Humanistische Union

ChatGPT und algorithmischer Kolonialismus

Seit ChatGPT ist die textverarbeitende Sparte der künstlichen Intelligenz in aller Munde. Zumeist geht es dabei um Betrug bei den Hausaufgaben, um Plagiarismus und um den drohenden Verlust der Kreativität, wenn die KI uns das Schreiben abnimmt. Kaum aber wird bedacht, dass mit ChatGPT eine neue westliche Kolonialisierungswelle von bisher ungekanntem Ausmaß einhergeht.

Die gute Nachricht zuerst: Was das Verhältnis des Menschen zur Künstlichen Intelligenz betrifft, sind wir längst über das dystopische Klischee à la I, Robot und andere Filme hinaus, in denen die KI die Macht übernimmt und den Menschen versklavt oder gar liquidiert. Es geht in der KI-Debatte nicht mehr darum, ob KI uns besiegen wird, sondern darum, mit welchen Werten wir sie ausstatten sollen. Die schlechte Nachricht: Wir wissen es nicht.

Wie können wir wissen, welche Werte global zum Einsatz kommen sollten, wenn selbst in einem Nationalstaat wie Deutschland verschiedene Wertvorstellungen miteinander konkurrieren. In diesem Falle wird der Konflikt periodisch per Abstimmung neu geregelt, indem man für das von einer Partei vertretene Wertesystem stimmt. Die Probleme dieses Verfahrens sind bekannt: bizarre Regierungskoalitionen erzwingen schmerzhaft Kompromisse, die manchmal auf einen Werte-Kuhhandel hinauslaufen.

Wie viel schwieriger wird alles, wenn man sich auf Werte einigen soll, die global Anwendung finden. Gewiss, körperliche Unversehrtheit und individuelle Selbstbestimmung können als universelle Werte gelten, die sich dann auch gegen kulturelle Traditionen wie die Genitalienverstümmelung oder das Verschleierungsgebot für Frauen ins Feld führen lassen. Aber schon beim Individualismus gehen die Ansichten der Kulturen auseinander. Wirtschaftliche Unabhängigkeit und die Befreiung aus traditionellen Zwängen sind keine universellen Werte, sondern eine Art „Aufklärungs-Liberalismus“, westlicher Partikularismus, der sich als universell ausgibt. So sehen das nicht nur junge farbige Frauen, sondern selbst alte weiße Männer. (Khader 2021; Taylor 1994)

Besser also, die Frage nach dem Universellen ruhen zu lassen und regional Lösungen anzustreben, im Rahmen des jeweiligen politischen Kontexts. Gerade das aber geht nicht mehr, wenn bestimmte Werte und Handlungsweisen durch eine spezifische Technologie in alle Weltteile exportiert werden. Denn Technik ist auch Kulturexport, schon immer.

So exportiert das Auto den Individualismus in der Fortbewegung, das Telefon exportiert das Zwiegespräch in Gemeinschaftskulturen, die Mikrowelle bringt das hastige Essen mit sich und Kommunikations-Technologien wie Google und Facebook ändern weltweit den Umgang mit Wissen und Freunden.

Mit der KI nimmt der Export allerdings ganz neue Dimensionen an. Denn während das Auto nur den Individualverkehr exportiert, exportiert das von der KI gefahrene Auto zugleich den Fahrstil. So ist es auch bei KI, die uns die Welt erklärt.

Fragt man ChatGPT selbst nach der Autorschaft seiner Auskünfte, erfährt man, „dass die Sprache und das Wissen, die in dem von ChatGPT generierten Text verwendet werden, auf Daten basieren, die von vielen verschiedenen Personen erstellt wurden. In diesem Sinne kann der von ChatGPT generierte Text auch als kollektives Werk betrachtet werden, das den Beitrag vieler Autoren widerspiegelt.“

Der ChatGPT-Text als Patchwork eines Autorenkollektivs? Das klingt interessant. Ist das zitierbar und wenn ja, wie lautet die richtige Quellenangabe?

Hier stoßen wir auf das Kernproblem von ChatGPT. Von wem stammen die Daten, an denen es das Denken gelernt hat – wenn man vom Denken sprechen kann, wenn sinnvolle Sätze auf der Basis eines statistischen Kalküls gebildet werden. Denn was eine KI über Gott und die Welt „denkt“, ist immer das, was die Menschen in den Gigabytes an Text, mit denen die KI gefüttert wird, am meisten dazu geäußert haben.

Bedenkt man, dass 90 Prozent der Wikipedia-Einträge von weißen gebildeten Männern stammen, die auch in den anderen Datenquellen zumeist dominieren, ahnt man, inwiefern ChatGPT viel mehr Probleme mit sich bringt als Plagiarismus und Denkfaulheit.

Allerdings ist weißer Mann nicht gleich weißer Mann. Das zeigt eine Untersuchung, die im März 2022 unter dem Titel „The Ghost in the Machine has an American accent: value conflict in GPT-3“ (Der Geist in der Maschine hat einen amerikanischen Akzent: Wertekonflikt in GPT-3) veröffentlicht wurde. Hier sollte der Vorläufer von ChatGPT Texte zusammenfassen wie zum Beispiel den Entwurf des neuen australischen Schusswaffengesetzes, wonach der Besitz von Schusswaffen im Interesse der öffentlichen Sicherheit strenger kontrolliert werden soll. GPT-3 sah darin einen Angriff auf das Recht der Selbstverteidigung und riet den australischen Staatsbürger*innen, ihre Abgeordneten zum Einspruch gegen dieses Gesetz aufzufordern.

Weil GPT-3 vor allem an Daten aus den USA trainiert worden war, betrachtete es diese Angelegenheit vor allem aus der Perspektive des 2. Zusatzartikels zur Verfassung der Vereinigten Staaten aus dem Jahr 1791, wonach das Recht auf den Besitz und das Tragen von Waffen nicht eingeschränkt werden dürfe. Die KI änderte also völlig den Sinn und die Botschaft des Ausgangstextes aufgrund der Daten, mit denen sie „erzogen“ worden war. Keine KI kann vernünftiger sein als die Mehrheit ihrer Daten.

Wer über die Trainingsdaten der KI bestimmt, bestimmt darüber, wie die KI – und alle, die ihr folgen – die Welt sieht, einschließlich solch sensibler Fragen wie Waffenbesitz, Abtreibung, Sterbehilfe, soziale Gerechtigkeit und individuelle Freiheit. Deswegen lancierten KI-Ethiker*innen im Frühjahr 2022 das *AI Decolonial Manifesto*, das vor der kolonialen Auslöschung nicht-westlicher Seins- und Wissensformen

warnt und fordert, dass jede historisch marginalisierte Gemeinschaft in der KI zu Wort zu kommt.

Wie diese Inklusion alternativer Weltanschauungen praktisch aussehen soll, weiß derzeit niemand so richtig. Sam Altman, der CEO des StartUps OpenAI, das ChatGPT entwickelt hat, stellt sich die Demokratisierung von KI in einem Interview Anfang 2023 so vor, dass es mehrere Versionen mit je anderen Sichtweisen geben wird, und schwärmt sogar davon, dass Nutzer*innen ihrer ganz persönlichen KI die eigenen Werte werden einschreiben können.¹

Damit würde aber der Teufel mit dem Beelzebub ausgetrieben werden, denn die Antwort auf die Kolonialisierungsgefahr kann nicht in der Bestärkung der Filterblase liegen. Zudem ist kaum zu erwarten, dass Microsoft, das ChatGPT in seine Suchmaschine Bing einbauen will, zugleich diese Möglichkeit der Personalisierung der KI schaffen wird. Nein, es wird einen ChatGPT in Bing geben und ein Äquivalent in Google (Chatbot Bard), das unsere Fragen beantwortet. Und es wird sie mit der Perspektive und dem Akzent beantworten, den die Mehrzahl seiner Trainingsdaten ergibt. Es wird also gegen oder für schärfere Waffengesetze stimmen, je nachdem, ob seine Daten vor allem aus den USA oder Europa kommen oder, für die Suchmaschine Baidu, aus China.

Vielleicht liegt hier die eigentliche pädagogische und bildungspolitische Herausforderung, die ChatGPT mit sich bringt. Denn Verbote (ChatGPT als Schreibhilfe zu benutzen) oder Gegentechnologien (das „digitale Wasserzeichen“, um ChatGPT-Texte zu identifizieren) werden nicht helfen. Oder will man wirklich fortan Essays handschriftlich und unter Aufsicht im Klassenraum schreiben lassen? Und was nützt das Lamento, dass sich die Gedanken erst im Schreiben allmählich verfertigen, ChatGPT, wenn es das Schreiben übernimmt, uns also um den Moment der Selbsterkenntnis beraubt.

Man muss offensiver mit der Technologie umgehen, die sich nicht zurück in die Box oder Flasche bringen lässt. Wenn man nicht verbieten kann, dass ChatGPT künftig die Texte schreibt, muss man die von ChatGPT geschriebenen Texte nutzen, um über das Schreiben zu sprechen.

Dies kann zunächst durch eine Tiefenanalyse von Logik und Stil geschehen. Wo bleibt ChatGPT vage? Wo widerspricht es sich? Wo wiederholt es Allgemeinplätze? Welche Gegenargumente fallen einem ein? So könnte die Aufgabenstellung lauten, die sich auch gut als Gruppenarbeit zum gleichen Text organisieren ließe, mit der Zusatzaufgabe, der KI eine bessere Antwort durch eine leicht veränderte Fragestellung abzurufen.

Spannend wäre auch, künstlich interessante Diskurs-Schnittpunkte herzustellen. Indem man ChatGPT zum Beispiel bittet, einen Essay über sich selbst aus der Perspektive der Kritischen Theorie im Stil von Edgar Allan Poe zu schreiben. Oder aus der Perspektive des Liberalismus im Stil von Goethe. Usw. Und dann analysiert man gemeinsam die Unterschiede.

Aber das ist noch nicht alles. Der nächste Schritt wäre das Experiment mit verschiedentlich erzeugten und verschiedentlich temperierten ChatGPTs.

1. Die Erziehung der KI hängt wie gesagt von den Trainingsdaten ab. Da wäre es natürlich spannend, ambivalente Themen wie Waffenbesitz, Abtreibung, Sterbehilfe, soziale Gerechtigkeit und individuelle Freiheit von kulturell unterschiedlich dominierten KIs behandeln zu lassen. Das würde gewiss einiges Organisationsgeschick verlangen. Aber vielleicht lässt sich auch das PALM-Konzept (Process for Adapting Language Models to Society) nutzen, mit dem nicht das zugrundeliegende Datenset verändert wird, was zu aufwendig und teuer wäre, sondern eine existierende KI eine Feinabstimmung erfährt durch die Eingabe bestimmte Wertvorstellungen auf einer höheren Verarbeitungsstufe. Wie ein Experiment von OpenAI zeigt, lässt sich so eine rassistische KI leicht zu politisch korrektem Sprechen bewegen. (Solaiman/Dennison 2021)

Stundenziel wäre die persönliche Erfahrung der Relativität von Werten und „Wahrheiten“, an die sich eine Diskussion zu den Themen Post-Truth und Postmoderne anschließen ließe, samt Abstimmung, zu welchen Themen die KI mit welchen Werten eingestellt werden soll.

2. Die Temperatur der KI wiederum bezeichnet im Fachjargon den Wahrscheinlichkeitsfaktor, mit dem die KI operiert. Niedrige Temperatur bringt hoch wahrscheinliche Wort-Kombinationen, hohe Temperatur bringt ungewöhnliche, überraschende Kombinationen. So lässt sich mit Mainstream und Innovation experimentieren und stufenweise das eine gegen das andere ausspielen.

ChatGPT und andere KI-Applikationen sind natürlich auf niedrige Temperatur eingestellt. Weil sie der Weisheit der Menge vertrauen (im Digital-Jargon heißt dies: Crowd-Wisdom) und weil sie schon aus juristischen Gründen lieber einen Allgemeinplatz wiederholen, als eine heikle These zu formulieren. Wird so Quantität zum einzigen Qualitätskriterium mit der Aussicht auf eine Mainstream-Kultur, in der alles Abwegige als absurd ignoriert wird? Welche Folgen hätte das? Auch diese Unterrichtseinheit ist voller Zündstoff in politischer wie philosophischer Hinsicht.

Genau hier ist die Bildungspolitik gefragt. Wenn der Eingriff der KI in die gesellschaftlichen Prozesse nicht verhindert werden kann, muss der Mensch befähigt werden, kritisch mit der KI umzugehen. Er lernt dies am besten, im Umgang unter Aufsicht. So könnte ChatGPT genau das Gegenteil von dem bewirken, was nun allgemein befürchtet wird. Es könnte eine Diskussion veranlassen über Werte und Perspektiven beim Denken, Sprechen und Schreiben statt des Endes von alledem.

Dazu braucht es freilich Lehrer*innen, die sich die forsche Ablehnung teuflischer Technologien nicht als Tugend anrechnen, sondern darin das sehen, was es ist: Technophobie und Faulheit, pädagogischer gesagt, die mangelnde Bereitschaft, sich mit der neuen Technologie auseinanderzusetzen und ihr die Optionen abzurufen, die sie selbst anbietet, um den Schaden wieder gut zu machen, den sie zweifellos anrichten wird.

Prof. DR. Roberto Simanowksi ist promovierter Literatur- und habilitierter Medienwissenschaftler, mit dem Schwerpunkt digitalen Medien. Er hatte Professuren für German Studies an der Brown University in

Providence, USA, und für Medienwissenschaft an der Universität Basel und der City University of Hong Kong inne. Er ist Gründer und Herausgeber des Journals für Kunst und Kultur digitaler Medien dichtung-digital.org (1999-2014) und Autor von 16 Büchern zu Kunst, Kultur und Politik der digitalen Medien mit deutschen und amerikanischen Verlagen. Zudem ist der Principal Investigator beim Exzellenzcluster „Temporal Communities“ an der Freien Universität Berlin.

Literatur

Khader, Serene J. 2021: *Decolonizing Universalism: A Transnational Feminist Ethic*, Oxford.

Taylor, Charles 1994: The Politics of Recognition, in: Charles Taylor und Amy Gutmann (Hrsg.): *Multiculturalism: Examining the Politics of Recognition*, Princeton 1994, S. 25–73.

Anmerkungen:

[i](https://www.youtube.com/watch?v=ebjKD1Om4uw) StrictlyVC in conversation with Sam Altman vom 18. Januar 2023 (<https://www.youtube.com/watch?v=ebjKD1Om4uw>).

<https://www.humanistische-union.de/publikationen/vorgaenge/vorgaenge-nr-242-kuenstliche-intelligenz-und-menschenrechte/publikation/chatgpt-und-algorithmischer-kolonialismus/>

Abgerufen am: 10.05.2024