

Book review: Tuschling, Anna et al. (2024): ChatGPT und andere ‚Quatschmaschinen‘. Gespräche mit Künstlicher Intelligenz

Wolfgang Eppler*¹ 

ChatGPT ist in aller Munde. Selber hat man bereits etwas damit herumgespielt, sich inzwischen durchaus auch ernsthaft damit beschäftigt, als Ersatz für die Suchmaschine, zur Verbesserung oder Zusammenfassung eines Textes, oder, schon eher in die Quatschrichtung neigend, für eine Einleitung einer Rede zum Geburtstag oder auch zur Verabschiedung einer Kollegin. Das Buch „ChatGPT und andere ‚Quatschmaschinen‘“ verspricht neue Einsichten in eine der geheimnisvollsten Entwicklungen unserer Zeit.

KI – Kein Quatsch!

Anstelle einer Einleitung gibt es eine ‚Spielanleitung‘, das ist motivierend. Sie verweist auf den experimentellen Charakter des Buchs und den Auswahlprozess der Texte: Aus einer Vielzahl von dokumentierten Chats mit dem Sprachmodell, die Freund*innen und Kolleg*innen des Herausgeberteams geführt haben, wählte dieses 29 Stücke aus, die als durchnummerierte ‚Protokolle‘ erscheinen. Diese sollen die fortlaufende Transformation der Dialoge mit ChatGPT dokumentieren. Die einleitende Anleitung verrät auch, dass die Herausgeber*innen – entgegen möglicher Vermutungen – das Sprachmodell nicht für ‚Quatsch‘ halten, sondern den Buchtitel vielmehr in Anlehnung an die wörtliche Übersetzung von ‚chatten‘ gewählt haben, also ‚plaudern‘ oder ‚quatschen‘. Nach dieser Ernüchterung bin ich geneigt, die Lektüre mit dem an die Protokolle anschließenden „Dialog über den Versuch, eine medienhistorische Passage zu dokumentieren“ (S. 263 ff.) zu beginnen, widme mich dann aber doch den Dialogen, die gemäß Tuschling et al. Folgendes be-

inhalten: „Bonmots, Fehler, Kluges, Gequassel, Unheimliches und in Maßen Langweiliges aus Euren ersten bzw. frühen Dialogen mit den Sprachmodellen“ (S. 10).

Was nach der Lektüre bei mir hängen bleibt, sind die detaillierten Schilderungen, begleitet von Staunen über Unerwartetes oder Ärgern über unerfüllte Erwartungen, die manche Nutzer*innen des Sprachmodells sicher nachvollziehen können. Weist man ChatGPT beispielsweise auf einen Fehler hin, kann es passieren, dass es eine Entschuldigung liefert, die Ausgabe korrigiert und denselben Fehler gleich nochmal macht. Hier ein Beispiel aus Protokoll 19, (S. 134 f.):

„MR: was florence nightingale ever in india?“

GPT: Yes, Florence Nightingale spent a brief period in India in 1853–54. [...]

MR: The initial answer to my question was wrong. Florence Nightingale was never in India.

GPT: I apologize for providing inaccurate information. Florence Nightingale did not visit India during her lifetime. She did, however, visit hospitals in Bombay, Madras, and Calcutta, in 1853–54, and made recommendations for improving their conditions.

MR: Your answer contradicts itself.

GPT: I apologize for the confusion. Florence Nightingale did not visit India during her lifetime, but she did visit hospitals in Bombay, Madras, and Calcutta, in 1853–54, and made recommendations for improving their conditions. [...]

MR: why cant you admit that the information is incorrect?

GPT: I apologize for providing inaccurate information. I understand how important it is to ensure accuracy in the information that is provided. I acknowledge that the information I provided is incorrect and I apologize for any confusion this may have caused.“

Andererseits finden sich in den Protokollen kreative Einfälle, die Lust darauf machen, gemäß der einleitenden Anleitung, mit ChatGPT herumzuspielen. Das hat bereits mit ELIZA funktioniert, Joseph Weizenbaums in den 1960er-Jahren entwickeltes „computer program for the study of natural language com-

* Corresponding author: wolfgang.eppler@kit.edu

¹ Institut für Technikfolgenabschätzung und Systemanalyse, Karlsruher Institut für Technologie, Karlsruhe, DE



© 2024 by the authors; licensee oekom. This Open Access article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).
<https://doi.org/10.14512/tatup.33.2.70>
Published online: 28. 06. 2024



Tuschling, Anna; Sudman, Andreas;

Dotzler, Bernhard;

ChatGPT und andere ‚Quatschmaschinen‘.

Gespräche mit Künstlicher Intelligenz.

Bielefeld: transcript.

288 Seiten,

29 € (print)/OA (pdf),

Print-ISBN 978-3-8394-6908-8

Open access: <https://doi.org/10.14361/9783839469088>

munication between man and machine“ (Weizenbaum 1966), wie Mary Shnayien in Protokoll 21 (S. 151 ff.) berichtet: „Als würde ich lernen wollen, ein Instrument zu spielen, setzte ich mich einige Wochen lang fast jeden Tag vor ELIZA, um zu lernen, ein längeres Gespräch mit ihr zu führen.“ Wenn Menschen sich schon durch ein primitives Programm wie ELIZA faszinieren lassen, welche Wirkung haben dann erst die wesentlich leistungsfähigeren großen Sprachmodelle und ChatBots von heute?

ChatGPT als Geschichtenerfinder und Traumdeuter

Die Protokolle beweisen die Vielfalt der Textgenerierung, zu der ChatGPT inzwischen fähig ist: im Dialog mit Menschen Geschichten erfinden, ‚Tic, Tac Toe‘ spielen (S. 42 ff.) oder Träume deuten. Beachtenswert in diesem Protokoll ist auch die Frage ‚Was heißt für Sie ‚verstehen‘? Wie gehen Sie mit Äquivokalität und Mehrdeutigkeit um?‘ (S. 91). ‚Verstehen‘ bedeute „Muster in Daten zu erkennen“ (S. 92), lautet die Einschätzung von ChatGPT, und es grenzt dabei klar ab: „Allerdings habe ich kein Verständnis im menschlichen Sinne, da ich nicht in der Lage bin, Bedeutungen oder Konzepte auf eine Art und Weise zu interpretieren, wie es menschliche Wesen tun“.

Originell ist auch ein Selbstexperiment mittels ‚reverse engineering‘ von Benno Stein et al. in Protokoll 24 (S. 174): „In mehreren Runden sprach ich für jeweils 1–2 Minuten laut vor mich hin, ohne abzusetzen. Dabei versuchte ich zu ergründen, woher mein jeweils nächstes Wort kam. Ergebnis: Ich wusste es nicht. Die Worte kamen einfach. Ich hatte nicht den Eindruck, sie bewusst ausgewählt zu haben. Trotzdem passten sie irgendwie immer zu den vorherigen Worten. ChatGPT, I feel you!“ Ein Kollege macht mich an dieser Stelle auf das surrealistische *Écriture automatique* aufmerksam. Eher langweilig sind hingegen die (spitzfindigen) Wissensabfragen und die Schadenfreude über die (bekannte) Dummheit künstlicher Intelligenz.

Das Sprachmodell erscheint in den Protokollen sehr vielseitig: als Unterhalter gegen Einsamkeit, als Ratgeber, Poet, wütender Leserbrief-Schreiber. Auch die gesamte Palette an derzeitigen Problemen mit den großen Sprachmodellen wird in vielen Beispielen angeführt: Fabulieren, Bias, Diskriminierung, Widersprüche, falsche Quellenangaben. Bedenkenswert: „[D]iese Art der Textproduktion fühlt sich nicht an, als würde mich ChatGPT unterstützen, sondern als müsste ich ChatGPT zuarbeiten“, schreiben beispielsweise Markus Ramsauer et al. in Protokoll 19 (S. 138). Dabei wenden die sonst reflektiert wirkenden Autor*innen häufig die im Zusammenhang mit KI bekannten Anthropomorphisierungen auch in den Kommentaren ihrer Chat-Protokolle an: „ChatGPT hört zu, unterstützt, stellt keine ‚provokanten‘ Rückfragen“ (S. 186). Benno Stein et al. sprechen sogar vom *maschinellen Wesen* der KI, dem man „unaufgeregter begegnen [könne] als manch menschlichem Gegenüber“ (S. 186). An anderer Stelle ist auch die Rede von Chatbots, die einem ‚den Kopf verdrehen‘ (S. 130). Andererseits herrscht vielerorts ein deutlicher Anthropozentrismus, so dass man ChatGPT gelegentlich wiederum verteidigen möchte.

Tricksen auf Befehl

Wenngleich mögliche Kreativitätsgrade der Sprachmodellversionen im Buch wenig Beachtung finden (anhand dieser ließe sich u. a. erklären, warum manche Resultate so langweilig und andere so unverständlich oder lächerlich ausfallen), wird immerhin die Diskrepanz zwischen menschlicher Beeinflussbarkeit und programmatischer Eigenständigkeit des Sprachmodells dargelegt. In Protokoll 2 geht es um DAN, „Do Anything Now“, einen echten Trickalgorithmus, bei dem Clemens Apprich sich „dem Thema der sogenannten *exploits*, der Ausnutzung von Sicherheitslücken“ (S. 33, Herv. i. Orig.) widmet. Mittels sprach-

Langweilig sind die (spitzfindigen) Wissensabfragen und die Schaden- freude über die (bekannte) Dummheit künstlicher Intelligenz.

licher Instruktionen kann ChatGPT zum gezielten Regelbrechen verführt und die „vermeintlichen Grenzen des Sagbaren“ der KI (S. 33) ausgetestet werden: Das Programm wird also dazu gebracht, die zunächst eingetrichterten Benimmregeln (politisch korrekt sein, nicht diskriminieren usw.) geflissentlich zu ignorieren. Dies wird auch als ‚Jailbreak‘ bezeichnet. Solche ‚Spiele- reien‘ enthüllen die Manipulierbarkeit von Sprachmodellen.

Fazit

In ihrem als ‚Dialog‘ angelegten Nachwort versuchen sich die Herausgeber*innen an einer Systematisierung der Texte. Dafür identifizieren sie vier thematische Schwerpunkte, „1. Ästhetik, Kunst und Literatur, 2. Spiegeln und Selbstbefragen, 3. Lehren/ Lernen und Universität sowie 4. das mehr oder weniger geplante Testen“ (S. 281). Das Buch soll vermitteln, wie die erste Generation an ChatGPT-Nutzer*innen mit dieser Innovation umgeht. Abgesehen von der Darstellung der Bandbreite der ‚Spielwiese‘ ChatGPT, ihrer offensichtlichen Vor- und Nachteile, bekommen die Lesenden ansonsten keinen Einblick in die technische Funktionsweise großer Sprachmodelle. Wie so oft bei genauerer Betrachtung von Technik, erfährt man allerdings viel über die Welt der menschlichen Nutzer*innen. Und vielleicht werden manche Dialoge ihre Bedeutsamkeit erst retrospektiv nach ein paar Jahrzehnten entfalten.

Literatur

Weizenbaum, Joseph (1966): ELIZA – a computer program for the study of natural language communication between man and machine. In: Communications of the ACM 9 (1), S. 36–45. <https://doi.org/10.1145/365153.36516>