

Sitzung des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung

Expertengespräch zum Thema "ChatGPT"

Prof. Dr. Judith Simon, Universität Hamburg

Deutscher Bundestag

Ausschuss für Bildung, Forschung
und Technikfolgenabschätzung

Ausschussdrucksache

20(18)108b

21.04.2023

Sehr geehrte Mitglieder des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung,

haben Sie vielen Dank für die Gelegenheit heute vor Ihnen sprechen zu dürfen.

1. Zur Einordnung von ChatGPT

ChatGPT hat die Welt im Sturm erobert. Im November auf den Markt geworfen, sind die Nutzungszahlen innerhalb weniger Monate explodiert. Zwei Aspekte sind hierfür von zentraler Bedeutung: die Breite der Anwendbarkeit sowie die einfache und unmittelbare Nutzbarkeit von ChatGPT und verwandten Produkten wie Dall-E. Nutzer:innen benötigen nahezu keine Vorkenntnisse und nur geringe technische Voraussetzungen um in Sekundenschnelle Texte, Bilder oder Videos in sehr hoher Qualität aber ohne Wahrheitsbezug zu produzieren. Die Bedeutung dieser Entwicklung ist nicht zu unterschätzen, ist Sprache doch zentrales Medium menschlicher Kommunikation, während Bilder und Videos wiederum entscheidende Bedeutung für Fragen der Evidenz haben.

Es besteht daher die Notwendigkeit, reale und drängende Gefahren in den Blick zu nehmen und sich nicht von Scheindebatten ablenken zu lassen. Zu den weitgehend fiktiven Problemen gehören die Diskussionen darüber, ob ChatGPT Anzeichen allgemeiner künstlicher Intelligenz zeige, bzw. über Verständnis oder gar Bewusstsein verfüge. Dies ist nicht der Fall: Die sprachlichen Ausgaben von ChatGPT basieren auf der rein statistischen Analyse und Produktion von Text ohne jedwede Form von Verständnis – auch wenn dies Nutzer:innen so erscheinen mag.

2. Das Problem der dreifachen Täuschung

Hier kommen wir jedoch zu einem zentralen Problem generativer KI: dem Problem der dreifachen Täuschung.

Zunächst geht es um mögliche Gefahren, wenn Nutzer:innen im Unklaren gelassen werden, dass sie mit einer Maschine und nicht mit einem Menschen interagieren. Allerdings ist dies nur das vordergründigste Täuschungsproblem.

Darüber hinaus besteht auch das Problem der *Täuschung darüber, über welche Fähigkeiten eine Software verfügt*. Denn auch wenn aktuelle KI Systeme weder über Verständnis noch gar über Bewusstsein verfügen, so kann dies Nutzer:innen so *erscheinen* – selbst wenn sie wissen, dass sie mit einer Maschine interagieren. Dies zeigen frühe Erfahrungen mit

Weizenbaum's Software ELIZA ebenso wie aktuelle Berichte über die Interaktion von Nutzer:innen mit ChatGPT. Die Interpretation von Fähigkeiten durch ein menschliches Gegenüber sagt nichts über die Funktionsweisen und Fähigkeiten der Maschine aus, wohl aber über die menschliche Tendenz zur Anthropomorphisierung von Technik. Mir geht es nun keineswegs darum, über diese menschliche Fehlzuschreibung zu spotten. Im Gegenteil: Es geht darum, vor der performativen Kraft der Simulation von Intelligenz, Verständnis oder gar Emotion und Empathie zu warnen, da sie uns verletzlich macht. Wir reagieren kognitiv und emotional in besonderer Weise auf Sprache und Bilder – und das macht die Macht dieser neuen Technologien aus.

Und drittens geht es um die mannigfaltigen Gefahren aufgrund *täuschender Resultate* solcher Systeme: diese reichen von gefälschten wissenschaftlichen Publikationen über die kriminelle Nutzung gefälschter Stimmen, um Angehörige zu betrügen, zu Deep Fakes für Zwecke der Propaganda. Natürlich sind Täuschung, Propaganda und Manipulation keine neuen Themen. Aber: Die Qualität und vor allem die Einfachheit, Voraussetzungslosigkeit (seitens der Nutzer:innen) und Geschwindigkeit, mit der Texte, Bilder oder Videos hergestellt und weiterverbreitet werden können, gekoppelt mit der menschlichen Neigung emotional auf diese Materialien zu reagieren, eröffnet eine völlig neue Dimension des Missbrauchs. Bei allen möglichen Chancen bergen ChatGPT und verwandte Systeme daher eine reale Gefahr für unsere Demokratie, da grundlegende Prozesse der Information und Kommunikation schnell, einfach und nachhaltig gestört werden können.

3. Chancen und Herausforderungen mit Blick auf den Bildungsbereich

All diese Entwicklungen stellen unsere Bildung vor Herausforderungen – Bildung sowohl in Bezug auf den *Umgang* mit diesen Technologien, als auch deren *Entwicklung*. Die schnelle Verbreitung von ChatGPT hat Universitäten und Schulen zunächst sehr unmittelbar vor die Herausforderung gestellt, Prüfungen möglichst betrugssicher zu gestalten. Die vordergründige Frage war: Wie kann weiterhin Fairness garantiert werden, wenn manche Studierende sich ihre Hausarbeiten von ChatGPT schreiben lassen und andere nicht? Hintergründig eröffnet ChatGPT aber auch die Möglichkeit – und Notwendigkeit – uns zu fragen, was Wesen und Wert von Bildung sind. Welche Fähigkeiten und Fertigkeiten müssen unter der Bedingung neuer technologischer Möglichkeiten weiterhin erlernt werden, welche müssen hinzukommen, welche werden vielleicht nicht mehr benötigt?

Orientierung zur Beantwortung dieser Fragen hat der Deutsche Ethikrat in der Stellungnahme „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“ vorgelegt. Unsere Kernfrage ist, wie KI so gestaltet und eingesetzt werden kann, dass die Handlungsoptionen und Autorschaft der unterschiedlichen Beteiligten erweitert und nicht vermindert werden.

4. Fazit

Was ist nun in Anbetracht der Herausforderungen durch generative KI zu tun? Zum einen wird hinterfragt werden müssen, wie mit KI im Allgemeinen und generativer KI im Speziellen als *Grundlagentechnologie* umgegangen werden muss. Aufgrund der weitreichenden gesellschaftlichen Auswirkungen von ChatGPT scheint klar, dass eine rein anwendungsbezogene Regulierung Gefahr läuft, erneut überrollt zu werden.

Darüber hinaus wird deutlich, dass disruptive Effekte extrem schnell zu beobachten sind, wenn KI zum *Massenprodukt* wird, indem einfach nutzbare Tools frei verfügbar gemacht werden: Hochperformante Technologien wie ChatGPT erfordern nahezu keine Kompetenz seitens derjenigen, die sie nutzen, um Produkte herzustellen – überfordern aber andererseits die Kompetenz der Rezipient:innen dieser Materialien, weil diese echte von falschen oder gezielt irreführenden Inhalten nicht mehr unterscheiden können.

Die erste naheliegende Lösung in Bezug auf die Täuschungsproblematik liegt in einer *Kennzeichnungspflicht*, welche im AI Act bereits vorgesehen ist. Eine solche Kennzeichnungspflicht ist m.E. notwendig, aber nicht hinreichend, um dem Problem der Täuschung zu begegnen. Weder wird sich dadurch verhindern lassen, dass Menschen Technologien Eigenschaften zuschreiben, die diese nicht besitzen, noch kann man kriminelle oder missbräuchliche Verwendung damit gänzlich ausschließen. Entsprechend bedarf es auch neuer Normen und Kompetenzen im Umgang mit diesen Systemen und der klaren Benennung der Möglichkeiten aber auch Limitationen neuer Technologien.

Darüber hinaus stellen sich Fragen in Bezug auf die Vor- und Nachteile und genauen Modalitäten der *Verfügbarkeit und Offenheit* solcher Systeme. Auch wenn der freie und einfache Zugang zu generativer KI die zuvor geschilderten Probleme mit sich bringt, so ermöglichen Open Source Lösungen bessere Überprüfbarkeit von KI. Hier muss also genau geprüft werden, welche Form von Offenheit die meisten Vorteile und die wenigsten Nachteile mit sich bringt. Die jetzige Form des einfachen und freien Zugangs zu opaken, proprietären Systemen scheint die denkbar schlechteste Kombination.

Neben diesen spezifischen Problemen müssen im Fall von ChatGPT natürlich auch alle klassischen Probleme datenbasierter KI adressiert werden, bspw. hinsichtlich systematischer Verzerrungen und Diskriminierung; mangelnder Transparenz, Nachvollziehbarkeit und Kontrolle; dem Schutz personenbezogener Daten oder auch hinsichtlich der prekären Arbeitsbedingungen derer, die Daten für KI-Systeme aufbereiten.