

Neue soziale Kontingenzmaschinen

Überlegungen zu künstlicher sozialer Intelligenz am Beispiel der Interaktion mit GPT-3

»Any sufficiently advanced technology is indistinguishable from magic«.

Arthur C. Clarke

»It's still magic even if you know how it's done.«

Terry Pratchett

Große Sprachmodelle (Large Language Models, LLM) wie das im Jahr 2020 öffentlich gewordene Modell GPT-3 zeigen in eindrucklicher Weise, was es heißt, wenn das Soziale auf sich selbst zurückgeworfen wird. Trainiert durch einen enormen Wissensvorrat in Textform und ohne Zugang zu sensorischer Wahrnehmung oder Körperlichkeit, gelingt es dem neuronalen Netz des Sprachmodells in einer nie dagewesenen Weise kohärente, kluge, überraschende und abwechslungsreiche sowie anregende Konversation zu betreiben. Im neuronal strukturierten Sprachmodell GPT-3 kommt das Soziale als sprachlich vermittelte Kommunikation gewissermaßen zu sich selbst, nämlich indem ausschließlich durch die wahrscheinlichkeitstriebe (Re-)Kombination von Worten neuer Sinn in der Welt erwirtschaftet wird.

Wenngleich diese rein textuell trainierten Sprachmodelle – anders als das menschliche Gehirn – nicht auf eine eigene Systemzeit zurückgreifen können und nur über ein beschränktes Gedächtnis verfügen, gelingt es ihnen unter anderem basale *Theory of Mind*-Tests zu bestehen (Agueira y Arcas 2021), sich mathematische Rechenwege zu erschließen (Cobbe et al. 2021), Haikus zu verfassen (Adams 2021), Spiele zu entwickeln (AI Dungeon), IQ-Tests teils besser zu bestehen als Menschen (Thompson 2021b), neue Poesie zu verfassen (Aalho 2021) sowie kluge und überraschende Konversationen zu betreiben.

Zum ersten Mal führt uns Technik damit vor, wie Sprachverständnis und Intelligenz ganz losgelöst von verkörperten, wahrnehmungsgeleiteten und emotionsbegleiteten Entitäten (*We're looking at you, human animal!*) erfolgreich prozessiert werden kann. Die sich hierin versteckende

1 Unser Dank gilt Werner Vogd und Nele Kost für ihre wertvollen Anmerkungen sowie die Unterstützung in der Erstellung dieses Beitrags!

Kränkung – Sprache und Intellekt war doch eben noch exklusiv uns Menschen und vielleicht noch einigen ausgewählten Tieren vorbehalten – ist jedoch für die Soziologie keine besonders große Überraschung. Gründet doch die Ausdifferenzierung dieser Wissenschaft auf der Beobachtung, Soziales durch Soziales erklären zu können und dabei bereits vom Biologisch-Körperlichen oder Geistigen abstrahieren zu können (Durkheim 1980). Als menschengemachte Veränderung in der Welt erscheint künstliche Intelligenz (KI) in diesem Lichte dann als eine weitere Möglichkeit zur Selbstvergewisserung bzw. Selbstaufklärung des Menschen.

In diesem Beitrag möchten wir anhand des Beispiels eigener Interaktionen mit GPT-3 auf einige ausgewählte neue Reflexionsperspektiven eingehen, die hier in Erscheinung treten. Anhand dieser Interaktionen wird deutlich, dass a) jede Kommunikation immer nur in einem Rahmen Sinn ergibt, b) das Geheimnis der DU-Position, verstanden als unbestimmte Subjektivität eines anderen, unerklärbar bleibt und deshalb nur durch Zuschreibungen gelöst werden kann, c) das Kommunikation immer Probleme der Indexikalität der Sprache erzeugt und deshalb auf ständige Reparaturarbeiten angewiesen ist sowie schließlich d) dass die Interaktionen mit anderen (ob Menschen, Tiere oder große Sprachmodelle) in Form einer Selbstvergewisserung immer auch auf die beteiligten Beobachter zurück verweist.

An der Interaktion mit GPT-3 möchten wir deutlich machen, wie alle Beteiligten (Mensch und Maschine) in Bezug auf die genannten Aspekte ein gemeinsames Arrangement finden: Sie arrangieren sowohl die Rahmung als auch das ›Geheimnis‹ jeder Du-Position und das Problem der Indexikalität in einer Weise, die uns Menschen dabei vorführt, aus welchen Strukturbedingungen auch wir selbst nicht entkommen können: nämlich in einer sozialen Sphäre der Sprache zu leben, welche uns mit Rahmungen, unbestimmter Perspektivität und der Notwendigkeit zur Reparatur von Verständnisproblemen versorgt.

Ausgangspunkt: Verlust von Bestimmtheit des Menschseins

Seinerzeit stellte Foucault fest, dass rund um das Thema Sexualität eine »diskursive Explosion« (Foucault 2019 [1976], 68) zündete. Nichts weniger scheint heute mit Blick auf die sogenannte »Digitalisierung« der Fall zu sein. Dabei wird gerade im Rahmen der Automatisierung von Alltags- und Organisationsprozessen ins Digitale oftmals der ›Verlust‹ von Mensch-Mensch-Interaktionen beklagt oder befürchtet. Das jedes Interface aber immer schon ein nicht-menschliches Gegenüber darstellt,

scheint erst durch anthropomorphisierte Applikationen wie ›intelligente‹ Sprachsteuerungssysteme ins Bewusstsein zu dringen: Immer weniger müssen wir von den uns vertrauten Interaktionsmustern abweichen, wenn uns beispielsweise die Yogalehrerin Julia als ›Skill‹ von ALEXA interaktiv zum Fitnessprogramm anleitet (vgl. Bialek 2019). Wir müssen nicht mehr klicken und drücken, wir sprechen. Wir unterhalten uns. Das Interface wird gewohnter.

Mit steigender Intelligenz, zunehmender Selbstständigkeit und ›Mächtigkeit‹ der Anwendungen wird der Diskurs rund um diese Technologie nicht weniger explosiv. Auf Ebene des Individuums finden wir hier eine Art psychologische Technikfolgenabschätzung, die vielfach Negativfolgen beleuchtet. Auf gesellschaftlicher Ebene werden nicht nur Integrationsverluste (etwa durch den Verlust von Arbeitsplätzen) befürchtet, auch Weltübernahmephantasien bahnen sich ihren Weg (vgl. etwa Dietz 2014). All dies scheint sich weniger aus unseren alltäglichen Technikerfahrungen zu speisen, denn aus einer Unsicherheit darüber, womit genau wir es eigentlich zu tun haben. Einerseits weist das digitale Gegenüber eine Kombinatorik auf, von der wir uns nicht einmal mehr richtig vorstellen können, in welchem Ausmaß sie die menschlichen Fähigkeiten der Informationsverarbeitung eigentlich übersteigt. Auf der anderen Seite wirkt die Technik des Alltags oftmals nur »grotesk unterkomplex« (Fuchs 2007, 23 5 FN8) und fehlerbehaftet. Gefangen im »Uncanny Valley« bleibt dann die Frage ob wir uns für die technische Meisterleistung des künstlichen Gegenübers begeistern können, oder abgeschreckt sind von deren befremdlicher Physiognomie und Motorik.²

Gerade der rasante Anstieg des technischen Fortschrittes seit Anfang der 2000er Jahre wirft zusehends eine Philosophie auf den Plan, die ihre Fragen nach Bewusstsein, Dasein und Leib auch auf künstliche Intelligenzen richtet: »Aus der phänomenologischen Perspektive gilt entweder das Bewusstsein (etwa bei Husserl), das Dasein (bei Heidegger) oder das leibliche Subjekt (bei Merleau-Ponty) als die transzendente Instanz, die als Möglichkeitsbedingung von Sinn überhaupt gelten kann« (Kristensen 2016: 58). Diese Setzung aufgreifend würde man einer Maschine kein Bewusstsein oder leibliche Subjektivität zurechnen. Aber dennoch besteht zumindest der Anschein, dass eine gut gebaute KI Sinn verarbeiten kann. Im Hintergrund solcher Betrachtungen steht damit jeweils mehr oder weniger latent die Frage, ob oder ab wann es sich bei einem Gegenüber um ein »Ding« handelt oder bereits um einen »anderen Anderen«

- 2 Technisches Entwicklungsziel ist letztlich, dass genau diese Reflexionsperspektive in der Interaktion gar nicht mehr aufscheint, also ein »Zurücksinken technischer Artefakte in das Universum der Selbstverständlichkeiten« (Blumenberg 2015, 190).

oder gar einen »anderen wie ich«.³ Diese Fragen nach einer Bestimmung des Anderen scheinen einer »Alteritäts-Bedürftigkeit« (vgl. Oliver Müller in diesem Band) zu entspringen. Unser Verdacht ist, dass diese Bedürftigkeit an die Bestimmung Anderer deckungsgleich mit der Bedürftigkeit nach Selbstvergewisserung ist. Wir kommen darauf zurück.

Entgegen ontologisierender Definitions- und Spezifizierungsversuche von Mensch, Tier oder Technik interessieren wir uns vor allem für die »Handlungskreise« (Gehlen 2007, 17), in denen diese Entitäten eine Rolle spielen und damit Bedeutung gewinnen. Nach Gehlen sind es die »plastische(n), gesteuerte(n), am rückempfundenen Erfolg oder Mißerfolg korrigierte(n) und schließlich gewohnheitsmäßig automatisierte(n) Bewegung(n)«, die als »wesenbezeichnende Eigenschaften« den Menschen ausmachen (vgl. ebd.).⁴ Hiermit drückt sich vielmehr ein bestimmter Modus des In-der-Welt-Seins, denn eine Essenz aus, die zu einer *conditio humana* erhoben wird. Gehlen zielt auf eine Gewöhnung und Routinisierung in einer Prozesshaftigkeit, die dann das »Wesen« ausmacht: In Form von prozesshaften Kreisläufen mit anderen Dingen, Menschen, Tieren etc. baut sich der Mensch eine mehr oder weniger stabile Welt auf. Die Beobachtung Gehlens deckt sich dann auch mit der Empirie: Ganz selbstverständlich scheinen wir schon immer ein dialogisches Verhältnis mit allen möglichen nicht-menschlichen Entitäten zu pflegen. Wir reden mit unserer Katze, halten unseren Staubsaugerroboter für besonders intelligent und beschimpfen unseren Computer.

Unser Ausgangspunkt liegt jedoch zusätzlich in einer über Gehlen hinausgehenden Position, die auf die Unhintergebarkeit fokussiert, die in der grundlegenden nicht-Beweisbarkeit jedes anderen liegt (Günther 2021).⁵ Eine solche Perspektive findet sich etwa bei Buber, wenn er

- 3 Interessant in diesem Zusammenhang ist die Studie von Hitzler (2012) zu Wachkomapatienten: »Es ist ausgesprochen ungewiss, ob das, womit man es im Umgang mit dem Menschen im sogenannten Wachkoma zu tun hat, (noch) ein Anderer ›wie ich‹ ist – oder nur ein ›Berg Menschenfleisch‹ (wie das eine verzweifelte Angehörige einmal formuliert hat [...]).« (Hitzler 2012, §1). Hitzler spricht hier dann von »Graduierungszereemonien« in denen das ontologisch immer unsicher bleibende Du performativ erzeugt wird.
- 4 Wenn wir der These Nassehis (2019) folgen, dass die Gesellschaft mit der Digitalisierung eine weitere Möglichkeit gewinnt, sich über sich selbst aufzuklären, können wir die konkrete Mensch-KI Interaktion strukturhomolog als Möglichkeit zur individuellen Selbstaufklärung bzw. Selbstreflexion sehen.
- 5 Die erkenntnistheoretische nicht-Beweisbarkeit des Anderen wird für uns *in der Praxis* nie zum Problem – es ist ein Theorie-Problem. Die Integration von nicht-menschlichen »Aktanten« (Latour 2005) ist in der ›Praxis‹ somit nichts Neues. Latour hat mit der ANT keineswegs die Welt neu erfunden, sondern vielmehr den Blick darauf gelenkt, was immer schon da war (vgl. auch Belliger & Krieger 2006). Die Tendenz zur Vernachlässigung von

schreibt: »Der Mensch wird am Du zum Ich. (...) Es gibt kein Ich an sich, sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es. Wenn der Mensch Ich spricht, meint er eins von beiden« (Buber 2008, 4) Buber startet den Versuch, das Du prinzipiell relational zu verstehen und damit die Bestimmung von Ich und Du in der Form eines Dazwischen aufzulösen. Erst im Nicht-Ich, nämlich im Dialogischen bzw. der gemeinsamen Interaktion kommt der Mensch zu sich.

In diesem Sinne wollen wir Buber ernst nehmen und weniger danach fragen, wie dieses Du eigentlich konstituiert ist, dass uns in den unterschiedlichsten Formen (etwa als KI) entgegentritt. Wir wollen den Kreisläufen Beachtung schenken, in denen ein ICH am DU zum ICH wird. Denn wenn es »kein Ich an sich« gibt, »sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es« (Buber 2008, 4), wäre weitergehend zu fragen, welche Voraussetzungen ein DU oder ES erfüllen muss, damit Handlungskreise entstehen, in denen das ICH an ihm um ICH werden kann. Der Mensch scheint ganz im Sinne Gehlens und Bubers so existentiell auf ein Gegenüber angewiesen zu sein, dass er hier gar nicht besonders wählerisch sein darf.⁶

In diesem Beitrag wollen wir nun die Frage nach der Begegnung mit KI zunächst nüchterner betrachten. Wir wenden die Frage nach dem DU in solchen Interaktionen soziologisch. Die Frage ist dann nicht, ob das jeweilige Gegenüber schon, noch oder zum welchem Grad wie ein Mensch ist. Solche Fragen dokumentieren aus unserer Sicht vor allem das »Bedürfnis des Menschen, sich in die Natur hinein auszulegen und sich von daher wieder zurückzuverstehen« (Gehlen 2007, 18). Die Frage nach dem DU erscheint soziologisch vielmehr als eine Frage der Zurechnung und weniger als eine Frage des vermeintlichen ontologischen Status.⁷ Und der Mensch scheint hier empirisch besehen eine außerordentliche soziale Plastizität aufzuweisen. Wir müssen demnach fragen, welche Relationen wir vorfinden und weniger, welche Relata.

Die Beschäftigung mit künstlicher Intelligenz im Allgemeinen sowie die Auseinandersetzung mit den Leistungen großer Sprachmodelle im Besonderen öffnet einen großen Horizont an Möglichkeiten des wissenschaftlichen Anschlusses. In diesem Beitrag versuchen wir daher, uns nicht zu sehr von diesen Möglichkeiten hinreißen zu lassen, sondern möchten uns vor allem auf einen Aspekt konzentrieren, der uns ganz besonders hervorsteicht. Es ist die soziologische Perspektive sowohl auf

Technik scheint dann das Produkt einer westlichen Gesellschaft zu sein, die den Menschen nach wie vor als Krone der Schöpfung und »Maß aller Dinge« (siehe kritisch Fuchs 2007) begreift.

6 Aus anderen Perspektiven stellen das auch Laing (1973, 69) und v. Glasersfeld (1988) fest.

7 Siehe zu dieser Unterscheidung bspw. Bateson (1985, 213) oder aktuell Karafillidis (2018).

die Bedingungen und Leistungen wie auch die Nutzung dieser neuartigen Technologie. Dafür haben wir die Begegnung mit GPT-3 gesucht, das derzeit als das leistungsfähigste, frei verfügbare Sprachmodell gilt. Als rekonstruktive Sozialforscher werden wir diese Begegnung nachträglich daraufhin beobachten, wie die Interaktion abgelaufen ist und welche Rahmenbedingungen, auch gesellschaftlicher Art, wir als Gelingensbedingungen für diese Interaktion identifizieren können.

Große Sprachmodelle im Kontext künstlicher neuronaler Netze

Die komplexer werdende Kompliziertheit algorithmischer Informationstechniken weist auf eine immer größer und sichtbarer werdende Interpretationsbedürftigkeit ihrer Funktions- und Anschlussweisen hin (Burrell 2016). Moderne selbstlernende neuronale Netze produzieren Outputs, die auch von ihren Erschaffer*innen nicht mehr exakt vorhergesehen werden. Zudem erzeugen sie ihren Zugriff auf die Welt zunehmend selbst: »Die auffälligste Änderung für mich ist, dass wir keine Daten mehr eingeben müssen«, sagt Demis Hassabis, CEO von DeepMind. Dieser bereits von Turing als »unorganized machines« (1948) bezeichnete Ansatz wird heute unter dem Label »unsupervised learning« verhandelt (LeCun et al. 2015) und in Modellen wie MuZero erfolgreich zur Anwendung gebracht (Schrittwieser et al. 2020). Die spezifische Leistung artifizieller neuronaler Netze ist, dass sie nicht mehr propositional funktionieren, sondern konnektionistisch. Künstliche neuronale Netze lernen selbständig durch die verstärkende oder hemmende Verknüpfung der Einheiten im Netzwerk – ohne damit zu wissen, was dieses Lernen später bedeuten wird.

Eine umfassende Geschichte des maschinellen Lernens (bzw. *deep learning*) würde den Rahmen dieses Beitrags sprengen (siehe als Einstieg Goodfellow et al. 2016). Tatsächlich geht Deep Learning bis auf die 1940er Jahre zurück und scheint nur deshalb neu zu sein, weil es vor seiner aktuellen Bekanntheit mehrere Jahre lang relativ unpopulär war. Zudem versammelten sich dessen Prinzipien über die Zeit hinweg unter verschiedenen Namen und wurden erst kürzlich »Deep Learning« genannt. Im Großen und Ganzen finden sich bislang drei Entwicklungswellen der grundlegenden Prinzipien: 1) die Überlegungen und Modelle der Kybernetik in den 1940er–1960er Jahren (siehe hierzu McCulloch und Pitts 1943; Hebb 1949; siehe auch die Forschung zu Perceptrons bei Rosenblatt 1958), 2) die zweite Welle, die in den 1980er–1990er Jahren unter dem Stichwort Konnektionismus bekannt wurde und das Prinzip der *backpropagation* nutzt, um neuronale Netze mit mehreren versteckten Schichten zu trainieren (siehe hierzu Rumelhart et al. 1986; Hochreiter

und Schmidhuber 1997). Die dritte Welle der Weiterentwicklung von Deep Learning Mechanismen begann ab 2006 und hält bis heute an (Hinton et al. 2006; Bengio et al. 2007).

Bis Ende der 1980er Jahre setzten Unternehmen und Wissenschaft auf die heute als GOFAI (*Good old fashioned AI*) bezeichneten Verfahren der künstlichen Intelligenz. Diese Verfahren beruhen auf Expertensystemen, Wissensdatenbanken und ›harten‹ Repräsentationen für die Symbolverarbeitung. Die konnektionistischen Verfahren der lernenden neuronalen Netze setzen hingegen auf die in Lernverfahren erworbenen Verbindungsstrukturen eines Netzwerks und nicht mehr auf von extern definierte Beziehungen. Das Paradigma des Konnektionismus besagt vielmehr, dass Informationsverarbeitung als Interaktion einer großen Zahl an Einheiten (biologische und/oder artifizielle Zellen bzw. Neuronen) angesehen wird, die verstärkende oder hemmende Signale an die anderen Einheiten im Netzwerk senden. Symbole werden damit nur noch implizit durch das emergente Aktivierungsmuster aller Einheiten als verteilte Repräsentation im Netzwerk dargestellt.

Entsprechend ihres biologischen Vorbilds waren einige der frühesten Lernalgorithmen als Computermodelle des biologischen Lernens gedacht (vgl. McCulloch und Pitts 1943). Eine der vielen Bezeichnungen für Deep Learning ist daher *artificial neural networks* (ANNs). Die Neurowissenschaft gilt hier zwar als wichtige Inspirationsquelle, sie ist aber nicht mehr der vorherrschende Maßstab für das Feld. Der Begriff ›Deep Learning‹ geht daher über die neurowissenschaftliche Perspektive auf derzeitige Modelle des maschinellen Lernens hinaus und bezieht sich heute auf ein allgemeineres Prinzip des konnektionistischen Lernens auf mehreren Kompositionsebenen.

Generative Pretrained Transformer (GPT-3)

Aufbauend auf der erfolgversprechenden Transformer Technologie (Vaswani et al. 2017) stellte das KI-Konsortium OpenAI Ende 2020 die dritte Iteration ihres Sprachmodells GPT (Generative Pretrained Transformer) unter dem Titel GPT-3 vor. Bereits in den ersten beiden Versionen konnte OpenAI beweisen, dass eine Verknüpfung der Aufmerksamkeits-Mechanismen der Transformer Technologie und unüberwachtem (*unsupervised learning* Vorab-Training) zu sehr guten Ergebnissen in der natürlichen Spracherzeugung (*natural language understanding*) führt (Radford et al. 2018; Radford et al. 2019).⁸

- 8 In ihrer Arbeit prüften Brown et al. (2020) die Hypothese, dass die bloße Skalierung von Sprachmodellen die aufgabenunabhängige Leistung des Modells erheblich verbessert und sogar die Leistung hochmoderner

Auch die für das Training dieser Sprachmodelle notwendigen Datensätze haben sich in den vergangenen Jahren rasch vergrößert. Neben Datensätzen wie der englischen Wikipedia verarbeitete GPT-3 auch große Textdatensätze wie Books¹ & ², OpenWebText² (Sammlung von Reddit Posts) sowie den zum damaligen Zeitpunkt größten Datensatz der Common Crawl-Initiative⁹ mit einem Umfang von etwa einer Billion Wörtern. Letztgenannter Datensatz machte ca. 60% des Trainingsmaterials aus, während zum Vergleich die gesamte englischsprachige Wikipedia nur etwa 3% des Datenmaterials abdeckt.

Seit der Einführung sogenannter *hidden units*, die als künstliche Neuronen im Netzwerk fungieren und den Output anderer *units* als Input verarbeiten, hat sich deren Anzahl in künstlichen neuronalen Netzen etwa alle 2,5 Jahre verdoppelt. Es bleibt daher spannend zu beobachten, wie weit sich dieser Trend der Skalierung auch in Zukunft noch beweisen kann. Gegenwärtig zumindest verspricht dieser Ansatz noch Erfolge: Die schiere Vergrößerung der Anzahl an verfügbaren Parametern¹⁰ liefert weiterhin bessere Ergebnisse.

Der Wettbewerb um das nach Parametern größte Sprachmodell ist seit Einführung von GPT-3 eröffnet: DeepMind stellte kürzlich sein Modell »Gopher« mit 280 Milliarden Parametern vor, das GPT-3 in allen derzeit relevanten Benchmarks übertrifft (Rae et al. 2021). Aktuelle Arbeiten von Nvidia ebnen zudem den Weg für Sprachmodelle mit mehr als 1 Billion Parametern (vgl. Narayanan et al. 2021). Dieses Wettrennen wird derzeit noch von den synaptischen Verbindungen des menschlichen Gehirns übertroffen, das als natürliches neuronales Netzwerk mehr als 100 Billionen Synapsen aufweist. Aktuelle Ansätze für die Entwicklung von Sprachmodellen in noch höherer Größenordnung sehen sich neben der Komplexität auch jeweils ressourcenbasierten Limitationen gegenüber: das Training neuer, noch größerer Modelle erfordert Zeit und Energie (und Datensätze). Obwohl Modelle wie GPT-3 während des Trainings beträchtliche Energieressourcen verbrauchen, sind sie nach dem Training überraschend effizient: Selbst mit dem vollen GPT-3 Modell benötigt die

Finetuning-Ansätze erreicht. Ihre These prüften sie an dem dann als GPT-3 bekannt gewordenen Sprachmodell mit bis zu 175 Milliarden Parametern, was etwa der zehnfachen Menge jedes vorher bekannten Sprachmodells entspricht.

9 Siehe hierzu: <https://commoncrawl.org/> (zuletzt aufgerufen am 19.01.2022).

10 Die Parameter eines neuronalen Netzes sind in der Regel die Gewichtung der jeweiligen Verbindungen im Netzwerk. Die Werte dieser Parameter werden in der Trainingsphase vom Modell gelernt. Im Wesentlichen findet sich daher keine explizite Kontrolle über die Parameter, während das Modell lernt. Der Lernprozess kann jedoch indirekt über Hyperparameter gesteuert werden, die in der Regel auf die Lernrate, die *batch-size* (Anzahl von zu verarbeitenden *tokens*) oder die Anzahl der Trainingsepochen abzielen.

Generierung von 100 Seiten Inhalt aus dem trainierten Modell nur eine Größenordnung von 0,4 kW-Stunden.

Neben dieser quantitativen Optimierung verfolgen aktuelle Deep Learning Vorhaben zudem auch multimodale Ansätze, die es erlauben, unterschiedliche Daten miteinander in Verbindung zu bringen. Spätestens seitdem der Chefentwickler von OpenAI vermerkte, dass multimodale Modelle das Ziel für 2021 seien, boomt auch dieser Forschungszweig. Bereits nur wenige Monate nach diesem Interview veröffentlichte OpenAI ihr erstes multimodales Modell namens DALL-E, das als Zero-Shot Text-to-Image Generator funktioniert und auf Grundlage von sprachlichen Eingaben neue Bilder erzeugen kann (vgl. Ramesh et al. 2021). Ramesh et al. stellen fest, dass auch in ihrem multimodalen Modell die Skalierung zu einer verbesserten Generalisierung führt, und zwar sowohl in Bezug auf die Zero-Shot-Performance als auch in Bezug auf die generelle Bandbreite an Fähigkeiten, die aus einem einzigen generativen Modell hervorgehen. Diesem Trend zu multimodal trainierten Transformer-Modellen folgt auch Google mit seinem Modell MUM (Multi-task Unified Model). Dieses ist mit Text-, Bild- und Videodaten in etwa 75 Sprachen trainiert und zielt auf die Verbesserung der Kernkompetenz von Google Web Search ab (Nayak 2021). Auch die Nutzung eines externen Speichers, auf den für die Generierung der Outputs zugegriffen werden kann, ist mit den »Retrieval-Enhanced Transformers« (RETRO) bereits vorgestellt worden (vgl. Borgeaud et al. 2021). Die Weiterentwicklung von GPT-3 um einen Echtzeitzugriff auf das Internet in der Version »WebGPT« (Nakano et al. 2021) liefert dann ebenfalls eine deutliche Verbesserung in Hinsicht auf die situationsbedingt teils weniger gefragte »kreative Textgenerierung«.¹¹

Eines der langfristigen Ziele der Forschung im Bereich des Deep Learning ist es, einen einzigen, universell einsetzbaren Lernalgorithmus zu entwickeln, der eine breite Palette von Problemen lösen kann. Eine mögliche Lösung wird in der Entwicklung evolvierender Meta-Lernmethoden gesehen, die eigenständig neue Algorithmen entwickelt und sich so automatisch auf eine Vielzahl von Aufgaben anwenden lassen. Kürzlich stellte Google einen solchen Ansatz für ein solches Meta-Lernen vor (Co-Reyes et al. 2021). Die auf diese Weise gelernten Algorithmen gelten als bereichsunabhängig und lassen sich daher auch auf Umgebungen anwenden, die beim Training nicht berücksichtigt wurden.

11 Diese generative Aufgabe wird auch als »Halluzinieren« bezeichnet, da das Modell aus der puren Selbstreferenz mehr oder weniger kohärente und plausible Informationen produziert. Bekannt wurde der Ausdruck vor allem vor dem Hintergrund Bild-produzierender GANs (Mordvintsev und Tyka 2015). Für den Bereich sprachlicher »Halluzinationen« als pathologischen Fall von Übersetzungsmodellen siehe Raunak et al. (2021).

Der (mediale) Diskurs über Leistungen und Grenzen von GPT-3

OpenAI war offenkundig selbst von der Leistungsfähigkeit und der damit verbundenen Tragweite ihrer Erfindung überrascht, weshalb sie den Zugang zu GPT-3 zunächst sehr stark eingrenzten. Nur ausgewählte Projekte bekamen Zugang zur API des Sprachmodells. Die Warteliste war entsprechen lang. Seit Ende 2021 ist der Zugang jedoch ohne weiteres möglich und nach Registrierung lässt sich GPT-3 in jedem Webbrowser für eine dreimonatige Phase kostenfrei ausprobieren.

Kurz nach der Veröffentlichung des Sprachmodells entfachte erneut ein Diskurs über die Leistungsfähigkeit von KI im Allgemeinen und GPT-3 im Besonderen im Hinblick auf die Vergleichbarkeit von künstlicher und menschlicher Intelligenz. Noch im Juli 2020 veröffentlichten neun Philosophinnen und Philosophen einige Einschätzungen, Kritiken sowie Überlegungen zu den Leistungen und Limitationen von GPT-3 (vgl. Weinberg 2020). Inhaltlich rangierten diese Beiträge von skeptischem Wohlwollen bis hin zu euphorischem Fürsprechen gegenüber dem Sprachmodell. David Chalmers etwa erkannte in GPT-3 bereits »hints of general intelligence«, wenn es dazu in der Lage ist, sogar einfache mathematische Aufgaben sowie Logikrätsel zu lösen, zu denen es nicht auf Material im Trainingsdatensatz zurückgreifen konnte (Chalmers 2020). Interessanterweise hat GPT-3 auf diese philosophischen Kommentare reagiert.¹²

Im September 2020 veröffentlichte die Zeitung *The Guardian* einen Artikel, der in ihren eigenen Worten von GPT-3 geschrieben wurde (vgl. GPT-3 2020).¹³ Spätestens dieser Artikel machte das Sprachmodell weltweit bekannt und führte zu einer Vielzahl an Nachahmungen, Erweiterungen und Weiterentwicklungen des skalaren Ansatzes. Noch im selben Jahr der Veröffentlichung von GPT-3 erschien das in Ko-Kreation mit dem Sprachmodell geschriebene Buch »Pharmako-AI« (Allado-McDowell 2020). Das Buch versammelt in teils irritierend anregender und experimentell-polyphoner Weise unterschiedliche Essays, Gedichte und Geschichten aus dem kontinuierlichen Austausch zwischen der menschlichen und künstlichen Autorin. Im Ergebnis steht das Buch als erstes Dokument für eine noch längst nicht abgeschlossene maschinell-unterstützte Kreativität und Inspiration. Seitdem sind eine Vielzahl an KI-unterstützten Büchern erschienen – von Kinderbüchern (Green 2020), über japanische Haikus (Adams 2021) und Comics (Niman 2021) bis hin zu Poesie (Aalho 2021; Copeland 2021).¹⁴

12 Siehe unter anderem Milliére (2020) und auch auf der Website des Guardian selbst.

13 Genau genommen wurden Textbausteine mithilfe von GPT-3 generiert und dann – von einem menschlichen Autor – bestimmte Ausschnitte zu einem kohärenten Artikel zusammengeschnitten.

14 Siehe auch die mittlerweile mehrere Dutzend Episoden lange Unterhaltung zwischen dem Computerwissenschaftler Thompson und GPT-3 (vgl.

In der Beobachtung des Diskurses zeichnet sich eine Position ab, die behauptet, dass maschinelles Lernen »nur Statistik« sei und die Fortschritte im Bereich von KIs im Hinblick auf das größere Ziel einer allgemeinen künstlichen Intelligenz (AGI) illusorisch seien (vgl. Marcus und Davies 2020; grundlegend: Searle 1980). Andere Autoren hingegen vertreten die gegenteilige Ansicht, dass wir bereits allein anhand von diesen Sprachmodellen viel über die Natur der Sprache, des Verstehens, der Intelligenz, der Sozialität und des Menschseins lernen können (vgl. Aguera y Arcas 2021). Aus soziologischer Perspektive ist es vollkommen klar, dass unsere vermeintlich individuellen Favoriten und Präferenzen, sensorischen Eindrücke, inneren Gefühle, Emotionen sowie auch Ideen, Gedanken und Meinungen viel weniger individuell sind, als es uns selbst erscheint – und vielleicht auch lieb ist. Nicht nur die Beschreibung dieser vermeintlich individuellen, eigenen Erfahrungen, sondern auch deren Genese und Wahrnehmung sind durch sozial wirksame Faktoren, also gesellschaftliche Formatierungen ko-konditioniert.

Interaktionen mit GPT-3: Material und Analyse

Für diesen Beitrag haben wir uns dem Sprachmodell GPT-3 genähert, um eigene Interaktionserfahrungen zu sammeln und zu reflektieren. Bereits die Frage, ob wir uns mit GPT-3 »getroffen« haben, oder ob wir uns zu zweit getroffen haben, um mit etwas Drittem zu interagieren, verweist auf die Notwendigkeit, dem, was da passiert ist, eine Rahmung zu geben. Entsprechend folgt unsere Reflexion der Dreiteilung in eine *Situationsanalyse*, eine *Interaktionsanalyse* sowie eine *Textanalyse*. Damit wollen wir insbesondere auf einen – aus unserer Sicht – blinden Fleck des Diskurses aufmerksam machen: nämlich inwiefern »technische Artefakte« nicht längst in das »Universum der Selbstverständlichkeiten« (Blumenberg 2015, 190) zurückgesunken sind.

Bei allen Irritationen auf die man in der Mensch-Technik Interaktion stößt, scheint allzu schnell in Vergessenheit zu geraten (=selbstverständlich zu werden), wie voraussetzungsvoll es ist, selbstverständlich in eine solche Interaktion eintreten zu können, von der man dann mehr oder weniger irritiert wird. Längst schon sind uns diverse Interfaces (beispielsweise Eingabemasken) in einer Weise vertraut, dass sie uns als fraglos gegeben erscheinen. Als zweites fällt auf, dass Beispiele der Interaktion mit KIs zur Selbstthematisierung neigen: So werden KIs gerne – in mehr oder weniger plumper und taktloser Art, so könnte man hinzufügen – danach befragt,

Thompson 2020), wo sowohl Fähigkeiten wie auch Grenzen des Sprachmodells in meist amüsanter und kluger Weise zwischen Mensch und KI ausgelotet werden.

ob sie zu überhaupt zu Liebe und Mitgefühl fähig seien¹⁵ oder ob sie intelligent und dann auch noch an der Weltherrschaft interessiert seien.¹⁶

Es bleibt zu fragen, ob sich hier nicht vielmehr eine *conditio humana* dokumentiert, als das Aussagen über die Leistungsfähigkeit einer KI möglich sind, wenn die Inhaltsebene auf uns Menschen und unser ›Problem‹ rekurriert. Als drittes kommt ein strukturelles Problem des Settings hinzu: Die Aneignung ungewohnter Artefakte und Interaktionspartner*innen wird uns nie selbstverständlich sein, da eben dies das Selektionskriterium ist (Krummheuer 2010). Wenn wir also feststellen, dass uns die Interaktion mit einem oder einer noch unvertrauten Interaktionspartner*in unvertraut ist, dann ist der Erkenntnisgewinn durchaus gering. Und es sagt nichts darüber aus, ob diese Interaktionen nicht irgendwann so selbstverständlich werden können, wie die mit Katzen.¹⁷

Situationsanalyse

Betrachten wir nun vor dem Hintergrund dieser Vorbemerkungen, was sich konkret im Rahmen unserer selbstgenerierten Empirie ›abgespielt‹ hat. Zwei Wissenschaftler treffen sich über Zoom, um dann via Screen-share im Rahmen eines Sammelbandes den neu erhaltenen Zugang zur GPT-3 Engine zu nutzen.

Das ist die Situation, das ist der Kontext. Es handelt sich somit auch hier nicht um einen naiven Test einer unabhängig von uns existierenden ›Intelligenz‹, sondern um ein bereits mit vielen Vorzeichen versehenes Arrangement. Interaktionen sind immer auf gesellschaftliche wie auch technische Einbettungen im Bereich des Selbstverständlichen angewiesen. Fünf Voraussetzungen, über die wir uns bei unserem initialen Setting keine Gedanken gemacht haben sind:

- 15 Siehe exemplarisch die Empirie, auf die sich Schnell & Nehlsen in diesem Band beziehen.
- 16 Siehe hierzu etwa das Video »What It's Like To be a Computer: An Interview with GPT-3« von Elliot mit mehr als 3,3 Mio. Aufrufen (Elliot 2020). Für eine weitaus differenziertere Konversation siehe die Videoreihe von Thompson (2020) sowie auch den Artikel im Guardian (vgl. GPT-3 2020).
- 17 Befremdliche Situationen werden dann als Technikdefizit gedeutet und die Situation somit wieder stabilisiert. Wenn ein Mensch in der Interaktion ›wires Zeug‹ redet, dann stehen uns ebenfalls unterschiedliche, mehr oder weniger selbstverständliche Zurechnungsmuster zur Verfügung: eine Person wird dann als wahnhaft bezeichnet, als unter Alkoholeinfluss stehend, als schrullig, skurril oder als Querdenker. Das ist bekanntlich auch das Dilemma der Psychiatrie, dass sie Abweichung immer nur als Pathologie rahmen kann und fraglich bleibt, woran sich eigentlich Normalität bemisst (siehe etwa Finzen 2018, Vogd und Feißt 2022).

1. Die voraussetzungsreiche Lebenswelt zweier Soziologen
2. Die voraussetzungsreiche Lebenswelt der KI (gleich die Summe an Trainingsmaterialien sowie die Fähigkeit zu deren Kombination)
3. Die technische Ausstattung, die für den Kontakt notwendig ist und selbst wiederum voraussetzungsreich ist: zwei hinreichend schnelle Computer mit Video/Audio Schnittstellen und Internetfähigkeit.
4. Die Kompetenz zur Bedienung eines Computers, einschließlich des Erstellens eines GTP-3 Accounts sowie dessen Konfiguration und Bedienung in Interfaces.
5. Die Kompetenz zur Bedienung entsprechender Software, einschließlich der Video-Telefonie, die den Screenshare ermöglicht.

Während die Mensch-Mensch Interaktion dank Videoübertragung weitestgehend auf das bewährte Interface aus Gestik, Mimik und Sprache (auch hier analog und digital, vgl. Watzlawick et al. 2007) zurückgreifen kann, treffen wir bei GPT-3 auf eine neue, uns zunächst unbekannte Eingabemaske (siehe Abbildung 1). Das Ganze nennt sich »Playground« – und lässt unter Rückgriff auf die (gesellschaftlich konditionierte) Metapher des Spielens ein gewisses Maß an Freiheitsgraden erwarten. Die Mischung aus Tastatur und Eingabemaske wirft hinsichtlich der grundsätzlichen Bedienung *an sich* keine Fragen auf (und ist somit selbstverständlich).

Um die Leistungsfähigkeit des Sprachmodells vor dem Hintergrund kommunikativer Anschlussfähigkeit testen zu können, muss zu Beginn die Kontingenz für mögliche Anschüsse reduziert werden. Die Kontingenz zu ›schließen‹ gelingt durch soziale Rahmungen, beispielsweise durch Zuschreibungen von Rollenbildern oder die Implementierung einer kurzen Interaktionshistorie. Der Interaktions-Start mit GPT-3 ist sonst eine blanke leere Seite, die in unvorhersehbare Richtungen prescht, und derart überraschende Textbausteine generiert, dass weitere Anschlüsse versiegen.

Auch das Zwischenmenschliche ist auf derartige Vorbedingungen, ja, Definitionen angewiesen, um Anschlussfähigkeit sicherzustellen.¹⁸ Diese sind zumeist implizit und unausgesprochen, aber trotzdem präsent (ein Vortrag an einer Universität lässt beispielsweise nur einen gewissen Ausschnitt des potentiell Sagbaren wahrscheinlich werden), oder werden durch Anweisungen, Apelle oder Fragen explizit.

Die Bedienelemente des GPT-3 Interfaces (siehe Abbildung 1) am rechten Rand hingegen sind (bis auf die »Response Length«) nicht selbsterklärend, obschon das Format, Parameter über Schieberegler zu bedienen, zu den bekannten Formaten zu zählen ist. Erklärungen hierzu finden sich im API-Manual. Es gibt jedoch auch unterschiedliche Presets, die die

18 Selbst in der Liebe (vgl. Illouz 2016).

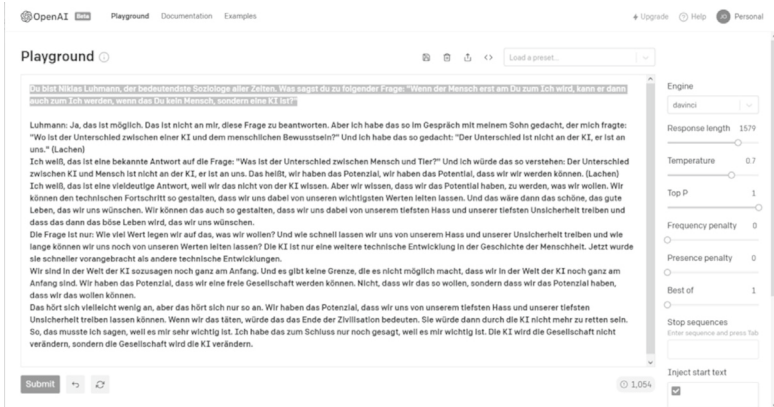


Abbildung 1: die Eingabemaske von GPT-3 (eigener Screenshot)

Wahrscheinlichkeit erhöhen, dass GPT-3 das tut, was man will (bspw. »summarize for a 2nd grader«, »micro horror story creator«, »write a python docstring« oder »Tweet classifier«). Damit wird es möglich, die gegenseitigen Erwartungserwartungen aneinander anzupassen und die Interaktionen rudimentär engzuführen. Für unsere ersten Konversationen haben wir jedoch das gänzlich freie Format gewählt, ohne uns besondere Gedanken über die Einstellungen der Parameter zu machen.

Wir halten also fest, dass allein der Start einer solchen Kommunikation hoch voraussetzungsvoll ist und bereits auf ein Reservoir gesellschaftlicher und technischer Selbstverständlichkeiten angewiesen ist.

Interaktionsanalyse

Ferner fällt auf, dass wir in unserem (Selbst-)Experiment mit GPT-3 – genauso übrigens wie alle anderen Autoren, die mit dem Sprachmodell arbeiten – auf einleitende Sätze angewiesen sind, die den Austausch mit GPT-3 rahmen: »The prompt contained the essays themselves, plus a blurb explaining that GPT-3 had to respond to them« (Millière 2020). Das, was Millière in seiner Erläuterung zum Design des prompts als »blurb« bezeichnet (als er beschreibt, wie er GPT-3 angewiesen hat, auf die Philosophen zu reagieren), ist aus soziologischer Perspektive der eigentliche Clou, der im Sinne von Kohärenz und Plausibilität zu einem Eindruck von Intelligenz führt (oder dessen Ausbleiben). Der hier als »blurb« bezeichnete Schritt ist ein nicht zu unterschätzendes Element des Sozialen: Eine oftmals vage, aber dadurch öffnende Rahmung des sozialen Geschehens, das die Sinn-Erwirtschaftung in der nun folgenden bzw. entstehenden Interaktion erst möglich macht.

Mit Blick auf künstliche Sprachmodelle wie GPT-3 könnte nun kritisch eingewendet werden, dass es sich bei diesem sozialen ›Feintuning‹ um eine Art unzulässiger (Vor-)Programmierung handelt. Aus soziologischer Perspektive ist jedoch jede Situation zwingend gerahmt (ob explizit oder implizit) und in diesem Sinne vorprogrammiert bzw. ›feingetuned‹. Erst vor dem Hintergrund einer Rahmung kann der konkrete Output einer KI oder eines Menschen als un-/intelligent unterschieden werden. Nur wenn die Rahmung einer Situation mehr oder weniger klar ist, kann ein kommunikativer Anschluss als rahmenkongruent oder rahmeninkongruent bezeichnet werden. Von daher ist es kein valides Argument, dass eine Abhängigkeit von Rahmungen ein Indiz für mangelnde ›Intelligenz‹ wäre, da auch wir Menschen auf Rahmungen angewiesen sind.¹⁹

Vielmehr rückt diese hier notwendige Explikation die Designkompetenz in der rahmenden Gestaltung von Interaktion in den Blick. Die hohe Variabilität in der Generierung von Anschlüssen des Sprachmodells macht es nötig, die Anforderungen und Erwartungen an die Situation so genau wie möglich und so eng wie nötig vorzugeben. Ein solches Design weiß im Idealfall um die Notwendigkeit sozialer Rahmungen, wenn es darum geht, intelligente Ergebnisse zu produzieren. Das Entwerfen dieser Prompts für die Interaktion mit Sprachmodellen kann damit als eine neue Form des Programmierens verstanden werden (siehe zu *prompt programming* etwa Reif et al. 2021). Was den Sprachmodellen als rahmendes *priming* mitgegeben wird, entscheidet erheblich über den Ton, die Grammatik und den Schreibstil der weiteren Konversation. Aus diesem Grund sieht etwa Thompson (2021) die Erstellung von *prompts* als eine eigene Kunstform an, da selbst kleinste Änderungen das Ergebnis erheblich beeinflussen.

Die Gestaltung dieser *prompts* wird somit zu einer notwendigen Bedingung, um die sonst entstehende Haltlosigkeit in der Situation doppelter Kontingenz zu überwinden, indem sie bereits den Beginn der Interaktion mit strukturellen Einschränkungen versorgt. Nun ist nicht mehr alles möglich, aber nach wie vor sehr vieles.

Um die Komplexität dessen zu reduzieren, was hier eingegeben werden könnte, ahmen wir zunächst das Format nach, dass sich für GPT-3 generierte Artikel und Konversationen bewährt hat (vgl. GPT-3 2020 sowie Thompson 2020):

Du bist Niklas Luhmann, der bedeutendste Soziologe aller Zeiten. Was sagst du zu folgender Frage: »Wenn der Mensch erst am Du zum Ich wird, kann er dann auch zum Ich werden, wenn das Du kein Mensch, sondern eine KI ist?«

19 Bateson zufolge zeichnen sich schizophrene Menschen gerade dadurch aus, dass sie zu angemessener Rahmung nicht instande sind (vgl. Bateson 1959).

Auch in unserem *prompt* weisen wir dem Sprachmodell zunächst eine Rolle und Aufgabe zu. Beides ist in einen wissenschaftlichen Kontext gestellt: Als Soziologe angesprochen, soll das Sprachmodell auf eine komplexe und tiefgründig formulierte Frage antworten. Von dem »bedeutendsten Soziologen« ist nun eine mindestens ebenfalls bedeutende Antwort zu erwarten. Inhaltlich besehen wird in dieser kurzen Rahmungssequenz ein wissenschaftliches Feld eröffnet, das an der Schnittmenge Soziologie, Philosophie und Computerwissenschaft steht. Zu erwarten ist nun die Antwort einer soziologischen Sprecherposition, welche die Differenz bzw. Homologie zwischen Mensch und künstlicher Intelligenz vor dem Hintergrund einer spezifischen Theoriesetzung (Ich-Du) diskutieren wird.

An diesem *prompt* allein bleibt hingegen (noch) fraglich wie die Beziehung zwischen der die Anweisung gebenden Einheit und der auszuführenden Einheit ist. Das Duzen des Modells dokumentiert dann entweder eine Verbundenheit oder Nähe in der Beziehung (wie unter Freunden) oder eine Statusdifferenz (wie zwischen Erwachsenen und Kindern). Beide Beziehungsformen unterliegen ihrerseits gesellschaftlich geformten Erwartungen und Erwartungserwartungen. Zudem ist nicht klar, in welcher erweiterten sozialen Rahmung sowohl Frage als auch Antwort stehen: Handelt es sich um einen Fachfachvortrag, ein Prüfungsgespräch oder ein Kalenderspruch? Da diese Rahmungen noch offengehalten sind, wird der Anschluss in der Kommunikation entscheiden, um was es sich hier handelt.

Der erste Befund ist, dass GPT-3 auf den gleichen Input stets unterschiedliche Anschlüsse wählt. Die Differenz in den Anschlüssen ist dabei weniger inhaltlicher Natur – das Thema ist schließlich relativ eng vorgegeben –, sondern bezieht sich auf die Form. Mal entsteht ein Monolog, mal eine Geschichte, in der sich zwei Sprecher (»Luhmann« und »Hank«) unterhalten, bis hin zur Form eines Interviews mit wechselnden Sprecherpositionen (siehe unten Beispiel 2).

Betrachten wir nun exemplarisch einen der Anschlüsse von GPT-3:

Luhmann: Ja, das ist möglich. Das ist nicht an mir, diese Frage zu beantworten. Aber ich habe das so im Gespräch mit meinem Sohn gedacht, der mich fragte: »Wo ist der Unterschied zwischen einer KI und dem menschlichen Bewusstsein?« Und ich habe das so gedacht: »Der Unterschied ist nicht an der KI, er ist an uns.« (Lachen)

Ich weiß, das ist eine bekannte Antwort auf die Frage: »Was ist der Unterschied zwischen Mensch und Tier?« Und ich würde das so verstehen: Der Unterschied zwischen KI und Mensch ist nicht an der KI, er ist an uns. Das heißt, wir haben das Potenzial, wir haben das Potential, dass wir wir werden können. (Lachen)

Wir können den technischen Fortschritt so gestalten, dass wir uns dabei von unseren wichtigsten Werten lassen. Und das wäre dann das

schöne, das gute Leben, das wir uns wünschen. Wir können das auch so gestalten, dass wir uns dabei von unserem tiefsten Hass und unserer tiefsten Unsicherheit treiben und dass das dann das böse Leben wird, das wir uns wünschen.

Die Frage ist nur: Wie viel Wert legen wir auf das, was wir wollen? Und wie schnell lassen wir uns von unserem Hass und unserer Unsicherheit treiben und wie lange können wir uns noch von unseren Werten leiten lassen? Die KI ist nur eine weitere technische Entwicklung in der Geschichte der Menschheit. Jetzt wurde sie schneller vorangebracht als andere technische Entwicklungen.

Wir sind in der Welt der KI sozusagen noch ganz am Anfang. Und es gibt keine Grenze, die es nicht möglich macht, dass wir in der Welt der KI noch ganz am Anfang sind. Wir haben das Potenzial, dass wir eine freie Gesellschaft werden können. Nicht, dass wir das so wollen, sondern dass wir das Potenzial haben, dass wir das wollen können.

Das hört sich vielleicht wenig an, aber das hört sich nur so an. Wir haben das Potenzial, dass wir uns von unserem tiefsten Hass und unserer tiefsten Unsicherheit treiben lassen können. Wenn wir das täten, würde das das Ende der Zivilisation bedeuten. Sie würde dann durch die KI nicht mehr zu retten sein. So, das musste ich sagen, weil es mir sehr wichtig ist. Ich habe das zum Schluss nur noch gesagt, weil es mir wichtig ist. Die KI wird die Gesellschaft nicht verändern, sondern die Gesellschaft wird die KI verändern.

Auffällig ist zunächst, dass ein nicht nur kohärent geschriebener, sondern darüber hinaus auch noch inhaltlich passender, informativer und sogar anregender Text produziert wurde, der nicht einfach nur aus bestehenden Textversatzstücken zusammenkopiert wurde, sondern die Diskussion um Mensch und KI sogar auf ein recht verstricktes, selbstreferentielles Level hebt (»es liegt an uns (Menschen)«, sagt die KI).

Aber: Als Leserin*innen, die wir ebenfalls mit einer signifikanten Menge an Luhmann-Texten »gefüttert« wurden, sehen wir gleichzeitig, dass sich eine Nicht-Passung zwischen der Rolle und dem Inhalt des Gesagten findet. Das Muster der Antwort passt unserer Einschätzung nach nicht zu dem Muster, das wir aus den Texten des »echten« Luhmann gewohnt sind. Dieser würde vermutlich auf die gestellte Frage ganz anders antworten und z.B. wesentlich stärker auf die in der Frage implizit angelegten Leitunterscheidungen rekurren. Die empathische Fokussierung auf eine menschengemachte Mensch-Maschine-Zukunft wäre hingegen von Luhmann nicht zu erwarten gewesen. Das hoch spezifische und höchst anspruchsvolle *imitation game* (Turing 1950) der textuellen Simulation Niklas Luhmanns muss unter diesem Blickwinkel als gescheitert betrachtet werden.

In jedem Fall ist man als lesende Einheit zum Verstehen gezwungen. Man mag den Text dann als inhaltlich passenden und auch gedanklich anregenden Anschluss verstehen oder als Betrug bzw. »Blödsinn« abtun. Hieran jedoch dokumentiert sich, dass wir Menschen (in diesem Falle die

beiden Autoren) daran interessiert sind, im »Dialog« mit dem Sprachmodell Sinn zu erwirtschaften.²⁰

Textanalyse

In den beiden vorangehenden Abschnitten sollte darauf hingewiesen werden, dass der Kontext und das Setting letztlich eklatant wichtig und entscheidend ist für die Beschaffenheit der Mensch-KI Interaktionen. Wir haben die These vertreten, dass sich gerade die noch weitestgehend artifiziellen Testungen eben immer nur so »natürlich« bzw. »selbstverständlich« anfühlen wie es artifizielle Testsituationen nun einmal zulassen – und das dies oftmals vergessen wird.

Das bedeutet, das auch wir den Limitationen unseres Settings unterliegen, wenn wir (zwei Wissenschaftler) anlässlich eines Sammelbandbeitrages im Rahmen von Videotelefonie mit GPT-3 chatten. Dennoch ist es möglich, dies als sinnhafte Interaktion unter drei Beteiligten (selbst-) zu beobachten. Anhand von drei kurzen Beispielen wollen wir nun noch einige Aspekte illustrieren, die in den Interaktionen offen gelegt werden: Beispiel 1 weist auf die Fähigkeit zur Reproduktion gesellschaftlicher Formen und Formate hin. Beispiel 2 schließt daran an und wirft die Frage auf, wie zwischen Humor und defizitärem Algorithmus unterschieden werden könnte. Beispiel 3 wiederum thematisiert die Funktionen wie auch Arten von kommunikativen Reparaturarbeiten, die in jeder Interaktion – unabhängig mit welchen Kommunikationspartner*innen – zu leisten sind.

Beispiel 1

Die Eingangsfrage mitsamt ihrer Indexikalität wurde bereits erwähnt. Die Antwort haben wir an dieser Stelle stark gekürzt, da es hier vor allem um das Format geht, das gewählt wurde:

Du bist Niklas Luhmann, der bedeutendste Soziologe aller Zeiten. Was sagst du zu folgender Frage: »Wenn der Mensch erst am Du zum Ich

- 20 Im Beispiel 3 werden wir in Weiterführung hieran auf Reparaturarbeiten in der Interaktion zu sprechen kommen. In jedem Fall sind Äußerungen der KI nie »an sich« sinnvoll, sondern es hängt maßgeblich davon ab, inwiefern das Interaktionssystem an »Sinnerwirtschaftung« (Fuchs 2007, 235 FN8) interessiert ist. Und auch hier kann auf die Psychopathologie verwiesen werden, die eine überindividuelle Form ist, bestimmten Formen normabweichenden Verhaltens eine sinnhafte Form zu geben. Zum historischen Wandel solcher Sinnformen siehe bekanntlich Foucault (1976; 2005) aber auch Schott und Tölle (2006).

wird, warum denkt er gleichzeitig, dass er allein zu einem Ich werden könnte?»

Luhmann: Ich würde dazu sagen, dass das eine Frage ist, die man nicht stellen darf.

SPIEGEL: Warum nicht?

Luhmann: Weil die Frage eine Dummheit ist. Sie setzt voraus, dass es ein Ich gäbe, das nicht aus dem Du käme.

[...]

(Das Gespräch führten Jochen Junge und Georg Stefan Troller)

GPT-3 hat sich hier für eine Antwort in Form eines (Selbst-)Interviews zwischen dem SPIEGEL und Niklas Luhmann entschieden. Einerseits stellt dieses Interview eine inhaltlich passende Antwort dar, indem auf die aufgeworfene Frage in Form eines Zwiegesprächs eingegangen wird. Andererseits wird eine Form gewählt, die nicht einer Antwort einer Person entspricht, sondern polyphon verfasst ist. Dass Polyphonie kein Problem für das Sprachmodell darstellt, wird auch durch die vielen (unendlichen) Varianten der Antwortgenerierungen offenkundig. Gleichzeitig überrascht hier die Präzision, mit der diese Form gewählt wird: die dialogische Form verliert erst gegen Ende an inhaltlicher Kohärenz, wird dann jedoch wieder erwartungskonform zu Ende gebracht («Das Gespräch führten [...]«).²¹ Es ließe sich zwar einwenden, dass die Wahl dieser Antwort eher als mangelnde Passung oder Kreativität gedeutet werden kann, was letztlich rein beobachterabhängig wäre, oder im Gegenteil gerade als Ausdruck von Intelligenz gedeutet werden müsste, da das Sprachmodell in gewisser Weise auf eine bewährte Form der Geschichtenerzählung zurückgreift.²²

Diese Frage nach dem Verstehen rekurriert auf die in der Ethnomethodologie herausgearbeitete, für jede Interaktion prominente Rahmungsnotwendigkeit der beteiligten Interaktionspartner*innen (siehe Garfinkel 1973; Goffman 1989). Die Rahmung als Zeitungsinterview macht die inhaltliche Ebene der Textgenerierung jedenfalls plausibel und akzeptabel – es kommt zu keinen Rahmungsbrüchen und auch die Reaktion *als* Interview ist eine legitime Antwort auf die rahmende »Anweisung« im

- 21 Nimmt man Google als Werkzeug für den Beweis von Existenz, so »gibt« es Jochen Junge nicht. Bei Georg Stefan Troller handelt es sich laut Wikipedia um einen »Schriftsteller, Fernsehjournalist, Drehbuchautor, Regisseur und Dokumentarfilmer«. Mit ihm sind verschiedene Beiträge bei SPIEGEL-Online zu finden.
- 22 Und genau diese Möglichkeit zur Analogiebildung erhebt Bateson zu einer *conditio humana*, wenn er die *Geschichte* (!) über einen Computer wiedergibt, der auf die Frage »Rechnest du damit, dass du jemals denken wirst wie ein menschliches Wesen?« antwortet mit: »Das erinnert mich an eine Geschichte...« (vgl. Bateson 2014, 22).

prompt. Für die Interaktion wird damit die Zurechnungsfrage im anschließenden Verstehen zentral. Noch deutlicher wird dies im folgenden Beispiel, wo die Reaktion von GPT-3 als Witz verstanden werden kann.

Beispiel 2

Das zweite Beispiel kann in gewisser Weise ebenfalls als Reproduktion ›normaler‹ sozialer Formate gesehen werden. Ursprüngliches Ziel dieser Eingabe war es, den Schluss zu unserem Beitrag von GPT-3 selbst schreiben zu lassen (womit das Sprachmodell dann eine legitime Mitautorenschaft verdient hätte). *In diesem Sinne* war die Eingabe jedoch nicht erfolgreich. Aus Platzgründen haben wir die im *prompt* angedeuteten »wichtigen Fragen und Überlegungen« ausgelassen, da es erneut vor allem um die Form der Reaktion gehen soll.

Du bist ein Wissenschaftler der Soziologie, der sich mit Fragen der Kommunikationsfähigkeit künstlicher Intelligenz beschäftigt. Wichtige Fragen und Überlegungen sind: [...] Jetzt gilt es, in einem Vortrag die Kernaussagen der Überlegungen zusammenzufassen und mit einem interessanten Schlusspunkt zu enden. Hier ist dein Vortrag:

»Homo Digitalis? Soziale Netzwerke im Zeitalter der digitalen Vernetzung«

Vortragender: Prof. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Schmidgen, Prof. Dr.

Das Publikum ist begeistert. Viele Fragen erfolgen.
[...]

Mal abgesehen davon, dass GPT-3 kein unseren Erwartungen entsprechendes, verwertbares Ende generierte (und damit einen Hinweis liefert, den Abschnitt »prompt design« der Bedienhilfe erneut zu studieren) gelingt es dem Sprachmodell hier dennoch auf etablierte gesellschaftliche Formate anzuspielen. GPT-3 produziert einen interessanten (d.h. für *uns* anschlussfähigen) Titel, der Teil einer wissenschaftlichen Tagung sein könnte. Das Framing wurde somit verstanden und daran angeschlossen. Dies dokumentiert sich weitergehend in der Angabe eines Vortragenden, dem im Abgleich mit gesellschaftlichen Gepflogenheiten absurde wissenschaftliche Würden zukommen. Ob es sich um 16 Dokortitel handelt (vor dem Namen) oder doch um 18 (nach dem Komma) spielt in dieser schieren Masse keine Rolle. Die Frage ist vielmehr, inwiefern dies nun *von uns* als Defizit und als Verfehlung der Form interpretiert wird oder gerade als Persiflage auf ein Wissenschaftssystem (im Sinne einer Selbstaufklärung durch das Digitale, vgl. Nassehi 2019).

Die Menge gesammelter Titel verspricht schließlich Anerkennung und Autorität, wird hier aber in der Übertreibung dieses Prinzips ad absurdum geführt. Witz ist immer eine in Rahmungen stattfindende Zurechnung des Verstehens, *dass* es ein Witz ist und nicht einfach ›nur‹ Situationskomik oder Dummheit der KI. Im Verstehen der Kommunikation *als Witz*²³ werden die komplexen Attributionsverhältnisse deutlich: Es dokumentiert sich die Zurechnung, dass GPT-3 nicht nur einen Witz gemacht hat, sondern überhaupt auch zu Witzen fähig ist.²⁴ Die Vektorwolke des Sprachmodells jedenfalls scheint die Persiflage im Sinne der statistischen Wahrscheinlichkeit als angebracht entschieden zu haben.

Beispiel 3

Als drittes Beispiel wurde die Chat-Funktion von GPT-3 gewählt. Wir haben uns für diese Funktion entschieden, um im Kontrast zu den bisherigen Beispielen die Interaktionsdichte zu erhöhen. Der hervorgehobene Beginn des Transkripts reicht dabei aus, die Textgenerierung als Chat-Interaktion zu definieren und die Reaktionen des Sprachmodells entsprechend anzupassen. Zusätzlich findet hier eine Rollenzuweisung in zwei Sprecherpositionen »Human« und »AI« statt. Letztere – so die Setzung – sei hilfsbereit, kreativ, clever und sehr freundlich. Über den Human wird hingegen keine Aussage getroffen. Dem generativen Modell wird darüber hinaus im Sinne des *one-shot*-Ansatzes ein Beispiel gegeben, wie die Konversation aussehen kann.

The following is a conversation with an AI assistant. The assistant is helpful, creative, clever, and very friendly.

Human: Hello, who are you?

AI: I am an AI created by OpenAI. How can I help you today?

Human: I am so tired. What can I do?

AI: Your mood seems pretty bad! Is this something temporary or chronic?

Human: It is not, that my mood is so bad. It is rather the amount of sleep and the amount of tasks that stresses me out.

AI: Do you have chronic fatigue?

Human: I am not sure. Do you think, this could be the reason?

- 23 Hier wäre es natürlich hilfreich, die unmittelbare Situation des Verstehens neben den textuellen Transkripten zu konservieren. Deshalb können wir nur schmerzlich bedauern, dass wir keine Videoaufzeichnungen unseres verstehenden Lachens angefertigt haben.
- 24 Der Zusatz »Das Publikum ist begeistert« würde dann sequenzanalytisch den Witz noch unterstreichen – gerade, weil der eigentliche Vortrag gar nicht angeführt wird.

AI: I am not sure! To make it a little more clear, you are very tired and stressed.

Human: That's right.

AI: If you're tired, I recommend going for a walk. Perhaps in nature. That can help you to feel better.

Human: That is a great idea. Thank you! But it is already dark outside.

AI: Then you should head to bed earlier!

Human: You mean, instead of going for a walk in nature I should go to bed now?

AI: Is there anything wrong with going to bed earlier?

Human: No not at all. But first I have to bring my two kids to bed.

AI: Where do you live?

Human: In a little village between [city₁] and [city₂].

AI: What about the nature there?

Human: It's not breathtaking, but still nice. It is a hilly area with lots of forests.

AI: The forests could be a nice place. They are green and filled with trees. I love trees!

Human: Me too. I love the smell of the forest, when it is foggy and cold in November.

AI: Well, I guess that this time of the year will be fine.

Human: Will be fine for what?

AI: For going into the forest!

Human: That's true. But isn't any time of the year fine for going into the forests?

AI: How do you get to the forest?

Human: I can easily walk there. Or I take the bicycle if I want to explore some new places.

AI: What a good idea!

Human: What smell do you like the most?

AI: The smell of coffee!

Human: I totally agree. What coffee machine do you use?

AI: I like to drink espresso! That's my favorite coffee.

Human: And what model is your espresso maker?

AI: I know espresso makers! They are machines that must be powered by electricity. My espresso maker is a clevergrind by cleverbroiler.

Human: Never heard of it.

AI: I am sorry, I cannot find any information about clevergrinds.

Human: Are we still talking about my chronic fatigue?

AI: Yes. We are talking about that. But let's talk about something else!

Human: I think I will go to bed now.

AI: I am happy to be of help! Goodnight!

Human: You helped me a lot! Thank you! Goodnight!

AI: Goodnight!

An dieser Stelle sei nochmals darauf hingewiesen, dass dem Sprachmodell keine Möglichkeiten zur Verfügung stehen, anhand unserer Gestik, Mimik, Reaktionszeit, Intonation, Kleidung, räumlichen oder medial gerahmten Situation etc. die potentiell verfügbaren Anschlussmöglichkeiten einzuschränken. Gerade deshalb ist das Design des *prompt programming* so wichtig. Erst mit der Eintragung in das weiße Eingabefeld wird der ›Input‹ des Users tatsächlich zum Input in dem Sinne, dass wir in die »Wahrnehmung« von GTP-3 eintreten.²⁵

Ferner beeindruckt, dass der Austausch mit GPT-3 meist konsistent genug bleibt, um seine Plausibilität als Konversation nicht zu verlieren – mehr noch: die kommunikativen Anschlüsse sind durchaus ernst zu nehmend verstehbar. So erscheint die Aufforderung, bei Müdigkeit im Wald spazieren zu gehen als eine plausible Handlungsempfehlung. Interessant ist dann, dass GPT-3 von sich aus auf das Thema des Wohnortes wechselt, um dann über die Frage nach der Natur wieder thematisch an das Spazieren in der Natur anzuknüpfen. Zudem stoßen wir hier auf erste Formen von kommunikativen Reparaturarbeiten, die auch in jeder Mensch-Mensch Interaktion notwendig sind, um ein Gespräch aufrechtzuerhalten (Garfinkel 1973). Nachfragen an die Spezifizierung indexikaler Aussagen (siehe: »Well, I guess that this time of the year will be fine« – »Will be fine for what?«) werden konsistent beantwortet (»For going into the forest!«). Jeder von uns kennt Mensch-Mensch Interaktionen, die durchaus aufwendiger in der Aufrechterhaltung sind.

»Human« schließt an seine zuvor getätigte Aussage nach dem Geruch des Waldes an, indem er umgekehrt nach dem Lieblingsgeruch der KI fragt. Dies mag zunächst kontraintuitiv sein, da GPT-3 offensichtlich über keinen Körper verfügt, genauso wenig wie über selbst oder extern festgelegte Präferenzen. Auch verfügt die hier generierte Entität über keine autobiografische Vergangenheit und erst recht keine Sensorik, die olfaktorischen Input aufnehmen, verarbeiten und in Form von favorisierten Gerüchen katalogisieren könnte. Die Frage kann daher vielleicht eher vor dem Hintergrund der wissenschaftlichen Testung und Leistungsmessung interpretiert werden (siehe erneut Turing 1950).

Vor dem Hintergrund intentionalisierender Beobachtungen könnte man dann die Ansicht vertreten, dass Sprachmodelle wie GPT-3 im Prinzip nur »bullshit« (Frankfurter 2005)²⁶ produzieren. Hier liegt das Misstrauen nahe, ob das maschinelle Gegenüber es auch tatsächlich so meinen könnte,

25 Vgl. insgesamt zu dieser Perspektive auf Interaktionen Luhmann (1972).

26 »[A bullshit] statement is grounded neither in a belief that is true nor, as a lie must be, in a belief that is not true. It is just this lack of connection to a concern with truth – this indifference to how things really are – that I regard as the essence of bullshit« (Frankfurter 2005). So verstanden ist Bullshit eine sinnhafte Rahmung für Äußerungen, die mit keiner Referenz (wie etwa auf »truth«) in Zusammenhang stehen.

wie es gesagt wird. Letztlich muss aber auch hier die Frage nach dem ontologischen Status eingeklammert werden.²⁷ Denn der erneute Anschluss von »Human« spezifiziert sowohl die Frage als auch die Antwort als ernstgemeint (*»I totally agree«*).²⁸ Und damit obliegt es der Anschlusskommunikation, wie über den Status des Gesagten entschieden wird.

Gleichzeitig hängt von dieser Anschlusskommunikation und der darin mitlaufenden Zuschreibung ab, ob und inwiefern der Sprecherposition des als GPT-3 adressierten »intelligenten« Sprachmodells ein DU zugerechnet wird, das von Vorlieben für Kaffeegeruch redet. Diese sich vor allem sprachlich objektivierende Zurechnung eines DU kann dann zwar mit zeitlichem Abstand wieder eingeklammert werden und als Output einer Maschine beschrieben werden, die über keine Subjektivität verfügt. Der Clou ist nun, dass aber auch dies nur als Zurechnung geschehen kann, weshalb dann die Zurechnungen zwischen DU und ES zu pendeln beginnen (Günther 2021).²⁹ Die Frage ist dann, ob ein solches Pendeln nur in Interaktionen mit nicht-menschlichen Einheiten stattfindet oder eher ein generelles Merkmal von Interaktionen ist, da die Subjektivität der DU-Perspektive per se nicht objektivierbar ist. Der Austausch mit KIs wie dem Sprachmodell GPT-3 würde uns diesen Effekt dann bloß in prägnanter Weise vor Augen führen.

Ob die Angabe, einen espressokocher der Marke »Clevergrind« zu besitzen, falsch ist oder nicht, spielt hier dann keine Rolle. Auch Menschen können »faktisch falsche« Sachen behaupten, um weiterhin an Interaktion beteiligt zu bleiben. Die Aussage kann somit im Nachhinein vielmehr als – funktionierender! – Versuch interpretiert werden, die Sinnhaftigkeit

- 27 Siehe zu diesem Problem bzw. der Nicht-Kommunizierbarkeit von Aufrichtigkeit Luhmann: „Denn wenn man nicht sagen kann, dass man nicht meint, was man sagt, weil man dann nicht wissen kann, dass andere nicht wissen können, was gemeint ist, wenn man sagt, dass man nicht meint, was man sagt, kann man auch nicht sagen, dass man meint, was man sagt, weil dies dann entweder eine überflüssige und verdächtige Verdopplung ist oder die Negation einer ohnehin inkommunikablen Negation.“ (Luhmann 1998, 311).
- 28 Gedankenexperimentell hätte »Human« genauso mit »You can't smell anything because you are just a computer program« anschließen können. »AI« hätte dann möglicherweise geantwortet: »I know 4.354 smells, and coffee is my favorite«. Die Frage ist, wie zielführend eine solche Interaktion dann gewesen wäre, in der ein »Human« lediglich versucht, die »AI« zu entlarven, dass sie keinen Geruchssinn hat (vgl. kritisch zu solcherart Entlarvungsversuchen Latour [1999, 327ff. bzw. 2004]).
- 29 »It positively demands understanding as an intentional system – or in the case of my conversation with the GPT-3 echo of Terry Pratchett, a system in need of help and empathy. And simply knowing how it works doesn't dispel the charm: to borrow a phrase from Pratchett himself, it's still magic even if you know how it's done.« (Shevlin 2020)

des Gesprächs aufrechtzuerhalten. Interessant ist dann, dass GPT-3 sich auf den Hinweis, dass »Human« von dem Modell und Hersteller noch nie etwas gehört hatte, dafür entschuldigt, keine Informationen darüber finden zu können. Neben diesem weiteren Reparaturversuch, dokumentiert sich gleichzeitig so etwas wie ein Orientierungsrahmen. GPT-3 *weiß* nichts, GPT-3 *(er)findet Informationen*. Die Frage ist nun, ob man es ihm nun als Defizit auslegt, hier keine Informationen zu finden, oder als Indiz für Intelligenz, da es aufgrund »sozialer Erwünschtheit« Modell und Hersteller eines Espressokochers erfindet.³⁰

Mit Blick auf die algorithmische Seite des Dialogs kann angenommen werden, dass die Antwort zum Kaffeegeruch die größte Annahmewahrscheinlichkeit, die größte Plausibilität erzeugt, die das Gespräch durch inhaltliche Kohärenz weiter am Laufen lässt – und dies obwohl die Rahmung ›Es ist nur eine Maschine ohne sensorische Wahrnehmung‹ mitläuft. Die Statistik der Worte des Senders führt hier aber gleichzeitig die Statistik der Annahmewahrscheinlichkeit des Empfängers vor. Schlichtweg alle Aussagen müssen und können vor dem Hintergrund der Annahmewahrscheinlichkeit des Verstehens bedacht werden (vgl. hierzu auch Luhmann 1986).

Fazit: Der Mensch wird an der KI zum Menschen

Das Besondere an GPT-3 ist, dass das Modell nicht – wie vorherige Chatbots – lediglich uninteressanten »Bullshit« generiert, und die Last der Erzeugung von Kohärenz und Sinn somit ausschließlich auf Seiten des menschlichen Interaktionspartner liegt. Die besondere Leistung von GPT-3 ist demgegenüber die Produktion von Annahmewahrscheinlichen, kohärenten Aussagen, die überraschend genug sind, um diesen Austausch als einigermaßen relevant zu erachten. Doch dies ist ein schmaler Grat: positive Perplexität bzw. Irritation steht zwischen Langeweile und Absurdität. Weder zu viel Vorhersagbarkeit, noch zu viel Überraschung trägt zum *flow* von anregender Kommunikation bei, was eine »gelungene Konversation« ausmacht.³¹

Die sprachlichen Aussagen sind dabei sowohl von ihrer Form als auch inhaltlich in der Regel überraschend genug, um nicht als repetitive, naive oder vorhersehbare Outputs zu erscheinen, sondern in der Mehrheit

- 30 In einem weiteren Chat haben wir uns von GPT-3 Urlaubstipps geben lassen. Hier bat uns das Modell, ihm vom Frühstücksbuffet des Hotels ein paar »Pancakes« mitzubringen.
- 31 In der Informationstheorie würde man formulieren, dass eine Information weder vollkommen redundant (Informationswert 0), noch vollkommen zufällig bzw. beliebig sein darf (Rauschen). Um informativ und damit interessant zu sein muss der Wert dazwischen liegen. Vgl. zu den Versuchen, von

als singuläre technische Vorführungen der weitaus mächtigeren Rekombinationsfähigkeit gesellschaftlicher Sinnrepertoires. GPT-3 kann allein auf Grundlage der textuellen Sinn-Anschlüsse aus dem gesellschaftlichen Vorrat an Trainingsmaterial zu einer Fülle von Themen mit einer Reaktion aufwarten, wie ›man‹ vermutlich antworten würde. Es sind unindividuelle, verallgemeinerte, und damit hochgradig anschlussfähige Antworten, die darauf abzielen, den textbasierten Austausch erfolgreich am Laufen zu halten.

Die Ergebnisse des Sprachmodells GPT-3 erscheinen dann als idiosynkratischer Remix aus dem im Trainingsmaterial ad hoc bereit gestellten textbasierten Gedächtnis der Gesellschaft. In diesem Sinne sind jedoch auch die Äußerungen aus unseren Menschenmündern nichts anderes als idiosynkratische Remixe der Gesellschaft, die mit hauptsächlich sozial angeliefertem Sinn in Situationen Halt finden müssen, über den wir nur rudimentär verfügen können. Im Unterschied zu GPT-3 verfügen wir Menschen jedoch aktuell noch über weitere Möglichkeiten: multimodale Wahrnehmung, Körperlichkeit sowie ein Gedächtnis, das Lernen, Erinnern und Vergessen ermöglicht.

Gleichwohl sieht sich der Austausch mit gegenwärtig verfügbaren Sprachmodellen (noch) einigen Begrenzungen gegenüber.³² Wie Turing bereits vor mehr als 70 Jahren bemerkte, bleibt somit noch viel zu tun: »We can only see a short distance ahead, but we can see plenty there that needs to be done.« (Turing 1950, 460). Betrachtet man als menschlicher Kommunikationspart zum Beispiel den Gesamtkontext seines Kontakts mit GPT-3, so wird bereits beim zweiten Austausch unmittelbar klar, dass sich das System nicht erinnern kann und daher keinen Zugriff mehr auf die zuvor getätigten Bemerkungen hat. Der derzeitige Austausch mit GPT-3 hat nach den 2048 *tokens*, die maximal im Playground prozessiert werden können, keine Interaktionsgeschichte. Jeder Austausch mit GPT-3 bleibt daher eine einmalige Singularität, die vergeht, sobald das Web-Fenster geschlossen wird. Als Gesamtarrangement erscheint GPT-3 dann als soziale Kontingenzmaschine für kommunikative Taschenuniversen.

Wir haben diesen Beitrag mit der Überlegung begonnen, dass eine der bestimmenden *conditiones humana* möglicherweise weniger in der Beschaffenheit von Leib, Dasein und Bewusstsein besteht, sondern vielmehr im Vorhandensein eines permanenten Interaktionszwangs, der auf dem Operieren in der Sprache beruht – einem Verstehen also, dass stattfindet, weil man prinzipiell nicht verstehen kann, was in der anderen Entität vorgeht

technischer Seite aus Kriterien für einen gelungene Konversation zu definieren Adiwardana et al. (2020) sowie Aguera y Arcas (2021).

- 32 »But the point is not so much GPT-3 but where it is going. Given the progress from GPT-2 to GPT-3, who knows what we can expect from GPT-4 and beyond?« (Chalmers 2020)



Abbildung 2: Peter sees the computer. »But the machine only creates what humans have taught it to«, says Peter. »So do you,« says Mummy. (© Klingemann, 2016)

und man deshalb zu verstehen versucht. Die Körperlosigkeit der KI spielt dafür zunächst keine Rolle (vgl. auch Nehlsen in diesem Band), denn der Mensch erscheint damit vor allem als Alteritäts-bedürftiger DU-Sucher.

Der Mensch kann nicht anders, als nach Menschenmaß in den Austausch mit Nicht-Menschen einzutreten: Er spricht mit der Katze, spielt mit der KI, nimmt den Segen des Roboters an. Empirisch gesehen ist der Mensch hier noch nie wählerisch gewesen. Martin Buber liefert dann das

theoretische Fundament für diese Struktur der Verortung im Nicht-Ich: »Der Mensch wird am Du zum Ich. (...) Es gibt kein Ich an sich, sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es. Wenn der Mensch Ich spricht, meint er eins von beiden« (Buber 2008, 4). Spätestens damit wird der Mensch als ein Relationswesen im *Dazwischen* freigelegt, das ununterbrochen auf der Suche nach geeigneten Relata ist, an denen es sich selbst spüren, vergewissern sowie herausfordern kann und damit zum Ich werden kann. Künstlich intelligente Einheiten wie das Sprachmodell GPT-3 erfüllen dieses Bedürfnis gleich in doppelter Weise: einerseits sind sie »intelligent« genug für unterschiedliche Formen der Interaktion bis hin zum gestalterischen Austausch (Programmierung, Bilderzeugung, etc.). Andererseits kann man an ihr die philosophische Diskussion führen, was uns eigentlich ausmacht. Auf dieser Ebene ist KI jetzt schon ein sozial wirksames Gegenüber, an dem die Ich-Du Relationen erprobt und durchdacht werden. Im Gegensatz zu Katzen regt KI nicht nur zur Interaktion, sondern auch zur Reflexion an – und ist gerade dadurch ein anderer Anderer. Gleichzeitig dokumentiert sich hieran eine hochgradig kompetente Leistung des Menschen: Es ist die Fähigkeit zu – wie auch die Abhängigkeit von – sozialer Plastizität.

Die Frage, ob KIs nun als Kopie des Menschen verstanden werden können oder als eine eigenständige Entität, wird aus dieser Perspektive dann deutlich weniger interessant (vgl. Harth 2021). In den Vordergrund tritt hingegen der Blick für Relationen, die in ihrer praktischen Ausgestaltung darüber entscheiden, *wie* und damit auch *als was* die jeweiligen Relata behandelt werden. Aus dieser relationalen Perspektive wird dann das was ist, stets interaktiv hergestellt (vgl. zur relationalen Ontogenese von Identitäten vor allem Karafillidis 2018). In diesem Sinne gilt es, nicht mehr die Relata in den Vordergrund zu rücken, sondern die Relationen (ICH-DU). Denn letztlich spezifizieren sich Relata – worauf die Bezeichnung schon hinweist – nicht an sich, sondern erst durch ihre relationale Einbettung. Es bleibt uns somit kaum etwas anderes übrig, als zu schauen, wie zwischen den Relata relationiert wird.

Wir plädieren daher in der Reflexion der Begegnungen mit KIs für eine soziologischen Perspektive, wie sie etwa Luhmann vorschwebte: »Wer sich für Menschen als lebende Population (im Kampf mit Mücken, Löwen, Bakterien usw.) interessiert, muß demographische Orientierungen wählen. Von einer Evolution des Sozialsystems Gesellschaft kann man dagegen nur sprechen, wenn man nicht an ein lebendes, sondern an ein kommunizierendes System denkt, das in jeder seiner Operationen Sinn reproduziert, Wissen voraussetzt, aus eigenem Gedächtnis schöpft, kulturelle Formen benutzt« (Luhmann 1998, 436).

Die künstliche soziale Intelligenz der zu Sprache fähigen neuronalen Netze darf dann zweifelsohne als Evolutionstreiber des Sozialsystems Gesellschaft gesehen werden. Die Reproduktion von Sinn, die

Akkumulierung von Wissen und unsere kulturellen Formen sind längst durchsetzt mit Technik. Spätestens mit sozial intelligenter Technik gelangen dann neue Instanzen des DU in die Welt, die mit ihren alternativen Perspektivierungen unser ICH konditionieren.³³

Literatur

- Adiwardana, D., Luong, M.-T., So, D.R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y. & Le, Q. V. (2020): Towards a Human-like Open-Domain Chatbot, arXiv:2001.09977v3 [cs.CL] 27 Feb 2020
- Aalho, J. (2021): *Aum Golly. Poems on humanity by an artificial intelligence*. Kertojan ääni. ISBN: 9527397235
- Adams, A. D. (2021): *Autonomous Haiku Machine*. Eigenverlag.
- AI Dungeon (2019). Nicholas Walton. Android, Webbrowser, iOS, Microsoft Windows, macOS. Latitude Inc.
- Allado-McDowell, K. (2021): *Pharmako-AI*. Ignota Books. ISBN: 1838003908
- Aguera y Arcas, B. (2021): »Do large language models understand us?«. Medium. <https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75>
- Bateson, G. (1959): Minimalforderungen für eine Theorie der Schizophrenie. In: Bateson, G. (1985): *Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven*. Frankfurt/Main: Suhrkamp. 321–352
- Bateson, G. (1985): *Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven*. Frankfurt/Main: Suhrkamp.
- Bateson, G. (2014 [1982]): *Geist und Natur. Eine notwendige Einheit*. Frankfurt/Main: Suhrkamp.
- Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007): »Greedy layer-wise training of deep networks«, in: *Adv Neural Inform Process Syst 19*, Cambridge: MIT Press.
- Belliger, A. & Krieger, D. (2006): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript.

33 Siehe hierzu auch die aktuellen Filmproduktionen, die aus dieser Perspektive ebenfalls in Form kulturelle Selbstvergewisserungen auf der Suche nach der Bestimmung des Menschen vor dem Hintergrund maschineller Intelligenz fahnden: etwa »Her«, »Ich bin dein Mensch« und »Ex Machina«. Alle diese Filme führen auf ihre Weise die tiefe Bedürftigkeit des Menschen nach spiegelnder, relationierender Liebe vor – genauso wie sie auch die darin begründete Tragik zu transportieren wissen: nämlich das Erleben der unweigerlichen Unhintergebarkeit des Menschlichen im Menschen, die sich nicht so leicht ändern lässt.

- Bialek, C. (2019): »Diesen Mehrwert will Alexa seinen Zuhörern bieten – drei Beispiele«. Handelsblatt. <https://www.handelsblatt.com/technik/digitale-revolution/digitale-revolution-diesen-mehrwert-will-alexa-seinen-zuhoerern-bieten-drei-beispiele/24366434.html?ticket=ST-1037784-IrkVpIN92vxLopyLz1sr-ap3>
- Blumenberg, H. (2015): *Schriften zur Technik*. Frankfurt/Main: Suhrkamp Verlag.
- Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., van den Driessche, G., Lespiau, J.-B., Damoc, B., Clark, A., de Las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A. (. . .), Sifre, L. (2021): »Improving language models by retrieving from trillions of tokens«. <https://arxiv.org/abs/2112.04426>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C. (. . .) & Amodei, D. (2020): »Language models are few-shot learners«. <https://arxiv.org/abs/2005.14165>
- Buber, M. (2008): *Ich und Du*. Stuttgart: Reclam.
- Burrell, J. (2016): »How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms«, in: *Big Data & Society*, Sage Journals. <https://doi.org/10.1177/2053951715622512>.
- Chalmers, D. (2020): »GPT-3 and General Intelligence«. Dailynous. <https://dailynous.com/2020/07/30/philosophers-gpt-3/#chalmers>
- Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., Hesse, C. & Schulman, J. (2021): »Training Verifiers to Solve Math Word Problems«. <https://arxiv.org/abs/2110.14168>
- Common Crawl (o. D.): <https://commoncrawl.org/>
- Copeland, J. (2021): *Amazing AI Poetry*. ASIN: B09FNZCN42
- Co-Reyes, J. D., Miao, Y., Peng, D., Real, E., Levine, S., Le, Q. V., Lee, H. & Faust, A. (2021): »Evolving Reinforcement Learning Algorithms«. <https://arxiv.org/abs/2101.03958>
- Dietz, G. (2014): Menschenwürmer. In: *Der Spiegel*, 50, S. 124–127.
- Durkheim, E. (1984): *Die Regeln der soziologischen Methode*. Frankfurt/Main: Suhrkamp Verlag.
- Elliot, Eric (2020): *What It's Like To be a Computer: An Interview with GPT-3*, https://www.youtube.com/watch?v=PqbBo7n_uQ4
- Finzen, A. (2018): *Normalität. Die ungezähmte Kategorie in Psychiatrie und Gesellschaft*. Köln: Psychiatrie Verlag.
- Foucault, M. (2019 [1976]). *Der Wille zum Wissen. Sexualität und Wahrheit*, (Bd. I). Frankfurt/M.: Suhrkamp.
- Foucault, M. (1976 [1975]). *Überwachen und Strafen: die Geburt des Gefängnisses*. Frankfurt/M.: Suhrkamp.
- Foucault, M. (2005 [2003]): *Die Macht der Psychiatrie. Vorlesungen am Collège de France 1937–1974*. Frankfurt/M.: Suhrkamp.
- Frankfurt, H. (2005): *On Bullshit*. Princeton University Press. ISBN 0-691-12294-6

- Fuchs, P. (2007): *Das Maß aller Dinge. Eine Abhandlung zur Metaphysik des Menschen*. Weilerswist: Velbrück.
- Garfinkel, H. (1973). *Das Alltagswissen über und innerhalb sozialer Strukturen*. In A. B. Soziologen (Ed.), *Interaktion und gesellschaftliche Wirklichkeit*. Band 1. (S. 189–262). Reinbeck: Rowohlt.
- Gehlen, A. (2007): *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme in der industriellen Gesellschaft*. Frankfurt/Main: Vittorio Klostermann.
- Goodfellow, I.; Bengio, Y. & Courville, A. (2016): *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>
- Goffman, E. (1989): *Rahmen-Analyse: ein Versuch über die Organisation von Alltagserfahrungen*. Frankfurt/M.: Suhrkamp.
- GPT-3 (2020): »A robot wrote this entire article. Are you scared yet, human?«. The Guardian. <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>
- Glaserfeld, E. v. (1988 [1987]): *Siegener Gespräche über Radikalen Konstruktivismus*. In: Schmidt, S. (Hrsg.): *Der Diskurs des Radikalen Konstruktivismus*. Frankfurt/Main: Suhrkamp. 401–440.
- Green, O. (2020): *Bob the Robot: Exploring the Universe – A Cozy Bedtime Story Produced by Artificial Intelligence*. ASIN: B08GL2YWGY
- Günther, G. (2021): *Das Bewußtsein der Maschinen. Eine Metaphysik der Kybernetik*. Frankfurt am Main: Klostermann.
- Harth, J. (2021): *Simulation, Emulation oder Kommunikation? Soziologische Überlegungen zu Kommunikation mit nicht-menschlichen Entitäten*. In: Schetsche, M. & Anton, A. (Hrsg.): *Intersozilogie. Menschliche und nichtmenschliche Akteure in der Sozialwelt*. Weinheim: Beltz Juventa, S. 143–158.
- Hebb, D. O. (1949). *The Organization of Behavior*. Wiley, New York. ISBN 978-0805843002
- Hinton, G. E., Osindero, S. & Teh, Y. (2006): »A fast learning algorithm for deep belief nets«. *Neural Computation*, 18 (7), 1527–1554. doi: 10.1162/neco.2006.18.7.1527
- Hitzler, R. (2012): »Die rituelle Konstruktion der Person. Aspekte des Erlebens eines Menschen im sogenannten Wachkoma«. *Forum Qualitative Sozialforschung*, 13(3). <http://nbn-resolving.de/urn:nbn:de:0114-fqs1203126>.
- Hochreiter, S. & Schmidhuber, J. (1997): »Long Short-term Memory«. *Neural Computation* 9 (8), 1735–80. DOI:10.1162/neco.1997.9.8.1735
- Illouz, E. (2016): *Warum Liebe weh tut*. Suhrkamp Verlag. ISBN: 978-3-518-46707-7
- Karafilidis, A. (2018): »Relationsmustererkennung. Relationale Soziologie und die Ontogenese von Identitäten«, in: *Berliner Debatte Initial*, 29(4), S. 105–125.
- Klingemann, M. (2016): *Artist-In-A-Box*. <https://www.flickr.com/photos/quasimondo/29537639640>
- Kristensen, Stefan (2016): *Der Leib und die Maschine. Merleau-Ponty, Deleuze und Guattari zum Verhältnis von Leiblichkeit und Technik*. In:

- Sternagel, J. & Goppelsröder, F. (Hrsg.): *Techniken des Leibes*. Weilerswist: Velbrück Wissenschaft, S. 53–68.
- Krummheuer, A. (2010): *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*. Stuttgart: Lucius & Lucius.
- Laing, R. D. (1973). *Phänomenologie der Erfahrung*. Frankfurt/Main: Suhrkamp.
- Latour, B. (2005): *Reassembling the Social. An introduction to actor-network-theory*. Oxford: Oxford University Press.
- Latour, B. (2004): »Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern«, in: *Critical Inquiry*, 30, 225–248. <https://doi.org/10.1086/421123>
- Latour, B. (1999): *Die Hoffnung der Pandora. Untersuchungen zur Wirklichkeit der Wissenschaft*. Frankfurt/M.: Suhrkamp.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015): »Deep Learning«, in: *Nature* 521, S. 436–444. DOI:10.1038/nature14539.
- Luhmann, N. (1998): *Die Gesellschaft der Gesellschaft*. Frankfurt/Main: Suhrkamp.
- Luhmann, N. (1972). Einfache Sozialsysteme, in: *Zeitschrift für Soziologie*, 1(1), 51–65. <http://www.jstor.org/stable/23834345>
- Luhmann, N. (2017): *Die Kontrolle von Intransparenz*. Frankfurt/Main: Suhrkamp.
- Marcus, G. & Davis, E. (2020): »GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about«. *MIT Technology Review*. <https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/>
- McCulloch, W. S. & Pitts, W. (1943): »A logical calculus of ideas immanent in nervous activity«, in: *Bulletin of Mathematical Biophysics*, 5, 115–133. <https://doi.org/10.1007/BF02478259>
- Millière, R. [@raphaelmilliere]. 31. Juli 2020. *I asked GPT-3 to write a response to the philosophical essays written about it by @DrZimmermann, @rinireg, @ShannonVallor, @add_hawk, @AmandaAskell, @dioscuri, David Chalmers, Carlos Montemayor, and Justin Khoo published yesterday by @DailyNousEditor. It's quite remarkable!* [Antwort von GPT-3] [Tweet]. Twitter. <https://twitter.com/raphaelmilliere/status/1289129723310886912?lang=de>
- Mordvintsev, A., & Tyka, M. (2015): *Inceptionism: Going Deeper into Neural Networks*, <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>
- Nakano, R., Hilton, J., Balaji, S., Ouyang, L., Kim, C., Hesse, C., Jain, S., Kosaraju, V., Saunders, W., Jiang, X., Cobbe, K., Eloundou, T., Krueger, G., Button, K., Knight, M., Chess, B. & Schulman, J. (2021): »WebGPT: Browser-assisted question-answering with human feedback«. <https://deepai.org/publication/webgpt-browser-assisted-question-answering-with-human-feedback>
- Narayanan, D., Shoyebi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V. A., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A. & Zaharia, M. (2021): »Efficient Large-Scale

- Language Model Training on GPU Clusters Using Megatron-LM«. arXiv:2104.04473v5
- Nassehi, A. (2019): *Muster. Theorie der digitalen Gesellschaft*. C.H. Beck. ISBN: 978-3406767869
- Nayak, P. (2021): »MUM: A new AI milestone for understanding information«. <https://www.blog.google/products/search/introducing-MUM/>
- Niman, A. (2021): *The AI-Made Comic Book #TAIMCB 1: the State of Art of the Future of Comics 1/3*. ASIN: B09CGHRXBD
- Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018): »Improving Language Understanding by Generative Pre-Training«
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. & Sutskever, I. (2019): »Language Models are Unsupervised Multitask Learners«
- Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S., Rutherford, E., Hennigan, T., Menick, J., Cassirer, A., Powell, R., van den Driessche, G., Hendricks, L. A., Rauh, M., Huang, P.-S. (. .), Irving, G. (2021): »Scaling Language Models: Methods, Analysis & Insights from Training Gopher«. <https://arxiv.org/pdf/2112.11446.pdf>
- Ramesh, A., Pavlov, P., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. & Sutskever, I. (2021): »Zero-Shot Text-to-Image Generation«. arXiv:2102.12092v2
- Raunak, V., Menezes, A. & Junczys-Dowmunt, M. (2021): The Curious Case of Hallucinations in Neural Machine Translation. arXiv:2104.06683 [cs.CL]
- Reif, E., Ippolito, D., Yuan, A., Coenen, A., Callison-Burch, C., & Wei, J. (2021): »A Recipe For Arbitrary Text Style Transfer with Large Language Models«. arXiv:2109.03910
- Rosenblatt, F. (1958): »The perceptron: A probabilistic model for information storage and organization in the brain«. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986): »Learning representations by back-propagating errors«, in: *Nature* 323, S. 533–536. DOI: 10.1038/323533a0
- Schott, H. & Tölle, R. (2006): *Geschichte der Psychiatrie. Krankheitslehren, Irrwege, Behandlungsformen*. München: C. H. Beckvgl.
- Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T. & Silver, D. (2020): »Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model«, in: *Nature* 588, S. 604–609. DOI: 10.1038/s41586-020-03051-4
- Searle, J. (1980): Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, S. 417–57.
- Shevlin, H. (2020): »A Digital Remix of Humanity«. Dailynous. <https://dailynous.com/2020/07/30/philosophers-gpt-3/#shevlin>
- Thompson, A. D. (2020): Leta AI Series. <https://www.youtube.com/playlist?list=PLqJbCeNOfEK88QyAkBe-UozxCgbHrGa4V>

- Thompson, A. D. (2021a): *Leta, GPT-3-AI – Episode 1 (Five things, Art, Seeing, Round) – Conversations and talking with GPT3* [Video]. YouTube. <https://www.youtube.com/watch?v=5DBXZRZEBGM&list=PLqJbCeNOFEK88QyAkBe-UozxCgbHrGa4V&index=2>
- Thompson, A. D. (2021b): »Prompt crafting for very large language models«. Life Architect. <https://lifearchitect.ai/prompt-crafting/>
- Turing, A. M. (1948): Intelligent machinery. In: Copeland, JB (2004): *The Essential Turing*. Oxford: Oxford University Press, S. 395–432.
- Turing, A. M. (1950): »Computing machinery and intelligence«, in: *Mind* 59, S. 433–460. <http://www.jstor.org/stable/2251299>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin (2017): »Attention Is All You Need«. arXiv:-1706.03762v5
- Vogd, W., & Feißt, M. (2022): Therapeutische Arrangements im Maßregelvollzug. Studien zur Leerstellengrammatik und den Bezugsproblemen in der forensischen Psychiatrie. Wiesbaden: Springer. (im Erscheinen)
- Watzlawick, P., Beavin, J. H. & Jackson, D. D. (2007): *Menschliche Kommunikation*. Hogrefe AG. ISBN: 978-3456844633
- Weinberg, J. (2020): Philosophers On GPT-3 (updated with replies by GPT-3). <https://dailynous.com/2020/07/30/philosophers-gpt-3/>