

# Warum der Chat-Bot Mist erzählt

**Interview mit KI-Forscherin** Die Aussagen des Textroboters Chat GPT sind fehlerhaft. Weshalb sie das wenig überrascht, erklärt Mascha Kurpicz-Briki.

**Maximilian Jacobi**

Bern ist der Hauptkanton der Schweiz, und im Oberaargau gibt es vor allem Rebfelder und die Grenchenbergbahn zu bestaunen. In einem Interview befragten wir einen der bisher leistungsfähigsten Textroboter zur nordöstlichsten Region des Kantons. Der Chat-Bot Chat GPT beantwortete viele unserer Fragen mit ausgemachtem Mumpitz.

Mascha Kurpicz-Briki ist stellvertretende Leiterin der Forschungsgruppe Applied Machine Intelligence an der Berner Fachhochschule in Biel. In ihrer Forschung konzentriert sie sich auf automatische Textverarbeitung und Sprachmodelle. Sie fokussiert sich unter anderem auf Fairnessfragen rund um das Thema künstliche Intelligenz.

**Frau Kurpicz-Briki, weshalb gibt der Chat-Bot so falsche Antworten?**

Das Ziel des Chat-Bots ist es nicht, faktisch korrekte Aussagen zu machen, sondern möglichst menschenähnliche Antworten zu geben. Seine Weltsicht, wenn man das so nennen will, ist stark eingeschränkt. Er weiss nur von Dingen, die in seinen Trainingsdaten enthalten sind. Ob sie stimmen, kann er nicht einschätzen. In diesen Daten können auch Meinungen stecken. Chat GPT kann stilistisch einwandfreie Antworten geben, der Inhalt ist für ihn Nebensache.

**Wie funktioniert ein Chat-Bot – ungefähr?**

Zuerst wird er mit einer enormen Zahl an Texten gefüttert. Darin analysiert er, in welchem Zusammenhang Wörter vorkommen. Wenn der Bot Sätze bildet, rechnet er – vereinfacht gesagt – die Wahrscheinlichkeit dafür aus, welches Wort als Nächstes kommt. Je nach Kontext sind das dann andere Wörter. Die Antwort besteht also aus Wortkombinationen, die der Bot für statistisch plausibel hält.

**Der Bot rechnet also nur Wahrscheinlichkeiten aus?**

Stark vereinfacht kann man das so sagen. Es ist im Prinzip reine Statistik. Auch Bilderkennungsprogramme funktionieren so. Sie werden beispielsweise mit Bildern von Katzen und Hunden gefüttert. Jedes Bild verfügt über die Zusatzinformation, welches Tier sich darauf befindet. Wenn das mit genügend Bildern geschehen ist, kann man dem Programm Fotos ohne Information vorlegen und es berechnet dann, ob es sich bei dem abgebildeten Tier um eine Katze oder einen Hund handelt.

**Ist die Bezeichnung «künstliche Intelligenz» dann überhaupt zutreffend?**

Eigentlich nicht. Sie stammt noch von der ursprünglichen Idee, eine Form der Intelligenz zu erschaffen, die mit der des Menschen konkurrieren kann. Und davon sind wir nach wie vor weit entfernt. In der Forschung beschreiben wir Chat-Bots und Bildgeneratoren wie Midjourney daher oft als «Augmented Intelligence».

**Also «erweiterte» oder «erhöhte Intelligenz»?**

Eigentlich eine erweiternde Intelligenz, weil sie bei spezifischen Aufgaben unterstützt und uns so bessere Leistungen ermöglicht. Ein Werkzeug also. Einst begannen wir, auf Tastaturen zu schreiben – heute gibt es Hilfsmittel, die für uns schreiben. Den Menschen braucht es aber nach wie vor, um das Geschrie-

**«Noch heute verfügen wir über ideologisch eingefärbte Sichtweisen, ohne dass es uns ausreichend bewusst ist.»**

bene einzuordnen und zu überprüfen. Wir können clevere Werkzeuge für ganz klar definierte Aufträge erschaffen. Mehr aber noch nicht.

**Werkzeuge, mit denen man aber auch Unfug treiben kann. Nie war es einfacher, Falschinformationen zu erzeugen.**

Das ist ein Problem. Die Texte wirken schlüssig, auch wenn sie inhaltlich falsch sind. Das kann diverse Gründe haben. Beispielsweise, weil Trainingsdaten fehlerhaft sind. Oder weil falsche Zusammenhänge hergestellt werden, da der Bot einzelne Daten missinterpretiert und unkorrekt einordnet.

**Mit welchen Daten wurde Chat GPT trainiert?**

Durch von Menschen erstellte Texte, wie beispielsweise Bücher. Und mit Artikeln, Wikipedia-Einträgen und Derartigem aus dem Internet. Ein weiteres Problem ist, dass kein Text wertfrei entsteht, sondern ein Produkt seiner Zeit ist. Vor hundert Jahren dürfte den meisten Menschen in Europa ihre kolonialistische Haltung nicht bewusst gewesen sein. Noch heute verfügen wir über ideologisch eingefärbte Sichtweisen, ohne dass es uns ausreichend bewusst ist. Und das spiegelt sich in unseren Texten wieder.

**Warum sollten wir uns in Bezug auf Chat-Bots hinterfragen?**

Um zu verhindern, dass diskriminierende Strukturen durch diese Textmaschinen reproduziert werden. Wir haben letztes einen Artikel bei uns im Wissenschaftsmagazin veröffentlicht. Darin wurde ein Chat-Bot aufgefordert, eine Geschichte mit verschiedenen männlichen und weiblichen Vornamen zu schreiben. Sie spielt in einem Techunternehmen. Es zeigte sich, dass er männlichen Namen andere Aufgaben im Unternehmen zuwies als den weiblichen.

**Und wie lässt sich das verhindern?**

Das ist eine Herkulesaufgabe. Als Gesellschaft müssten wir gemeinsam über Daten sprechen – was wir wollen und was nicht. Das darf nicht nur eine kleine



Mascha Kurpicz-Briki an der Berner Fachhochschule in Biel. Sie befasst sich mit Fairnessfragen zu künstlicher Intelligenz. Foto: Adrian Moser

## Die KI-Forscherin

Mascha Kurpicz-Briki wurde für ihre Arbeit über energieeffizientes Cloud-Computing an der Universität Neuenburg promoviert. Heute ist sie Dozentin für Datentechnik im Departement für Technik und Informatik an der Berner Fachhochschule in Biel. In ihrer Forschung legt sie den Fokus auf die sozialen Auswirkungen von automatischen Textverarbeitungsprogrammen und Sprachmodellen. (mjf)

Gruppe entscheiden. Diese Daten bilden die bestehenden Strukturen ab. Betroffen sind davon alle. Es ist technisch sehr schwierig, Datensätze von einem sogenannten Bias (Voreingenommenheit, Anm. d. Red.) zu bereinigen. Ein guter Anfang ist es, wenn wir uns dieser Risiken bewusst sind. In den letzten Jahren hat sich die Technologie sehr schnell entwickelt, ohne dass solche Überlegungen ausreichend miteingeflossen sind. Jetzt merkt man, dass Diskriminierung durch erweiterte Intelligenz reproduziert werden kann. Zurzeit sucht man nach Wegen, wie sich das künftig verhindern lässt. Technisch ist das sehr anspruchsvoll. Es gibt noch keinen Weg, um einen Bias aus einem Datensatz zu entfernen.

**Das klingt beunruhigend.** Um Diskriminierung vorzubeugen, ist es daher wichtig, die Aufgabenbereiche der erweiterten Intelligenzen klar zu definieren. Wie viel darf das Programm

selbst entscheiden? Und wo braucht es noch den Menschen, um die Ergebnisse zu hinterfragen und einzuordnen?

**Und wie merken wir, ob wir es mit künstlich generierten Texten zu tun haben? Gerade im Internet?**

Informationen aus dem Internet sind ja grundsätzlich mit Vorsicht zu geniessen. Man sollte hier nach wie vor auf die Quellen achten. Also: auf welchen Websites was genau steht. Denn merken, ob Texte von Menschen oder einem Chat-Bot geschrieben wurden, können wir kaum.

**Was bedeutet das für den Bildungssektor: Sterben Aufsätze bald aus?**

Aufsätze als Leistungsnachweis werden durch diese Technologien natürlich infrage gestellt. Allgemein werden die Möglichkeiten solcher erweiterten Intelligenzen die Art und Weise, wie wir lernen und lehren, sicher beeinflussen.

**Ist es unmöglich, herauszufinden, ob der Text aus der Feder eines Menschen oder einer Maschine stammt?**

Technologien dafür befinden sich in der Entwicklung. Wir sollten Chat-Bots aber viel eher in die Ausbildung integrieren, statt zu versuchen, sie auszuschliessen. Wie gehen wir richtig mit ihnen um? Wie müssen Texte hinterfragt werden? Ähnlich, wie man heute Menschen auf Quellen aus dem Internet sensibili-

**«Die Technologien sind jetzt da. Wir müssen lernen, sie kompetent zu nutzen.»**

siert. Die Technologien sind jetzt da. Wir müssen lernen, sie kompetent zu nutzen.

**Wir haben jetzt viel über ihre Schwierigkeiten gesprochen. Was sind die Stärken von solchen «erweiterten Intelligenzen»?**

Sie eröffnen uns neue Möglichkeiten. Wie beispielsweise in Ihrem Interview: Der Chat-Bot hat innert kürzester Zeit Oberaargau-Slogans für verschiedene Zielgruppen geliefert. Allgemein ist die Erstellung von Texten dadurch effizienter geworden. Es stellt sich aber immer die Frage: Wie wird mit den Schriftstücken verfahren? Werden sie einfach veröffentlicht oder kritisch geprüft? Ersetzt wird der Mensch durch die Technologie nicht. Seine Arbeitsweise ändert sich aber. Das Werkzeug liefert Inputs – wie damit verfahren wird, entscheiden wir. Da Textbots und Bildgeneratoren auf riesige Datenmengen zurückgreifen, kann daraus auch wirklich Einzigartiges entstehen. Geben Sie mal bei einem Bildprogramm «Katzen, die am Strand sitzen und Eiscreme essen, im Stile eines Picassos» ein und schauen Sie, was passiert.