

Pressemitteilung

Ethik in der Künstlichen Intelligenz:

Liefert die KI eigentlich korrekte und gerechte Antworten?

Hof - Die rasante Entwicklung der Künstlichen Intelligenz (KI) bringt nicht nur technologische Fortschritte, sondern auch komplexe ethische Fragen mit sich. Insbesondere bei generativer KI, wie Sprach- oder Bildgeneratoren, rückt das Thema der voreingenommenen Ergebnisse (Bias) in den Mittelpunkt der Diskussion. Prof. Dr. René Peinl, Marc Lehmann und Prof. Dr. Andreas Wagener vom Institut für Informationssysteme der Hochschule Hof (iisys) haben diese Problematik nun analysiert und kommen zu spannenden Erkenntnissen.

Bias in KI-Modellen bezeichnet die Neigung, Ergebnisse zu liefern, die einseitig oder von menschlichen Vorurteilen geprägt sind. „Diese Verzerrungen entstehen oft durch die Daten, mit denen die Modelle trainiert werden, sowie durch deren algorithmische Verarbeitung. In Studien wird häufig stillschweigend davon ausgegangen, dass eindeutig definiert ist, was eine „korrekte“ oder „unvoreingenommene“ Antwort darstellt“, so Prof. Dr. René Peinl. Doch die gesellschaftliche Realität zeigt: Solche Definitionen sind in Wahrheit oft sogar außerordentlich umstritten.

Wer entscheidet über „korrekte“ Antworten?

In der Praxis gibt es keinen Konsens darüber, was als „richtige“ oder „gerechte“ Antwort gelten sollte. Themen wie gendersensible Sprache, der menschengemachte Klimawandel oder die Gleichstellung Homosexueller sind gesellschaftlich zum Teil hoch umstritten. „Wenn ein KI-Modell zu einer Frage eine scheinbar voreingenommene Antwort liefert, stellt sich die Frage, ob dies tatsächlich ein Ausdruck von Bias ist – oder einfach die statistisch wahrscheinlichste Antwort“, erläutert Prof. Dr. Andreas Wagener.

Beispiel: Ein generiertes Bild eines „bayerischen Mannes“ zeigt häufig einen Mann in Lederhose mit Bierkrug. Diese Darstellung mag klischeehaft wirken, spiegelt jedoch eine kulturelle Symbolik wider, die für viele Menschen eine klare Aussage vermittelt. Ein Mann im Anzug oder Jogginganzug würde diesen Zusammenhang weniger deutlich machen.

Technische Grenzen

Ein Großteil der scheinbar oft voreingenommenen Ergebnisse entsteht durch die Qualität des Modells und durch die Eingaben selbst. „KI-Modelle müssen oft Entscheidungen treffen, wenn Eingaben vage oder nicht ausreichend spezifiziert sind. „So könnte die generische Eingabe „Kuh“ dazu führen, dass ein Modell vorwiegend Kühe auf einer Wiese oder im Stall generiert – auch dies ist ein Beispiel für einen „Bias“, wenngleich wohl für einen durchaus gewünschten“, so Marc Lehmann. Hinzu kommt, dass unklare Aufgabenstellungen die Modelle dazu zwingen, wahrscheinliche Varianten zu wählen. Eine Verbesserung der Modellergebnisse erfordert daher präzisere Eingaben und eine detailliertere Betrachtung der statistischen Verteilung.

Mögliche Lösungsansätze

Die Forscher der Hochschule Hof haben verschiedene Lösungsansätze zur Minimierung von Bias untersucht, dabei jedoch keine universelle Lösung gefunden. Die Spaltung in westlichen Gesellschaften erschwere es zusätzlich, Modelle so zu gestalten, dass sie allgemeiner Akzeptanz entsprechen. In einigen Fällen kann die Verteilung innerhalb der Grundgesamtheit als Orientierung

dienen. „Beispielsweise sollten Bildgeneratoren Männer und Frauen bei geschlechtsneutralen Berufsbezeichnungen gleichermaßen darstellen, um vergangene Benachteiligungen nicht zu wiederholen“, schlägt Prof. Dr. René Peinl vor.

Berücksichtigung von Minderheiten

In anderen Fällen ist es jedoch nicht sinnvoll, Gleichverteilung anzustreben. So sind z. B. 2 % der deutschen Bevölkerung homosexuell. Ein Modell, das bei generischen Eingaben wie „happy couple“ jedes vierte Bild als homosexuelles Paar darstellt, würde die statistische Realität stark überrepräsentieren. Stattdessen sollte ein KI-Modell explizite Eingaben wie „gay couple“ korrekt umsetzen und entsprechende Bilder generieren.

Landesspezifika: Ein praktikabler Kompromiss?

Ein weiterer Vorschlag der Forscher ist die Einführung landesspezifischer „Defaults“. So könnte die Eingabe „Mann“ in China zu einem asiatisch aussehenden Mann führen, in Nigeria zu einem dunkelhäutigen und in Deutschland zu einem kaukasischen Mann. Diese Anpassungen würden kulturelle und demografische Unterschiede berücksichtigen, ohne diskriminierend zu wirken.

Fazit: Die Balance zwischen Präzision und Neutralität

Die Forschung zeigt, dass die Entwicklung unvoreingenommener KI-Modelle eine enorme Herausforderung darstellt. Es gibt keine einfachen Antworten, da viele Probleme auf gesellschaftliche Uneinigheiten zurückzuführen sind. Eine mögliche Lösung ist es, Modelle so zu gestalten, dass sie klare Eingaben akkurat umsetzen und landesspezifische Kontexte berücksichtigen. Doch selbst diese Ansätze erfordern eine fortlaufende Diskussion und Anpassung, um ethischen und technischen Anforderungen gerecht zu werden.

Pressekontakt:

Rainer Krauß, Hochschulkommunikation / PR
Alfons-Goppel-Platz 1, 95028 Hof
Telefon: 09281/409-3006
E-Mail: pressestelle@hof-university.de

Über die Hochschule Hof:

Für die Hochschule Hof stehen ihre aktuell über 3800 Studierenden an erster Stelle. Alle Studienangebote werden kontinuierlich angepasst, um die Studierenden fit für die Welt von morgen zu machen. Praxisorientierung, Internationalisierung und intelligente Ressourcennutzung stehen im Fokus von Lehre und Forschung an der Hochschule Hof. Im Bereich Internationalisierung legt die Hochschule einen Schwerpunkt auf Indien. Im Hinblick auf das Thema intelligente Ressourcennutzung stehen Wasser- und Energieeffizienz im Vordergrund. Das breitgefächerte und interdisziplinäre Studienangebot reicht von Wirtschaft über Interdisziplinäre und innovative Wissenschaften bis hin zu Informatik und Ingenieurwissenschaften.

Der Campus Münchberg bietet durch eng mit der Wirtschaft verzahnte Textil- und Designstudiengänge eine in Deutschland einmalige Ausbildung. Am Lucas-Cranach-Campus in Kronach ist ein innovativer Studienort entstanden, an dem man sich mit globalen und regionalen Zukunftsthemen beschäftigt – hier geht es um Schwerpunkte wie Innovative Gesundheitsversorgung. Am Lernort Bamberg finden u.a. Pflegestudiengänge für Berufserfahrene und im Bereich Erststudium mit monatlicher Vergütung, statt. Am Standort Selb wird den Studierenden der Studiengang Design &

Mobilität angeboten.

Studierende mit Berufserfahrung finden an der Studienfakultät für Weiterbildung ebenso den passenden Studiengang an der Hochschule Hof. Die berufsbegleitenden Angebote, die mehrheitlich in Blended Learning Einheiten stattfinden, reichen vom Einzelmodul über Zertifikatslehrgänge bis zum Bachelor- und Masterstudiengang. Ein neues Kompetenzzentrum Digitale Verwaltung soll insbesondere deutsche Behörden und Institutionen auf dem Weg hin zu bürgerfreundlichen und effektiven Services begleiten und unterstützen. Studierende mit StartUp- oder Gründungsinteresse werden durch das Digitale Gründerzentrum Einstein1 am Campus der Hochschule beraten und gefördert.

Die angewandte Forschung an der Hochschule Hof sichert die Aktualität des Wissens für die Lehre und entwickelt nützliche Lösungen, die in der Wirtschaft zum Einsatz kommen. Durch die Einrichtung von Kompetenzzentren und Instituten an der Hochschule profitieren auch die hochfränkischen Unternehmen. Die Schwerpunkte der sechs Forschungsinstitute liegen auf den Bereichen Biopolymerforschung, Informationssysteme, Materialwissenschaften, Wasserstoff- und Energietechnik, nachhaltige Wassersysteme sowie Wirtschafts- und Organisationsforschung. Zudem ist das Fraunhofer-Anwendungszentrum Textile Faserkeramiken TFK am Campus Münchberg angesiedelt und entwickelt u.a. neue Anwendungen für die Luft- und Raumfahrt sowie für die Automobilindustrie. Das an die Hochschule Hof angegliederte Bayerisch-Indische Zentrum für Wirtschaft und Hochschulen **BayIND** koordiniert und fördert darüber hinaus die Zusammenarbeit zwischen Bayern und Indien.

Die moderne Hochschule Hof ist nicht nur architektonisch offen gestaltet, sie bietet auch ein freundliches und familiäres Umfeld. Die Studierenden wissen dies zu schätzen und wählten die Hochschule im Jahr 2023 und 2024 zur „Beliebtesten Hochschule Deutschlands“ (lt. Studienportal [studycheck.de](https://www.studycheck.de)).