



Leseprobe aus De Florio-Hansen, KI-Tools für den Unterricht,
ISBN 978-3-407-63310-1 © 2024 Beltz Verlag, Weinheim Basel
[http://www.beltz.de/de/nc/verlagsgruppe-beltz/
gesamtprogramm.html?isbn=978-3-407-63310-1](http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-407-63310-1)

Inhalt

Vorwort	8
----------------------	---

Teil I: Die Bedeutung von KI-Sprachmodellen für Schule und Unterricht	9
--	---

1. Funktionsweise und Nutzungsmöglichkeiten von KI-Sprachmodellen	11
1.1 Warum Verbote keine Lösung sind	11
1.2 Zur Funktionsweise von KI-Sprachmodellen	13
1.3 Auswirkungen auf Gesellschaft und Berufswelt	18
1.4 Lehren und Lernen mit KI-Sprachmodellen	24
2. Zielführende Prompts verfassen	30
2.1 Definition	30
2.2 Beispiel für eine Generierung	30
2.3 Hilfen beim Erstellen von Prompts	33
2.4 Weitere Beispiele für Prompts	36
2.5 Dem Ziel mit Geduld näherkommen	39
2.6 Zero-Shot Prompting und Few-Shot Prompting	44

Teil II: Zum Umgang von Lehrpersonen und Lernenden mit KI-Sprachmodellen	45
---	----

3. Zum Leistungsspektrum von KI-Sprachmodellen	47
3.1 Ziele von KI-Sprachmodellen	47
3.2 Kurzgeschichten im Dialog verfassen	48
3.3 Ein Lerntagebuch führen	50
3.4 Unterstützung bei der Bildung von Kleingruppen	53
3.5 Vorschläge für den Umgang mit Unterrichtsstörungen	55

3.6	Die Kommunikation mit einzelnen Lernenden verbessern	58
3.7	Angst und Stress bei Prüfungen reduzieren	60
4.	Grundlegende Anwendungsmöglichkeiten für den Unterricht	62
4.1	Voraussetzungen für die Nutzung von KI-Sprachmodellen und LGAIMs (<i>Large Generative AI Models</i>)	62
4.2	Was Lehrkräfte bei der Vorbereitung und Durchführung von Unterricht bedenken sollten	63
4.3	Eine vermeintlich leichte Aufgabe für ChatGPT	64
4.4	ChatGPT und die Lernenden testen	69
4.5	Kann ChatGPT Mathematik?	71
4.6	Eine Probe aufs Exempel	74
4.7	Prompts zum Ausprobieren	77
5.	Weitere Nutzungsmöglichkeiten	79
5.1	Multiple-Choice-Tests erstellen	79
5.2	Versuch, mit ChatGPT eine Mindmap zu generieren	81
5.3	Visuelle Lernhilfen	83
5.4	Sich über Unterrichtsstrategien und -techniken informieren	85
5.5	Fragen will gelernt sein	87
6.	Die Lernenden mit KI-Sprachmodellen vertraut machen	91
6.1	Grundlagen	91
6.2	Die etwas andere Einführung	93
6.3	Prompting für Lernende	96
6.4	Folgeprompts nutzen	98
6.5	Einen individuellen Schreibstil entwickeln	99
	Teil III: Der Hype geht weiter – wie damit umgehen?	103
7.	Neuere KI-Sprachmodelle sowie weitere Generative Systeme ..	104
7.1	Von OpenAI zu ClosedAI	104
7.2	Veränderungen und Verbesserungen von ChatGPT	105

7.3	KI-Sprachmodelle	107
7.4	Large Generative AI-Models: mehr als Sprache	114
7.5	Absehbare Entwicklungen	116
8.	Was wir tun können	118
	Glossar	120
	Literaturverzeichnis	124

Teil I:

Die Bedeutung von KI-Sprachmodellen für Schule und Unterricht

In den folgenden beiden Kapiteln gehe ich in gebotener Kürze auf die Funktionsweise sowie Möglichkeiten und Grenzen von KI-Sprachmodellen sowie vorhersehbare Entwicklungen ein. **Kapitel 1** führt in wichtige Grundlagen ein, deren Kenntnis für eine sinnvolle Nutzung von KI-Sprachmodellen in Schule und Unterricht unerlässlich ist. Ausgehend von einem fiktiven Gespräch zwischen zwei Lehrpersonen (**1.1**) geht es im folgenden Abschnitt (**1.2**) um die technologischen Voraussetzungen, die es ermöglichen, mithilfe einer KI (wie beispielsweise ChatGPT) menschenähnliche Sprache zu generieren, besser gesagt: mit einem Chatbot in einen Dialog einzutreten. Bei vielen Generierungen sind inhaltliche Verzerrungen, aber auch sprachliche Unzulänglichkeiten nicht zu übersehen. In **1.3** geht es um erwartbare Auswirkungen der sich rasch weiterentwickelnden KI auf Gesellschaft und Arbeitsplätze. Was müssen Schülerinnen und Schüler lernen, um in ihrem späteren (Berufs-)Leben sinnvoll mit KI-Systemen umzugehen? Im letzten Abschnitt (**1.4**) werden einige Unzulänglichkeiten der Generierungen an konkreten Beispielen aufgezeigt, bei denen auch mehrere Prompts, sogenannte Folgeprompts, nicht zu einem zufriedenstellenden Ergebnis führten. Wenn Schülerinnen und Schüler sich auf ChatGPT oder eine andere KI-Anwendung stützen, sind zudem Fragen des Copyrights und des Datenschutzes zu bedenken. Deshalb müssen die Lernenden angehalten werden, ihre Quellen in geeigneter Form zu benennen. Die wichtigste Möglichkeit, Täuschungsversuche zu unterbinden oder zumindest einzuschränken, besteht in einer grundlegenden Änderung der Aufgabenstellungen.

In **Kapitel 2** geht es um die schwierigste, aber auch höchst interessante Aufgabe, ohne die von einem KI-Sprachmodell nichts zu erfahren ist, nämlich die sogenannte Eingabeaufforderung, den Prompt. Nach einer Definition (**2.1**) wird zunächst an einem Beispiel gezeigt, welches das Hessische Kultusministerium in ei-

ner Handreichung ausführlich kommentiert hat, was man beim Prompting bedenken muss. Auch wenn man noch so präzise formuliert, muss der Output nicht korrekt sein. Das kann an den Verzerrungen und Unzulänglichkeiten liegen, die von den Daten herrühren, mit denen ChatGPT trainiert wurde. Man kann nicht oft genug darauf hinweisen, dass ChatGPT ausschließlich mit englischsprachigen Daten trainiert wurde. Das wird deutlich an einem Beispiel im Zusammenhang mit Wolfgang Borcherts Kurzgeschichte: »Nachts schlafen die Ratten doch« (2.2). Verschiedene wissenschaftsorientierte Publikationen benennen wichtige Grundregeln. Sie sind aber nur eingeschränkt brauchbar, weil sie selten auf Deutsch vorliegen und sich nur in Ausnahmefällen auf schulischen Unterricht in deutschsprachigen Ländern beziehen. Das folgende englischsprachige Beispiel zu Harper Lees »To Kill a Mockingbird« eignet sich jedoch sehr gut für den Englischunterricht (2.3). Es ist ratsam, beim Prompting (gleichgültig ob auf Deutsch oder Englisch) möglichst kleinschrittig vorzugehen (2.4). Das zeigt sich auch bei einem Vergleich zwischen der Suchmaschine Google Drive und ChatGPT zur Unterrichtsmethode *Flipped Classroom*, die häufig bei der Nutzung von KI-Sprachmodellen eingesetzt wird (2.5). Einen Gegensatz zum üblichen Prompting, das auch Zero-Shot Prompting genannt wird, bildet das Few-Shot Prompting, bei dem KI-Sprachmodelle mit knappen Antworten reagieren (2.6).

1. Funktionsweise und Nutzungsmöglichkeiten von KI-Sprachmodellen

1.1 Warum Verbote keine Lösung sind

Als Susanne B. das Lehrerzimmer betritt, hält sie nach ihrem Kollegen Frederik M. Ausschau, denn sie hat etwas wegen einer Klassenarbeit mit ihm zu besprechen. Glücklicherweise ist neben dem Kollegen noch ein Sitzplatz frei. Aber was ist das? Der sonst so frohgemute Frederik blickt höchst unglücklich drein.

Susanne B.: Hallo, was ist los? Du siehst richtig unglücklich aus. Bist du krank?

Frederik M.: Nein, aber ich mache mir Gedanken ...

Susanne B.: Gedanken ... worüber?

Frederik M.: Dieser Chatbot regt mich auf.

Susanne B.: Du meinst ChatGPT? Warum regt dich das Tool auf?

Frederik M.: Am liebsten wäre mir, es würde an Schulen ganz verboten. Dann bräuchte ich mir keine Gedanken zu machen.

Susanne B.: Meiner Meinung nach wäre das der falsche Weg. Aber ich gebe zu, dass die Nutzung sehr zeitaufwendig ist, obwohl immer wieder von Zeitersparnis geredet und geschrieben wird.

Frederik M.: Nicht nur das. Die Schülerinnen und Schüler lernen das Schreiben nicht, das heißt, sie bilden keinen individuellen Stil aus.

Susanne B.: Da hast du sicher recht; aber dem müssen wir entgegenwirken, und das erreichen wir nicht, wenn wir die Nutzung im Unterricht oder beim Erledigen der Hausaufgaben verbieten. Hast du dir einmal überlegt, was passieren würde, wenn ChatGPT und ähnliche KI-Modelle aus der Schule verbannt würden?

Frederik M.: Darüber sollen sich die anderen Gedanken machen, die Eltern zum Beispiel.

Susanne B.: Das bringt doch nichts.

Frederik M.: Was befürchtest du denn?

Susanne B.: Zunächst einmal würde das Schummeln weiter zunehmen.

Frederik M.: Gibt es denn keine Detektoren?

Susanne B.: Doch, aber sie funktionieren nur höchst eingeschränkt. Wenn wir erreichen wollen, dass unsere Schülerinnen und Schüler einen in-

dividuellen Stil ausbilden, müssen wir ihnen zeigen, in welchen Fällen KI-Sprachmodelle nützlich sein können und wo solche statistischen Systeme an ihre Grenzen stoßen.

Frederik M.: Mich stören auch die vielen unpassenden und falschen Outputs, die in wohlklingender Sprache daherkommen.

Susanne B.: Richtig! Folglich müssen wir den Schülerinnen und Schülern aufzeigen, warum solche Modelle Grenzen haben und wie sie sie erkennen können. Vor allem aber müssen sie lernen, dass ein KI-Sprachmodell eine technische Errungenschaft ist, dessen Leistungen nicht mit menschlicher Intelligenz verglichen oder gar gleichgesetzt werden können.

Frederik M.: Vermutlich hast du recht. Aber wie soll man solche Einsichten fördern, wenn es Dutzende von Blogposts und YouTube-Videos gibt, die ChatGPT und ähnliche Tools über den grünen Klee loben?

Susanne B.: Ich gebe zu, dass das viel Zeit kostet, aber es geht ja nicht nur um das Hier und Jetzt. Schließlich ist es auch unsere Aufgabe, die Lernenden auf die Arbeits- und Berufswelt von morgen vorzubereiten.

Frederik M.: Da hast du sicher recht. Aber wie?

Susanne B.: Lass uns im Gespräch bleiben!

In ihrer kurzen Unterredung sprechen die beiden Lehrkräfte – direkt oder indirekt – drei wichtige Gesichtspunkte an, die mit der Nutzung von KI-Sprachmodellen (*Large Language Models*, LLMs) beim schulischen Lehren und Lernen verbunden sind. In den folgenden Abschnitten gehe ich kurz auf diese Aspekte ein, denn sie bilden die Grundlage für eine sinnvolle Verwendung von KI-Systemen im Unterricht.

Zunächst behandle ich in gebotener Kürze die technischen Grundlagen von KI-Sprachmodellen: Sich einzuarbeiten, um die Tools sinnvoll nutzen zu können, bedarf einer gewissen Anstrengung und ist mit Zeitaufwand verbunden. Dabei spielt es keine große Rolle, ob sie gratis zu haben oder ob sie kostenpflichtig sind. Experten wie Ethan Mollick gehen davon aus, dass Lehrpersonen fünf bis zehn Stunden investieren müssen, bevor sie angemessen mit den Tools im Unterricht arbeiten können (Mollick; Blogpost: One useful thing 17.03.2023). Damit ist nicht die vergleichsweise einfache Einrichtung eines KI-Sprachsystems auf dem eigenen Laptop oder Smartphone gemeint. Vielmehr geht es darum, das Leistungsspektrum der Tools für die Verbesserung der Lernergebnisse und die Persönlichkeitsentwicklung der Schülerinnen und Schüler kennenzulernen. Das erreicht man nur, wenn man dem Bot passende Anweisungen, sogenannte Prompts, geben kann (vgl. Kap. 2). Darüber hinaus darf man die vielfältigen Unzulänglichkeiten umfangreicher KI-Modelle nicht außer Acht lassen.

In dem kurzen Meinungsaustausch zwischen den beiden Lehrpersonen wird – über die technischen Voraussetzungen hinaus – der erwartbare Einfluss der Tools auf die Gesellschaft und insbesondere auf die Berufswelt angesprochen. Ist es sinnvoll, KI-Sprachmodelle wie ChatGPT mit der Erfindung des Buchdrucks oder des Internets zu vergleichen? Werden die Berufe, auf die sich Schülerinnen und Schüler heute vorbereiten, in Zukunft noch existieren? Werden sie überhaupt einen Arbeitsplatz finden? Vor allem aber: Was sollten sie lernen, um in einer von künstlicher Intelligenz dominierten Welt zu bestehen?

Daran schließen sich erste Überlegungen in Bezug auf die Nutzung der Systeme für schulisches Lehren und Lernen an. Ein konkretes Generierungsbeispiel vermittelt nicht nur erste Einblicke in die Funktionsweise von ChatGPT, sondern belegt auch die aus meiner Sicht größte Unzulänglichkeit des Bots, nämlich über dem Wortreichtum die inhaltlichen Aspekte zu vernachlässigen. Ein weiteres Beispiel gibt erste Hinweise auf Nutzungsmöglichkeiten nicht nur für Lehrkräfte, sondern auch für Schülerinnen und Schüler. Daraus ergeben sich Fragen des Urheberrechts, des Datenschutzes sowie des Umgangs mit Täuschungsversuchen (vgl. Rieck 2023). Hauptsächlich aber geht es um die längst fällige Neuorientierung des Unterrichts: weg von reproduktiven Aufgaben, die KI-Sprachmodelle erledigen können, hin zu Aktivitäten, bei denen kritisches Denken und persönliche Stellungnahmen im Vordergrund stehen.

1.2 Zur Funktionsweise von KI-Sprachmodellen

Der Begriff »Künstliche Intelligenz« taucht erstmals im Zusammenhang mit dem *Dartmouth Sommer Research Project on Artificial Intelligence* auf, einer Tagung, die 1956 am Dartmouth College in New Hampshire stattfand. Eine der Grundannahmen der Initiatoren hat auch heute noch Gültigkeit: KI stützt sich auf die Annahme, dass grundsätzlich alle Aspekte des Lernens und andere Merkmale der Intelligenz so genau beschrieben werden können, dass eine Maschine zur Simulation dieser Vorgänge gebaut werden kann (vgl. auch zum Folgenden De Florio-Hansen 2020, S. 42 ff.). Die Auseinandersetzung mit der Frage, ob es möglich ist, menschliche Intelligenz in Maschinen nachzubilden, hat zuvor schon Alan Turing (1912–1952), den Urheber des Turing-Tests, beschäftigt (vgl. De Florio-Hansen 2020, S. 42–47).

Trotz beeindruckender Fortschritte gibt es bisher keine Ausprägung von KI, welche die kognitiven, emotionalen und sozialen Fähigkeiten des Menschen aufweist, und es ist fraglich, ob es sie jemals geben wird. Das Problem dabei ist nicht die in Science-Fiction häufig thematisierte Machtübernahme durch KI-gestützte Systeme (*AI takeover*), sondern unser Umgang mit ihnen. Bereits 2019 (deutsch 2020) hat der bekannte KI-Forscher Stuart Russell, Professor für Informatik an der

University of California in Berkeley, ein hilfreiches Buch mit dem Titel *Human Compatible – Künstliche Intelligenz und wie der Mensch die Kontrolle über superintelligente Maschinen behält* veröffentlicht.

Schaut man sich den Output von KI-Sprachmodellen wie ChatGPT an, ist es nicht verwunderlich, dass viele User Bewusstsein, Empathiefähigkeit und Denkvermögen mit den Tools verbinden. So hochentwickelt sie auch sein mögen: KI-Sprachmodelle beruhen auf mathematischen Berechnungen in Form statistischer Modelle. Sie haben keinerlei Weltverständnis und kennen die Bedeutung der Wörter und Texte nicht, die sie generieren. Darauf weist unter anderem Stefan Buijsman immer wieder hin (Buijsman 2021, S. 15).

Seit den 1950er Jahren hat sich die KI-basierte Forschung trotz einiger Rückschläge kontinuierlich weiterentwickelt. Ab 2010 ist ein exponentielles Wachstum zu verzeichnen. Das ist unter anderem auf die immer ausgeklügelteren Algorithmen zurückzuführen, welche die Grundlage heutiger KI-Systeme bilden (zu den Fachtermini vgl. Glossar).

Ein Algorithmus ist eine eindeutige Handlungsanweisung zur Lösung eines Problems. Folglich beruht ein Kochrezept auf einem Algorithmus, obwohl der Terminus in der Regel für mathematische Instruktionen bei der Programmierung von Computern verwendet wird. Eine wissenschaftsorientierte Definition von Algorithmus lautet:

1. Eine präzise, das heißt in einer festgelegten Sprache abgefasste, endliche Beschreibung eines allgemeinen Verfahrens unter Verwendung elementarer Verarbeitungsschritte zur Lösung einer gegebenen Aufgabe.
2. Lösungsverfahren in Form einer Verfahrensweisung, die in einer wohldefinierten Abfolge von Schritten zur Problemlösung führt (vgl. Gabler Wirtschaftslexikon 2023).

So präzise die Definitionen auch sein mögen: Algorithmen sind immer nur so gut wie die Daten und Anweisungen, auf denen sie basieren.

Darüber hinaus beruhen die bedeutenden Fortschritte des letzten Jahrzehnts auf dem sogenannten maschinellen Lernen (*Machine Learning*). Diese Art von KI erkennt in Daten, mit denen das System trainiert wurde, Muster und kann diese Erkenntnisse anschließend auf neue Daten anwenden, ohne dass ein Programmierer eingreifen muss. Vereinfacht ausgedrückt, berechnet ein KI-System wie ChatGPT auf der Grundlage eines eingegebenen Textsegments, welches Wort bzw. welche Wörter mit hoher Wahrscheinlichkeit auf den Input folgen, und verarbeitet sie zum Output.

Ein Grund für die Erfolge des maschinellen Lernens besteht darin, dass es sich mehr und mehr auf die Nachbildung von neuronalen Netzen (*Neural Networks*) des menschlichen Gehirns stützt. Neuronen können Informationen von außen oder von anderen Neuronen aufnehmen und modifiziert an andere Neuronen

weiterleiten oder als Endergebnis ausgeben. Zwischen der Eingabe und der Ausgabe können verschiedene Schichten liegen, die – je nach Anwendung – das Endergebnis bestimmen.

Im Zusammenhang mit künstlichen neuronalen Netzwerken ist häufig von *Deep Learning* (etwa »tief gehendes Lernen«) die Rede. Bisher wurde nicht festgelegt, wie viel Schichten zwischen der Eingabe und der Ausgabe liegen sollten, damit maschinelles Lernen als *Deep Learning* gelten kann. Je zahlreicher die zwischen Ein- und Ausgabe aufeinander folgenden Schichten sind, umso schwerer ist es für Menschen, auch für KI-Experten und -Expertinnen, die Lösung nachzuvollziehen.

Die rasante Verbesserung der genannten Bereiche hat in den letzten Jahren die Entwicklung von *Large Generative AI Models (LGAIMs)* ermöglicht, das heißt von KI-Systemen, die Ergebnisse in verschiedenen Bereichen (z. B. Text, Bild, Video, Audio; vgl. Kap. 7) generieren. Eine spezielle Form stellen *Large Language Models (LLMs)* dar, die auf dem *Natural Language Processing (NLP)* beruhen. Gebräuchlich sind auch die Termini Computerlinguistik oder *Language Technology*. NLP – das Verfahren ist bereits um 1950 entstanden – ist eine KI-Methode, die sich mit der Analyse, dem Verständnis und der Generierung von natürlicher Sprache beschäftigt. Inzwischen geht es darum, Rechner so zu programmieren, dass sie große Mengen natürlicher Sprachdaten mündlich und/oder schriftlich verarbeiten und analysieren können. Der beeindruckende Output von KI-Sprachmodellen wie ChatGPT beruht auf Techniken mathematischer Optimierung, die während des Trainings immer wieder angewendet werden, bis das System den Input in wünschenswerten Output transformiert. Die Konvertierung wurde und wird vor allem dadurch unterstützt, dass immer größere Datenmengen und leistungsstärkere Computer zur Verfügung stehen. Trotz der Fortschritte stellt die Vielfalt menschlicher Sprache, einschließlich Ironie und Mehrdeutigkeiten – man denke nur an das Wort »Gericht« –, auch weiterhin eine große Herausforderung dar.

Häufig ist auch von *Foundation Models* (etwa: »Grundlagen-Modelle«) die Rede. Sie verfügen über vielfältige Verwendungsmöglichkeiten und liegen den meisten spezialisierten Systemen zugrunde. Zum Vorgehen von OpenAI, dem Unternehmen, das ChatGPT geschaffen und in Umlauf gesetzt hat, schreibt Helen Toner in einem Blogpost vom Mai 2023, dass ChatGPT das KI-Sprachmodell GPT-3.5 zugrunde liegt, dieses Foundation Model jedoch von OpenAI an die Nutzung in einem Chatbot-Setting angepasst wurde (<https://time.com/6280533/ai-chatbots-improv-machines/>; letzter Zugriff August 2023).

Die bisher kurz erläuterten Grundlagen gelten für (fast) alle LLMs. Die meisten werden über das *Unsupervised Learning* (Unüberwachtes Lernen) hinaus auch durch *Supervised Learning* (Überwachtes Lernen) optimiert. Beim Überwachten Lernen werden ChatGPT Unterhaltungen vorgelegt, in denen menschliche Trainer beispielsweise beide Rollen übernehmen: die des Users und die des Chatbots.

Man kann sich fragen, welche Vorstellungen die menschlichen Trainer zugrunde legen bzw. wie unvoreingenommen sie tatsächlich sind. Was also macht ChatGPT so besonders? Obwohl OpenAI nicht alle Verfahren zur Optimierung seiner Sprachmodelle offengelegt hat (es gibt kein Handbuch oder auf *Peer-Review* gegründete wissenschaftliche Veröffentlichungen), haben der CEO Sam Altman und weitere Experten mehrfach darauf hingewiesen, wodurch sich ChatGPT von anderen LLMs unterscheidet.

Ausführliche Erläuterungen gibt Thang Long Phan in einem Blogpost mit dem Titel »Wie ChatGPT mit Hilfe von Reinforcement Learning optimiert wird« (<https://dida.do/de/blog/wie-chatgpt-mit-hilfe-von-reinforcement-learning-optimiert-wird>; letzter Zugriff August 2023). Durch *Reinforcement Learning from Human Feedback* kann ein zugrunde liegendes KI-Sprachmodell durch die Verbindung mit einem weiteren KI-System deutlich verbessert werden. Das zusätzliche KI-System, mit dem ChatGPT optimiert wurde, ist ein sogenanntes Reward Model (»Belohnungsmodell«). Durch das Reward Model wird festgelegt, welche Antworten menschliche Nutzerinnen und Nutzer in bestimmten Fällen bevorzugen. In der Regel besteht die Auswahl, an der sich das Reward Model orientiert, aus einer vorgegebenen Liste von Möglichkeiten.

Dieser Vorzug von ChatGPT bringt einen entscheidenden Nachteil mit sich: Nicht die inhaltliche Wahrheit einer Generierung, sondern die wohlklingende Formulierung hat für das System Priorität. Hinzukommt, dass die kostenfreie Version von ChatGPT keinen Zugriff auf das Internet hat und die Trainingsdaten mit dem Jahr 2021 enden. Außerdem teilt sich der Chatbot zahlreiche Unzulänglichkeiten mit fast allen LLMs: Wenn die Systeme überfordert sind, kommt es zu sogenannten *Hallucinations*, zuverlässig klingenden Aussagen, die durch die Trainingsdaten nicht gerechtfertigt sind. Die umfänglichen Trainingsdaten der meisten LLMs sind außerdem von einer westlichen Sicht auf die Welt geprägt. Der Bot generiert immer das, was dem statistischen Mittelwert entspricht: Lässt man eine Liebesgeschichte generieren, geht es immer um ein heterosexuelles Paar. OpenAI ist zwar bemüht, sexistische sowie rassistische Diskriminierungen zu unterbinden. Fragt man aber geschickt genug, sind diese Mängel immer noch vorhanden.

Außerdem können KI-Sprachmodelle in vielen Fällen keine Quellen für ihre Generierungen benennen. Letzteres führt bei ChatGPT dazu, dass das System auf Nachfrage bisweilen Quellen erfindet: Es benennt Publikationen einschließlich Autoren und ISBN-Nummern, die nicht existieren. Bei den konkreten Unterrichtsbeispielen in Teil II haben wir Gelegenheit, uns mit den Einschränkungen auseinanderzusetzen.

Festzuhalten gilt der von den Anbietern und in wissenschaftlichen Publikationen mehr oder weniger ausführlich behandelte Hinweis: Jede Generierung eines KI-Sprachmodells bedarf der Überprüfung anhand von eigenem Wissen sowie von zuverlässigen Quellen. Auf der Startseite zu ChatGPT heißt es:

Limitations¹

- May occasionally generate incorrect information
- May occasionally produce harmful instructions or biased content
- Limited knowledge of world and events after 2021

Einer der zahlreichen Posts von OpenAI (vgl. openai.com) trägt den Titel *Educator Considerations for ChatGPT*. Darin gehen die Autoren detailliert auf Vorurteile und Stereotypen sowie auf zahlreiche Risiken ein, die mit der Nutzung des Chatbots verbunden sind. Mit Fehlinformationen und Antworten, die mit Vorurteilen behaftet sind, beschäftigt sich auch das Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen in seinem im März 2023 veröffentlichten Handlungsleitfaden *Umgang mit textgenerierenden KI-Systemen*: »Die produzierten Inhalte dürfen also nicht unreflektiert übernommen, sondern müssen von den Lernenden kontrolliert werden. Dies ist nur möglich, wenn die Lernenden ausreichend eigenes Wissen über den Gegenstand erworben haben. Damit gewinnt Medienkompetenz noch einmal an Bedeutung. Gerade die Fähigkeit, Fake News von Fakten auf der Grundlage eines gesicherten Wissens zu unterscheiden, wird immer wichtiger« (MSB NRW 2023, S. 4).

Fitzpatrick und seine Mitautoren, die den genannten Handlungsleitfaden in englischer Übersetzung als Zusammenfassung ihres Buches verwenden, empfehlen für die Überprüfung und Verbesserung von KI generierten Inhalten das Verfahren EDIT (Fitzpatrick u. a. 2023, S. 100 ff.):

Evaluate your AI output content for language, facts and structure.

Determine accuracy and corroborate with source.

Identify biases and misinformation in output.

Transform content to reflect adjustments and new findings.²

1 »Beschränkungen

- Kann gelegentlich falsche Informationen erzeugen
- Kann gelegentlich schädliche Anweisungen oder voreingenommene Inhalte produzieren
- Begrenztes Wissen über die Welt und die Ereignisse nach 2021« (*eigene Übersetzung*)

2 »Evaluieren (Überprüfen) Sie den Inhalt des Outputs in Bezug auf Sprache, Fakten und Struktur.

Determinieren (Bestimmen) Sie die Genauigkeit und untermauern Sie sie mit Quellen.

Identifizieren (Erkennen) Sie Verzerrungen und Fehlinformationen im Output.

Transformieren (Verändern) Sie den Inhalt durch Anpassungen und neue Einsichten.«

(*eigene Übersetzung*)

1.3 Auswirkungen auf Gesellschaft und Berufswelt

Selbst wenn man die Nutzung von KI-Sprachmodellen beim fachlichen Lehren und Lernen mit Skepsis betrachtet, sollte man die Behandlung Generativer AIs im Unterricht keinesfalls aussparen. Jede Lehrperson hat die Aufgabe, die Schülerinnen und Schüler über fachliche Ziele und Inhalte hinaus auf ihr künftiges Leben, vor allem die Berufswelt, vorzubereiten. Dazu bedarf es verschiedener Einblicke in KI-Systeme, vor allem in LLMs wie ChatGPT.

In ihrem jüngsten Buch widmet die Wissenschaftsjournalistin Manuela Lenzen KI-Sprachsystemen ein Kapitel mit der Überschrift *Groß, größer, Sprachmodelle*. Der einleitende Abschnitt bietet eine Diskussionsgrundlage für den Unterricht der meisten Schulfächer: »Sprache ist, zumindest beim Menschen, ein zentraler Aspekt von Intelligenz. Zudem müssen künstliche Systeme, die für uns von Nutzen sein sollen, mit uns kommunizieren können. Wenn es nicht anders geht, in formalen Sprachen via Tastatur, aber viel lieber doch einfach in gesprochener natürlicher Sprache, so, wie wir eben auch mit Menschen sprechen. Hier leisten die Sprachmodelle Erstaunliches. Sie schreiben Texte oder fassen bestehende Literatur zusammen, sie erstellen Programmcodes und man kann mit ihnen immer überzeugendere Dialoge führen. Sie sind für einen Gutteil der jüngsten medienwirksamen Erfolge der KI verantwortlich. Vor allem seit der Veröffentlichung von ChatGPT, das von dem US-Forschungsunternehmen OpenAI entwickelt wurde, sind die Medien voller beeindruckender Beispiele ihrer Texte. Nicht nur Laien, auch Forscherinnen und Forscher, die diese Systeme testen, wundern sich bisweilen, wieviel diese Systeme aus ihren Trainingsdaten lernen können« (Lenzen 2023, S. 42 f.).

Je nach Kenntnisstand und Erfahrung der Lernenden kann man ihnen einen von ChatGPT generierten Text vorlegen, der sich in irgendeiner Form auf einen im Unterricht behandelten Themenaspekt bezieht. Dabei sollte man exotische Vorgaben, mit denen viele Autorinnen und Autoren ihre Texte ausschmücken, möglichst vermeiden (z. B. Create a story about a time-travelling banana; <https://stayweird.com/craziest-chatgpt-prompts/>; letzter Zugriff August 2023). Der Themenaspekt sollte klar, aber nicht zu eng gefasst sein, um zeigen zu können, dass die wohlklingenden Texte zwar nicht unbedingt falsche Informationen enthalten, aber häufig sehr oberflächlich sind.

Es gilt, die Inhalte stärker in den Blick zu nehmen. Bisher wird die Qualität der von KI-Sprachmodellen wie ChatGPT generierten Texte hauptsächlich danach beurteilt, ob man dem Output anmerkt, dass er von einer Maschine und nicht von einem Menschen erstellt wurde. In vielen Fällen ist dies sekundär. Die Frage nach der inhaltlichen Angemessenheit sollte Vorrang haben, nach dem Motto: *form follows function*.