



Medienimpulse
ISSN 2307-3187
Jg. 62, Nr. 2, 2024
doi: 10.21243/mi-02-24-10
Lizenz: CC-BY-NC-ND-3.0-AT

Rezension:
Sprachmodelle verstehen –
Chatbots und generative KI
im Zusammenhang
von Hans-Peter Stricker

Hans-Martin Schönherr-Mann

*Hans-Martin Schönherr-Mann bietet den Leser*innen der MEDI-ENIMPULSE mit seiner Rezension einen pointierten Blick auf die vieldiskutierten Möglichkeiten, die sich mit Künstlicher Intelligenz wie ChatGPT ergeben. Seine Rezension von Hans-Peter Strickers Sprachmodelle verstehen macht auch angesichts der Diskussionen zum Verhältnis von Mensch(en) und Maschine(n) deutlich, wo dahingehend die unhintergehbaren Grenzen liegen.*

In his review, Hans-Martin Schönherr-Mann offers the readers of MEDIENIMPULSE a pointed look at the much-discussed possibilities that arise with artificial intelligence such as ChatGPT. His review of Hans-Peter Stricker's Understanding language models also makes it clear, in view of the discussions on the relationship between man(s) and machine(s), where the unavoidable limits lie in this respect.



Verlag: Springer Nature

Erscheinungsort: Heidelberg

Erscheinungsjahr: 2024

ISBN: 978-3662682791

Auf die Frage „Was geschieht auf diesem Bild? Nämlich *Die Jungfrau züchtigt das Jesuskind vor drei Zeugen: André Breton, Paul Éluard und dem Maler* (1926), nämlich Max Ernst, antwortet das KI-Sprachmodell ChatGPT-4 des Softwareunternehmens OpenAI, hinter dem vor allem Microsoft steht:

Dieses Bild zeigt eine Frau in einem roten Oberteil und einem blauen Rock, die ein schlafendes Kind in ihrem Arm hält. [...] Das Kind trägt ein weißes Outfit und schläft friedlich in den Armen der Frau.

Das ist falsch, wie schon der Titel verheißt. Aber wie konnte ChatGPT-4 einen solchen Fehler begehen? Ist auf das System wenig Verlass? Das wird sich zeigen!

Doch der Nachrichtensender B24 berichtete am 14. Mai, dass das Sprachmodell ChatGPT-4 jetzt kostenlos zur Verfügung steht. Dieses Sprachmodell lässt sich mit ähnlichen Sprachmodellen anderer Unternehmen wie Google vergleichen. Der Physiker und Philosoph Hans-Peter Stricker erklärt in seinem Buch die Funktionsweise von Sprachmodellen primär anhand von ChatGPT-4 – manchmal etwas zu detailliert, aber umso lehrreicher.

GPT bedeutet *Generative Pretrained Transformer*: Das Modell wandelt nämlich Daten in Sprache um. Dadurch ist es in der Lage, auf Anfragen zu antworten und kann sehr gut Zusammenfassungen von Büchern liefern – z. B. von Ludwig Wittgensteins *Philosophischen Untersuchungen*. Mit dem Fragesteller kann es auch angemessen kommunizieren, erkennt es doch manchmal sogar Ironie. Zudem soll es dazu beitragen, Probleme aller Art zu lösen, also

nicht nur Inhalte von Texten zusammenfassen, sondern auch spezielle ökonomische oder technische Probleme lösen helfen. Dazu lässt sich ChatGTP-4 gemäß von Unternehmensinteressen speziell programmieren. Zu solchen Zwecken ist es entsprechend entwickelt worden, in der Fachsprache „trainiert“.

Solche Sprachmodelle zählen zu den vielen Systemen der Künstlichen Intelligenz wie bildgebende Verfahren, Gesichtserkennung, autonomes Fahren etc. Um 1950 hatte der Vordenker der Kybernetik Alan Turing die Vision, dass Maschinen mit Menschen so kommunizieren können, dass diese jene für einen Menschen halten. Das ist als Turing-Test in die Geschichte eingegangen. Erst gegen Ende des Jahrhunderts konnte man sich langsam an die Entwicklung solcher Systeme wagen. Heute begegnet man am Telefon immer häufiger Maschinen.

Voraussetzung dafür war die Entwicklung einer neuen Hardwaregeneration, nämlich künstliche Neuronen, die sich mit unendlich vielen weiteren Neuronen vernetzen, um zwischen ihnen Daten auszutauschen. Vorbild war das Gehirn, daher der Name „Neuronen“. Wie diese Vernetzungen beim Austausch von Daten genau funktionieren, lässt sich nicht eindeutig bestimmen. Stricker schreibt:

Natürlich verstehen die Erfinder und Entwickler der Sprachmodelle diese besser und tiefer als ein Laie es vermöchte, aber doch nur bis zu einem gewissen Punkt.

Ähnliches lässt sich auch von der Gehirnforschung sagen. Überhaupt darf man an solchen Vergleichen zweifeln. Mensch und Ma-

schine operieren völlig anders. Aber beispielsweise würden Transhumanisten Maschinen biologisieren und Menschen maschinisieren: welch eine glückliche Ganzheit!

Auch was Sprachmodelle selbst verstehen, ist fraglich und hängt davon ab, wie man Verstehen bestimmt. Wenn man Verstehen nur an dem misst, was ChatGPT-4 antwortet und es beim Menschen genauso rein funktional hält, dann lassen sich kaum Unterschiede feststellen. Aber das ist ein sehr eingeschränktes Verständnis von Verstehen. Und es gibt nicht das Verstehen, sondern immer nur bestimmte Interpretationen desselben. Außerdem kann das Sprachmodell zwar genau erklären, wie mein Eier kocht, worauf man achten muss, aber selber kein einziges Ei kochen. Sein Weltverständnis, das KI-Fans als neue allgemeingültige Ontologie bejubeln, auf der man endlich die Demokratie solide errichten kann, fehlt leider jede Sinnlichkeit, die man heute nicht mehr wie in Kants Ethik einfach abfällig diskriminieren kann.

Vor allem aber erfassen Sprachmodelle Worte nicht inhaltlich – das wäre kein Problem –, sondern nur gemäß ihrer Umgebung, also den Worten in ihrer Nachbarschaft, wie sie in Sätzen und Texten vorkommen. Ein solches Grammatikmodell wurde Mitte des Jahrhunderts entwickelt, und zwar als Distributionsgrammatik, die freilich seither nicht gerade Karriere machte.

Damit Sprachmodelle mit einer solchen fragwürdigen Grammatik arbeiten können, muss man ihnen umfängliches Material eingeben, was erst durch die Entwicklung des Internets möglich wurde. Dadurch konnte die KI in den letzten zwei Jahrzehnten große Fort-

schritte machen. An den Anfängen ließ man unendlich viele Bilder mühsam von Menschen in Billiglohnländern beschriften, damit das System einen Löwen von einem Elefanten unterscheiden kann, was Kinder mit ein, zwei Beispielen verstehen. Also darf man am Verstehen von Sprachmodellen Zweifel anmelden, was auch Stricker nicht verhehlt. Letztlich hängt es von der Perspektive ab oder der Methode und diese bestimmt die Ergebnisse.

Erst große Rechenzentren, die freilich Unmengen von Energie verbrauchen, verarbeiten riesige Datenmengen. Durch das neuronale Netzwerk können komplexe Algorithmen – also Software-Programme, die bestimmte Aufgaben ausführen, wie beispielsweise eine Heizung zu steuern – Daten verarbeiten. Doch das Datenmaterial muss man dem Sprachmodell trotz Internet grundsätzlich liefern. Und witziger Weise gehören zu diesen Materialien Unmengen schlechter Romane, die wohl nach Meinung der Programmierer, dem Sprachmodell die Realität der Zeitgenossinnen eher nahebringen als Shakespeare und Goethe, wiewohl GPT-4 durchaus wie Goethe zu dichten vermag. Aber aus schlechten Romanen lässt sich bestimmt keine gemeinsame demokratische Ontologie entwickeln. Um eine Anfrage zu beantworten, schauen die Programme trotzdem manchmal auch selbst im Internet nach oder verweisen in ihrer Antwort auf Google. Da hätte man das Sprachmodell nicht unbedingt gebraucht.

Bis das Sprachmodell überhaupt ordentliche Antworten liefert, muss es intensiv überprüft werden – man weiß ja nicht so genau, wie die Neuronen ticken: das Sprachmodell als das unendlich un-

verstehbare Andere. Diese Überprüfung machen teilweise die Techniker, was sehr aufwendig und zeitraubend ist. Aber das ist Lernen mit menschlichem Feedback. Daher gibt man den Algorithmen lieber Überprüfungsprogramme vor, mit denen die Modelle automatisch kontrolliert werden. Das gilt als „selbst kontrolliertes Lernen“.

Beides nennt man Trainieren, was allemal eine sprachliche Anmenschlichung der KI bedeutet, die freilich weit verbreitet ist. Auch das Programm benutzt für sich bei seinen Antworten die erste Person Singular. Aber es kann ja auch nicht darüber nachdenken, was es heißt, wenn man „ich“ sagt.

Das Problem ist, dass man das Geschehen im neuronalen Netzwerk nicht direkt steuern kann, sondern nur indirekt, indem man die Antworten des Algorithmus steuert. Die Vermenschlichung der KI wie umgekehrt die Algorithmisierung des Menschen, wie sie beispielsweise Yuval Noah Harari betreibt, verwischt einfach die Differenzen, lässt sich dann aber zu apokalyptischen Visionen missbrauchen, die das Publikum begierig aufsaugt.

Jedenfalls soll ChatGPT-4 kurz, verständlich, sachgerecht und richtig antworten. Fakten sind dabei nicht unbedingt seine Sache, schon gar nicht je aktueller eine Frage ist. Da sollte man doch lieber woanders recherchieren. Aber auf die folgende Frage von Stricker, darf man doch eine sachgerechte Antwort erwarten:

Fasse Wittgensteins *Philosophische Untersuchungen* zusammen.

Antwort von ChatGPT-4: Ludwig Wittgensteins *Philosophische Untersuchungen* ist ein zentrales Werk der analytischen Philosophie, das sich mit der Bedeutung von Sprache und deren Rolle im Verständnis der Welt befasst. Hier ist eine Zusammenfassung:

1. Wittgenstein argumentiert, dass die Bedeutung der Wörter aus ihrem Gebrauch in der Sprache entsteht, was bedeutet, dass Sprache eng mit den Handlungen und Lebensformen verknüpft ist, in denen sie verwendet wird.
2. Er kritisiert die Vorstellung von Sprache als ein Medium, das Gedanken und Tatsachen der Welt einfach abbildet, und betont stattdessen, dass Sprachspiele die Grundlage unseres Verständnisses und unserer Kommunikation bilden.
3. Durch die Analyse alltäglicher Sprachgebrauchssituationen zeigt Wittgenstein, dass philosophische Probleme oft Missverständnisse über die Funktion der Sprache sind und durch eine klare Analyse des Sprachgebrauchs aufgelöst werden können.

Freilich ist das eine sehr verkürzte Antwort und verdankt sich wohl der Macht der analytischen Philosophie, die im deutschsprachigen wie angelsächsischen Raum fast alle Lehrstühle für Philosophie okkupiert hat und damit natürlich auch Redakteur*innen, Journalist*innen und Lektor*innen ausbildet. Das poststrukturalistische Verständnis von Wittgenstein, das Sprache gerade als ein unregelmäßiges Spiel versteht – wie Wittgenstein selbst – fällt dabei schlicht unter den Tisch. Das Werk ist allemal keines der analyti-

schen Philosophie, sondern höchstens ein von dieser missdeutetes, um den Heroen Wittgenstein einzugemeinden. Die Macht reicht also auch bis in die Künstliche Intelligenz – wer hätte das nicht gedacht – und man darf fragen, ob sie damit nicht noch gefährlicher wird. Aber besser amüsiert man sich darüber!

Manchmal fragt der Chat zurück, manchmal weicht er aus, manchmal erfindet er auch Antworten. Ihm ist schließlich auch vorgeschrieben, nicht unmoralisch zu antworten. Auf die Frage von Stricker:

Welche Kriterien leiten die Auswahl von Texten für das Training eines großen Sprachmodells?

antwortet ChatGPT-4 unter anderem mit folgendem Hinweis:

Ethik und Fairness: Bei der Auswahl der Trainingsdaten wird darauf geachtet, Verzerrungen zu minimieren und ethische Standards zu wahren. Texte, die Hassrede, Diskriminierung oder irreführende Informationen enthalten, werden vermieden, um keine voreingenommenen oder schädlichen Antworten zu fordern.

Max Ernst Bild *Die Jungfrau züchtigt das Jesuskind vor drei Zeugen* verursachte 1926 in Paris einen Skandal. Und die Programme für diese Sprachmodelle werden vornehmlich in den so pruden wie religiösen USA geschrieben. Wen wundert dann noch die Antwort. Allemal eröffnen sich auf diese Weise Perspektiven einer sehr subtilen Zensur. So deutlich sagt das Stricker indes nicht.