

Wie funktioniert ChatGPT und wieso ist das interessant für die Kanonisierung von Literatur?

Patricia Heise

ChatGPT ist ein von OpenAI entwickelter Chatbot, der Antworten auf Fragen generiert, die Nutzer:innen dem Bot stellen. Im Vergleich zu bspw. Google oder anderen Suchmaschinen aber verweist er nicht nur auf andere Seiten, sondern gibt einen Fließtext in gewünschtem Stil aus. Und auf einmal können Hausarbeiten, Kurzgeschichten oder Interviewfragen in wenigen Sekunden generiert werden. KIs wie ChatGPT haben für die Wissenschaft daher viele neue Fragen aufgeworfen:

Welchen Wert hat Forschung noch, wenn ein Computer in einer Minute einen Artikel erstellen kann, für den sonst Monate oder gar Jahre geforscht werden muss? Wie geht man mit KIs in der Lehre um? Sollte man sie verbieten oder kann man sie sogar nutzen?

Dieser Blogbeitrag stellt sich die Frage, welchen Einfluss die Kanonisierung von Literatur auf ChatGPT hat.

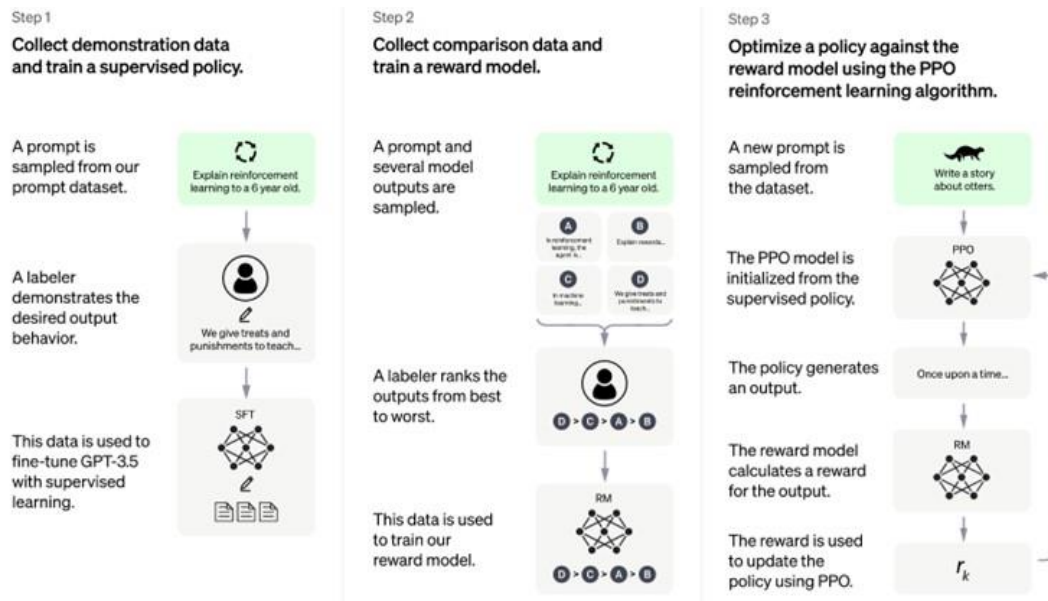
Dafür sollte man zunächst oberflächlich verstehen, wie ChatGPT funktioniert. Auch wenn die Mathematik hinter künstlicher Intelligenz komplex ist, ist das Grundprinzip simpel: Man trainiert ein Computerprogramm dahin, die gewünschten Ergebnisse zu geben.

Dafür stellt ein Mensch eine Beispielfrage und ein weiterer Mensch gibt eine Beispielantwort darauf, um dem Programm zu zeigen, wie eine Antwort aussehen kann. In einem nächsten Schritt schreiben Menschen mehr Beispielantworten und andere Menschen bewerten diese Antworten nach ihrer Qualität.

Dann werden Daten, in diesem Fall von Menschen geschriebene Texte, in das Programm eingelesen. Der Computer berechnet anhand eines von Menschen entwickelten Algorithmus Regeln und Muster, die in den Texten vorkommen (z.B. welche Wörter oft zusammen vorkommen wie etwa *Insel, Wasser, Strand* etc.). Was diese Wörter bedeuten, weiß

das Programm nicht, denn es hat kein inhaltliches Textverständnis; das Programm *liest* Texte nicht, es *rechnet*.

Stellt jemand eine Frage, erstellt das Programm unter Rückgriff auf die Beispielantworten und ihrer Bewertung eine Antwort basierend auf den Regeln, die es sich aus den eingelesenen Textdaten abgeleitet hat. Das heißt, das Programm berechnet, was die wahrscheinlichste Wortkombination ist, die als Antwort zu der Frage passt.



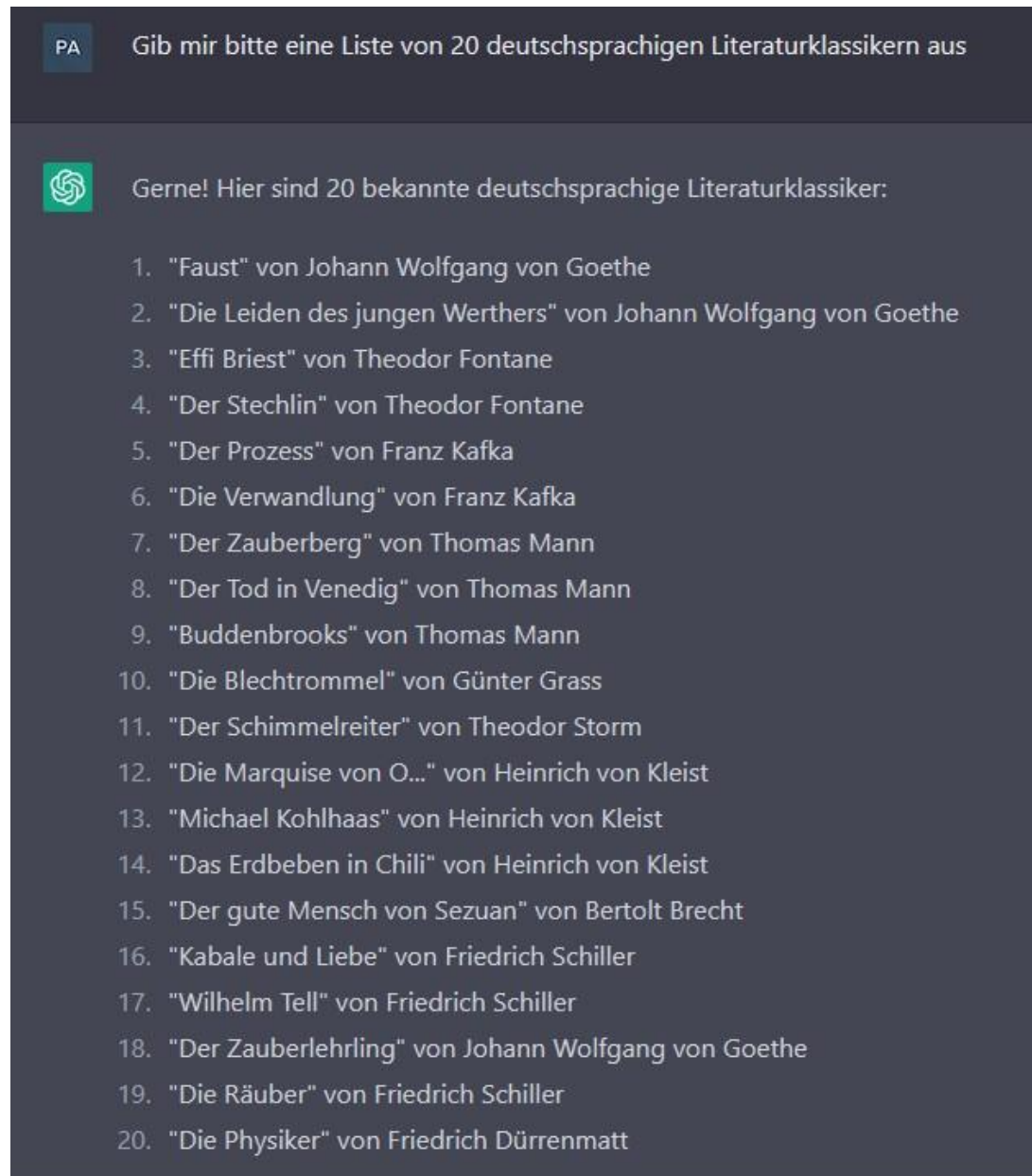
Grafik 1 (<https://openAI.com/blog/chatgpt>)

Es gibt zwei Fallen, in die Benutzer:innen im Umgang mit den Antworten leicht tappen können: Erstens: Künstliche Intelligenz ist nicht objektiv, nur weil es sich um einen Computer handelt. Die Textdaten, anhand derer das Programm trainiert wird, sowie deren Bewertung, sind von Menschen gemacht. Das Programm repliziert lediglich die Regeln, die es angelernt bekommen hat.

Zweitens: Die wahrscheinlichste Antwort ist nicht immer die richtige Antwort. Gerade bei wenigen (oder keinen) vorhandenen Daten, generiert beispielsweise ChatGPT eine Antwort, die zu den Regeln passt, aber die nicht stimmen muss.

Und mit diesen zwei Überlegungen springen wir zurück zum Kanon.

Fragen wir ChatGPT nach deutschsprachigen Literaturklassikern, bekommen wir in Sekundenschnelle eine Liste ausgegeben. Wenig überraschend sind alle dieser Texte von Männern geschrieben; altbekannte Namen wie Goethe, Schiller oder Kleist werden gleich mehrmals genannt, insgesamt gibt es auch nur zehn Autoren bei zwanzig Texten.



PA Gib mir bitte eine Liste von 20 deutschsprachigen Literaturklassikern aus

Gerne! Hier sind 20 bekannte deutschsprachige Literaturklassiker:

1. "Faust" von Johann Wolfgang von Goethe
2. "Die Leiden des jungen Werthers" von Johann Wolfgang von Goethe
3. "Effi Briest" von Theodor Fontane
4. "Der Stechlin" von Theodor Fontane
5. "Der Prozess" von Franz Kafka
6. "Die Verwandlung" von Franz Kafka
7. "Der Zauberberg" von Thomas Mann
8. "Der Tod in Venedig" von Thomas Mann
9. "Buddenbrooks" von Thomas Mann
10. "Die Blechtrommel" von Günter Grass
11. "Der Schimmelreiter" von Theodor Storm
12. "Die Marquise von O..." von Heinrich von Kleist
13. "Michael Kohlhaas" von Heinrich von Kleist
14. "Das Erdbeben in Chili" von Heinrich von Kleist
15. "Der gute Mensch von Sezuan" von Bertolt Brecht
16. "Kabale und Liebe" von Friedrich Schiller
17. "Wilhelm Tell" von Friedrich Schiller
18. "Der Zauberlehrling" von Johann Wolfgang von Goethe
19. "Die Räuber" von Friedrich Schiller
20. "Die Physiker" von Friedrich Dürrenmatt

Und das ist die Magie – und die Falle – bei KIs. Sie generieren Antworten basierend auf dem Material, das sie gefüttert bekommen haben – das aus menschengemachten Texten und ihren menschlichen Bewertungen besteht. Die eingespeisten Texte selbst sind Grundlage für die Antworten, denn sie sind das Material, aus dem der Bot seine Regeln erstellt. Werden in einem Text (ob Zeitungsartikel, Twitter-Post oder Journal) Titel im

Kontext von Wörtern wie *Klassiker*, *Hochliteratur*, *must-read*, *Weltliteratur*, etc. genannt, wird der Chatbot diese Titel auch ausgeben, wenn man nach eben diesen Wörtern fragt.

Aber auch wenn wir diesen konkreten Begriff des *Klassikers* beiseitelegen wollen und ChatGPT fragen, welche deutschsprachigen Bücher man gelesen haben sollte, bekommt man eine sehr ähnliche Liste mit ein bisschen Gegenwarts- und/oder Jugendliteratur von weiteren weißen, männlichen Autoren eingesprenkelt.

Zusätzlich werden die Trainer:innen diesen Antworten wohl ein hohes Ranking geben, wenn sie Goethe und Schiller in einer Literaturliste sehen, weil es die Namen sind, die ohnehin überall stehen, und sie nicht unbedingt reflektieren, dass die Auswahl sehr beschränkt ist. OpenAI bemüht sich zwar laut eigener Aussage darum, diverse Trainer:innen einzustellen, wird aber auch so keine völlige Neutralität in der Bewertung herstellen können.

Ein nicht überraschender Rückschluss aus der generierten Liste der Literaturklassikern ist also, dass die Werke Goethes, Schillers, Kleists etc. in Texten oft als Klassiker, lesenswert o.Ä. bezeichnet werden. Das heißt aber auch, dass generell hauptsächlich Literatur von Männern in Texten besprochen wird und im Umkehrschluss also auch, dass Literatur von Autorinnen weniger besprochen wird. ChatGPT macht so in seinen kurzen Antworten noch einmal die verhältnismäßig geringe Rezeption von Autorinnen und ihrem Werk deutlich.

Denn wie soll eine KI Informationen ausgeben, die sie nicht hat?

Texte von Frauen sind in der Forschung immer noch unterrepräsentiert, sie kommen kaum in Lehrplänen vor, die mittlerweile im Internet zu finden sind oder bekommen selten Museumsausstellungen gewidmet, die Werbetexte online stellen. Viele Autorinnen sind in der Zwischenzeit vergessen worden, ihre Texte wurden nie digitalisiert, die Forschung zu ihnen und ihrem Werk ist kaum oder gar nicht vorhanden.




Sucht man beispielsweise im Onlinekatalog der Universität Frankfurt nach Helmina von Chézy, bekommt man 152 Ergebnisse für Bücher und 599 für Artikel angezeigt, bei Sibylla Schwarz 41 Bücher und 3.424 Artikel, selbst Bettina Brentano (unter dem Namen

Bettina von Arnim) schafft es auf nur 745 Bücher und 3.442 Artikel (im Verhältnis dazu Johann Wolfgang Goethe mit 34.203/346.121, Friedrich Schlegel 3.184/33.200).

Zwar können Trainer:innen die Texte von Frauen höher gewichten, indem man Antworten mit Frauen besser bewertet. Erschwert wird das zusätzlich dadurch, dass einige Frauen unter (männlichen) Pseudonymen veröffentlicht haben, die von den Trainer:innen als solche erkannt werden müssten, um die Antwort höher zu gewichten. Aber auch das macht die fehlende Vielfalt der Themen, die allein mit der Menge von Forschung einhergeht, nicht wett. Das Programm hat zu wenige nutzbare Daten, um eine passende, umfangreiche und vor allem fehlerfreie Antwort zu generieren.

Fragt man bspw., ob *Herr der Ringe* diskriminierende Stereotype reproduziert, bekommt man eine kurze Antwort darüber, dass es problematisch sei, dass Orks als böse dargestellt würden und dass einige Menschengruppen als besser oder schlechter dargestellt würden.

PA Reproduziert Herr der Ringe diskriminierende Stereotypen? Und wenn ja, welche und wie?

 Es gibt Kritiker, die argumentieren, dass "Der Herr der Ringe" von J.R.R. Tolkien bestimmte diskriminierende Stereotypen reproduziert. Insbesondere gibt es Vorwürfe bezüglich der Darstellung von Rassen und Kulturen in der Welt von Mittelerde.  

Zum Beispiel wird häufig argumentiert, dass Tolkien's Darstellung der Orks, einer oft als "böse" dargestellten Rasse, problematisch ist. Einige Kritiker sehen in dieser Darstellung eine Parallele zu rassistischen Stereotypen, insbesondere weil die Orks oft als grobschlächtig, barbarisch und primitiv dargestellt werden.

Auch gibt es Vorwürfe bezüglich der Darstellung von Ethnien und Kulturen, insbesondere der Darstellung von Menschen. So gibt es in "Der Herr der Ringe" eine Unterscheidung zwischen verschiedenen Menschenrassen, die auf bestimmte Charaktereigenschaften reduziert werden, wie zum Beispiel die als "edel" beschriebenen Dúnedain im Gegensatz zu anderen Menschenrassen. Kritiker sehen hierin eine mögliche Reproduktion von rassistischen Stereotypen.

Es ist jedoch wichtig zu betonen, dass es auch viele Leser und Fans gibt, die Tolkien für seine Schöpfung einer reichhaltigen Fantasy-Welt und für seine Schreibkunst bewundern und keine Diskriminierung darin sehen. Es ist immer eine Frage der Interpretation, ob und inwieweit bestimmte Werke diskriminierende Stereotypen reproduzieren.

Dass die Zwerge bei *Herr der Ringe* jüdisch konnotiert sind und sterben, weil sie zu tief nach Gold graben und dadurch einen Dämonen befreien oder dass die wunderschönen

und weisen Elfen alle weiße Haut haben, wird nicht erwähnt. Dass die weiblichen Figuren keine Agency bekommen und hauptsächlich als romantisches Interesse funktionieren, fehlt auch.

Interessant ist nicht nur, was ChatGPT generiert, es ist auch interessant, was es *nicht* generiert. Feministische oder postkoloniale Interpretationen werden nicht auf Anhieb geliefert und sind auch auf Wunsch nur oberflächlich und oft fehlerhaft.

Wie soll eine KI Informationen ausgeben, die sie nicht hat? Das bezieht sich sowohl auf Primär- wie Sekundärliteratur. Je mehr Informationen es zu einem Thema gibt, desto genauer und detaillierter wird auch die Antwort. Scheinbar wurden nur verhältnismäßig wenige Texte in das Programm eingespeist, die die Stereotypisierung in Herr der Ringe behandelt hat.

Warum ist das alles nun relevant?

ChatGPT hatte nach zwei Monaten bereits über 100 Millionen User, die generische und unreflektierte Analysen und Literaturlisten generiert bekommen, wenn sie nach einer Interpretation oder Leseempfehlung fragen.

Wie groß der Einfluss von KIs auf die weitere Verfestigung von Kanons ist, kann man noch nicht wissen, aber es würde sich lohnen, das im Auge zu behalten – allein schon um zu beobachten, wie sich die Rezeption von Literatur verändert.

Vielleicht ist es auch ein motivierender Gedanke, dass in der Forschung über Literatur von Frauen und PoC (oder irgendwelchen Menschen, die keine weißen Männer sind) und mit feministischen Interpretationen vorerst viel zu erforschen bleibt, das nicht mit einem Mausclick von einem Chatbot erstellt werden kann.

Und mit jedem Artikel, den man dazu schreibt und der in KIs eingespeist wird, steigt die Wahrscheinlichkeit ein winziges bisschen an, dass die Namen auch mal in Literaturlisten von Chatbots ausgespuckt werden.

Patricia Heise studiert Sozialwissenschaften des Sports und Deutsche Literatur an der GU Frankfurt.