

GENERATIVE KI: ZWISCHEN WERKZEUG UND KOMMUNIKATIONSPARTNER

Benedikt Zönnchen
Hochschule München University of Applied Sciences
Munich Center for Digital Sciences and AI

5. April 2024

1 EINLEITUNG

Im November bot sich mir die Gelegenheit, in zwei Impulsvorträgen das Feld der *Generativen Künstliche Intelligenz* kurz zu beleuchten. Der erste Vortrag setzte sich mit dem Einsatzgebiet in der sog. *Extended Reality XR für den Mittelstand* (2023) auseinander, und der zweite widmete sich den wandelnden Herausforderungen und Chancen für die *Hochschulbildung* thematisiert im *Dialogforum Generative KI und Hochschule* (2023). Ich sprach von den verschiedenen Möglichkeiten, mit Hilfe der Generative KIs neue mediale Inhalte erzeugen—von Text-zu-Bild, Bild-zu-Bild bis hin zu Text-zu-Video Generatoren. Zudem gab ich eine Prognose darüber ab, in welchem Ausmaß diese Werkzeuge beide Domänen verändern werden, einschließlich der Frage, wie sich möglicherweise die Menge und Qualität der erzeugten medialen Inhalte verändern wird.

Generative KI ist mittlerweile zu einem zentralen öffentlichen Diskussionspunkt herangewachsen, umgeben von Debatten über ihre gesellschaftlichen Auswirkungen, ethische Bedenken, die Disruption des Arbeitsmarkts und natürlich die Veränderung der Bildungslandschaft. Die außerordentlichen Fähigkeiten dieser Technologie beeindrucken und verängstigen viele Menschen. Da diese undurchsichtigen Systeme unglaublich verführerische Text generieren, vermuten selbst vereinzelt Ingenieur:innen so mancher Sprachmodelle einen „Geist in der Maschine“. Es ist die Rede von einer „Ära bewusstseinsfähiger Terminatoren“. Derartige Spekulationen über die angeblich gefährliche Intelligenz der Maschine vernebelt jedoch den Blick auf das Wesentliche.

Vor diesem Hintergrund zielten meine Mini-Vorträge darauf ab, einen neuen, bereits aufgegriffenen Blickwinkel vorzustellen, den ich für besonders aufschlussreich in der Debatte um Generative KI halte. Dieser Artikel widmet sich genau diesem Blickwinkel. Er schlägt möglicherweise eine Brücke zwischen der aktuellen Technologie, ihrer Anwendung und möglichen Auswirkung.

2 WAS IST GENERATIVE KI?

Die folgende, informelle Beschreibung soll eine Intuition vermitteln. Für eine umfassende Übersicht verweise ich auf (Gozalo-Brizuela & Garrido-Merchan, 2023).

Wenn wir heute von Generativer KI sprechen, beziehen wir uns in der Regel auf IT-Systeme, die auf tiefen generativen Modellen basieren. Genauer gesagt, sprechen wir über *Maschinelles Lernen* oder noch genauer über *Tiefes Lernen*. Das definierende Merkmal dieser Modelle ist ihre Fähigkeit, neue Instanzen¹ zu erzeugen, die gemäß einer „erlernten“ Wahrscheinlichkeitsverteilung plausibel sind. Das bedeutet, es ist statistisch sehr schwierig zu unterscheiden, ob die generierte Dateninstanz aus den Trainingsdaten stammt oder nicht. Die Modelle generieren nicht einfach eine Instanz aus den Trainingsdaten², sondern ziehen diese von einer Wahrscheinlichkeitsverteilung. Diese Wahrscheinlichkeitsverteilung wird durch die Trainingsdaten, die Modellarchitektur und dem Trainingsprozess—kurz, durch das trainierte Modell—modelliert. Die neuen Instanzen sind somit weder beliebig noch exakt determiniert. Die Modelle sind *tief*, da sie auf einem neuronalen Netz basieren, das aus mehreren Schichten von miteinander verbundenen Einheiten oder Neuronen besteht.

Im Unterschied zu Modellen die eine Eingabe klassifizieren sollen—also beispielsweise berechnen, ob es sich auf einem Bild um eine Katze oder einen Hund handelt—generieren diese Modelle neue Katzen- bzw. Hundebilder. Für die Unterscheidung ist es nicht notwendig, die gesamte Verteilung der Pixel von Hunde- bzw. Katzenbildern zu kennen. Es reicht aus, wenn die Grenzen zwischen Hund- und Katzenbildern „gelernt“ werden. Für die Generation wird hingegen eine Annäherung der Wahrscheinlichkeitsverteilung „gelernt“. Es ist deshalb einleuchtend, dass die Generation deutlich schwieriger ist als die Unterscheidung. Dies spiegelt sich in der benötigten Datenmenge, der Trainingszeit und der Größe der Modelle wider.

Heutige Sprachmodelle, wie ChatGPT, gehören zu einer Klasse von generativen Modellen, die in diesem Artikel mein besonderes Interesse wecken, da sie Text erzeugen. Wenn wir viele Datensätze aus der modellierten Verteilung eines dieser Modelle ziehen, erhalten wir Texte, deren (empirische) Verteilung so aussieht, als stammen sie direkt von der modellierten Verteilung. Im Falle von Chatbots wird die gesamte Verteilung durch unsere Eingabe bedingt, das heißt, unsere Eingabe bildet den Anfang des erzeugten Textes und bestimmt somit, was folgen kann. Zudem wurden die Verteilungen, die die Chatbots modellieren, so *ausgerichtet*, dass sie sich als Assistenten eignen: Auf eine Frage folgt gewöhnlich keine weitere Frage, sondern eine Antwort. Interessant ist auch, dass für jedes Wort³ ungefähr die gleiche Rechenarbeit geleistet wird. Deshalb kann es sinnvoll sein, das Modell ruhig etwas mehr *schwätzen* zu lassen.

Anhand der Arbeitsweise, in Form der sukzessiven Vorhersage des nächsten Worts, verbunden mit einem immer ähnlichen Rechenaufwand, lässt sich schwer erkennen,

¹Die Ausgabe des Modells.

²Die Daten durch die das Modell optimiert wird.

³Eigentlich generiert das Modell Tokens und ein Token repräsentiert unter Umständen nur einen Teil eines Wortes.

dass diese Sprachmodelle ähnlich operieren wie unser „Geist“. Menschen überlegen gewöhnlich, welches Konzept oder welche Idee sie ausdrücken wollen, und finden dann einen Weg dorthin, d. h., eine wörtliche und damit abstrakte Repräsentation. Diese Repräsentation ist diskret, um auch über ein mit Rauschen behaftetes Medium kommunizieren zu können. Wenn wir stattdessen wie ein Sprachmodell reflexartig antworten, befinden wir uns in einem anderen Modus—im schnellen Denken, wie [Kahneman \(2011\)](#) es beschreibt. Das gleiche System ist aktiv, wenn wir einen Ball fangen—wir tun es einfach (unterbewusst), ohne dabei eine Differenzialrechnung aufzustellen oder zu lösen. Böse Zungen könnten nun behaupten, dass Sprachmodelle auch deshalb so intelligent wirken, weil wir selten zum langsamen Denken übergehen. Sie sind viel besser darin, unüberlegte Antworten zu liefern.

Es war schlussendlich sehr überraschend, dass es Forscher:innen durch das Prozessieren bzw. Modellieren von Sprache mithilfe von gigantisch großen Modellen gelang, Systeme zu bauen, die erstaunliche Fähigkeiten aufweisen, die wir intuitiv als „intelligent“ klassifizieren. Einer der Durchbrüche kam mit GPT-2 ([Radford et al., 2019](#)), als deutlich wurde, dass das „Lernen“ von Sprache eine Art des „*Multitask-Lernens*“ sein könnte. Das überraschte selbst die Entwickler:innen bzw. Forschenden. Im letzten Jahr wurde dann mit GPT-4 von „Sparks of artificial general intelligence: Early experiments with GPT-4“ ([Bubeck et al., 2023](#)) gesprochen. In der genannten Veröffentlichung werden die außerordentlich beeindruckenden Fähigkeiten aufgeführt, die von Programmierkenntnissen und mathematischen Kompetenzen bis hin zum Komponieren von Musik reichen. Allerdings ist hier Vorsicht geboten, denn wie [Bender und Koller \(2020\)](#) anmerken, ist das Prozessieren von Sprache nicht gleichbedeutend mit dem Verstehen von Sprache. Dennoch ist festzuhalten, dass diese Systeme im Laufe der Jahre den „Alltagsverstand“ für sprachliche Denkaufgaben übertroffen haben. Sie lösen Aufgaben, von denen viele dachten, sie könnten nur von einer Maschine bewältigt werden, die in dem Sinne denkt, wie wir es üblicherweise Menschen zuschreiben.

Wenn [Bender und Koller \(2020\)](#) richtig liegen, wovon ich ausgehe, und diese Modelle im Sinne des menschlichen Verstandes, eben nicht *verstehen*, finden wir dann einen besseren Begriff für das, was sie tun?

3 NUR EIN WERKZEUG?

In den angesprochenen Vorträgen vertrat ich folgende These:

Systeme der Generativen KI können auf einem Spektrum von Werkzeug bis Kommunikationspartner angeordnet werden.

Diese Idee ist durchweg inspiriert von Elena Esposito und ihrem Buch *Artificial Communication* ([Esposito, 2022](#)). Ich erinnerte mich an ihren Beitrag, als ich dem Vortrag von Björn Ommer auf der re:publica 2023 lauschte. Ommer vertrat die sinnige Ansicht, dass diese neue Technologie als Werkzeug zu sehen sei, eine Ansicht, die die meisten in der Technologiesphäre vertreten (siehe, z. B., [Epstein et al., 2023](#)).

Als Randnotiz sei erwähnt, dass sich Ommer heute darum bemüht, klugen Algorithmen und kleinen Modellen, und damit auch der Demokratisierung, eine Chance zu geben (Brühl, 2023)—eine Bemühung, die ich nur unterstützen kann.

Der Grund für die Kategorisierung der Generativen KI als ein *Werkzeug* scheint stimmig: Wir sollten der Vermenschlichung der Maschine entgegenwirken. Metaphern wie „Künstliche Intelligenz“, „Halluzinieren“, „Lernen“ und „Lehren“ verschleiern die Tatsache, dass es sich um eine konstruierte, unbelebte Maschine handelt. Insbesondere kann die Wahrnehmung menschenähnlicher Handlungsfähigkeit die Anerkennung der Schöpfer:innen, deren Arbeit den Ausgang des Systems unterliegt, untergraben und von der Verantwortung der Entwickler:innen und Entscheidungsträger:innen ablenken, wenn diese Systeme Schaden verursachen. Wir sprechen daher über Generative KI als Werkzeug zur Unterstützung kreativer Gestalter:innen anstatt eines Akteurs, der fähig ist eigene Absichten oder Urheberschaft zu hegen (Epstein et al., 2023). Das bedeutet natürlich nicht, dass wir unser Verständnis von Urheberschaft und geistigem Eigentum nicht anpassen oder dieses Eigentum mit neuen Regulierungen schützen müssen. Wenn wir Generative KI als Werkzeug verstehen, liegt der Schluss nahe, dass sie ähnliche Auswirkungen wie zuvor die Kamera und andere Werkzeuge haben wird. So wie die Kamera das Porträtzeichnen teilweise überflüssig machte, aber zugleich neue Möglichkeiten schuf, wird Generative KI aktuelle Arbeitsweisen durch neue Arbeitsweisen ersetzen—so die allgemeine These.

Epstein et al. (2023) betonen, dass diese Sichtweise wenig Spielraum für die „kreative“ Maschine übrig lässt, was wiederum die Frage aufwirft, ob wir den Menschen nicht auf ein zu hohes bzw. einzigartiges Podest stellen. Sind künstliche oder biologische Prozesse, wie die Bildung einer Ameisenstraße, von der Kreativität ausgeschlossen? Zumindest ist es schwer zu leugnen, dass viele Prozesse, die ohne unser Zutun ablaufen, eine gewisse Intelligenz aufweisen. Auch ist die starke Betonung der Individualität der Künstler:innen unverträglich mit systemischen Denkrichtungen, die insbesondere für eine nachhaltige Entwicklung als wesentlich angesehen werden (vergleiche, z. B., Pasqualino, Jones, Monasterolo & Phillips, 2015; Skene, 2020).

Wenn wir in der Zeit zurückblicken, können wir immer wieder Ängste feststellen, die durch technologische Entwicklungen ausgelöst wurden. Ein interessantes Beispiel ist die Einführung des Alphabets, welches einige damalige Denker:innen als „die Zerstörung des Gedächtnisses“ angesehen hatten. Rückblickend klingt das heute ziemlich absurd. In gewisser Weise behielten die Kritiker:innen jedoch recht. Die Menschen verlernten die Fähigkeit, sich epochale Geschichten zu merken. Aber sie lagen auch falsch, denn das Gedächtnis ist nicht verschwunden. Es wurde vielmehr vom „Geist“ aufs Papier gebracht; vom Menschen, als biologisches Wesen, in seine Umgebung.

Der moderne Mensch befindet sich in einem Netzwerk aus überlebenswichtigen Werkzeugen und sozialen, biologischen sowie ökologischen Systemen. Eine Klassifizierung bzw. Trennung zwischen dem atomaren Individuum als Verbindung von Körper und „Geist“ und seiner Umwelt scheint immer schwieriger zu sein. Mit einer systemischen Perspektive betrachtet, kann Technologie stattdessen als symbiotisch angesehen werden: Gesellschaften irritieren die Evolution der Technologie, und diese irritiert wie-

derum die Evolution der Gesellschaft.

Ist also die Generative KI doch nur ein weiteres Werkzeug? Tut sie es der Schrift und der Fotokamera gleich?

4 KÜNSTLICHE KOMMUNIKATION

Rückwirkend könnte man sagen, dass das Dilemma des Begriffs *Künstliche Intelligenz* mit dem sogenannten *Turing-Test* begann. Dieser Test sieht vor, dass eine Maschine oder ein Algorithmus als „intelligent“ angesehen werden kann, wenn Menschen nicht unterscheiden können, ob es sich um eine Maschine (bzw. Algorithmus) oder einen Menschen handelt. Der Test orientiert sich an Intelligenztests und betrachtet die KI als Blackbox. Es sei erwähnt, dass Turing in seiner Originalfassung drei Spieler:innen vorsah: Eine Person mit Eigenschaft A, eine Person mit Eigenschaft B und eine Fragestellerin, die herausfinden muss, welche der Personen die Eigenschaft A bzw. B besitzt. Dabei wird die Rolle der Person A unter Umständen von einer Maschine übernommen (Turing, 1950). Die Frage, die sich stellt, ist: Verändert sich die Erfolgsquote signifikant, wenn Person A von einer Maschine imitiert wird?

Wir wissen durch den Test nicht, ob diese Intelligenz simuliert wird (schwache KI) oder echt ist (starke KI)—was auch immer *echt* bedeuten mag. Viele würden behaupten, dass heutige Sprachmodelle diesen Test bestehen können. Demnach wären diese Modelle intelligent. Zugleich würden die wenigsten Entwickler:innen dieser Systeme behaupten, dass diese Algorithmen tatsächlich verstehen oder denken können, auch wenn die Fachterminologie des *Maschinellen Lernens* dies suggeriert. Solche Behauptungen sucht man außerhalb von Werbetexten vergeblich. Im Gegenteil, Forschende und Entwickelnde geben explizit bekannt, dass sie nicht versuchen, menschliche Intelligenz zu erschaffen. Nach Esposito (2022) sowie Rosengrün (2021) sind sie gerade deshalb so erfolgreich, weil sie von diesem Pfad abgewichen sind. Sie folgen damit dem Beispiel der Ingenieur:innen der Flugzeuge, die davon abgekommen sind, Vögel zu konstruieren.

Weshalb Sprachmodelle zwangsläufig an ihre Grenze stoßen begründen Browning und LeCun (2022) in einem philosophischen Aufsatz der viele alte Weisheiten aufwärmt. In gewisser Weise ist es der Irrweg des frühen Wittgenstein (1889–1951), den der späte Wittgenstein wieder ausräumt. Wittgenstein, aber auch viele andere Denker:innen des 19. und 20. Jahrhunderts, wie etwa Russell (1872–1970) und Frege (1848–1925) dachten, dass Sprache alles ausdrücken kann was wir wissen können, d. h. dass theoretisch alles Wissen in einer Bibliothek stehen könnte. Wittgenstein revidierte mit seinen *Sprachspielen*⁴ sein eigenes Projekt und später wird es Hubert Dreyfus (1929–2017) sein, der die frühen KI-Entwickler:innen mit mäßigem Erfolg darauf aufmerksam machen wird⁵ (siehe, z. B., Dreyfus & Dreyfus, 1986). Man kann wohl sagen, dass

⁴Wittgenstein beschreibt sprachliche Äußerungen als komplexe Handlungen, bei denen über die Kenntnis der Wörter hinaus auch das Verständnis für deren Funktion (in der echten bzw. konstruierten Welt) in einem bestimmten Zusammenhang erforderlich ist.

⁵Dreyfus bedient sich der Phänomenologie von Husserl, Heidegger und Merleau-Ponty.

Dreyfus einen Beitrag zum Scheitern der Versuche in der symbolischen KI geleistet hat.

Mit den künstlichen neuronalen Netzen hat man genau diesen symbolischen Weg teilweise verlassen, befindet sich jetzt jedoch wieder an einem ähnlichen Punkt. So argumentieren [Browning und LeCun \(2022\)](#):

„Alle Darstellungsschemata beinhalten eine Komprimierung von Informationen über etwas, aber was in der Komprimierung beibehalten oder weggelassen wird, variiert. Das Darstellungsschema der Sprache hat Schwierigkeiten mit konkreteren Informationen, wie der Beschreibung unregelmäßiger Formen, der Bewegung von Objekten, dem Funktionieren eines komplexen Mechanismus oder der nuancierten Pinselarbeit eines Gemäldes.“

– ([Browning & LeCun, 2022](#))

Dieser Gedanke klingt wie ein phänomenologisches Argument, nach dem wir Lebenswelt und Intelligenz nicht so einfach trennen können.

Anstatt eine riesige Datenbank logisch verbundener Sätze zu nutzen, wie es die symbolische KI tat, „lernen“ Sprachmodelle die Rolle von Wörtern, aber nur im Bezug auf andere Wörter und nicht im Bezug auf deren Funktion in einer vom Menschen konstruierten, kulturell geprägten Gesellschaft. Es ist erstaunlich und aufschlussreich zu sehen, wie weit man damit bisher gekommen ist. Dennoch glauben viele, dass wegen dieser linguistischen Limitierung eine KI, die tatsächlich planen kann, früher oder später Zugriff auf ein adaptives *Weltmodell* haben muss, um Situationen „durchzuspielen“ zu können. Wenn Dreyfus recht behält, ist das unmöglich. Er glaubte nicht an ein Weltmodell in unserem Kopf, sondern dachte, dass nur die Welt selbst das gesuchte Modell sein kann. Folgt man jedoch einer konstruktivistischen Auffassung⁶, wie sie etwa Niklas Luhmann (1927–1998) vertritt, besteht eine große Chance, dass sich Dreyfus in diesem Punkt irrt. Ob wir ein solches Weltmodell konstruieren können und dieses zu einer intelligenten Maschine führt, wird die Zukunft zeigen. Möglicherweise braucht es eine Kombination aus symbolischer und datengetriebener Modellierung, wobei die Fähigkeit symbolische Manipulationen durchzuführen, unter Umständen „gelernt“ werden kann ([Marcus, 2022](#))—jedoch nicht, indem man, wie es zuvor versucht wurde, mit Symbolen beginnt.

Wenn also heutige Maschinen und Chatbots nicht intelligent sind bzw. nicht denken, was tun sie dann? An dieser Stelle bringt Esposito Luhmanns Kommunikationsbegriff⁷ ins Spiel. Dieser ist konstruktivistisch definiert: Luhmann zufolge müssen Kommunikationspartner:innen nicht notwendigerweise einander verstehen. Vielmehr wird das psychische System (der „Geist“) der Empfängerin durch den Inhalt der Kommunikation irritiert. Der Sinn der Kommunikation wird von den Empfänger:innen konstruiert und ist nicht objektiv irgendwo „da draußen“. Das, was ein Text mir kommuniziert,

⁶Die Realität erscheint als Resultat einer Konstruktion der Beobachter:innen. Was natürlich nicht bedeutet, dass es die Realität nicht gäbe.

⁷Luhmann fasst den Begriff sehr breit, z. B. ist auch eine Zahlung oder eine Benotung eine Art der Kommunikation. Alle sozialen Systeme, wie etwa das Wirtschaftssystem oder das Bildungssystem, sind für ihn Kommunikationssysteme.

kann sich deutlich von dem unterscheiden, was die Verfasserin des Texts übermitteln wollte. Zudem kann die Interpretation beispielsweise von Literaturkritiker:innen als sinnvoller angesehen werden und über das hinausgehen, was die Autorin ursprünglich ausdrücken wollte. Die Interpretation kann weit von der ursprünglichen Intention abweichen. Für eine talentierte Künstlerin ist es daher nicht unbedingt erforderlich, den großartigen Sinn der eigenen Kunst zu verstehen, da dieser durch die Beobachter:innen konstruiert wird.

Deshalb führt Esposito für scheinbar intelligente Maschinen den Begriff der *Künstlichen Kommunikation* ein, um zu beschreiben, dass eine Generative KI nicht intelligent ist, sondern die Fähigkeiten erlangt hat, an Kommunikation⁸ teilzunehmen, ohne zu verstehen, was kommuniziert wird—genauso wie das Papier ein Gedächtnis darstellt, welches nicht denken kann. Damit umgeht sie den schwer zu definierenden Begriff der Intelligenz.

Im Begriff *Künstliche Kommunikation* schwingt deutlich weniger Magie mit. Es ist weniger Entzauberung seitens der Entwickler:innen der besagten Systeme und anderer Instanzen, wie uns Lehrenden, nötig. Auch glaube ich, dass sich mit dem Begriff *Kommunikation* besser erklären oder verstehen lässt, weshalb diese effektiv sein kann, obwohl einer der Kommunikationspartner den *Sinn*, den die andere Kommunikationspartnerin aus der Kommunikation *macht*, nicht begreifen kann. Kurz gesagt, es scheint für eine effektive Kommunikation unerheblich, ob sie vom Sender und der Empfängerin verstanden werden kann. Wichtig ist lediglich, ob der Inhalt für die Empfängerin Sinn ergibt. Eigentlich klingt das sehr einleuchtend, gerät aber im Diskurs ins Hintertreffen, wenn wir davon sprechen, „dass ChatGPT *verstehen* was ich meine“. Wir gehen davon aus, dass es verstehen muss, da es andernfalls keine Antwort geben könnte, aus der ich Sinn konstruieren kann. Den Begriff „*verstehen*“ sollte man deshalb, wenn überhaupt, nur metaphorisch nutzen.

Zusätzlich vermeiden wir durch den Ausdruck *Künstliche Kommunikation* dem Menschen zu viel Kontrolle über das Werkzeug zuzusprechen, und präzisieren, womit wir es zu tun haben und worauf wir uns möglicherweise vorbereiten müssen: Weder die KI, als übermenschlich intelligentes Wesen die Welt an sich reißen möchte, noch das vollständig kontrollierbare Werkzeug, stellt meiner Meinung nach eine besonders hilfreiche Vorstellung von der Generativen KI dar.

5 (AUSSER) KONTROLLE

Mit dem (unzuverlässigen) *Kommunikationspartner* kann man sich gegen den Begriff des *Werkzeugs* und des *intelligenten Wesens* stemmen. Was aber unterscheidet die Generative KI nun von einem Werkzeug?

Wie Esposito (2022) richtigerweise anmerkt, erwarten wir von einem Werkzeug, dass es funktioniert. Werkzeuge verschwinden aus unserer Wahrnehmung; sie werden zu einer Art Erweiterung unseres Körpers. Beim Schreiben dieses Textes denke

⁸Kommunikation kann über verschiedene Medien, z. B. Bilder oder Sprache geschehen.

ich nicht über die Tastatur nach, sondern konzentriere mich auf meine Gedanken und darauf, wie die einzelnen Buchstaben in meinem Textverarbeitungsprogramm auftauchen. Von einer Uhr erwarte ich, dass sie mir die richtige Uhrzeit anzeigt. Und von einer Fotokamera, erwarte ich, dass sie mir realistische Bilder erzeugt. Erst wenn die Uhr die falsche Uhrzeit anzeigt, das Drücken der Tasten keine Buchstaben mehr erscheinen lässt oder die Kamera lediglich schwarze Bilder erzeugt, wird mir das Werkzeug wieder bewusst. Werkzeuge werden sichtbar, wenn sie nicht so funktionieren, wie wir es erwarten.

Demgegenüber ist Kommunikation *kontingent*, also

„[...] etwas, was weder notwendig noch unmöglich ist; was also so, wie es ist (war, sein wird), sein kann, aber auch anders möglich ist. Der Begriff bezeichnet mithin Gegebenes (zu Erfahrendes, Erwartetes, Gedachtes, Phantasiertes) im Hinblick auf mögliches Anderssein; er bezeichnet Gegenstände im Horizont möglicher Abwandlungen.“ – (Luhmann, 1987, S. 152)

Die Definition passt besonders gut zum autoregressiven Charakter vieler Modelle. Beispielsweise verhält es sich bei Sprachmodellen so, dass jedes ausgegebene Wort alles, was noch kommen kann, einschränkt. Im Falle der Diffusionsmodelle, wie Dall-E, Stable Diffusion, schränkt jeder kleine Schritt der Reduzierung des Rauschens ein, was noch möglich ist.

Kommuniziere ich mit jemandem, weiß ich nicht, welche Antworten ich bekommen werde—ansonsten wäre die Kommunikation für den Informationsgewinn kaum von Nutzen. Auch weiß der Gegenüber nicht, wie ich reagieren werde, und wir wissen beide um diesen Umstand. Diese Situation bezeichnet Luhmann als *das Problem der doppelten Kontingenz*. Anders als bei einem Werkzeug gehe ich von dieser Kontingenz aus. Ich bin vielleicht irritiert, wenn mein Gegenüber etwas Merkwürdiges sagt. Und es gibt viele Regeln und Verhaltensweisen, um mit der doppelten Kontingenz klarzukommen, beispielsweise auf der rechten Straßenseite zu fahren. Die Möglichkeit des Missverständnisses ist aber immer präsent. Und doch sind die Antworten, so wie ich sie mir durch den Inhalt der Kommunikation konstruiere, auch nicht beliebig, d. h., ohne jedwede Einschränkung. Zum Beispiel kann ich nur aus bestimmten Sätzen Sinn erzeugen.

Analog sind generierte Bilder, z. B. produziert durch Midjourney, kein rein zufälliges Resultat. Gleichzeitig gibt es einen Bereich des Möglichen—eine mehrdimensionale Wahrscheinlichkeitsverteilung über einem latenten Raum, definiert durch die erlernten Modellparameter—den ich nicht verlassen kann. Für Nutzer:innen sind die Systeme undurchsichtig, wie auch die systeminternen Operationen der menschlichen Kommunikationspartner:innen für uns undurchsichtig sind.

Kann die Nutzerin sehr genau kontrollieren, welches Ergebnis ihr das KI-System liefert, wandert das System auf meiner imaginären Skala in Richtung *Werkzeug*. In der entgegengesetzten Richtung befindet sich der *Kommunikationspartner*. Wandern wir zu weit in diese Richtung, wird dieser unverständlich, und die Kommunikation schlägt

fehl. Dabei ist es wohl wichtig, zwischen Variationen und Überraschungen unterscheiden. Ich kann einen Algorithmus entwerfen, der für jeden Durchlauf ein anderes Bild erzeugt. Handelt es sich dabei jedoch lediglich um Variationen des gleichen Prinzips oder um die Generierung *unsinniger* Bilder, ist die Irritation entweder zu klein oder zu groß. Ähnlich wie ein Kunstwerk, erzeugt das Ergebnis Aufmerksamkeit, sofern es Sinn macht und trotzdem überrascht. Das gebannte Warten auf das Ergebnis des nächsten Prompts entsteht vielleicht durch diese Eigenschaft: Ich weiß nicht, was geschehen wird, und dennoch habe ich ein gutes Gefühl dafür, was nicht geschehen kann.

Beim Einsatz dieser Technologie bedeutet das für mich, dass wir nicht beides haben können: Das *Werkzeug* wird keine Überraschungen erzeugen—abgesehen von der Überraschung über seine Leistungsfähigkeit—und der *Kommunikationspartner* wird keine exakten Ergebnisse liefern. Je nach Anwendung und Einsatz können wir fragen, wo sich das System auf dieser Skala befinden sollte. Beispielsweise scheint es zunächst sinnvoll, dass Chatbots ausschließlich verlässliche Informationen liefern—Tatsachen eben! Aber funktioniert so eine effektive Kommunikation? Ist die Kontingenz der Kommunikation vielleicht sogar notwendig oder zumindest erwünscht? Wenn wir davon ausgehen, dass wir alle unseren eigenen Sinn erzeugen, dann sollte ein guter Kommunikationspartner dies berücksichtigen. Und wie verhält es sich für KI-Systeme im Kulturbereich? Sollten diese vielleicht eher einen unzuverlässigen Kommunikationspartner darstellen, der durch „Halluzinationen“ und andere Mittel der Beobachterin die Gelegenheit bietet, *Sinn zu erzeugen*, der die Kontingenz sichtbar macht und so bestehende Ordnungen in Frage stellt?

Ein weiterer wichtiger Zusammenhang, der in der Diskussion um Generative KI als reines Werkzeug oft zu kurz kommt, ist die besondere Stellung, die große Sprachmodelle einnehmen. Selbst wenn diese eben nicht verstehen, was am anderen Ende Sinn ergibt, so könnte es sein, dass zukünftig künstliche und soziale Systeme durch das Prozessieren von Sprache gekoppelt werden. Zum ersten Mal in der Menschheitsgeschichte wäre es möglich, fast fließend mit Maschinen zu kommunizieren. Diese Koppelung könnte der zwischen sozialen und psychischen Systemen ähneln. Die Schnittstelle zu diversen digitalen Systemen ist vermutlich bald, wenn gewünscht, als Sprachschnittstelle realisierbar und zwar mit all den Vor- und Nachteilen des kontingenten Charakters der Kommunikation. Eine solche Kopplung könnte auch bedeuteten, dass sich eine Symbiose bilden könnte, ähnlich der zwischen sozialen Netzwerken und deren Nutzer:innen.

Befreit vom Konkurrenzgedanken zwischen menschlicher und „künstlicher“ Intelligenz, können wir fragen, was wir von einem guten Kommunikationspartner erwarten. Da wäre zum Beispiel die *Anschlussfähigkeit*, d. h., mein Partner kommuniziert in einer Art und Weise, dass es für mich soweit Sinn ergibt, dass ich zumindest Rückfragen stellen kann. Von einer guten Kommunikationspartnerin erwarte ich auch, dass sie mir Erklärungen liefert, sodass ich nachvollziehen und verifizieren kann, was die Maschine getan hat. Es geht mir weniger darum, in das Gehirn meiner Gesprächspartnerin zu blicken und die genauen Gehirnoperationen zu verstehen, die sie dazu gebracht haben, das zu sagen, was sie gesagt hat. Wenn man große Sprachmodelle unter dieser

Prämisse betrachtet, sind sog. „Halluzinationen“ keine Fehler eines intelligenten Systems, sondern ein hilfreiches Mittel, um den Zweck der Kommunikation zu erfüllen: Der Algorithmus nimmt an der Kommunikation teil und versucht, diese aufrechtzuerhalten. Es ist ja in der Tat so, dass bei Nachfrage die Falschinformation gewöhnlich auffliegt, da das System eben nicht intelligent ist und seinen Fehler verschleiert oder auch nur versteht, dass es einen Fehler begangen hat.

Auch sollten wir beim Thema der *erklärbaren KI* beachten, dass wir beim Versuch, die internen Operationen zu verstehen, unter Umständen in einer Sackgasse landen. Aus solchen Erklärungen können, wie mir scheint, ohnehin nur Expert:innen Sinn erzeugen. Stattdessen könnte Kommunikation der Schlüssel sein, das heißt, Systeme müssen so gestaltet sein, dass Benutzer:innen aus der Kommunikation mit den Systemen Sinn erzeugen können. Die Systeme sollten sich selbst durch Kommunikation erklären können.

Selbstverständlich kann der konstruierte Sinn schädlich oder gefährlich sein, und mit Sicherheit bleibt das objektive Fenster in die Welt weiterhin ein unmögliches Ideal. Aus Luhmanns konstruktivistischer Sicht gibt es keinen objektiven Standpunkt, da Beobachtung immer selektiv sein muss und deshalb immer ein *blinder Fleck*⁹ entsteht. Daraus folgt sicherlich nicht, dass es keine besseren oder schlechteren Geschichten gäbe oder dass wir nicht in der Lage wären, mehrere Geschichten zu tolerieren. Um dieser Vielfalt Rechnung zu tragen, sollten unsere neuen künstlichen Kommunikationspartner unterschiedliche Verzerrungen—unterschiedliche Perspektiven mit unterschiedlichen blinden Flecken—kommunizieren.

Aus der fehlenden Intelligenz der KI-Algorithmen sollten wir nicht folgern, dass deren Einsatz unsinnig oder ineffektiv wäre, im Gegenteil. Ihr Erfolg, aber auch ihre Limitierungen gründen sich darauf, dass sie nicht wie Menschen denken. Deshalb glaube ich nicht, dass wir uns vor einer Superintelligenz zu fürchten brauchen. Vielmehr gibt es konkrete Herausforderungen, auf die sich die Forschung und die Gesetzeshüter:innen konzentrieren sollten. Probleme entstehen, da diese Algorithmen nicht intelligent sind.

Künstliche Kommunikation klingt zwar weniger bedenklich als Superintelligenz, bis uns klar wird, dass die gesamte Gesellschaft im Grunde genommen aus Kommunikation besteht.

„Gerade weil die Generative KI in die Kommunikation eingreift, müssen wir uns große Sorgen über ihre Auswirkungen machen.“
– (Esposito, 2024, S. 49)

Als Soziologin spricht Esposito insbesondere die *Desinformation* und die Verbreitung von *Fake News* an. Dabei macht sie deutlich, dass die Massenmedien bisher und auch in Zukunft kein objektives Fenster in die Welt bieten können. Stattdessen bauen sie

⁹Ein metaphorischer Begriff. Wann immer wir auf etwas blicken, müssen wir alles andere ausblenden. Wir brauchen stets eine Perspektive und auf diese Perspektive können wiederum die Blicke fallen. Jede Wahrnehmung beruht auf Bedingungen die sich eben dieser Wahrnehmung entziehen. Die Welt als Ganzes kann niemals wahrgenommen werden.

eine eigene spezifische Welt auf, die zur *Referenzwelt der Öffentlichkeit* wird. Diese ist nicht beliebig und hält die Gesellschaft in einem rastlosen Zustand. Die Gefahr ist nicht nur, dass falsche Informationen verbreitet werden, sondern dass die Referenzwelt zerbröckelt, sodass niemand mehr plausibel davon ausgehen kann, informiert zu sein und einer Gemeinschaft von Menschen anzugehören, die sich auf dieselben Nachrichten bezieht (Esposito, 2024, S. 66). Einerseits muss die offene Gesellschaft mehrere Welten tolerieren; andererseits wäre eine Gesellschaft, in der alle Individuen in ihrer eigenen Welt leben, wohl keine Gesellschaft mehr.

Esposito beendet ihr Buch mit einer nüchternen Prognose: Was wir in der Gegenwart als Neuheit erkennen, ist in gewisser Weise bereits alt. Radikale Veränderungen lassen sich nicht vorhersehen. Sie rät uns, in der Geschichte zurückzublicken, zu Zeiten, in denen sich die Kommunikation verändert hat—zur Entwicklung der Sprache, der Schrift und des Buchdrucks. Sie vermutet, dass die Rolle und Perspektiven der Menschen unverzichtbar bleiben, aber an Priorität verlieren werden, da es vielleicht nicht mehr notwendig ist, sie bei der Kommunikation zu berücksichtigen (Esposito, 2024, S. 79).

„Der Beitrag des Menschen bleibt unverzichtbar: Das Material, aus dem Daten abgeleitet werden, wird und muss weiterhin von Menschen produziert werden.“ – (Esposito, 2024, S. 79)

LITERATUR

- Bender, E. M. & Koller, A. (2020, Juli). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (S. 5185–5198). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.463
- Browning, J. & LeCun, Y. (2022). *AI and the limits of language*. <https://www.noemamag.com/ai-and-the-limits-of-language/>. (Accessed: 2024-02-16)
- Brühl, J. (2023). *Wie ein Münchner KI-Professor gegen den Größenwahn der Branche kämpft*. <https://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-ki-neurips-konferenz-bjoern-ommer-1.6318009>. (Accessed: 2024-02-16)
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. *Dialogforum Generative KI und Hochschule*. (2023). https://www.hm.edu/lehren/generative_ki/dialogforum.de.html. (Accessed: 2024-02-16)
- Dreyfus, H. L. & Dreyfus, S. E. (1986). From Socrates to expert systems: The limits of calculative rationality. In C. Mitcham & A. Huning (Hrsg.), *Philosophy and technology ii: Information technology and computers in theory and practice* (S. 111–130). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-4512-8_9
- Epstein, Z., Hertzmann, A., Herman, L. M., Mahari, R., Frank, M. R., Groh, M., ... Russakovsky, O. (2023). Art and the science of generative AI. *Science*, 380, 1110 - 1111. Zugriff auf <https://api.semanticscholar.org/CorpusID:259095707>
- Esposito, E. (2022). *Artificial communication*. The MIT Press. doi: 10.7551/mitpress/14189.001.0001
- Esposito, E. (2024). *Kommunikation mit unverständlichen Maschinen*. Residenz Verlag.
- Gozalo-Brizuela, R. & Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A state of the art review of large generative ai models*. arXiv. doi: 10.48550/ARXIV.2301.04655
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.

- Luhmann, N. (1987). *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Suhrkamp.
- Marcus, G. (2022). *Deep learning alone isn't getting us to human-like AI*. <https://www.noemamag.com/deep-learning-alone-isnt-getting-us-to-human-like-ai/>, note = Accessed: 2024-02-16.
- Pasqualino, R., Jones, A., Monasterolo, I. & Phillips, A. (2015, 23. 7). Understanding global systems today – a calibration of the world3-03 model between 1995 and 2012. *Sustainability*, 7 (8), 9864-9889. doi: 10.3390/su7089864
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. Zugriff auf <https://api.semanticscholar.org/CorpusID:160025533>
- Rosengrün, S. (2021). *Handbuch der Künstlichen Intelligenz* (G. Görz, U. Schmid & T. Braun, Hrsg.). Berlin, Boston: De Gruyter Oldenbourg. doi: doi:10.1515/9783110659948
- Skene, K. R. (2020, oct). No goal is an island: The implications of systems theory for the sustainable development goals. *Environment, Development and Sustainability*, 23 (7), 9993–10012. doi: 10.1007/s10668-020-01043-y
- Turing, A. M. (1950, 10). I.–Computing machinery and intelligence. *Mind*, LIX (236), 433-460. Zugriff auf <https://doi.org/10.1093/mind/LIX.236.433> doi: 10.1093/mind/LIX.236.433
- XR für den Mittelstand*. (2023). <https://www.ihk-muenchen.de/xr-mittelstand/>. (Accessed: 2024-02-16)