. Die meisten Daten sind deshalb auf Englisch und darüber hinaus in der Hand großer Player, die eine gewisse Kontrolle darüber haben. Alle Daten in KI-Systemen sind ein Stück weit verzerrt. Wenn man sie vor dem Hintergrund der Stereotypen anschaut, finden sich in allen KI-Trainingsdaten problematische Aussagen. Klassischer Sexismus,

Rassismus, alles, was man sich so an Stereotypen vorstellen kann. Das kommt nicht nur durch sehr explizit sexistische Aussagen, wie etwa: „Alle Frauen sollten Familien haben.“ Sondern auch durch Zusammenhänge, die auf Basis unserer gemeinsamen Geschichte und unserer Perspektiven auf die Welt entstanden sind. Nehmen wir z.B. die Wikipedia: Das sind relativ neutrale, enzyklopädische Texte. Dort gibt es über 1,5 Mio. Biografien berühmter Wissenschaftler\*innen, Autor\*innen und Forscher\*innen. Aber nur ein ganz kleiner Teil davon behandelt Frauen. Und wenn wir unsere KI-Systeme auf diesen Daten trainieren, dann greifen diese Sprachmodelle – auf denen auch ChatGPT basiert – diese verzerrten Zusammenhänge auf. So werden Stereotype auch aus vermeintlich neutralen Texten kodiert – und das gilt auch für andere potenziell problematische Aspekte. Wenn wir jetzt je nach soziotechnischem Umfeld so ein System einsetzen, dann können verschiedene Dinge passieren: Zum einen kann ein repräsentationsbezogener Schaden entstehen. Das bedeutet, dass die Systeme die Stereotypen immer weiter reproduzieren und sich dann auch weiter manifestieren in der Gesellschaft.

Es kann aber auch ein allokatonsbezogener Schaden durch solche Stereotype entstehen. Das heißt, dass eine KI Entscheidungen trifft, die Individuen, die zu bereits marginalisierten Gruppen gehören, weniger Ressourcen zuweisen. Da gab es z.B. 2018 einen berühmten Fall einer KI-basierten Recruiting Engine eines großen Unternehmens, die Frauen eine entsprechende Gruppe von Jobs nicht zugetraut hat.<sup>1</sup>

#### *Gibt es Ansätze, diese Probleme zu lösen?*

Der erste Punkt, den wir dazu klären müssen, ist auch gesamtgesellschaftlich: Was wollen wir eigentlich? Wir haben in den letzten Jahren große Fortschritte gemacht, um diese stereotypischen Verzerrungen zu messen und zu verringern. Vor allen Dingen sehe ich die Fortschritte darin, dass die Problematik allgemein akzeptiert ist, mehr als noch vor einigen Jahren. Aber das Problem ist nicht gelöst, was auch daran liegt, dass die Frage nach den Regularien nicht gelöst ist. Welche Aussagen einer KI sollen korrigiert werden? Was sind Stereotype, die uns wichtig sind? Welche Individuen und marginalisierten Gruppen müssen wir schützen? Und wie werden die Mechanismen, auf die man sich einigt, implementiert?

#### *ChatGPT wurde bereits in manchen Papers als Co-Autor angeführt. Wer trägt die Verantwortung für die generierten Inhalte? Werden wir das öfter sehen?*

Ich glaube nicht, dass wir das jetzt öfter sehen werden. Am Anfang war es ganz interessant, einfach, um zu sehen, wie die Verlage reagieren, und Diskussionen auszulösen. Cambridge University Press hat vor wenigen Tagen ihren Ethik-Kodex für Autor\*innen erweitert und

festgelegt, dass jegliche Nutzung von KI-Software angegeben werden muss und KI nicht als Co-Autor\*in geführt werden darf. Letztendlich ist man als Autor\*in verantwortlich für das, was man produziert – und verantwortlich für eine wissenschaftliche Arbeit kann erst mal nur ein Mensch sein. Das heißt, ChatGPT sollte als Werkzeug gesehen werden und nicht als Mitautor\*in. Viele urheberrechtliche Fragen sind ja auch noch nicht geklärt. Wenn ich so ein Tool einsetze, muss ich mir ganz klar darüber sein, dass ich komplett die Verantwortung für den Output trage. Und auch dafür, dass keine Plagiate erzeugt werden. Anders ist es aktuell nicht lösbar.

#### *Was erwartet uns im Bereich „KI“ in den nächsten Jahren?*

KI ist, global betrachtet, eines der größten Forschungsfelder mit sehr viel Industriebeteiligung. Was die technischen Entwicklungen angeht, wird es auf jeden Fall immer bessere und umfangreichere Systeme geben. Wir werden immer mehr Modalitäten, d.h. geschriebene und gesprochene Sprache sowie statische und bewegte Bilder, auf denen unsere Kommunikation basiert, in KI-Systeme integrieren.

Daran grenzt auch die Frage des „Embodiments“: Mit der Zeit werden wir gewisse, vielleicht auch humanoide, Formen haben, in denen uns KI-Systeme begegnen. Das finde ich spannend, weil so neue Modalitäten möglich sind, mit denen diese Systeme die Welt erfahren, wie z.B. Tastsinn.

KI wird auf jeden Fall unsere Gesellschaft transformieren. Allem voran die Art und Weise, wie wir arbeiten. Es wird wahrscheinlich Verschiebungen geben: Wir werden viele Aufgaben einfach von KI-Systemen übernehmen lassen. Zum Beispiel, wie vorhin gesagt, kann ich mir erste Ansätze holen, meine Schreibblockaden vielleicht überwinden, Texte verbessern lassen – sofern ich mir sicher bin, dass die Informationen darin stimmen. Dann brauche ich mich nicht eine Stunde hinsetzen und über jedes Wort nachdenken. Tätigkeiten und Arbeiten, die „higher-level tasks“ sind, wie z.B. das Treffen strategischer Entscheidungen, werden dann entsprechend wichtiger für uns – bzw. haben wir dann mehr Zeit für sie.

---

Das Interview wurde erstmalig am 21. März 2023 auf [wissenschaftskommunikation.de](https://wissenschaftskommunikation.de) veröffentlicht.

#### **Anmerkung:**

**1 Dastin, J.:** Amazon scraps secret AI recruiting tool that showed bias against women. In: Reuters.com, 11.10.2018. Abrufbar unter: <https://www.reuters.com>

Paul Sutter ist studierter Linguist und schreibt u.a. für [wissenschaftskommunikation.de](https://wissenschaftskommunikation.de).