Die Nicht-Vernunft der Chatbots

Was macht auf Large Language Models beruhende Künstliche Intelligenz philosophisch interessant?

Sybille Krämer

Abstract: *Humans and machines are constitutively different; but at the same time, technology is a genuine dimension of human existence. What does this ambivalence mean for the interpretation of contemporary Large-Language Models-based artificial intelligence? An anthropomorphizing interpretation should be avoided, as – this is the thesis – artificial intelligence is becoming a cultural technique that forms a dimension of digital literacy. The arguments are developed with reference to a phenomenon that is based on the written character of training data in the scale of entire collective memories. From the perspective of tokenization, the fundamental difference between human language interpretation and algorithmic token statistics becomes obvious: Humans and machines realize completely different ways of using language. What does this mean for issues such as epistemic blackboxing or the ethical trustworthiness of chatbot-generated texts? One thing is clear: artificial intelligence systems belong to the genre of ›non-reason‹ (*Nicht-Vernunft*). The exercise of reason and unreasonableness remains the prerogative of their human designers and users.*

Keywords: *anthropomorphism; chatbots; digital literacy; tokenization; Large Language Models (LLM); distributional semantics; cultural technique of flattening; Digital Humanities; trust*

1. Die konstitutive Andersartigkeit des Technischen

Seit es einen Diskurs über Künstliche Intelligenz gibt, artikuliert sich ein drängendes, immer wieder auftauchendes Ansinnen, das, was Maschinen können, in eine anthropomorphisierende Perspektive einzurücken. ›Anthropomorphisierend‹ heißt dabei: etwas, das aus menschlichen Zusammenhängen – seien diese biologisch oder kulturell – vertraut ist, auf die Technik als ein anzustrebendes Leistungsmuster zu projizieren. Damit werden Mensch und Maschine auf eine Vergleichsebene gerückt, die beide in ein Wettrennen einspannt, bei dem es natur-

gemäß zwei Optionen gibt: die Maschine wird – jetzt schon oder auf die Länge der Zeit – die Menschen übertreffen, ihn gar ersetzen, was für viele, wenn auch nicht für alle, apokalyptisch als Kontrollverlust und Entmachtungsszenario imaginiert wird. Oder es wird eine Domäne unersetzbarer Einzigartigkeit des Menschen identifiziert, in die einzudringen oder diese gar zu erobern allem, was maschinenhaft ist, grundständig verwehrt bleibt.

Schon der Turingtest (Turing 1950) – zumindest so, wie er von vielen verstanden wurde – evozierte die Idee einer auszutestenden Ununterscheidbarkeit von Mensch und Maschine. Wie umgekehrt die kritische Diskussion künstlicher Intelligenz gerne die situierte Verkörperung in einer Lebensform (Dreyfus 1972) oder die Beschränkung auf rein syntaktische Operationen (Searle 1980; Searle 1999) als unüberschreitbare Demarkationslinie gezogen hat.

Doch es gibt auch einen dritten Weg jenseits der Ersetzbarkeit respektive Unersetzbarkeit des Menschen durch technische Apparate. Er besteht in einem veränderten Ausgangspunkt: Mensch und Maschine als kategorial verschiedenartig anzusehen. Es ist daher ein Kategorienfehler, das, was die Maschine leistet, am Vorbild menschlicher Befähigung zu qualifizieren (Crockett 2024). Was immer die Maschine bewerkstelligt, auch dann, wenn das Ergebnis dem Menschenwerk ähnelt, macht sie auf eine ganz andersartige Weise. Allerdings: das ist nur die eine Seite. Denn die Pointe dieses Gedankens konstitutiver Andersartigkeit besteht darin, dass diese Grunddifferenz zugleich die Bedingung der Möglichkeit ist, dass Mensch und Technik in ein enges Interaktionsverhältnis treten können. Ein Verhältnis, das anthropologisch so tief verwurzelt ist und zugleich so weit reicht, dass die Dimension technischer Operativität genuin zur ›Natur‹ des Menschen gehört und eine allen seinen Praktiken inhärente Dimension verkörpert. Die Diversität von Mensch und Technik grundiert und eröffnet erst deren produktives Zusammenspiel.

Wir sind immer zugleich auch ein anderer. Das ist der methodische Ausgangspunkt der folgenden Reflexionen über die zeitgenössischen Chatbot Technologien Künstlicher Intelligenz. Also über eine Technik, die erstaunlicher- oder auch irritierenderweise Texte zu produzieren vermag, die häufig – wenn auch nicht immer – von Texten, die Menschen erzeugten, nicht mehr zu unterscheiden ist.

2. Worum es geht

Eine Familie von Algorithmen bzw. Sprachmodellen macht gegenwärtig Furore (Durt et al. 2023), die als game-changer oder iPhone Moment charakterisiert werden und – da die Chatbots GPT3/4 auch öffentlich verfügbar sind – von Millionen von Nutzer:innen gebraucht, sowie in fast allen gesellschaftlichen Bereichen von Bildungsinstitutionen, über den Journalismus bis zu Stammtischrunden mit Ausdauer kommentiert werden. Firmen (Microsoft, Google sowie die chinesischen

Baidu und Alibaba) arbeiten mit Hochdruck an diesen Systemen; auch daran, sie mit Suchmaschinen und Office-Lösungen zu verknüpfen. Kurzum: eine steilere Aufmerksamkeitskurve als jene, seit ChatGPT3 2022 von Open AI veröffentlicht wurde, hat Künstliche Intelligenz noch kaum erlebt; und dies trotz einer bis in die 50er Jahre des 20. Jahrhunderts zurückreichenden Vergangenheit, die reich war an – gerne mit saisonalen Metaphern bezeichneten – Auf- und Abschwüngen.

Was an dieser neuesten Entwicklung ist philosophisch aufschlussreich? Und was ist überhaupt das Sujet, ein irgendwie bemerkenswerter Kern dieser Entwicklung?

Es sind verschiedene Termini, die dabei kursieren: Im weiteren Sinne wird von Synthetischen Medien und Generativer Künstlicher Intelligenz gesprochen. Damit ist gemeint, dass Künstliche Intelligenz eingesetzt wird um Texte, Bilder, Animationen oder Quellcode zu erzeugen, die in dieser Form und Konstellation gerade nicht als Digitalobjekte im Netz existierten; daher vom System nicht einfach übernommen bzw. adaptiert, vielmehr synthetisch erzeugt werden. Keine Kopien und Plagiate werden produziert, vielmehr Originale, allerdings indem empirisch vorliegende Datenschnipsel oder Datenteilstücke – die Token – synthetisch kombiniert werden. Hinzu kommt ein entscheidendes weiteres Attribut: die Instruktionen und Anfragen – ›prompts‹ genannt – sind in Alltagssprache verfasst, setzen keine Programmierkenntnisse voraus und sind daher – wird das System öffentlich zur Verfügung gestellt – auch von allen einsetzbar, welche die entsprechende App bzw. den Link herunterzuladen und die eigene Emailadresse anzugeben bereit sind.

In einer eingeschränkteren, nur auf die Erstellung von *Texten* fokussierten Sicht geht es um Künstliche Neuronale Netze, die mit hunderttausenden von Textkorpora, Billionen von Worten trainiert werden. Indem als Grundlage das digitalisierte kulturelle Gedächtnis bevorzugt englischsprachiger Kulturen, riesige Büchersammlungen, sowie Webpages und Blogs dienen, entstehen sukzessive Large Language Models (LLMs), die in ihren intern aufgebauten Parametern weder für die Ingenieure, noch für ihre Nutzer außerhalb des je produzierten Outputs transparent sind.

Die Familie der Large Language Models hat viele zeitgenössische Anwendungen selbst auf dem Gebiet der Bildproduktion: ChatBot GPT in seinen verschiedenen Versionen ist nur eine davon. Dass Large Language Models nicht als Modelle für das menschliche Denken zu interpretieren sind, sei hier festgehalten (Mahowald et al. 2023). – Das also ist in grober Skizze das Panorama, vor dem sich unsere Überlegungen entwickeln.

Auch wenn philosophische Reflexionen – zumindest im deutschsprachigen Diskurs – oftmals orientiert sind darauf, mit Künstlicher Intelligenz verbundene technische Entwicklungen kritisch zu kommentieren, ist unsere Absicht eine andere: Unser Ziel ist diese Entwicklung zu *verstehen, zu begreifen,* vielleicht auch: darüber *aufzuklären*. Das allerdings ist nicht möglich, ohne zugleich einzusehen, um welches Verhältnis von Kontinuität *und* Bruch, Tradition *und* Disruption innerhalb der Evo-

lution unserer Medien es dabei geht. Denn auch wenn häufig ein revolutionäres Vokabular geboten erscheint, um den faszinierenden Leistungssprung gegenwärtiger Chatbots zu charakterisieren, hat diese computergenerierte produktive Kraft Vorläufer, an denen bereits zutage tritt, was sich gegenwärtig – wie durch einen Brennspiegel – verstärkt radikalisiert.

3. Was ist neu an der gegenwärtigen Chatbot Technologie?

Medien verändern sich in Schüben; ihr Einsickern in die Alltagspraktiken wiederum hängt von vielfältigen sozialen, also nicht-medialen Bedingungen ab und erstreckt sich meist über lange Zeiträume, um schließlich in ganz unterschiedlichen Medienkulturen zu resultieren. Es gibt also eine Kluft zwischen Medieninnovation und ihrer kulturtechnischen Diffusion. Doch das Beispiel der zeitgenössischen Chatbots ist von markant anderer Dynamik: die Medieninnovation und ihr Gebrauch – dessen Internationalisierung beflügelt wird durch das Übersetzungspotenzial der Maschinen, welche viele Muttersprachen akzeptieren – erfolgt nahezu in ›Echtzeit‹.

Dieses Schrumpfen des Zeitintervalls zwischen Innovation und massenweiser Nutzung verweist auf einen elementaren Tatbestand: die Komponenten, deren Zusammenwirken die Effekte der Synthetischen Medien hervorbringen – so einschneidend neuartig das alles auch erscheint – verfügen über eine meist gut sondierte Vorgeschichte. Allen voran geht es um die Technik Künstlicher Neuronaler Netze, die ein Thema ist schon seit den 40er Jahren des letzten Jahrhunderts (McCulloch/Pitts 1943). Selbstlernende und -optimierende Neuronale Netze zeigen dann seit ca. 2009 (Bengio 2009) überraschende Erfolge in Form selbstadaptiver Lernalgorithmen unter dem Stichwort des ›Deep Learning‹. Allerdings ist das ›Selbstlernen‹ eher Metapher, wenn nicht gar Mythos: Nicht nur werden die Künstlichen Neuronalen Netze trainiert anhand von Daten (Texten, Bildern, Videos), die Menschen mit kollektiver Intelligenz erzeugt, oft auch etikettiert und ausgezeichnet haben; sondern klar ist auch – dem höflich-rücksichtsvollen Kommunikationsgestus gegenwärtiger ChatGPTs unschwer ablesbar – dass Tausende von Clickworkern dafür sorgen, die ins Netz gestellten Chatbots nicht binnen kürzester Zeit zu Rassisten und Hetzern mutieren zu lassen – wie in der Vergangenheit allzu oft geschehen. Gleichwohl: Dass zeitgenössische Lernalgorithmen durch Training interne Modelle bilden, die meist im weiteren Gebrauch sich optimieren, ist eine bereits gut eingeführte Technik.

Auch der Umstand einer dialogisch anmutenden Interaktion mit der Maschine ist nicht neu. Joseph Weizenbaums Computerprogramm ELIZA (Weizenbaum 1966) ließ in den 70er Jahren des 20. Jahrhunderts die Wogen einer kritischen Diskussion von Künstlicher Intelligenz hochschlagen (Weizenbaum 2000[1972]): Denn dieses Programm wurde von einigen Nutzern als ein ›menschlicher‹ Kommunikati-

onspartner empfunden, der mit seinem ›Einfühlungsvermögen‹ psychotherapeutisches Resonanzvermögen nicht nur zeigte, sondern dieses gar übertreffe. Und das obwohl ELIZA lediglich auf geschickte Weise die jeweils zuletzt eingesetzten Worte in Fragen umformulierte – ohne nur irgendetwas wie ein Sinnverstehen dazu zu benötigen.

Das, was in der zeitgenössischen Chatbot Technologie und den ihr zugrundeliegenden Lernalgorithmen tatsächlich neu ist, realisiert sich vor allem auf der Ebene der *Datenarchitekturen* und des *Datenumgangs* und bezieht sich auf zwei Komponenten: (i) auf die exorbitant große Datengrundlage, die als Trainingsreservoir dient und diese seit ca. 2017 (ii) kombiniert mit einer ›Transformer-Technologie‹, welche sich auf ›aufmerksamkeitsgesteuerte‹ Lernverfahren des Systems bezieht – übrigens ein durch und durch fehlgeleiteter psychologisierender Ausdruck!

Zu (i): Die jede Vorstellungskraft sprengende Größe ist schnell charakterisiert. Die internen LLMs sind das Resultat von ca. 570 Gigabyte Texteingabe. Die Parameterzahl, welche die Modelle ausbilden und Voraussetzung ihres Inputanalysepotenzials sowie ihrer synthetischen Leistungen sind, umfassen 175 Milliarden, in anderen Ansätzen sind es schon weit mehr. Für das menschliche Vorstellungsvermögen – von unseren operativen oder gar kontrollierenden Fähigkeiten erst gar nicht zu sprechen – geht es um Größen angesiedelt jenseits aller Vorstellungskraft. Quantität schlägt definitiv um in bemerkenswerten Qualitätszuwachs. Dass das Quantitative hier eine neue Bedeutung erhält – je mehr Trainingsdaten, umso besser die Leistung frei nach dem Motto: ›mehr bringt mehr‹ –, wird von vielen vermerkt.

Zu (ii): Genau auf der Schwelle dieses Umgangs mit unvorstellbar großen Datenmengen, ist die Transformer-Technologie entscheidend, welche Rechenschritte eines Systems zu reduzieren hilft (Vaswani et al. 2017). Das zu entwickelnde Modell erschließt die ›Bedeutung‹ von Worten aus den Kontexten, in denen diese jeweils positioniert sind. Das für sich ist nicht neu und aus der Linguistik bekannt als Distributionelle Semantik, aber auch schon philosophisch sondiert mit der Idee, dass die Bedeutung eines Wortes der Inbegriff seiner Verwendungsweisen und d.h. auch seiner Wortkontexte und Wortnachbarschaften ist. Bei riesigen Datenkorpora steigt allerdings die Anzahl der zu den Wortkontextermittlungen nötigen Rechenschritte unermesslich. Transformer-Technologie nun ist ein Mechanismus, der mit dem operiert, was psychologisch als ›Aufmerksamkeit‹ bekannt ist und – seiner anthropomorphen Hülle entkleidet – datentechnisch lediglich Folgendes meint: Wenn das Wort ›Bank‹ zu disambiguieren ist im Sinne von ›Ruhebank‹ oder ›Finanzinstitut‹, dann heißt ›mit Aufmerksamkeit‹ zu lernen, dass das System nicht *alle*, sondern nur bestimmte Worte, die in der Vorgeschichte auftauchen, dann zur Spezifizierung dieses Wortes überhaupt in Betracht zieht. Also: Worte wie ›Park, Bäume, Spazierengehen etc.‹ bildet das eine und ›Gebäude, Stadt, Kunden, Kreditvergabe, Vorstand…‹ das andere Wortfeld, auf welche die Aufmerksamkeit zu richten ist, um den Verwendungskontext von ›Bank‹ zu ermitteln. Solche Worte mit Aufmerksamkeits-

index werden im internen Modell mit hohen Zahlen belegt bzw. dargestellt, denn das Modell arbeitet immer – daher das Wort ›Transformer‹: ›Umwandler‹ – mit der Übertragung von Texten in Vektorräume, letztlich in *Zahlendarstellungen,* mit denen gerechnet wird, um das plausibelste nächste Wort vorzuschlagen. Was statistisch dann ermittelt wird, ist die Verteilung von Elementen in den Punkte-Populationen in Vektorräumen, in welche die Daten zu transformieren sind. Deshalb spielt das Räumliche, also Nähen und Distanzen zwischen den Worten – letztlich zwischen Datenpunkten – eine so grundlegende Rolle. Dabei ist klar, dass die Plausibilität dessen, was der Chatbot produziert, sich an der Eloquenz menschlicher Kommunikation ausrichtet – und nicht an menschlicher Intelligenz, Denkkraft oder gar ›Wahrheit‹. Elena Esposito (Esposito 2022) schlägt daher vor, dass Künstliche Intelligenz als apparative Künstliche Kommunikation zu verstehen sei.

Das ist natürlich alles eine allzu grobschlächtige Beschreibung. Doch geht es nur darum deutlich zu machen, dass die Neuartigkeit der zeitgenössischen Chatbots in der Dimension des operativen Datenumgangs liegt, verbunden mit sich beständig steigernder Rechenkraft der Hardware.

Doch wir richten nun unseren Fokus auf eine Dimension, die vielleicht so selbstverständlich ist, dass sie in Kommentaren kaum reflektiert wird: Und das ist der grundständige *Schriftcharakter* der Ein- und Ausgaben. Wobei ›Schriftcharakter‹ auch die prinzipielle Übertragbarkeit akustischer Ereignisse in Notation einschließt.

4. Operativität der Token

Es sind nicht einfach Worte, die das Operationsfeld der LLMs bilden, sondern Token. Token sind bedeutungslose kurze Buchstabenzusammenstellungen, wobei auch der Leerraum vor oder hinter einem Buchstaben zum Token zählen kann. Diese ›Leerräume‹ sind ein typisches Phänomen der Schriftbildlichkeit (Giertler/Köppel 2012), nicht unserer Lautlichkeit: die selbstverständlichen Pausen im Sprechen – zumeist solche zum Atemholen – sind der diskreten Zwischenräumlichkeit im Schriftbild, also den regulären Leerstellen und Lücken im Weißraum der Texte, gerade nicht kongruent.

Doch Digitalität bedarf der unterscheidenden Disjunktivität, also der Diskretheit und eindeutigen Differenzierbarkeit zwischen den Zeichen. Zu digitalisieren heißt, etwas, das relativ kontinuierlich und in dieser Hinsicht ›analog‹ ist, in diskrete Elemente zu zerteilen, die codiert und arbiträr miteinander kombinierbar sind (Gramelsberger 2023: 90). Wobei das, was als kontinuierlich und das, was als diskret gilt, relativ sind. Wird ein Drucktext digitalisiert, so nimmt er den Platz des Analogen ein und diese Rollenzuweisung des funktionell ›analogen Parts‹ erfolgt auch bei der Sequenzialisierung bereits digitalisierter Texte in eine Abfolge von Token. Wir

vernachlässigen an dieser Stelle, dass diese Token dann wiederum als Zahlen dargestellt und bearbeitet werden, sondern halten nur fest: Ein- und Ausgabe erfolgen in natürlicher Sprache, doch die maschinelle Verarbeitungsbasis sind Tokenstrukturen der Schrift. Die ›Tokenisierung‹ – ein in diesem Zusammenhang neu geprägtes Wort – bildet die Voraussetzung der generativen Künstliche Intelligenz.

Doch warum ist das bemerkenswert? Besteht unser Sprechen nicht ebenfalls darin, Phoneme – gemäß der Linguistik die kleinsten bedeutungslosen, aber Bedeutungen spezifizierenden Sprechelemente – aufeinanderfolgend hervorzubringen im akustischen Fluss der Rede? Doch eben dies täuscht: Wir können durch die Aneinanderreihung von Phonemen keinen natürlichsprachlichen Ausdruck erzeugen, so wie Buchstabenfolgen einen Text ergeben (Lüdke 1969). Vielmehr erweist sich das Phonem als ein linguistisches Konstrukt und Abkömmling des Graphems: entstanden aus der Projektion der Buchstabenstruktur auf die mündliche Sprache (Stetter 1997). Derrida hatte recht mit seiner zuerst einmal irritierenden Annahme, dass die Schrift der Sprache vorausginge (Derrida 1988). Denn erst die alphabetische Transkription des Sprechens löst die kommunikative Gesamtäußerung mit ihrem typischen Zusammenspiel von Prosodie, Gestik, Mimik, Verbalität und Deixis auf, isoliert und präpariert den verbalen Strang durch Transkription und erzeugt das Sprachliche als ein solitäres Objekt, das als eigenständige Entität – eben als ›die Sprache‹ – erst beobachtbar wird.

Halten wir fest: Die Bedingung der Möglichkeit, dass Künstliche Intelligenz in der zeitgenössisch generativen Form wirksam werden kann, ist die im Schriftcharakter der Datenkorpora gründende ›Tokenisierung‹.

Diese Token-Perspektive verkörperte eine Dimension der Schriftsprachlichkeit, zu der Menschen gewöhnlich keinen Zugang haben. Eine verschriftete Sprache zeigt Regelmäßigkeiten in der Häufigkeit, der Verteilung und Zusammenstellung der Buchstaben. Andrei Markov (Markov 1912) berechnete früh schon stochastisch Buchstabensequenzen in der russischen Literatur. Doch die Mathematik dieser Regelmäßigkeiten spielt für die Kulturtechnik alphabetischer Literalität, also in der Perspektive menschlicher Praktiken, keine Rolle – jedenfalls außerhalb der Kryptologie, die Geheimschriften zu decodieren hat. Denn selbstverständlich kann die Aufdeckung der mathematischen Buchstabenrelationen der entscheidende Schlüssel sein in der Entzifferung einer Geheimschrift. Doch gewöhnlich bleiben diese mathematisch identifizierbaren und analysierbaren Muster unterhalb der Oberfläche dessen, was beobachtbar und den Sprechenden, Schreibenden und Lesenden bewusst ist. Sinnhaftes Sprachverständnis und berechenbare Sprachstatistik scheinen einmal mehr den konstitutiven Unterschied von Mensch und Technik aufzurufen.

Und doch ist diese definitive Unterscheidbarkeit nun zu relativieren. Die Differenz zwischen human verstehbarem Sinn und maschinell berechenbarer Statistik als zwei unterschiedlichen Erschließungsformen von Sprache, zeigt nur die ›halbe

Wahrheit‹. Denn es gibt einen Punkt, wo beide sich berühren und auch dieses Zusammenspiel gründet wiederum im Schriftcharakter der Sprachdaten. Wenn – wie Wittgenstein, aber auch pragmatische Sprachtheorien es nahe legen – die Bedeutung eines Wortes sein Gebrauch ist (Durt et al. 2023), dann kommt die Verteilung von Worten in den Schriftbildern von Texten ins Spiel, denn diese Distribution ist signifikant für Wortbedeutungen. Worte sind charakterisierbar durch diejenigen Worte, die sie begleiten. Wie nah beisammen oder wie entfernt voneinander Worte in Textkorpora vorkommen, kann in ihren räumlichen Entfernungen jeweils gemessen und analysiert werden. Tritt ›Bank‹ in räumlicher Nachbarschaft zu ›Park‹ und ›Spazierengehen‹ etc. oder eher zu ›Kreditkonditionen‹ und ›Bilanzen‹ auf?

In einem viel radikaleren Sinne als Philosophen, Sprachwissenschaftler und Informatiker dies vermutet hätten, können aus diesen auf Textoberflächen in Erscheinung tretenden Wortrelationen – allerdings nur, sofern das Korpus der Sprachdaten genügend groß ist – maschinell Schlüsse gezogen, also Informationen mit dem Status eines ›Weltwissens‹ abgeleitet werden. Und dies, obwohl diese Informationen nicht explizit in den Trainingsdaten vorgelegen haben und ein LLM basiertes System – gewöhnlich, aber das wird sich ändern – auch keinen Zugang zur Außenwelt hat. Die Systeme können implizite Strukturen in Textkorpora erkennen und explizit machen – unter Umständen diese auch herbei phantasieren – kraft ihres synthetischen Vermögens. Wir kommen auf die halluzinatorischen Kräfte der LLMs noch zurück.

Die Linguistik der Distributionellen Semantik hat diesen Blickwinkel der Kontextabhängigkeit von Wortbedeutungen bereits entfaltet und damit eine Bedeutungstheorie vorgeschlagen, die unabhängig ist von einer denotativen, externen Weltreferenz (Rieger 1991). Worte haben eine ähnliche Bedeutung, sofern sie in ähnlichen Wortumfeldern vorkommen (Firth 1957). Und auch dabei ist klar, dass die Distributionelle Semantik Sprache in ihrer schriftsprachlichen Version zur Grundlage machen muss: Schriften sind räumliche Anordnungssysteme (Ehlich 2012). Deshalb erst kann mit der Schrift die Idee eines Sprach*raumes* entstehen, in dem Verteilungen, Positionierungen, Entfernungen etc. analysierbar werden.

Halten wir fest: Die embryonale Digitalität (Krämer 2022) des alphanumerischen Zeichenraumes, welcher Alphabete ebenso umfasst wie das dezimale Positionssystem, bildet eine wesentliche Grundlage gegenwärtiger Chatbot-Techniken und markiert eine notwendige Bedingung der zeitgenössischen synthetischen Künstlichen Intelligenz. Es ist zugleich der Faden, den die gegenwärtige Technik mit der Tradition und Historie der Kulturtechniken der Literalität verbindet. Neu ist dabei, dass das tradierte Nadelöhr zur Maschineninstruktion, das darin besteht, Programmiersprachen, also formale Schriften einsetzen zu müssen, bei Interaktionen mit den LLM basierten Chabots entfällt. Durch die Möglichkeit natürlichsprachlichen In- und Outputs kann diese Technologie den Status einer Kulturtechnik annehmen, die nicht nur in Expertenkreisen, sondern im

Alltag angekommen ist und überdies durch jede bestätigte Anwendung sich optimiert. Gerade die Effizienz, mit der die zeitgenössische Künstliche Intelligenz als ›Interaktionspartner‹ auf der Operationsbasis von Token agiert zeigt, dass jedweder Anthropomorphismus fehl am Platze ist: Ein Chatbot versteht nicht, was er kommuniziert. Allerdings sollten wir nicht vergessen, dass auch Menschen sich miteinander verständigen können, ohne sich – im emphatischen Sinne – verstehen zu müssen.

5. Der Computer – eine Oberflächentechnologie

Wir wollen uns der grundständigen Differenz zwischen Menschen und Computer noch einmal in anderer Perspektive annähern. Der Terminus ›Computer‹ sei hier eine Chiffre für alle möglichen Varianten digitalisierter Algorithmensysteme inklusive entsprechender Hardwareapparaturen. Die unvorstellbar großen Datenreservoire, welche Computer auf Muster hin durchsuchen und aus denen auch neue Muster synthetisiert werden können, verkörpert eine – um es mit einem Wortmonster auszudrücken – ›Oberflächenbezugnahmetechnologie‹. Denn es gibt einen Zusammenhang zwischen ›ein Muster sein‹ und ›Oberflächen inskribiert zu sein‹. Selbst so komplexe Figuren wie komponierte Musik können in Teilaspekten ihrer Performanz als Schriftmuster auf den Notenblättern der Partituren dargestellt werden. Überdies können Muster den Status von Spuren annehmen und in der Forensik der Spur übertrifft das Analysevermögen von Computern jeden Menschen. Um eine etwas schiefe optische Analogie zu bemühen: Computer sind Teleskope und Mikroskope in Datenuniversen. ›Schief‹, weil bei diesen Metaphern der synthetische, generative Aspekt entfällt.

Doch unser Fokus hier ist ein anderer: Zwar scheint wieder der Anlass gegeben, die Unterschiedlichkeit von Menschen und Maschine zu betonen: Die Maschine vermag die umfangreichen, inskribierten und illustrierten Oberflächen sozialer Gedächtnisse zu durchmustern, während Menschen bei den nur wenigen Texten, die sie lesend zu erschließen in ihrem Leben überhaupt in der Lage sind, auf die Hermeneutik einer Tiefeninterpretation angewiesen sind. Und doch wäre es falsch der ›Oberflächlichkeit‹ computergenerierter Verfahren die ›Tiefe‹ hermeneutischer Interpretation seitens der Menschen entgegenzuhalten (Krämer 2023b). Wir sind zwar sozialisiert mit dem Narrativ einer Rhetorik, das tiefgründiges Denken erwünscht und fruchtbar, auf die Oberflächen orientiert zu bleiben dagegen diskriminiert, wenn nicht gar als ›oberflächlich‹ tabuisiert. Doch menschliche Erkenntnisarbeit – und nicht nur diese – ist undenkbar ohne den Einsatz von Bildern, Schriften, Diagrammen, Graphen oder Karten, mithin inskribierter und illustrierter artifizieller Flächen (Krämer 2016; Krämer 2022b). Alle Wissenschaften, viele Künste, komplexe Architektur und Technik sowie die Verwaltung großer Organisationen

sind nicht machbar ohne eine ›Kulturtechnik der Verflachung‹. In der Ubiquität des Smartphones kulminiert diese Entwicklung. Die Projektion in die Zweidimensionalität bleibt kein schlichtes Reduktions- oder Abstraktionsverfahren, sondern verkörpert eine kulturstiftende und epistemische Produktivkraft (Krämer 2016).

Dieser Einsatz der artifiziellen Flächigkeit ist die ›Ebene‹, auf der sich Mensch und Maschine treffen. Eine Ebene, die den Nährboden operativer Kontrollierbarkeit zugleich bereitstellt, wie sie ihn auch wieder zunichtemachen kann.

Alles was körperlich ist, inklusive unsere leiblich-räumliche Situierung, ist durch drei senkrecht aufeinander stehende Achsen charakterisierbar: oben/unten, rechts/links, vorne/hinten. Notgedrungen bleibt im dreidimensionalen Raum erst einmal für menschliche Wesen verborgen und außer Kontrolle, was sich hinter ihnen befindet: Die Erfindung inskribierter und illustrierter Flächen mit ihren zwei Anordnungsrichtungen: oben/unten und rechts/links amputiert genau diese dritte, unübersichtliche, die verborgene Tiefendimension und schafft oder suggeriert einen Raum perfekter Übersicht und Kontrolle, erst recht wenn dieser – daher der Siegeszug des Papiers – überaus mobil ist (Krämer 2022a). Nicht zufällig hat das erste niedergeschriebene Computerprogramm von Ada Lovelace 1843 die Form einer Tabelle (Lovelace 1843; Krämer 2015): In horizontalen Zeilen einerseits und vertikalen Kolumnen andererseits entblättern sich die verschiedenen Maschinenzustände im Akt des Berechnens einer bestimmten Zahlenfolge (in diesem Falle der Bernoulli Zahlen). Mit dieser tabellarischen Instruktion wurde die ›analytical engine‹, die Charles Babbage als ersten Universalcomputer auf Papier entwarf – im Prinzip – von Ada Lovelace in eine konkrete Maschine für spezielle Zwecke verwandelt.

Der Computer ist also nicht nur eine forensische, er ist eine diagrammatische Maschine (Mackenzie 2017), die von zweidimensionalen Inskriptionen zehrt, auch wenn diese wiederum in die Linearität von Bitfolgen zu verwandeln sind. Die Turingmaschine, die Alan Turing als mathematische Präzisierung des Algorithmenbegriffs entwickelte und mit der er auf grundständige Weise ausbuchstabiert, was es heißt einem Algorithmus zu folgen, hat die Form einer Tabelle bzw. Tafel, in welche die vier grundlegenden Rechenschritte einer Maschine, nämlich die Beschriftung eines Arbeitsfeldes, das Nach-rechts-gehen, das Nach-links-gehen, sowie das Stoppen in zweidimensionaler Anordnung notiert ist.

Die Stationen computeraffiner Diagrammatizität sind hier nicht abzugehen. Doch in dem Augenblick, in dem die Datenkonvolute inskribierter Flächen zu Massendaten werden, die als Trainingsgrundlage selbstadaptierender Lernalgorithmen fungieren, ändert sich etwas grundsätzlich. Ursprünglich ist der epistemische Gebrauch artifizieller Flächigkeit auch eine Strategie der Schaffung von Übersicht, Transparenz und Kontrolle. Doch genau dies ist angesichts der gesammelten Mannigfaltigkeit kollektiver digitaler Textproduktion im Umfang kultureller Ge-

dächtnisse nicht mehr zu realisieren. Jorge Louis Borges ›Unendliche Bibliothek‹ (Borges 2013) liefert dazu ein passendes Bild.

Und so kommen mit den Large Language Models die Dunkelkammern unübersichtlicher Tiefenregionen zurück. Und nicht erst mit diesen. Für die vernetzten Nutzer:innen, die vor den Bildschirmen Texte und Bilder anschauen und bearbeiten, ist immer schon klar, dass die rhizomartige Verknüpfung und Interaktion zwischen den Computern, Algorithmen und Protokollen, die sich im Hintergrund der Bildschirme vollzieht, ihnen ebenso entzogen bleiben, wie die ins Unendliche sich vervielfältigenden Navigationsrouten, welche die Web-links eröffnen und die doch immer nur in den ersten Passagen individuell aktiviert werden können: Dass so unüberschaubar viele Wissenszugänge anzusteuern sind, wird untrüglich zum Anzeichen für all das, was zugleich *nicht* begangen, liegen gelassen wird und also gewusst werden kann. Doch nun gilt dieser Entzug von Wissen, einfach weil keine Übersicht und Kontrolle der intern entwickelten Modelle mit ihren in die Millionen gehenden Parameter überhaupt möglich ist, auch für die Ingenieure selbst.

Davon zeugt ein interessantes Phänomen: die Unvorhersehbarkeit eruptiver Leistungssteigerung im Zuge des Trainings der LLMs. Wenn Systeme mit einer fortschreitenden Erweiterung der Massendatenumfänge trainiert werden, zeigen sie über lange Phasen keinen mathematisch über dem Zufall liegenden Zugewinn hinsichtlich der anzutrainierenden Fähigkeiten. Doch dann zeigt sich wie in plötzlicher Eruption, ein steiler Zuwachs in der Ausübung der anvisierten Fähigkeit: die Lernkurve steigt steil (Wei et al. 2022) empor. Der Zeitpunkt der Emergenz dieser signifikant neuen Qualität auf Grundlage der Steigerung des Umfangs der Trainingsdaten, bleibt für die Beteiligten unvorhersehbar.

Diese Unvorhersehbarkeit gilt auch für jene Faktoren, die nicht einfach als Leistungssteigerung zu verbuchen sind, sondern umgekehrt die Risiken zeitgenössischer Chatbots markieren. In der technischen Literatur zu den LLMs werden viele solcher Risikobereiche sondiert. Wei et al. (Wei et al. 2022) heben drei Risiken hervor: Wahrhaftigkeit, Vorurteilsbehaftetheit, Toxikalität. Wobei unter ›Toxikalität‹ verstanden wird, dass die Reaktionen eines Chatbots auf Prompts gerade nicht hilfreich, nicht frei von Beleidigungen und nicht geprägt durch Respekt ausfallen. Für alle drei Bereichen hat sich gezeigt, dass mit Steigerung des Umfangs der Trainingsdaten, auch das Vorkommen dieser unerwünschten Eigenheiten von Chatbots zunehmen kann. Zugleich wird mit Hochdruck daran gearbeitet, wie dieses so unerwartete wie kommerziell desaströse Verhalten von Chatbots zu verändern ist. Überdies werden die Diagnosen dieser Risiken begleitet von Entwürfen für detaillierte Umgangsregeln für die Chatbot Nutzung (s. Antoniak et al. 2023).[1]

1 Antoniak et al. schreiben: »While the output of LLMs often looks very convincing, we recommend that you ask yourself the following questions before trusting it[:] Is this an appropriate use of an LLM, given the limitations of LLMs and the risks of my intended application? [/]

Dies macht klar, dass das Blackboxing, das zu einer Dimension der großen Sprachmodelle geworden ist, keineswegs identifiziert werden darf mit dem generellen Verlust an Kontrolle gegenüber den Systemen Künstlicher Intelligenz. Nur erstreckt die Kontrollierbarkeit sich auf den Output, nicht auf die interne Modellbildung. Gerade der Umstand, dass die gegenwärtigen Chatbots nicht mehr umkippen in Maschinen der Hetze und Hassrede, legt davon Zeugnis ab, in welchem Maße Regulierbarkeit und Kontrolle möglich sind. Kontrolle ist keine technische, sondern eine politische Frage.

Halten wir fest: Die ins Vielfache angewachsene Zugänglichkeit des kollektiven Wissens – von den Suchmaschinen bis hin zu den Chatbots – wird erkauft durch einen Zuwachs des Nichtwissens, durch das Blackboxing bezüglich interner Funktionsbereiche der Technologie. Und dies ist kein Betriebsunfall, der demnächst zu bereinigen wäre, sondern ist strukturell. Die Proportionalität zwischen Wissenszuwachs und Nichtwissen bildet nur das Echo jener Proportionalität, bei der die Erhöhung der Datenvolumina als Trainingsgrundlage, tatsächlich zu einer Erhöhung der Leistungen führt.

6. Künstliche Intelligenz wird zur Kulturtechnik

Unabhängig der Auf- und Abschwünge Künstlicher Intelligenz (Floridi 2020) galt doch zumeist, dass der Werkzeugkasten ihrer Verfahren für solche Domänen eingesetzt wurde, die als Expertenwissen zählten. Zwar änderte sich das, als mit Spamfiltern, Gesichts- und Spracherkennung, Smartphonefotografie, Marketingprofiling etc. auch die technischen Zurüstungen des digitalen Alltags durchdrungen werden von Künstlicher Intelligenz – doch diese bleiben zumeist ›stumme‹ Hintergrundverfahren, weder als spezifische KI-Technologie zugänglich noch überhaupt den Nutzern bewusst. Doch nun vollzieht sich darin eine signifikante Wende: Indem Chatbots der jüngsten Generation öffentlich nutzbar sind und das mit natürlichsprachlichen Eingaben, avanciert Künstliche Intelligenz zu einer sich im Alltag verankernden Kulturtechnik. Sie wird zu einer Facette der ›Kulturtechnik digitaler Literalität‹.

Is this an appropriate use of an LLM, given my own vulnerabilities or the vulnerabilities of people using the LLM? [/] Am I ok with my prompts being stored and shared with others? Is there any private information(medical history, finances) in my prompts? [/] Have I checked the accuracy of the output? Does the output contain information that I didn't ask for? [/] Am I asking the kind of questions where giving credit would be important, and if so, am I be able to identify the authors of the model's output so that I can credit them? [/] Does the output contain any opinions or advice, and if so, am I ok with my own opinions being influenced on this topic? [/] Do I have enough distance from the LLM, or am I interacting with the LLM as if it were a person (or encouraging others to interact with the LLM as if it were a person)?«. (Antoniak et al. 2023)

Was bedeutet ›Kulturtechnik‹? Um die Jahrtausendwende fand sich in Berlin eine Gruppe von acht Wissenschaftler:innen zusammen,[2] die das Helmholtz-Zentrum für Kulturtechnik an der Humboldt-Universität sowie einen neuartigen interdisziplinären Forschungsschwerpunkt ›Bild, Schrift, Zahl‹ begründeten (Krämer/Bredekamp 2003). Das Ziel war eine Umorientierung und Neujustierung im gängigen Kulturkonzept: Kultur sollte nicht länger primär als geistiges Phänomen begriffen werden, ausgerichtet an den Gipfelpunkten abendländischer Kunst und Bildung. Vielmehr galt es die konstituierende und konstruktive Rolle der Materialität, Medialität und Technizität alltäglicher Praktiken als kulturstiftende Dimensionen und zivilisatorische Partizipationsmöglichkeit zu begreifen und zu rekonstruieren.

Eingerückt in diesen Horizont befinden wir uns in einem Umbruch, bei dem die alphanumerische Literalität übergeht in die digitale Literalität und dieser Wandel schließt sukzessive Verfahren Künstlicher Intelligenz als genuine Facetten zeitgenössischer digitaler Kulturtechniken mit ein.

Dass Künstliche Intelligenz den Status einer Kulturtechnik anzunehmen beginnt, zeigt sich untrüglich darin, dass Institutionen wie Universitäten und Schulen darüber reflektieren und debattieren müssen, wie der Umgang mit dieser Technologie in Bildung und Ausbildung zu organisieren ist, wie das Prüfwesen sich verändern wird und welche Kontrollmöglichkeiten überhaupt zu erschließen und praktisch angewendet werden können. Fragen über Fragen, deren Antworten schwierig zu finden sein werden.

Auf einen weiteren Aspekt sei noch aufmerksam gemacht. Die Geisteswissenschaften haben ein Selbstbild gepflegt, das Interpretation und Hermeneutik gerne als Königsweg und Alleinstellungsmerkmal (ver)klärte (Krämer 2018). Doch solche Sicht ist problematisch, wenn nicht gar: falsch. Denn auch die Geisteswissenschaften haben von Anbeginn – denken wir an historische Datierungen, an Seitenzahlen, Konkordanzen, Werkkataloge – nicht nur mit Zahlen und Daten, sondern ›buchstäblich‹ auch mit Objekten und Materialien zu tun, welche gefunden, gesammelt, geordnet, ausgezeichnet, annotiert, verglichen, archiviert etc. werden müssen. Ohne dieses ›Handwerk des Geistes‹ wäre geisteswissenschaftliche Interpretation nicht möglich.

Es ist keine Frage, dass sich dieses Handwerk der Gelehrtenarbeit unter digitalen Bedingungen ändert. Die meisten Geisteswissenschaftler:innen sind – um es im Jargon zu sagen – ›analog unterwegs‹; sie setzen also gerade nicht die datengetriebenen Verfahren der Digital Humanities ein. Gleichwohl ist akademische Arbeit ohne das ›Abc‹ digitaler Grundoperationen kaum mehr praktizierbar. Das Lesen

2 Horst Bredekamp, Kunstgeschichte; Jochen Brüning, Mathematik; Wolfgang Coy, Informatik; Friedrich Kittler, Kulturwissenschaften; Sybille Krämer, Philosophie; Thomas Macho; Bernd Mahr, Informatik; Horst Wenzel, Mediävistik.

und Schreiben am Bildschirm, Kommunikation über Emails, gemeinsame Arbeit an Files, Nutzung von Suchmaschinen, digitalisierter Editionen und Quellen bestimmen den Alltag nahezu aller geisteswissenschaftlichen Arbeit.

Nun ist zu erwarten, dass diese digitale Literalität sich unter dem Einfluss jüngster Chatbots verändern wird. So wie heute Suchanfragen im Netz eine unerlässliche Dimension geisteswissenschaftlicher Arbeit sind, so werden in Zukunft Assistenzarbeiten von ChatGPT als da sind Literaturlisten zusammenstellen, Kurzinfos über Buchinhalte/Aufsätze geben, abstracts schreiben etc., viele Bereiche geisteswissenschaftlicher Arbeit unterstützen. Dass es um ›unterstützen‹ und nicht um ›ersetzen‹ geht, ist bedeutsam. Denn das Damoklesschwert über der akademischen Nutzung GPT produzierter Texte besteht in deren Fiktionalität – oft als deren Bereitschaft zu halluzinieren beschönigt. Ohne Überprüfung von Wahrheit und Faktizität, ist nachhaltige Chatbot-Aussagekraft nicht zu haben. Allerdings wird an der Sollbruchstelle ›Wahrheitsbezug‹ intensiv gearbeitet. Nur: gilt solche Unsicherheit bezüglich der Wahrheit eines Textes nicht letztlich für *jeden* Text? Doch die menschliche Kultur hat zur Bewältigung dieses Problems einen ›epistemischen Service‹ hervorgebracht. Was dieser mit dem Problemfeld der Chatbot Halluzinationen zu tun hat, sei im letzten Schritt des Essays sondiert.

7. Das Problem des Vertrauens

In 95% dessen, was wir wissen, verlassen wir uns auf Worte, Schriften und Bilder anderer. Vielleicht hören wir das nicht gerne, vielleicht ist uns das gar nicht bewusst: Dass wir unsere Überzeugungen eigenhändig auch rechtfertigen könn(t)en – wie es die Philosophie vorgibt bzw. erwartet – gilt nur für allzu wenige Wissensbereiche (Krämer 2023a). Zwar machte es historisch Sinn, dass die europäische Aufklärung in ihrem Versuch Erkenntnis von der kirchlichen, aber auch politischen Nabelschnur zu lösen, das Individuum stärkte; und also die Einzelnen ermutigte sich ihres eigenen Verstandes und ihrer Urteilskraft zu bedienen. Doch die Kollektivität unserer Intelligenz ist ebenso unumgänglich wie die Sozialität von Erkenntnis und Wissen. In vielen Hinsichten müssen wir dem, was andere sagen, schreiben oder zeigen, vertrauen. Erst dieses Vertrauen in Personen wie auch in Institutionen bereitet den Boden, aus dem ein Wissen durch die Worte anderer sich entwickeln kann.

Als Menschen neigen wir dazu denen zu vertrauen, die uns am ähnlichsten sind. Nicht zufällig waren es zu Beginn der Neuzeit gerade die ›Gentlemen‹, welche angesichts der Nicht-Reproduzierbarkeit naturwissenschaftlicher Experimente eingesetzt wurden, um deren Befunde zu bezeugen und zu verbürgen. Wir sind epistemisch abhängige Wesen (Krämer 2017): Das Vertrauen in andere ist also keineswegs nur praktisch, sondern auch epistemologisch von Belang. Dem Erkennen, geht das Anerkennen voraus.

Doch wie verhält es sich mit der Vertrauenswürdigkeit im Umgang mit Chat GPT erzeugten Texten? Hier zeigt sich ein Problem, das keineswegs auf Künstliche Intelligenz beschränkt ist, sondern immer schon maschinelle Textproduktionen und die Interaktion mit Technik z.B. in der Robotik begleitete. Es ist die Ambivalenz von Misstrauen einerseits und Übervertrauen andererseits in das, was eine Maschine, was Algorithmen tun.

Übervertrauen ist das Phänomen, wenn Menschen einem technischen System jenseits seiner aktual vorhandenen Kapazitäten trauen. Verstärkt wird Übervertrauen durch den psychologisch naheliegenden Mechanismus aus gelungenen vergangenen Erfahrungen auf die Zukunft zu schließen. Und dies, obwohl bezüglich vergangener ›guter‹ Erfahrungen keineswegs klar ist, ob deren Gelingen in der Richtigkeit und Angemessenheit eines Tuns wurzelte, oder einfach nur ›Glück‹ war. In der Robotik, insbesondere beim Einsatz sozialer Roboter, ist dieses Problem gründlich untersucht (Bahner 2008).

Die Eleganz, Natürlichkeit und Plausibilität, mit der die Large Language Models basierten Kommunikationsassistenten auf Eingaben reagieren, kann nicht nur einer anthropomorphisierenden Deutung der Chatbots als ›verstehenden Entitäten‹ Vorschub leisten, sondern kann auch massiv Phänomene wie eben das Übervertrauen befördern. Doch spätestens hier wird klar, mit welcher Skepsis und Distanz den Chatbots dann zu begegnen ist, wenn ihre Aussagen für ›bare Münze‹ gelten, sobald sie also als Instrumente akademischer oder pädagogischer Arbeit eingesetzt werden. Chatbots sind keine Suchmaschinen und keine Internetportale enzyklopädischer Information. Es ist unabdingbar die Variabilität der Informationsquellen des Internets zu nutzen, sobald Wahrheit im Spiele ist.

›Wahrheit‹ als ein Kriterium akademischer Arbeit zu etablieren, gehörte zum Telos der Europäischen Aufklärung. Und auch wenn die neuzeitliche Aufklärung sich selbst durch ihre kolonialen Verstrickungen ein Stück weit desavouierte, bleibt die Überprüfung des Wahrheitsbezuges ein durch sie auf die Agenda gesetztes epistemisches Potenzial, das auch ein Kernstück der Digitalen Aufklärung zu bleiben hat. Künstliche Intelligenz kann weder denken und erst recht nicht vernünftig oder unvernünftig sein. Ihre Systeme gehören dem Genre der ›Nicht-Vernunft‹ an. Vernunft und Unvernunft zu praktizieren bleibt ein Vorrecht ihrer menschlichen Konstrukteure und Nutzer.

Literatur

Antoniak, M. et al. (2023): Using Large Language Models With Care. How to be mindful of current risks when using chatbots and writing assistants, in: *AI2 Newsletter*, 2023(7). [https://blog.allenai.org/using-large-language-models-with-care-eeb17b0aed27] (Zugriff: 08.07.2023).

Bahner, J.E. (2008): Übersteigertes Vertrauen in Automation: Der Einfluss von Fehlererfahrungen auf Complacency und Automation Bias (TU Berlin Doctoral Thesis). [https://depositonce.tu-berlin.de/items/287a4685-ca1d-4972-94c5-beba68a10612] (Zugriff: 08.07.2023).

Bengio, Y. (2009): Learning Deep Architectures for AI, in: *Foundations and Trends in Machine Learning*, 2(1), 1–127.

Borges, J.L. (2013): Die unendliche Bibliothek, Frankfurt a.M.: Fischer.

Crockett, J. (2024): How to Raise Your Artificial Intelligence. A Conversation with Alison Gopnik and Melanie Mitchell May 31, 2024. [https://lareviewofbooks.org/article/how-to-raise-your-artificial-intelligence-a-conversation-with-alison-gopnik-and-melanie-mitchell/] (Zugriff: 08.06.2024).

Dreyfus, H. (1972): What Computers Can't Do, New York: MIT Press.

Durt, C. et al. (2023): Against AI Understanding and Sentience. Large Language Models, Meaning, and the Patterns of Human Language Use. [http://philsci-archive.pitt.edu/21983/ preprint].

Derrida, J. (1988): Signature Event Context, in: Ders. (Hg.), Limited Inc, Evanston: Northwestern University Press, 1–23.

Ehlich, K. (2012): Schrifträume, in: Krämer, S. et al. (Hg.), Schriftbildlichkeit, Berlin: Akademie Verlag, 39–60.

Esposito, E. (2022): Artificial Communication. How algorithms produce social intelligence, Cambridge (MA): MIT Press.

Firth, J.R. (1957): A synopsis of linguistic theory 1930–1955, in: *Studies in Linguistic Analysis*, 1957, 1–32 [Reprinted in: F.R. Palmer (Hg.) (1968), Selected Papers of J.R. Firth 1952–1959, London: Longman].

Floridi, L. (2020): AI and its New Winter. From Myths to Realities, in: *Philosophy & Technology*, 2020(33), 1–3.

Giertler, M.; Köppel, R. (Hg.) (2012): Von Lettern und Lücken. Zur Ordnung der Schrift im Bleisatz, München: Fink.

Gramelsberger, G. (2023): Philosophie des Digitalen zur Einführung, Hamburg: Junius.

Krämer, S. (2015): Wieso gilt Ada Lovelace als die »erste Programmiererin« und was bedeutet überhaupt »programmieren«?, in: Dies. (Hg.), Ada Lovelace. Die Pionierin der Computertechnik und ihre Nachfolgerinnen, München: Fink, 75–90.

Krämer, S. (2016): Figuration, Anschauung, Erkenntnis. Grundlinien einer Diagrammatologie, Berlin: Suhrkamp.

Krämer, S. (2017): Epistemic Dependence and Trust. On witnessing in the first, second, and third-person perspectives, in: Krämer, S.; Weigel, S. (Hg.), Testimony/Bearing Witness. Epistemology, Ethics, History, and Culture, London: Rowman & Littlefield, 247–259.

Krämer, S. (2018): Der »Stachel des Digitalen« – ein Anreiz zur Selbstreflexion in den Geisteswissenschaften? Ein philosophischer Kommentar zu den Digital Huma-

nities in neun Thesen, in: *Digital Classics Online,* 4(1). [https://doi.org/10.11588/dco.2018.0].

Krämer, S. (2021): Reflections on »operative iconicity« and »artificial flatness«, in: Wengrow, D. (Hg.), Image, Thought, and the Making of Social Worlds, Freiburger Studien zur Archäologie & Visuellen Kultur Bd. 3, Heidelberg: Propylaeum, 252–272.

Krämer, S. (2022a): Kulturgeschichte der Digitalisierung. Über die embryonale Digitalität der Alphanumerik, in: *APuZ: Aus Politik und Zeitgeschichte*, 72(2022), 10–17.

Krämer, S. (2022b): The Artificiality of the Human Mind: A Reflection on Natural and Artificial Intelligence, in: Nagl-Docekal, H.; Zacharasiewicz, W. (Hg.), Artificial Intelligence and Humane Enhancement, Berlin/New York: de Gruyter, 17–32.

Krämer, S. (2023a): Bearing Witness as Truth Practic: The Twofold – Discursive and Existential – Character of Telling Truth in Testimony, in: Jones, S.; Woods, R. (Hg.), The Palgrave Handbook of Testimony and Culture, Cham: Palgrave Macmillan, 23–38.

Krämer, S. (2023b): Should we really »hermeneutise« the Digital Humanities? A plea for the epistemic productivity of a »cultural technique of flattening« in the Humanities, in: *Journal Cultural Analytics*, 7(4). [https://doi.org/10.11588/dco.2018.0].

Krämer, S.; Bredekamp, H. (Hg.) (2003): Bild Schrift Zahl, Kulturtechnik Bd. 1, München: Fink.

Lovelace, A.A. (1843): Notes by A.A.L. (Augusta Ada Lovelace), Taylor's Scientific Memoirs, London, Vol. iii, wieder gedruckt in: Morrison P.; Morrison E. (Hg.) (1961), Charles Babbage and His Calculating Enginges. Selected writings by Charles Babbages and Others, New York: Dover Publications; dt. Version: Grundriß der von Charles Babbage erfundenen Analytical Engine, aus dem Französischen übersetzt und kommentiert von Ada Augusta Lovelace, in: Dotzler, B. (Hg.) (1996): Babbages Rechen-Automate. Ausgewählte Schriften, Computerkultur Bd. 6, Wien/New York: Springer, 666–731.

Lüdke, H. (1969): Die Alphabetschrift und das Problem der Lautsegmentierung, in: *Phonetik*, 20(2-4), 147–176.

Mackenzie, A. (2017): Machine Learners. Archaeology of Data Practice, Cambridge (MA): MIT Press.

Mahowald, K. et al. (2023): Dissociating Language and Thought in Large Language Models. A Cognitive Perspective. [https://doi.org/10.48550/arXiv.2301.06627].

Markov, A.A. (1912): Wahrscheinlichkeitsrechnung, Leipzig: Teubner.

McCulloch, W.; Pitts, W. (1943): A logical calculus of the ideas immanent in nervous activity, in: *Bulletin of Mathematical Biophysics*, 5, 115–133.

Rieger, B.B. (1991): On Distributed Representations in Word Semantics (Report). ICSI Berkeley 12–1991. CiteSeerX 10.1.1.37.7976.

Searle, J.R. (1980): Minds, Brains, and Programs, in: *The Behavioral and Brain Sciences*, 3(3), 337–356.

Searle, J.R. (1999): Chinese Room Argument, in: Wilson R.A.; Keil, F. (Hg.), The MIT Encyclopedia of the Cognitive Sciences, Cambridge (MA): MIT Press, 115f.

Stetter, C. (1997): Schrift und Sprache, Frankfurt a.M.: Suhrkamp.

Turing, A. (1950): Computing Machinery and Intelligence, in: *Mind*, 59(236), 433–460.

Vaswani, A. et al. (2017): Attention Is All You Need. [https://arxiv.org/abs/1706.03762].

Weizenbaum, J. (1966): ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine, in: *Communications of the ACM*, 9(1), 36–45.

Weizenbaum, J. (2000[1972]): Die Macht der Computer und die Ohnmacht der Vernunft, Frankfurt a.M.: Suhrkamp.

Wei, J. et al. (Hg.) (2022): Advances in Neural Information Processing Systems. [https://openreview.net/forum?id=_VjQlMeSB_J] (Zugriff: 08.07.2023).