

Vektorisierte Praktiken: Ein linguistisches Verständnis von Large Language Models

Noah Bubenhofer, Zürich

Abstract

Warum ist textgenerierende künstliche Intelligenz (KI) in der Lage, orthographisch und grammatikalisch korrekten Text zu produzieren, der semantisch kohärent und pragmatisch angemessen ist? Warum kann damit ein Chatbot erstellt werden, der recht gut die Kommunikation mit einem Menschen simuliert? Der Beitrag möchte aus linguistischer Sicht erklären, warum das so ist: Es werden ausgefeilte statistische Methoden verwendet, die Performanzdaten von Sprache sehr gut modellieren können.

Doch was kann eine textgenerierende KI nicht? Um diese Frage zu beantworten, hilft es, den Praktikenbegriff einzuführen, der zentrale Aspekte von Kommunikation beschreibt: Darunter fallen die Aspekte von Körperlichkeit und Räumlichkeit, die bei menschlicher Kommunikation eine wichtige Rolle spielen. Doch spielen diese Aspekte auch eine Rolle in Systemen künstlicher Intelligenz? Welche Form von Körper haben sie und welcher Art ist der Raum, in dem die Kommunikation mit ihnen stattfindet?

Zusätzlich betont der Beitrag, dass sogenannte „Artificial-Intelligence-Literacy“ gefördert werden muss, also die grundlegende Kompetenz, verantwortungsvoll und produktiv mit KI umgehen zu können. Im Beitrag werden Elemente dieser Kompetenz hinsichtlich eines performanzorientierten Sprachbegriffs definiert. Mit dieser Kompetenz wird es möglich, sowohl die Simulation von Kommunikation durch KI als auch die zugrundeliegenden gesellschaftlichen und sprachtheoretischen Implikationen zu verstehen und hinterfragen zu können.

1. Einleitung

Textgenerierende künstliche Intelligenz ist in der Lage, orthographisch und grammatikalisch korrekten Text zu produzieren, der semantisch kohärent und pragmatisch angemessen ist sowie Weltwissen einbezieht, um in Interaktion mit Menschen zu treten. Aus linguistischer Sicht ist es keine große Überraschung, dass dies so ist, zumindest wenn man sich sprachtheoretisch auf die bereits lange existierenden Ideen beruft, die in der Performanz von Sprache eine wichtige Rolle sehen. Es kann eine Denktradition vom Humboldt über Wittgenstein, Harris und Firth bis in jüngere Theorien zu Performativität von Sprache festgestellt werden, die solche Ideen postuliert. Mehr oder weniger davon abgekoppelt ereignete sich in der Computerlinguistik bereits vor mehreren Jahrzehnten eine Wende weg von regelbasierten Ansätzen zur Verarbeitung natürlicher Sprache hin zu statistischen Methoden. Heute können nun bereits alte Vorstellungen darüber, wie Sprache eigentlich modelliert werden müsste, in der Praxis ausprobiert werden – und das erleben wir aktuell mit generativer KI.

Dies bedeutet nun auf der einen Seite, dass viele theoretische Annahmen über Sprache empirische Evidenz erhalten. Dies ist selbstverständlich noch kein Beweis dafür, dass „menschliche Sprachverarbeitung“ genau so funktioniert wie eine KI, aber die Modellierung

von Sprache als distributionelles Performanzphänomen ist zumindest plausibler als die Modellierung als regelbasiertes System, weil es schlicht viel besser funktioniert.

Trotz dieser – wenn man so will – Erfolgsgeschichte performanzorientierter Sprachtheorie stellen sich aus linguistischer Sicht eine Reihe von Fragen: Welche Voraussagen darüber, wie Sprache „funktioniert“, können diese Sprachtheorien zusätzlich noch machen? Und welche Eigenschaften von Sprache werden in heutigen textgenerierenden KI-Anwendungen (noch) nicht modelliert?

Im folgenden Beitrag möchte ich aufzeigen, dass eine Reihe von Aspekten in den aktuellen KI-Anwendungen noch mehr oder weniger vernachlässigt werden, die jedoch unbedingt zur Idee performanzorientierter Sprachmodelle gehören. Ein wichtiger Aspekt ist der der Praktiken. Damit verbunden sind Körperlichkeit, Räumlichkeit, Historizität und Multimodalität. Daraus ergeben sich mit Blick auf heutige KI-Anwendungen einige Probleme, die bislang unterschätzt werden. Im vorliegenden Beitrag möchte ich die beiden Aspekte Körperlichkeit und Räumlichkeit vertieft betrachten.

Zunächst muss ich jedoch in aller Kürze erklären, wie textgenerierende KI funktioniert und welche Fähigkeiten sie heute hat. Dies mache ich jedoch nicht aus technischer, sondern aus linguistischer Sicht. Dies ist unabdingbar, um zu verstehen, wie performanzorientierte Sprachtheorien mit der statistischen Modellierung von Sprache zusammenhängen. Die Erklärung wird deshalb im Detail technisch ungenau sein, zielt jedoch darauf ab, konzeptionell aus linguistischer Sicht zu verstehen, wie und warum generative KI so funktioniert, wie sie funktioniert.

2. Wie und warum funktioniert textgenerative KI?

Wenn wir wissen wollen, was „Maske“ bedeutet, reicht es, sich ein Kollokationsprofil anzuschauen, das auf einer großen Textdatenmenge berechnet worden ist. Dies ist die Grundhypothese, die seit Firths Prägung des Begriffs „Collocation“ (Firth 1957, 194) korpuslinguistische Zugänge zu Semantik geleitet hat. Die in Tabelle 1 abgebildete Liste zeigt die Kollokatoren zu „Maske“, also die Ausdrücke, die statistisch gesehen signifikant häufig im Kontext von „Maske“ auftreten. Es sind Bestandteile des Kompositums sichtbar („FFP2-“), vor allem aber auch Adjektive, die typischerweise unmittelbar davor auftreten („schützender“, „schwarze“, „weiße“, „blutverschmierte“). Interpretativ lässt sich auf Basis eines Kollokationsprofils die Semantik von Lexemen beschreiben, was schon lange sowohl für lexikographische, semantische bis hin zu diskurslinguistischen Zwecken genutzt wird (Sinclair 1991; Evert 2009; Steyer 2013; Bubenhofer 2017).

Tabelle 1: Auszug aus DeReKoVecs (2024-IB) zu „Maske“, syntagmatische Beziehungen, <https://corpora.ids-mannheim.de/derekovecs?word=Maske&cutoff=500000&n=100&N=2000>

#	w'	max(a)	<a)	$\Sigma a / \Sigma w'$	$\perp(a/c)$	$\Sigma a / \Sigma w$	Kollokat (W2V)
1	+++++x+++++	0.985	0.817	1.647e-4	6.471e-4	6.543e-5	FFP2-
2x....	0.984	0.098	1.646e-4	1.646e-4	2.269e-5	schützender
3	+++++x+++++	0.978	0.828	1.635e-4	6.552e-4	6.756e-5	FFP2
4x....	0.974	0.097	1.629e-4	1.629e-4	3.189e-5	schwarze
5	+++++x+++++	0.992	0.963	1.625e-4	8.082e-4	7.199e-5	FFP-2-Maske
6x....	0.972	0.097	1.625e-4	1.625e-4	3.772e-5	ohne
7x....	0.968	0.097	1.619e-4	1.619e-4	3.471e-5	weiße

#	w'	max(a)	(a)	$\Sigma a/\Sigma w'$	$\perp(a/c)$	$\Sigma a/\Sigma w$	Kollokat (W2V)
8x.....	0.965	0.097	1.614e-4	1.614e-4	1.360e-5	xyxHTMEyxyOhne
9x.....	0.964	0.096	1.613e-4	1.613e-4	2.739e-5	blutverschmierte
10x.....	0.962	0.096	1.609e-4	1.609e-4	2.054e-5	schützende
11x.....	0.961	0.096	1.608e-4	1.608e-4	2.252e-5	selbstgenähte
12x.....	0.956	0.096	1.599e-4	1.599e-4	3.243e-5	aufgesetzter
13x.....	0.955	0.096	1.597e-4	1.597e-4	2.060e-5	schützenden
14	+++++x+++++	0.994	0.976	1.595e-4	8.233e-4	7.296e-5	Mundschutz
15x.....	0.951	0.095	1.591e-4	1.591e-4	2.786e-5	genähte
16x.....	0.951	0.095	1.590e-4	1.590e-4	3.788e-5	schwarzer
17x.....	0.950	0.095	1.589e-4	1.589e-4	1.975e-5	entstellende
18x.....	0.946	0.095	1.582e-4	1.582e-4	2.003e-5	waschbare
19x.....+	0.944	0.170	1.579e-4	2.019e-4	4.278e-5	hautfarbenen
20	+++++x+++++	0.991	0.966	1.576e-4	8.095e-4	7.227e-5	FFP2-Maske
21x.....	0.938	0.094	1.569e-4	1.569e-4	1.604e-5	Henry
22x.....	0.938	0.094	1.568e-4	1.568e-4	3.195e-5	schwarzen
23x.....	0.937	0.094	1.567e-4	1.567e-4	2.977e-5	blutverschmierten
24x.....	0.930	0.093	1.554e-4	1.554e-4	1.488e-5	isolierende
25	+++++x+++++	0.991	0.863	1.550e-4	6.736e-4	7.057e-5	Alltagsmaske
26x.....	0.925	0.092	1.546e-4	1.546e-4	1.670e-5	durchgestrichene
27x.....	0.924	0.092	1.545e-4	1.545e-4	1.985e-5	pechschwarze
28x.....	0.924	0.092	1.545e-4	1.545e-4	2.991e-5	neonfarbene
29x.....	0.921	0.092	1.540e-4	1.540e-4	2.111e-5	aufgesetzte
30x.....	0.920	0.092	1.537e-4	1.537e-4	1.618e-5	Schwarze
31x.....	0.920	0.092	1.537e-4	1.537e-4	1.616e-5	scharlachrote
32x.....	0.918	0.092	1.534e-4	1.534e-4	2.379e-5	furchteinflößende

Ein pragmatischer Blick auf Sprache steht hinter der Überzeugung, dass „Kollokationen“ eine geeignete Analysekategorie ist, um Sprachgebrauch zu fassen. Genauer: Die Idee, typische Muster im Sprachgebrauch als Mittel der Kontextualisierung zu verstehen. So zeigte etwa Feilke, wie wenige Wortformen wie „Durchzug von“, „weitgehend trocken“ und „vereinzelt“ reichen, um sofort die Textsorte des Wetterberichts zu evozieren (Feilke 2003, 211). Das ist so, weil das Schreiben von Wetterberichten eine Textroutine geworden ist, die eine spezifische soziopragmatische Relevanz aufweist: Die Beschreibung des Wetters und wie es sich in den kommenden Tagen entwickeln könnte, ist in der ausdifferenzierten modernen Gesellschaft eine professionelle Tätigkeit, die deswegen Handlungsrouninen entwickelte, um diese Tätigkeit auszuführen. Wir als Leser:innen erkennen an den typischen sprachlichen Mustern den professionell erstellten Wetterbericht und können ihn von einer Bauernregel („Donner'ts im Dezember gar, kommt viel Wind im nächsten Jahr.“¹) unterscheiden. Mit dieser engen Verquickung von gesellschaftlichem Handeln und Sprachgebrauch wird Sprache verstanden in ihrer „soziopragmatischen Funktion im Sprechen“ (Feilke 1996, 15–32, ausführlich im Kontext der Fachgeschichte: 2016). Das Berechnen von Kollokationen ist also nicht einfach ein praktisches Tool, um Semantik zu

¹ Vgl. https://de.wikipedia.org/wiki/Liste_von_Bauernregeln (letzter Zugriff: 5. November 2024).

modellieren, sondern kann als Ausdruck für die Bedeutung von sprachlichem Handeln gelesen werden.

Die Berechnung von Kollokationen gehört zwar zu den nach wie vor genutzten „Klassikern“ der Korpusanalyse, methodisch ist jedoch seit einiger Zeit mehr möglich. Während Kollokationen auf die syntagmatische Ebene abheben und die typischen Umgebungen von Ausdrücken systematisch berechnen, sind sie gleichzeitig die Basis für eine paradigmatische Sichtweise auf diese Ausdrücke. Die Grundidee liegt darin, dass Ausdrücke mit ähnlichen Kollokationsprofilen eine semantische Ähnlichkeit aufweisen. Einen wichtigen Ansatz dieser paradigmatischen Sicht ist die word2vec-Methode, die sog. „Word Embeddings“ berechnet (Mikolov et al. 2013). Diese Methode gehört zur Familie der verschiedenen Ansätze im Bereich der distributionellen Semantik (Lenci 2018), die alle davon ausgehen, dass sich die Semantik von Ausdrücken über die statistische Auswertung ihres Gebrauchs ermitteln lässt. Grundlegend für diese Methoden ist es, das Distributionsverhalten von Ausdrücken in großen Textmengen als Vektoren zu beschreiben. Tabelle 2 zeigt einen Ausschnitt von Types (Zeilen), für die das Ko-Vorkommen mit allen anderen Types in den Daten (Spalten) erfasst ist.

Tabelle 2: Matrix zur Erfassung des Kollokationsverhaltens von Ausdrücken in einem Korpus (Fantasiewerte)

	Zeichen	uneingeschränkt	Tür	Schweiz	international	Perron	Bahnhof	...
Solidarität	0.9	0.8	0	0.7	0.6	0	0	
Loyalität	0.6	0.7	0	0.3	0.4	0	0	
Bahn	0.1	0	0.1	0.4	0.5	0.8	0.9	
...								

Daraus lässt sich nun für jeden Type das semantische Profil in Form eines Vektors ausgeben; hier am Beispiel von „Solidarität“:

Solidarität = { 0.9, 0.8, 0, 0.7, 0.6, 0, 0, ... }

Diese Vektoren können grafisch dargestellt werden, sofern die Anzahl Dimensionen auf zwei oder drei reduziert wird. Abbildung 1 zeigt für die drei Types „Bahn“, „Solidarität“ und „Loyalität“ die Vektoren, allerdings reduziert auf das Ko-Vorkommen mit „uneingeschränkt“ (x-Achse) und „Zeichen“ (y-Achse). Die semantische Nähe bemisst sich nun am Winkel zwischen den Vektoren: dieser ist zwischen den Vektoren für „Solidarität“ und „Loyalität“ viel kleiner als zwischen den Vektoren „Solidarität“ und „Bahn“.

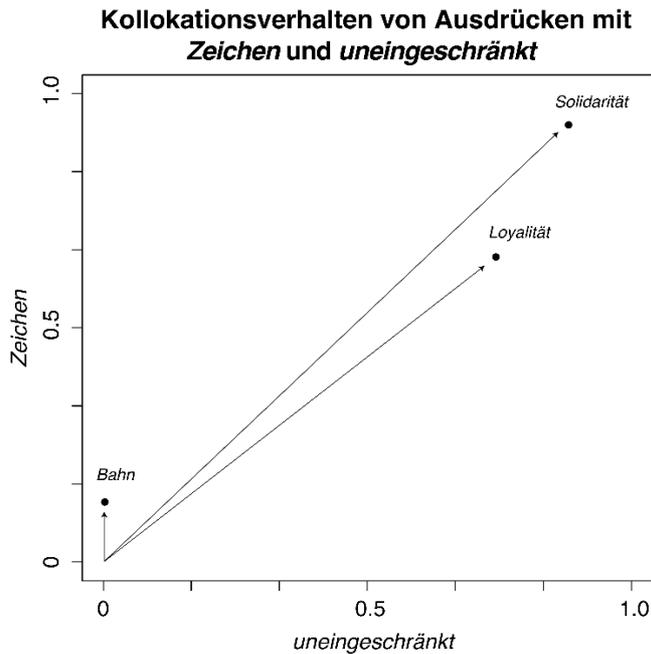
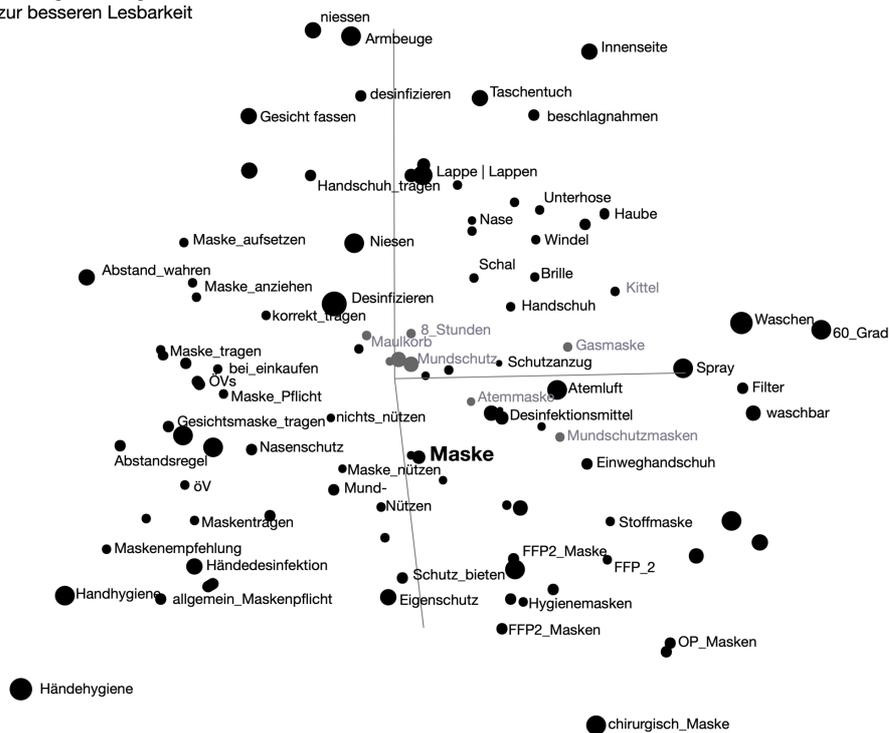


Abbildung 1: Vektordiagramm mit Fantasiewerten

Die Distanz zwischen den Vektoren kann problemlos auch für hochdimensionale Vektoren berechnet werden, auch wenn sie nicht mehr visualisiert werden können. So ergibt sich durch die Berechnung der Vektoren für alle Types in einem Textkorpus ein Vektorraum, aus dem Abbildung 2 wiederum einen kleinen Ausschnitt zeigt: Für die Darstellung musste der Vektorraum wiederum auf drei Dimensionen reduziert werden (die Visualisierung kann in der digitalen Form gedreht und gezoomt werden) und es ist ein Ausschnitt rund um „Maske“ sichtbar. In der daneben dargestellten Liste sind die semantisch ähnlichsten Ausdrücke zu „Maske“ ebenfalls aufgeführt.

Redesign der Originaldaten zur besseren Lesbarkeit



Search	Distance	COSINE
Maske		
Nearest points in the original space:		
Schutzmaske	0.392	
Gesichtsmaske	0.547	
Atemschutzmaske	0.576	
Mundschutz	0.595	
FFP2_Maske	0.608	
Hygienemasken	0.610	
Maskentragen	0.682	
FFP2-Masken	0.726	
FFP2	0.758	
tragen	0.760	
Maulkorb	0.762	
Trage Tragen	0.763	
nichts_nützen	0.772	
Stoffmaske	0.773	
Mund-	0.798	
Nützen	0.804	
Nasenschutz	0.807	
Maske tragen	0.810	
Maske_nützen	0.811	
OP-Masken	0.818	
FFP_2	0.828	
Tragepflicht	0.840	

Abbildung 2: Semantischer Raum rund um «Maske» in einem Korpus aus Zeitungsartikeln und «alternativen Medien» zum COVID-19-Virus. Es werden semantisch ähnliche Wörter zu «Maske» angezeigt.

Solche Vektorräume repräsentieren zudem semantische Relationen: Vektoren können miteinander addiert und subtrahiert werden. Dies ermöglicht es, semantische Korrespondenzen zu berechnen, z.B. in der Gleichung:

Schweiz : Bern = Frankreich : ?

Dafür werden die Vektoren für „Bern“ und „Frankreich“ addiert und „Schweiz“ subtrahiert (also der semantische Schweiz-Anteil von der Beziehung Bern-Frankreich abgezogen), was zu einem Vektor führt, in dessen nächster Nähe wahrscheinlich „Paris“ stehen wird.² Dies funktioniert nicht nur für Nomen, sondern auch für grammatikalische Verhältnisse („sagen“ : „gesagt“ = „sprechen“ : ?).

Diese sehr vereinfachte Darstellung von Word Embeddings soll einen ganz grundsätzlichen Einblick geben, wie plausibel die „distributionelle Hypothese“ (Lenci 2018, 152) ist. Diese geht davon aus, dass Performanzdaten alles enthält, um Sprachgebrauch automatisiert vorauszusagen. Selbstverständlich haben sich die Methoden stark weiterentwickelt; so fällt sofort auf, dass das oben dargestellte Verfahren simplizistisch ist, weil es davon ausgeht, dass jeder Type durch genau einen Vektor repräsentiert werden kann, also Lexeme nur eine semantische Lesart aufweisen. Dies ist selbstverständlich eine absurde Annahme, sodass neuere Methoden automatisch verschiedene Vektoren für denselben Type annehmen, wenn sich systematisch unterschiedliche Verwendungsweisen zeigen („Bank“ als Sitzmöbel oder als Institution).

Zudem können weitere Aspekte „vektoriert“ werden: Types können nach typischen Positionen in Sätzen differenzierte Vektoren erhalten, oder aber anstelle von Wörtern können größere oder kleinere Einheiten vektorisiert werden.

Weiter muss betont werden, dass die Berechnung der Vektoren nicht in einem schlichten Auszählen des Kovorkommens der linguistischen Einheiten besteht, sondern komplexe Verfahren des Machine Learnings zum Einsatz kommen. Entscheidend ist dabei eine Technologie, die durch den originellen Titel des Artikels, in dem die Methode erstmals beschrieben worden ist, auffällt: „Attention is all you need“ (Vaswani et al. 2017). Sie schlägt die sog. „Transformer“-Methode vor, mit der die oben beschriebene Vektorisierung der Wörter flexibler und abhängig von verschiedenen, sehr großen Kontexten, geschieht. Trotz der Komplexität: Das Ergebnis ist eine statistische Modellierung des distributionellen Verhaltens sprachlicher Einheiten, ein sogenanntes vortrainiertes („pre-trained“) „Sprachmodell“.

Es ist jedoch notwendig, dieses Sprachmodell an eine bestimmte Aufgabe anzupassen. Dies wird als sog. „Finetuning“ beschrieben, bei dem ein überwachtes maschinelles Lernen genutzt wird, um beispielsweise typische Frage-Antwort-Muster zu lernen. Damit kann dann ein Chatbot angeboten werden, der auf Fragen antworten kann. Ein anderes Finetuning kann darin bestehen, Zusammenfassungen von Volltexten als Trainingsdaten zu nutzen, um das Sprachmodell auf die Aufgabe des Zusammenfassens zu trainieren. Ebenfalls eine wichtige Anwendung ist die maschinelle Übersetzung, bei der bereits übersetzte Texte die Trainingsdaten darstellen.

² Mit bestehenden Modellen können solche Vektorkalkulationen online auf folgender Seite durchgeführt werden: <http://vectors.npl.eu/explore/embeddings/en/calculator/#> (letzter Zugriff: 5. November 2024).

Die Kombination von Sprachmodell und Finetuning führt also zu einem System, das auf menschlichen Input hin Text generieren kann. Das Feedback der Menschen, die das System benutzen, kann zudem wiederum verwendet werden, um das Finetuning zu verbessern.

Wir haben damit ein ganz grobes Verständnis, wie textgenerierende KI funktioniert. Terminologisch lassen sich folgende Ebenen unterscheiden:

- Sprachmodell / (Large) Language Model / (L)LM: Das mittels einem maschinellen Lernverfahren erstellte statistische Modell, basierend auf einer umfangreichen Textdatenmenge; Grundprinzip ist die Vektorisierung der „Token“ (normalerweise Wörter / laufende Wortformen) in Abhängigkeit ihres distributionellen Verhaltens.
- Chatbot, Übersetzungssystem etc.: Dienst, der auf einem Sprachmodell basiert und durch Finetuning entstanden ist.
- Als Kombination dieser beiden Elemente ist von (text-)generierender KI, „genAI“ o.ä. die Rede: Also ein Sprachmodell, das für einen bestimmten Zweck angepasst („finetuned“) worden ist und auf Input Text generieren kann.

Selbstverständlich können nicht nur Textdaten, sondern auch Bild-, Audio- und Videodaten auf ähnliche Weise vektorisiert werden. Dies ist nötig für multimodale generierende KI wie z.B. Bildgeneratoren. Diese stehen allerdings nicht im Fokus des vorliegenden Beitrags.

Allgemein spreche ich von „(text-)generativer KI“, „KI-Systemen“ oder „KI-Tools“ und meine damit die ganze Palette an Systemen, die irgendeine Form von generierender Funktion auf Basis von Sprach- bzw. multimodalen Modellen anbieten.

3. Welche Fähigkeiten zeigen sich?

Generative KI, wie sie in Chatbots zur Anwendung kommt, verfügt auf sprachlicher Ebene über eine Reihe von Fähigkeiten. Es lohnt sich, sich einen Moment vor Augen zu halten, welche Fähigkeiten dies sind, die sich alleine aus distributionellem Wissen über Sprachgebrauch ergeben:

- Morphologie und Syntax: Die generierten Texte sind morphosyntaktisch (Flexion, Konjugation) und syntaktisch normgerecht. Es ist nicht lange her, dass z.B. Systeme der maschinellen Übersetzung immer wieder bei komplexeren syntaktischen Strukturen scheiterten. Man denke z.B. an computerlinguistische Tools wie „Syntax-Parser“, die die Aufgabe haben, die syntaktischen Strukturen von natürlichsprachlichen Texten zu identifizieren. Gleichgültig, ob es sich um regelbasierte oder statistische Verfahren handelt, es bleibt das Problem, Daten zu verarbeiten, die syntaktisch komplexer oder normferner als erwartet sind, für die also entweder die definierten Regeln nicht passen oder aber die Trainingsdaten für das maschinelle Lernen von den zu verarbeitenden Daten zu stark abweichen. Diese Parser der älteren Generation sind also meist nicht „robust“ genug, um mit der Sprachgebrauchswirklichkeit umgehen zu können. Dies ist mit den LLMs gelöst, weil sie viel mehr Daten nutzen, um zu lernen, dabei aber auch viel größere Kontexte berücksichtigen können.
- Semantik: Die textgenerierende KI kann semantisch kohärente Texte produzieren, die zudem nicht nur in sich schlüssig sind, sondern auch als kohärent gegenüber den vorherigen Gesprächszügen. Auch die semantische Ebene war lange ein schwieriger Aspekt für maschinelle Methoden. Man behalf sich mit (manuell erstellten) Ontologien oder in jüngerer Zeit schon mit distributionellen Methoden. Wie oben

bereits betont, ist die Idee der Kollokationsanalyse schon alt und es zeigte sich bereits lange das Potenzial, Semantik so in den Griff zu kriegen.

- Pragmatik: Die Reaktionen von Chatbots auf menschlichen Input sind auch bezüglich pragmatischer Aspekte kohärent. Illokutionen wie FRAGEN, AUFFORDERN, BITTEN etc. werden adäquat benutzt, auch Präsuppositionen oder Implikaturen werden erkannt.
- Textwissen: Der Einbezug so großer Kontexte beim Lernprozess, aber auch bei der Textgenerierung, ermöglicht auch die Modellierung von Textwissen. So können verschiedene Textsorten erkannt und realisiert werden, was sich auch im Fall der Programmiersprachen bemerkbar macht; so kann Programmcode, bei dem viel weniger Variation möglich ist, generiert werden.
- Diskurswissen: Daraus ergibt sich dann auch, dass gesellschaftliche Diskurse ebenfalls in den Daten repräsentiert sind und deshalb auf entsprechenden Input auch auf Diskursebene reagiert werden kann. Es können für bestimmte Diskurse typische Argumentationsfiguren oder Topoi produziert werden.

In der Kombination dieser Fähigkeiten ergibt sich eine Art „Simulationskompetenz“: Es ist ein soziolinguistischer Gemeinplatz, dass Menschen sich über ihr sprachliches Verhalten gesellschaftlich situieren und positionieren. Alle die angesprochenen linguistischen Ebenen spielen dabei eine Rolle. Textgenerierende KI ist demnach in der Lage, das passende Register und die adäquaten Sprachhandlungen in Abhängigkeit von diskursivem Wissen zu wählen, um sprachlich-kommunikative Rollen zu imitieren. So kann ein Chatbot gebeten werden, sich in die Rolle einer Tutorin in einem Lernprozess, einer Psychologin oder aber auch eines guten Freundes zu begeben. Selbstverständlich ergibt sich daraus auch das Potenzial, textgenerierende KI manipulativ einzusetzen und z.B. Kommunikation in sozialen Medien mit einer bestimmten Agenda zu unterwandern.

Es ist wichtig sich zu vergegenwärtigen, dass es sich bei den angesprochenen Fähigkeiten nicht um Wissen handelt, von dem die KI weiß, dass sie darüber verfügt. Das LLM hat nicht das morphosyntaktische Wissen, dass die 2. Pers. Sg. Präsens von „wissen“ „weiß“ ist. Das LLM kann jedoch sehr gut voraussagen, dass nach „du“ die Zeichenfolgen „weiß“ sehr viel wahrscheinlicher ist als die Zeichenfolge „weiß“. Gleiches gilt für die Reproduktion von Diskurswissen: Als Antwort auf den Prompt zu definieren, was „Demokratie“ bedeutet, schlägt ein LLM nicht in einer Enzyklopädie nach und generiert daraus eine Definition, sondern generiert die wahrscheinlichsten Zeichenfolgen, mit denen normalerweise auf solche Fragen geantwortet wird. Es ist aber problemlos möglich, sich mit einem Chatbot über die Konjugation von Verben zu unterhalten und dieser kann auch eine Tabelle aller Formen des Konjugationsparadigmas erzeugen, allerdings auch nur dadurch, dass im Sprachgebrauch genug metasprachliche Äußerungen vorhanden sind, die genau das tun: Also Beispiele für solche Tabellen.

Dass also diese Fähigkeiten alleine über statistisches Lernen von Mustern im Sprachgebrauch modellierbar sind, ist aus linguistischer Sicht sprachtheoretisch interessant. Es macht m.E. sehr deutlich, dass linguistische Theorien wie z.B. eine Grammatik oder die Morphologie nichts anderes als mögliche Modellierungen von Sprachgebrauch sind. Sie haben das Ziel, ein Modell vorzuschlagen, das in der Lage ist, Voraussagen über bestimmte Aspekte von Sprachgebrauch zu treffen. Es hat sich so einigermaßen bewährt, ein Kasussystem anzunehmen oder unterschiedliche Funktionen wie Nomen oder Verb zu unterscheiden, es gäbe aber auch andere mögliche Modellierungen. Ein LLM stellt nun eine solche alternative Modellierung von Sprachgebrauch dar, die einige Vorteile ausnutzt, die Maschinen

gegenüber Menschen haben: So kann in recht kurzer Zeit ein ziemlich großer Ausschnitt von Sprachgebrauch systematisch überblickt werden und die statistische Modellierung kann sehr komplex sein. Die Modellierung, wie sie eine von Menschen vorgeschlagene Grammatik entwirft, ist diesbezüglich wahrscheinlich „eleganter“ im Sinn von sparsamer definiert, dafür auch nicht ganz so perfekt in der Vorhersage.

4. Was fehlt?

Kommunikative Praktiken

Die Modellierung von den im Abschnitt zuvor beschriebenen linguistischen Ebenen mit Sprachmodellen führt also zu einem neuen Blick auf Sprachtheorie. Die bereits älteren distributionalistischen und performanzorientierten Konzepte der Linguistik können jedoch gleichzeitig als theoretische Grundlage für die heutigen Sprachmodelle angesehen werden. Allerdings sehen wir bislang nur einen Teil dieser performanzorientierten Konzepte umgesetzt und es muss deshalb gefragt werden, was denn noch fehlt. Ich möchte zur Beantwortung dieser Frage auf einen spezifischen Aspekt von Performanz fokussieren, nämlich Praktiken im Verständnis der Praxistheorie.

Die aus der Soziologie kommende Praxistheorie (Reckwitz 2003) kann als eine Art Gegenreaktion auf den Linguistic Turn verstanden werden: Während letzterer die Bedeutung von Sprache hervorhebt und mit dem Diskursbegriff argumentiert, um auf die sprachliche Verfasstheit von gesellschaftlichen Strukturen zu verweisen, betont die Praxistheorie wiederum gerade Nicht-Sprachliches: Körper, Materialität und Raum.

Deppermann et al. (2016) haben jedoch gezeigt, dass die Linguistik wiederum sehr gut den Praktikenbegriff integrieren und auf „kommunikative Praktiken“ zuspitzen kann. Die Gesprächslinguistik zeigt etwa seit geraumer Zeit, dass sprachliche Interaktion mit Körperlichkeit, Materialität und Räumlichkeit zusammengedacht werden muss. Und Hausendorf (2022) zeigt, wie Räume ein Abbild von sprachlichen Routinen sind und diese wiederum mitprägen (vgl. z.B. den Hörsaal, das Parlament, den Kirchenraum oder den Schalter). Wie stark die Einbettung kommunikativer Äußerungen in einen bestimmten durch körperliche Anwesenheit geprägten Raum beeinflusst ist, wird in der Kommunikation mit einem browserbasierten Chatbot deutlich: Er sieht mir meinen Gefühlszustand nicht an, mit dem ich meine Frage genervt in die Tasten hämmere und fühlt auch nicht die räumliche Situation, in der ich mich befinde. Diese Kontextfaktoren beeinflussen jedoch stark meine kommunikative Interaktion.

Ich möchte nun den Fokus auf zwei wichtige Elemente kommunikativen Praktiken legen, deren Bedeutung die Praxistheorie für menschliche Interaktion postuliert: Körperlichkeit und Räumlichkeit. Diese Elemente müssten demnach auch wichtig sein, wenn wir über die Zukunft von KI nachdenken und antizipieren wollen, welche Ebenen von Sprache und Kommunikation zu den bestehenden noch modelliert werden, um menschliche Interaktion nachzuahmen. Ich beginne mit dem Element der Körperlichkeit.

Körperlichkeit

Es ist interessant, dass die Geschichte der humanoiden Robotik untrennbar mit künstlicher Intelligenz verknüpft ist. Bis zum jüngsten Durchbruch der LLMs in Form von Chatbots standen Roboter in Form von Humanoiden im Zentrum zumindest des öffentlichen Interesses (Brommer/Dürscheid 2021). Auf Webseiten wurden Avatare eingesetzt, die

versuchten, automatisiert Fragen zu beantworten, jedoch nicht viel mehr als bessere Suchmaschinen waren.

Der Chatbot der Firma „OpenAI“ erscheint uns jedoch zunächst als Browserfenster, also fast jeglicher Körperlichkeit entledigt. Es ist selbstverständlich naheliegend, den Chatbot mit einem Roboter zu kombinieren. Doch das wäre — wenn wir aus praxistheoretischer Perspektive Körperlichkeit ernst nehmen — zu wenig weit gedacht. Denn aus dieser Perspektive ist ein Mensch nicht einfach die Kombination von Körper und Sprachfähigkeit, sondern das sprachliche Handeln ist untrennbar mit Körperlichkeit verbunden in dem Sinne, dass sprachliche Äußerungen zusammen mit körperlichem Handeln eine bestimmte Praktik konstituieren (z.B. die einfache Praktik des Grüßens).

Für KI-Anwendungen bedeutet das, dass entweder 1) Körperlichkeit in Kombination mit Sprache in den grundlegenden Lernprozess einbezogen werden müsste oder 2) sich aus den LLMs, die sich wie heute auf Textdaten stützen, eine andere Art der Körperlichkeit ergibt, als wir es für menschliche Kommunikation annehmen würden.

Im Prinzip ist die erste Option einfach, denn die Verdatung von Sprachgebrauch in Form von Vektoren stellt ein universelles Format dar: Alles Mögliche kann in Vektoren konvertiert werden, wie wir das am Beispiel von Bildern und Audio/Video-Daten bereits sehen. Die Vektorrepräsentationen können dabei auch miteinander kombiniert werden, so dass ein holistischeres Modell entsteht. Daher böte es sich an, alle Facetten menschlicher Kommunikation zu vektorisieren. Praktisch gibt es dafür einige Hürden zu nehmen, da menschliche Kommunikation in ihrer Komplexität von Interaktion und Multimodalität weniger Datenspuren hinterlässt als Kommunikation in Form von geschriebener Sprache. Zwar gibt es genügend Videodaten, die bereits heute als Trainingsdaten für multimodale LLMs verwendet werden, jedoch zeigt die Gesprächslinguistik, wie kompliziert die vollständige Erfassung von Gesprächen ist und beispielsweise Kameraaufnahmen aus verschiedenen Perspektiven notwendig sind, um multimodale Interaktion komplett zu erfassen. Es wird mit zukünftigen technischen Entwicklungen auf alle Fälle einfacher, solche Aufnahmen zu erstellen (entscheidend sind beispielsweise Kameras, die Tiefeninformationen registrieren und so einfach dreidimensionale Aufnahmen machen können), allerdings wird es noch eine Weile dauern, bis diese in genügend großer Menge vorhanden sein werden. Ohnehin stellen sich dabei gravierende rechtliche (Datenschutz, Recht am eigenen Bild) und ethische Probleme.

Es könnte also sein, dass derart komplexe LLMs nicht so schnell entstehen werden, deshalb ist die zweite Option relevant, nämlich zu fragen, welcher Art denn die „Körperlichkeit“ primär textbasierter, ggf. durch AV-Daten ergänzter LLMs ist. Um dies abzuschätzen, ist es sinnvoll, den Aspekt der Räumlichkeit mit in die Überlegungen zu integrieren.

Räumlichkeit

Die Medienwissenschaft, vor allem auch die Medienlinguistik, hat bereits darauf hingewiesen, dass die digitale Gesellschaft neue Formen der Kopräsenz erlaubt: Kopräsenz ist nicht nur möglich durch Interaktion am gleichen Standort (z.B. in Form eines Gesprächs in Präsenz), sondern auch digital vermittelt, und zwar in viel größerer Intensität, als es das Telefongespräch seit über hundert Jahren tut. So zeigt Dang-Anh im Anschluss an Knorr Cetina und Latour wie ein soziales Medium wie Twitter zu einer synthetischen Situation für Protestkommunikation wird, in dem Interaktionen an verschiedenen Orten synthetisiert und

auf einer Sehfläche kombiniert und dadurch erst zum wahrnehmbaren Protest wird (Dang-Anh 2019, 56). Auch bei Couldry und Hepp findet sich der Begriff der „tiefen Mediatisierung“ (Couldry/Hepp 2017), die damit betonen wollen, wie „analoges“ und „digitales“ Selbst miteinander verschmelzen. Und Nassehi sieht aus systemtheoretischer Sicht in der Digitalisierung ein soziales System, das alle anderen sozialen Systeme verdoppelt (Nassehi 2019, 141; vgl. für eine ausführliche Diskussion Bubenhofer/Dreesen 2022). Ohne hier ins Detail zu gehen, machen diese Überlegungen deutlich, dass in der ausdifferenzierten digitalen Gesellschaft die Grenzen zwischen hier und jetzt mit dort und vorher/nachher verschwinden: Körperlichkeit und Räumlichkeit müssen also als deutlich fluider verstanden werden. In einer immersiven VR-Welt wird ein Körperbewusstsein konstruiert, das den Körper als gleichzeitig in mehreren Räumen präsent verstehen kann.

Nun agieren wir aber mit einem Chatbot über ein Interface, das wir auf unserem Computer im Hier und Jetzt benutzen. Das Gegenüber ist zudem weder einfach eine medial vermittelte andere Person an einem anderen Ort, sondern ist ein Sprachmodell, das potenziell ganz viele Gegenüber an vielen anderen Orten simulieren kann. Das gleiche Sprachmodell kommuniziert mit vielen verschiedenen Personen gleichzeitig, erscheint jedoch in unterschiedlichen Formen: als Assistent, um eine Reise zu organisieren, als Psychotherapeutin, als Datenanalyst, als Programmierhilfe, als Freundin etc. Gleichzeitig kann das Feedback dieser Personen laufend verarbeitet und ins Sprachmodell zurückgespeist werden, wovon wiederum alle menschlichen Gesprächspartner profitieren.

Das Sprachmodell beruht zudem auf den sprachlichen Äußerungen von Personen in der Vergangenheit und zunehmend der Gegenwart, wenn die Modelle neu gerechnet werden.

Es kann nun eingewandt werden, dass wir Menschen ähnlich funktionieren, weil unsere Sozialisation ja ebenfalls über Sprachinput geschehen ist und wir dieses Wissen nutzen, um in neue Dialoge zu treten. In Analogie zum Sprachmodell hätten aber diese Menschen alle die gleiche Sozialisation durchlaufen, kannten alle die gleiche, jedoch sehr breite Palette an sprachlichen Routinen und Praktiken und würden aber einfach je nach Gesprächspartner eine andere Rolle spielen.

Wenn nun aber ein humanoider Roboter gebaut wird, der ein Sprachmodell nutzt, um „intelligent“ zu kommunizieren, wird das pure Gegenteil suggeriert: Ein (mechanischer) Körper mit einer sprachlichen Intelligenz erscheint als ein Selbst an einem spezifischen Ort und simuliert eine „klassische“ Körperpräsenz als Einheit von Körper und Raum. Der Roboter ist aber eigentlich alles andere als das, er ist also das deutlichste Beispiel für die Diskrepanz von simulierter Körperpräsenz und tatsächlicher Fluidität von Körper und Raum.

Diese Fluidität zeigt sich jedoch auch im personalisierten Interface mit einem Chatbot: Alle Einstellmöglichkeiten suggerieren Individualisierung (persönliches Log-in, Definition eines spezifischen Kontextes etc.). Die Beziehung zwischen Mensch und Chatbot wird sogar durch eine Authentifizierung definiert, mit der genau soziale Selbstverortung (ich bin tatsächlich das behauptete Ich und habe berechtigten Zugang) versprochen wird (Vogel 2022).

Bereits 1993 kritisierte Kittler mit Blick auf Software, dass grafische Benutzeroberflächen „eine ganze Maschine ihren Benutzern entziehen“ (Kittler 1993, 233), weil sie nur bestimmte Operationen zulassen. Mit dem Chatbot oder dem humanoiden Roboter verschwindet jedoch die ganze Maschinerie des statistischen Lernens von Mustern in menschlicher Kommunikation hinter der Simulation einer individuellen Kommunikation mit der Maschine.

Zum Schluss möchte ich nun darlegen, welche Konsequenzen dies im Umgang mit textgenerierender KI hat und welche Kompetenzen notwendig sind, um einen kritischen Umgang damit zu entwickeln.

5. Fazit: AI Literacy

Aus den bisherigen Überlegungen ergeben sich nun folgende Erkenntnisse:

- Indem Sprachgebrauch statistisch modelliert wird, ist generative AI in der Lage, verschiedene kommunikative Rollen zu simulieren.
- Es ist also die performative Seite von Sprache, die der Schlüssel zu dieser Simulation ist. „Performativ“ verweist einerseits auf die Gebrauchseite, andererseits auf den Handlungscharakter von Sprache. Weil mit Sprache soziopragmatisch gehandelt wird, sind die sprachlichen Muster im Sprachgebrauch als Spiegel der Gesellschaft so wertvoll.
- Die performanzorientierten Sprachtheorien fassen jedoch Kommunikation als Praktik auf und postulieren deshalb, dass Körperlichkeit und Räumlichkeit ebenfalls wichtige Aspekte soziopragmatischen Handelns sind. Diese beiden Aspekte müssen deshalb auch in unserer Konzeption von textgenerierender KI mitgedacht werden.
- So kann einerseits danach gefragt werden, wie KI mit Körperlichkeit und Räumlichkeit angereichert werden kann. Dies geschieht offensichtlich durch den Bau von Robotern, weniger offensichtlich, dass suggeriert wird, eine individuelle Beziehung zur KI aufbauen zu können, z.B. über personalisierte Softwareinterfaces und Authentifizierungspraktiken.
- Andererseits kann gefragt werden, welcher Art von Körperlichkeit und Räumlichkeit Sprachmodelle repräsentieren: Es sind Formen fluider Körperlichkeit und Räumlichkeit, weil Sprachmodelle auf kommunikativem Handeln ganz vieler Menschen beruhen. Gleichzeitig werten sie unser individuelles Handeln mit den Sprachmodellen wiederum aus und stellen das Musterhafte daran wiederum dem Gesamtsystem zur Verfügung.
- Während also die Softwareinterfaces individuelle Praktiken mit einer KI suggerieren, basieren die dahinterliegenden Sprachmodelle auf einer Konzeption von fluider Körperlichkeit und Räumlichkeit.

Dieser Widerspruch ist grundsätzlich nicht neu, sondern zeigt sich bei der Interaktion mit Computern schon lange (Stichwort: grafische Benutzeroberfläche). Die Folgen des Widerspruchs sind m.E. jedoch bei KI-Systemen drastischer, weil die Interaktion mit der KI *sprachlich* erfolgt und nicht über ein stark formalisiertes System von Befehlen: Unabhängig davon, ob die Interaktion mit einem klassischen Computer über Zeilenkommandos oder über eine grafische Benutzeroberfläche erfolgt, immer benutze ich ein vorstrukturiertes und recht starres System von Befehlen, die über Tastatureingabe, Mausclicks, Fingerwische o.ä. ausgelöst werden.

Ein sprachlicher Dialog mit einer KI ermöglicht jedoch eine andere Form von Interaktion: Ich könnte mein Tablet anreden mit:

„Hey, Mist, ich muss jetzt endlich diesen Text über KI fertigschreiben, aber ich habe gerade so überhaupt keine Idee!“

Die im Betriebssystem eingebaute KI könnte antworten mit:

„Ja, verstehe ich, du hattest ja auch eine anstrengende Woche mit den vielen Terminen. Schau, ich öffne mal den Text, ich habe hier mal einen Vorschlag, wie deine Argumentation fortgeführt werden könnte, was denkst du? Ich lese dir das mal vor.“

Es ist offensichtlich, dass diese Form von Interaktion eine individuelle Beziehung simulieren kann, die auch Körperlichkeit (hier: gesprochensprachliche Interaktion, Auswertung stimmlicher Qualitäten) und Räumlichkeit (hier: spezifisches räumliches Setting, z.B. im Zug) miteinbeziehen kann.

Es wird auch deutlich, dass die Maschine in solchen Settings nicht einfach ein Werkzeug ist, sondern eine Form von Agency hat, also handelt. Die Maschine wird zum Partizipanden, wie dies Steinhoff (im Druck; 2023) im Anschluss an den Partizipandenbegriff bei Hirschauer (2016) beschreibt. Es ist naheliegend, dass sich daraus neue Praktiken der Mensch-Maschinen-Interaktion entwickeln, bei der das „Aktivitätsniveau“ (Steinhoff 2023, 9) der Maschine sehr hoch ist.

Um nun einen kritischen und reflektierten Umgang mit KI zu entwickeln, ist m.E. „AI-Literacy“ notwendig, also die Kompetenz, verantwortungsvoll und produktiv mit KI umgehen zu können. Es sollte im Anschluss an die bisherige Argumentation deutlich geworden sein, dass diese AI-Literacy weit über Kompetenzen, KI-Tools richtig anzuwenden, hinausgeht. Wichtig ist auch nicht eine „Promptingkompetenz“, denn die eigentliche Maschinerie der Maschine wird nicht nur hinter grafischen Benutzeroberflächen verschwinden, sondern in den Tiefen von Betriebssystemen verborgen sein. Gleichzeitig wird uns KI sowohl in einer individuellen Körperlichkeit erscheinen, die eine exklusive Beziehung zwischen Mensch und Maschine suggeriert, als auch völlig unsichtbar in einer fluiden Körperlichkeit und Räumlichkeit. Medien- bzw. zeichentheoretisch kann das mit einer „transparent immediacy“ (Bolter/Grusin 2000, 272) oder der „Transparenz“ der Medialität (Jäger 2010, 317) umschrieben werden.

Um literat bezüglich KI zu werden, sind also Praktiken notwendig, die – ebenfalls im Anschluss an Jäger und Bolter/Grusin – das „Störungs-Prinzip“ ausnutzen, das Zeichen immer eingeschrieben ist. Im Folgenden nenne ich einige Aspekte von AI-Literacy und möglichen „Störpraktiken“, die als Form der Sensibilisierung und Kritik aufgefasst werden können.

- *Was ist ein Sprachmodell?*

Zentral ist das Verständnis dafür, was ein Sprachmodell ist, was dies sprachtheoretisch bedeutet und wie generierende KI grundsätzlich funktioniert. Dabei rate ich, eine *linguistische*, nicht eine *technische* Erklärung dafür zu finden.

- *Tool oder Partizipand?*

KI-Tools können unterschiedlich konzeptualisiert werden: Als Instrument ohne oder als Partizipand mit Agency. Dies prägt unsere Form der Interaktion mit KI und beide Formen haben ihre Berechtigung. Wichtig ist jedoch zu verstehen, dass diese Konzeptionen von verschiedenen Akteuren mit unterschiedlichen Zielen nahegelegt werden. Daher muss die jeweilige Konzeption hinterfragt werden, beispielsweise indem das KI-Tool gerade nicht oder gerade doch als Partizipand behandelt wird. Und die Aufgabenteilung zwischen KI und Mensch muss entsprechend reflektiert werden (vgl. dazu Brommer/Rezat im Druck).

- *Körperlichkeit und Räumlichkeit von KI?*

Es lohnt sich zu hinterfragen, welche Körperlichkeit und Räumlichkeit uns in Sprachmodellen oder generell KI-Tools begegnet – explizit und implizit. In einem eigenen Experiment implantierte ich ein Interface zu OpenAIs „ChatGPT“ in einen

Roboter in Katzenform (vgl. Abbildung 3). Der sprachliche Input des Menschen wird dabei in einen Prompt an ChatGPT überführt, wobei der Prompt den Chatbot dazu bringt, eine Antwort auf den menschlichen Input als Katze, in Form von Bewegungen, zu formulieren. Das Sprachmodell wird also gezwungen, eine eigentlich verbale in eine körperliche Antwort zu übersetzen und mit dem Roboterkatzenkörper auszudrücken (Bubenhofner 2024). Die Bastelei führt automatisch zu einer Reflexion darüber, inwiefern Spuren körperlichen Handelns in Texten zu finden sind oder wie sich das Interface von ChatGPT und unsere Interaktion damit verändert, wenn das Interface eine Roboterkatze ist.

Diese Liste ist nicht abschließend, sondern stellt Grundelemente dar, die ein Fundament von AI-Literacy darstellen können.



Abbildung 3: Roboterkatze „CatGPT“, vgl. für ein Video: <https://www.bubenhofner.com/sprechtakel/2023/10/04/catgpt-wenn-sich-ein-sprachmodell-bewegt/> (Bild: NB)

Ebenso fruchtbar wird jedoch die Frage danach sein, welche Erkenntnisse die Linguistik aus den Erfahrungen mit textgenerierender KI ziehen kann. Diese Erfahrungen sind die empirische Basis, um wiederum die Sprachtheorien zu testen. Sind die Grundannahmen performanzorientierter Sprachtheorien haltbar? Oder bieten andere Theorien, etwa stärker kompetenzorientierte wie die Generative Grammatik, die besseren Erklärungsmodelle?

Nicht zuletzt wird der Vergleich menschlicher sprachlicher Interaktion mit maschinell simulierter in allen Kombinationen (Mensch-Mensch, Mensch-Maschine, Maschine-Maschine) wichtig sein, um nicht nur im wissenschaftlichen Kontext diese Interaktionen empirisch zu untersuchen, sondern auch im Alltag kritisch zu reflektieren.

6. Literatur

- Bolter, Jay David/Grusin, Richard (2000): *Remediation: Understanding New Media*. Cambridge, Mass: MIT Press.
- Brommer, Sarah/Dürscheid, Christa (Hrsg.) (2021): *Mensch. Maschine. Kommunikation: Beiträge zur Medienlinguistik*. Tübingen: Narr Francke Attempto.
- Brommer, Sarah/Rezat, Sara (im Druck): *Mensch-KI-Interaktion beim Schreiben – Theoretische Überlegungen zur Modellierung des Schreibprozesses*. In: Weder, Mirjam/Bubenhofer, Noah (Hrsg.): *Schreiben mit KI in der Literatur und im Alltag*. Bielefeld: Transcript.
- Bubenhofer, Noah (2017): *Kollokationen, n-Gramme, Mehrworteinheiten*. In: Roth, Kersten Sven/Wengeler, Martin/Ziem, Alexander (Hrsg.): *Handbuch Sprache in Politik und Gesellschaft (= Handbücher Sprachwissen 19)*. Berlin, Boston: De Gruyter. S. 69–93. DOI: 10.1515/9783110296310-004.
- Bubenhofer, Noah (2024): *Grammatik*. In: *Zeitschrift für Medienwissenschaft* 1 (30), S. 53–55. DOI: 10.25969/mediarep/21973.
- Bubenhofer, Noah/Dreesen, Philipp (2022): *Kollektivierungs- und Individualisierungseffekte*. In: Gredel, Eva (Hrsg.): *Diskurse – digital*. Berlin, Boston: De Gruyter. S. 173–190. DOI: 10.1515/9783110721447-009.
- Couldry, Nick/Hepp, Andreas (2017): *The mediated construction of reality*. Cambridge, UK: Polity Press.
- Dang-Anh, Mark (2019): *Protest twittern: eine medienlinguistische Untersuchung von Strassenprotesten (= Locating media Band 22)*. Bielefeld: Transcript.
- Deppermann, Arnulf/Feilke, Helmuth/Linke, Angelika (2016): *Sprachliche und kommunikative Praktiken: Eine Annäherung aus linguistischer Sicht*. In: Deppermann, Arnulf/Feilke, Helmuth/Linke, Angelika (Hrsg.): *Sprachliche und kommunikative Praktiken*. Berlin, Boston: De Gruyter. S. 1–24. DOI: 10.1515/9783110451542-002.
- Evert, Stefan (2009): *58. Corpora and collocations*. In: Lüdeling, Anke/Kytö, Merja (Hrsg.): *Corpus Linguistics (= Handbücher zur Sprach- und Kommunikationswissenschaft 29)*. Berlin, New York: Mouton de Gruyter. S. 1212–1248.
- Feilke, Helmuth (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt am Main: Suhrkamp.
- Feilke, Helmuth (2003): *Textroutine, Textsemantik und sprachliches Wissen*. In: Linke, Angelika/Ortner, Hanspeter/Portmann-Tselikas, Paul R (Hrsg.): *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*. Tübingen: Niemeyer. S. 209–230. (= Reihe Germanistische Linguistik).
- Feilke, Helmuth (2016): *2. Einführung: Sprache – Kultur – Wissenschaft*. In: Jäger, Ludwig/Holly, Werner/Krapp, Peter/Weber, Samuel/Heekeren, Simone (Hrsg.): *Sprache – Kultur – Kommunikation / Language – Culture – Communication. Ein internationales Handbuch zu Linguistik als Kulturwissenschaft / An International Handbook of Linguistics as a Cultural Discipline*. Berlin, Boston: De Gruyter. S. 9–36.
- Firth, John Rupert (1957): *Modes of Meaning*. In: *Papers in Linguistics 1934–1951*. London: Oxford University Press. S. 190–215.
- Hausendorf, Heiko (2022): *Over the counter*. In: Christmann, Gabriele B./Knoblauch, Hubert/Löw, Martina (Hrsg.): *Communicative Constructions and the Refiguration of Spaces*. London: Routledge. S. 194–224. DOI: 10.4324/9780367817183-14.

- Hirschauer, Stefan (2016): Verhalten, Handeln, Interagieren: Zu den mikrosoziologischen Grundlagen der Praxistheorie. In: Schäfer, Hilmar (Hrsg.): Sozialtheorie. Bielefeld: Transcript. S. 45–68. DOI: 10.14361/9783839424049-003.
- Jäger, Ludwig (2010): Intermedialität – Intramedialität – Transkriptivität. Überlegungen zu einigen Prinzipien der kulturellen Semiosis. In: Deppermann, Arnulf/Linke, Angelika (Hrsg.): Sprache intermedial: Stimme und Schrift, Bild und Ton. Berlin: de Gruyter. S. 301–324.
- Kittler, Friedrich (1993): Draculas Vermächtnis. Technische Schriften. Leipzig: Reclam.
- Lenci, Alessandro (2018): Distributional Models of Word Meaning. In: Annual Review of Linguistics 4 (1), S. 151–171. DOI: 10.1146/annurev-linguistics-030514-125254.
- Mikolov, Tomas/Chen, Kai/Corrado, Greg/Dean, Jeffrey (2013): Efficient Estimation of Word Representations in Vector Space. In: arXiv:1301.3781 [cs].
- Nassehi, Armin (2019): Muster: Theorie der digitalen Gesellschaft. München: C.H. Beck.
- Reckwitz, Andreas (2003): Grundelemente einer Theorie sozialer Praktiken / Basic Elements of a Theory of Social Practices: Eine sozialtheoretische Perspektive / A Perspective in Social Theory. In: Zeitschrift für Soziologie 32 (4), S. 282–301. DOI: 10.1515/zfsoz-2003-0401.
- Sinclair, John (1991): Corpus, Concordance, Collocation. Oxford: Oxford University Press.
- Steinhoff, Torsten (im Druck): Künstliche Intelligenz als Ghostwriter, Writing Partner und Writing Tutor. Zur Modellierung und Förderung von Schreibkompetenzen im Zeichen der Automatisierung und Hybridisierung der Kommunikation am Beispiel von ChatGPT. In: Albrecht, Christian et al. (Hrsg.): Personale und funktionale Bildung im Deutschunterricht. Theoretische, empirische und praxisbezogene Perspektiven. Stuttgart: Metzler.
- Steinhoff, Torsten (2023): Der Computer schreibt (mit). Digitales Schreiben mit Word, Whatsapp, ChatGPT & Co. als Koaktivität von Mensch und Maschine. In: MiDU - Medien im Deutschunterricht (1), S. 1–16. DOI: 10.18716/ojs/midu/2023.1.4.
- Steyer, Kathrin (2013): Usuelle Wortverbindungen: Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht. Tübingen: Narr Francke Attempto.
- Vaswani, Ashish/Shazeer, Noam/Parmar, Niki/Uszkoreit, Jakob/Jones, Llion/Gomez, Aidan N./Kaiser, Lukasz/Polosukhin, Illia (2017): Attention Is All You Need. arXiv. <https://arxiv.org/abs/1706.03762> (letzter Zugriff 26.11.2024).
- Vogel, Friedemann (2022): Identifizierung und Authentifizierung in digitalen Diskursen. In: Gredel, Eva (Hrsg.): Diskurse – digital. De Gruyter. S. 191–212. DOI: 10.1515/9783110721447-010.