in: Sprache im technischen Zeitalter 61, Nr. 245 (2023)

Hannes Bajohr

Artifizielle und postartifizielle Texte

Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben

Ich freue mich außerordentlich, in diesem Jahr die Walter-Höllerer-Vorlesung halten zu dürfen.[1] Wie Sie wissen, war ihr Namenspatron während seiner Zeit an der Technischen Universität Berlin für die Gründung einer Zeitschrift verantwortlich, die auch heute noch existiert. Ihr Titel beschreibt recht genau, was mich in dieser Vorlesung umtreiben wird: *Sprache im technischen Zeitalter*. In der ersten Ausgabe aus dem Jahr 1961 definiert Höllerer, welche Aufgabe eine Literaturwissenschaft erfüllen müsse, die auf der Höhe der Gegenwart und eben *im* technischen Zeitalter agiert: Sie solle keine Furcht vor der Technik haben oder sie als ihren natürlichen Feind ansehen; und sie solle sich den Ideologien der Technik nicht willfährig unterwerfen.[2] Beide Aspekte halte ich auch heute noch für robuste Leitlinien, an denen entlang sich die Frage diskutieren lässt: Wie ist es um die Sprache in jenem technischen Zeitalter bestellt, das wir *heute* bewohnen? Jenes Zeitalter nämlich, das durch den Aufstieg Künstlicher Intelligenz und Maschinellen Lernens geprägt ist – und noch viel mehr sein wird.

1961 war der Umfang, den Sprachtechnologien einmal annehmen würden, kaum absehbar – angestoßen war die Entwicklung dahin aber durchaus. Höllerer sah die allerersten Anfänge dessen, was man heute *natural language processing* nennt, und war hellsichtig genug, um auch der technischen Sprachverarbeitung Aufmerksamkeit zu wünschen. Unmittelbar auf sein programmatisches Vorwort folgte im ersten Heft der *Sprache im technischen Zeitalter* daher ein Aufsatz des österreichischen Computerpioniers Heinz Zemanek.[3] Sein Artikel wandte Höllerers zwei Leitlinien – keine Furcht vor der Technik zu haben und nicht ihren Ideologie aufzusitzen – ganz konkret auf die Sprachtechnologie der Übersetzung an: Damit Sprache *überhaupt* verarbeitet werden kann, muss man erst einmal annehmen, dass sie Regeln unterworfen ist, die einem Computer zumindest näherungsweise beigebracht werden können; wäre Sprache nur ein großes Mysterium, könnte man den Versuch gleich ganz bleiben lassen, den zu unternehmen bisher doch einige vorzeigbare Ergebnisse gezeitigt hat. Im selben Atemzug aber warnt Zemanek vor der Illusion *völliger* Automatisierbarkeit: Sprache ist komplex, situationsgebunden, oft mehrdeutig und immer Sache menschlicher Interpretation. Allein ihre Syntax zu automatisieren, was ein selbst bis heute nicht restlos gelöstes Problem ist, heißt noch lange nicht, auch ihre Bedeutung zu erfassen. Alle automatische

Sprachtechnologie ist so ein immer prekärer Balanceakt – zwischen der notwendigen Fiktion, Sprache sei automatisierbar, und der ständigen Mahnung, sie sei es in Wirklichkeit eben nicht.

Zemanek verdeutlicht das Problem an einem englischsprachigen Beispiel des Philosophen Yehoshua Bar-Hillel: *The box was in the pen*.[4] Da „pen" mindestens zwei Bedeutungen hat, kommen auch zwei Übersetzungen in Frage: *Die Kiste war im Laufstall*. Oder: *Die Kiste war im Stift*. Uns ist unmittelbar klar, dass einer dieser Sätze offensichtlich absurd ist, weil wir um gewöhnliche Größenverhältnisse wissen und darum, dass Stifte normalerweise kleiner sind als Kisten. Eine Software aber weiß das nicht. Sprachgebrauch setzt Intelligenz als umgehendes Weltverständnis voraus, das derartige Synonymien zu enträtseln vermag. Solange beides nicht zusammen gelöst ist, so Zemanek, bleibe eine hochqualitative, das heißt menschenähnlich gute Übersetzung ein „utopisches Ziel".[5] Schlägt man nun alle Vorsicht in den Wind und gibt sich der Verführung hin, die von lediglich hinreichend guten Ergebnissen ausgeht, läuft man Gefahr, aus der Fiktion automatisierbarer Sprache eine Ideologie werden zu lassen. Dann passiere es, dass „der ästhetische Eindruck des Resultats alle Zweifel einschläfert, [...] gleichzeitig aber schwierige Entscheidungen nicht anzeigt, sondern einfach trifft".[6] Das Ergebnis erscheint sinnvoll, ist es aber in Wirklichkeit nicht; die Macht über Entscheidungen wird dann im falschen Vertrauen auf die Kompetenz der Maschine kritiklos an sie übergeben.

Diese Mahnung gilt noch heute. Im Gegensatz zu 1961 ist Sprachtechnologie ungleich ausgereifter. Obwohl sie noch immer nicht intelligent sind – sie verstehen nicht tatsächlich, was sie tun –, vermitteln die allerneuesten KI-Modelle mehr denn je den Eindruck von Intelligenz. Dieses Erscheinen-Als betrifft die Art und Weise, wie Beobachterinnen Ausgaben interpretieren, wie sie ihnen gegenübertreten und von ihnen auf das dahinterstehende System zurückschließen. Und das ist nicht allein eine ideologische, sondern zugleich eine eminent ästhetische Frage – und damit sind wir wieder bei Höllerer, der der Literatur und ihrer Analyse die Funktion zuschrieb, Sprache als ästhetisches Konstrukt gerade in ihrer Interaktion mit Technik zu reflektieren. Denn die Sprache im technischen Zeitalter ist keine rein technische Angelegenheit. Sie ist ein soziales Phänomen, ein Bündel aus Sinn und Konnotationen, das kulturelle Praktiken und nicht zuletzt Rezeptionstraditionen bestimmt.

Hier möchte ich heute ansetzen und fragen, welche Auswirkungen die gegenwärtigen rapiden Fortschritte in der KI-Forschung auf den Umgang mit Sprache, genauer, auf unsere *Leseerwartungen* haben. Anders als Höllerer und Zemanek stehen wir heute bereits wirklich an der Schwelle, von Texten umgeben zu sein, die völlig künstlich hergestellt wurden – während wir zugleich bei unserem eigenen Schreiben immer weiter mit unseren Sprachtechnologien zusammenwachsen, so dass auch unsere Textproduktion mehr und mehr von

Assistenzsystemen unterstützt, erweitert und teilweise übernommen wird. Daher will ich – durchaus spekulativ, aber immer mit Blick auf den Stand der Technik – zwei Fragen diskutieren: Was geschieht, erstens, wenn wir neben *natürlichen* nun auch *artifiziellen Texten* ausgesetzt sind? Wie lesen wir einen Text, von dem wir nicht mehr sicher sein können, dass er *nicht* von einer KI geschrieben wurde? Und zweitens: In welche Richtung könnte diese Entwicklung gehen, wenn schließlich irgendwann die Unterscheidung zwischen natürlich und artifiziell selbst wieder hinfällig wird, so dass wir nach ihr gar nicht mehr fragen und stattdessen *post*artifizielle Texte lesen?

Die Standarderwartung an unbekannte Texte

Die Differenz zwischen artifiziellen und künstlichen Texten stammt nicht von mir. Etwa zur selben Zeit, als Höllerer in Berlin und Zemanek in Wien über die kulturellen und praktischen Seiten technischer Sprachverarbeitung nachdachten, führte in Stuttgart der Philosoph und Physiker Max Bense eine ganz ähnliche Unterscheidung ein. Im Aufsatz „Über natürliche und künstliche Poesie" aus dem Jahr 1962 macht er sich Gedanken speziell darüber, wie sich mit Computern hergestellte *Literatur* von der herkömmlichen, menschengeschriebenen unterscheidet. Bense konzentriert sich dabei auf die „Art der Entstehung"[7] eines Textes: Was geschieht auf Seiten von Autorinnen, wenn sie einen poetischen Text schreiben?

Für Bense ist das im Fall *natürlicher Poesie* klar: Damit ein Text Bedeutung tragen kann, müsse ein „personales poetisches Bewusstsein" ihn auch mit der Welt verknüpfen. Denn für Bense ist Sprache zu einem großen Teil durch „Ichrelation" und „Weltaspekt" bestimmt: Das Sprechen geht von einer Person aus, sie spricht sich also immer mit, ganz gleich, was sie sagt; und zugleich bezieht sie sich in ihrem Sprechen immer auf die Welt. Zusammengenommen: Das poetische Bewusstsein setze „Seiendes in Zeichen", also Welt in Text, und garantiere am Ende, dass das eine mit dem anderen in Verbindung steht.[8] Ohne dieses Bewusstsein wären die Zeichen und die Beziehung zwischen ihnen sinnlos. Damit wird bereits die Verbindung zur technischen Sprachverarbeitung sichtbar: Denn wie Zemanek anhand seines Übersetzungsbeispiels vorgeführt hat, trägt auch solcher Text keine Bedeutung – das Wort „pen" oder das Wort „box" sind dem System nur leere Symbole, operative Variablen, die auch völlig anders heißen könnten.

Genau diesen Fall beschreibt Benses zweite Kategorie, die *künstliche Poesie*. Damit meint er literarische Texte, die über die Ausführung einer Regel, eines Algorithmus hervorgebracht werden. Hier steht kein Bewusstsein mehr am Anfang, es gibt weder Bezug auf ein Ich noch auf die Welt. Stattdessen haben solche Texte einen rein „materialen" Ursprung – sie sind allein über

mathematische Eigenschaften wie Häufigkeit, Verteilung, Entropiegrad etc. zu beschreiben. Das Thema eines künstlich generierten Textes sei dann, selbst wenn seine Wörter zufällig für uns Dinge in der Welt bezeichnen sollten, nicht eigentlich mehr die Welt – sondern nur noch dieser Text selbst, als messbarer, berechenbarer, modellierbarer Gegenstand einer exakten Textwissenschaft. Entstammt die natürliche Poesie dem Reich der Verständigung, ist die künstliche eine Sache der Mathematik – sie will und kann nicht kommunizieren und spricht nicht mehr von einer geteilten menschlichen Welt.

Benses Stoßrichtung ist aber gerade nicht die Rettung einer romantischen Idee von unerklärlicher menschlicher Schaffenskraft. Im Gegenteil, „der Autor als Genie" ist bei ihm mausetot. Stattdessen will Bense wissen, was man von einem Text ästhetisch noch aussagen kann, wenn man von traditionellen Kategorien wie Bedeutung, Konnotation oder Referenz absieht. Die Antwort, die er vorstellt, ist seine Informationsästhetik: Sie berücksichtigt, streng positivistisch und in der Tradition von Claude Shannons und Warren Weavers Kommunikationstheorie, nur noch allein statistisch messbare Texteigenschaften. Künstliche Poesie ist dann, eben weil sie bedeutungslos ist, auch „reine Poesie" – sie kommt völlig ohne die Unterstellung eines Bewusstseins aus, ist selbstständiges ästhetisches Objekt, das immanent untersucht werden kann. Wie bei Zemanek wäre die Unterstellung, das textproduzierende System hätte Intelligenz, ein Fehler – und hier zudem gar ein ästhetischer Fauxpas.

Bense war selbst an mehreren Experimenten mit künstlicher Poesie beteiligt. Das bekannteste unter ihnen waren sicher die „Stochastischen Texte", die sein Schüler Theo Lutz 1959 auf dem Großrechner Zuse Z22 an der Universität Stuttgart angefertigt hatte und die als erstes Experiment mit digitaler Literatur im deutschen Sprachraum gelten.[9] „Stochastisch" sind diese Texte, weil sie nach einem Zufallsprinzip aus einer Sammlung von Vokabeln ausgewählt und zusammengesetzt wurden – dass diese Vokabeln aus Kafkas *Schloß* stammen, macht die Ausgabe allerdings kaum bedeutungsvoller. Sie enthalten Sätze wie: „NICHT JEDES SCHLOSS IST ALT. NICHT JEDER TAG IST ALT." Oder auch: „NICHT JEDER TURM IST GROSS ODER NICHT JEDER BLICK IST FREI." In Benses Zeitschrift *augenblick* druckte Lutz einige davon in Auswahl ab.[10]

Die „Stochastischen Texte" waren eines der ersten Beispiele für *natural language processing* in Deutschland und bewiesen, dass Computer nicht nur mathematische Operationen, sondern auch Sprache verarbeiten können. Überdies waren sie auch künstliche Poesie im Sinne Benses: So viele Variationen das Programm auch ausspuckt, kein Ich scheint sich hier auszusprechen, kein Bewusstsein dahinter und für die Bedeutung der Wörter einzustehen, die nur nach gewichteten Zufallsoperationen verkettet wurden. Dass der Computer selbst tatsächlich *Autor* dieses Textes sein könnte, erschien Lutz wie Bense jedenfalls absurd.[11] Aber beide wussten ja auch, wie er hergestellt worden war. Ob man ihm seinen künstlichen Ursprung selbst ansieht, er sich im

„ästhetischen Eindruck“ enthüllt, ist dagegen weniger klar; die Leser der Literaturzeitschrift *augenblick* kamen jedenfalls nicht in die Verlegenheit, diese Frage zu stellen. Ein begleitender Essay klärte sie in allen Details über seine Machart auf.

Als Lutz aber im Jahr darauf ein zweites Gedicht nach diesem Muster generierte (es trug den Titel „und kein engel ist schön“ – statt Kafka war nun Weihnachtsvokabular eingeflossen) und es in der Dezembernummer der von ihm geleiteten Jugendzeitschrift *Ja und Nein* veröffentlichte, fehlte jede Erklärung.[12] Allein der Autorinnenname „electronus“ hätte noch den Schluss darauf erlaubt, wer hinter diesem Text steckt; ansonsten stand das Gedicht kommentarlos auf Seite 3 unter den vermischten Meldungen, platziert wie andere Gedichte auch. Erst in der nächsten Nummer wurde aufgelöst, was gar nicht als Rätsel ersichtlich gewesen war: dass ein Computer den Text geschrieben hatte.

Offensichtlich hatte Lutz hier seinen Spaß: Zusammen mit einem Foto der Zuse Z22 und einem zweiten Gedicht „in der Handschrift des Dichters“ (nämlich als Fernschreiberausdruck) veröffentlichte er eine Reihe von Leserbriefen. Ihre Schreiber waren sich – ohne zu wissen, wie es entstanden war – recht uneins in der Bewertung des Gedichts: „Sie sollten sich vielleicht doch überlegen, ob Sie solchen modernen Dichterlingen die Spalten Ihres Blattes öffnen!“, beschwerte sich einer, ein anderer zeigte sich im Gegenteil avantgardistisch beeindruckt: „Endlich mal was Modernes!“ Und eine dritte Leserin war zumindest aufgeschlossen: „Ehrlich gesagt: Verstehen tu ich’s ja nicht, Ihr Weihnachtsgedicht. Aber irgendwie gefällt es mir trotzdem. Man hat den Eindruck, daß etwas dahintersteckt.“ Einzig eine aufmerksame und offensichtlich informierte Leserin erkannte, dass es sich um Computerdichtung handelte, und beglückwünschte die Zeitschrift zu ihrer kühnen Veröffentlichung.[13]

Was sich in diesen Reaktionen ausspricht, ist, was ich die *Standarderwartung an unbekannte Texte* nennen möchte. Das electronus-Gedicht war tatsächlich künstliche Poesie im Sinne Benses, ein artifizieller Text ohne Bedeutung und dahinterstehendes Bewusstsein. Doch weil sie diese Produktionsbedingungen nicht kannten, hielten seine Leser ihn für einen natürlichen Text und nahmen an, er sei von einem Menschen mit dem Ziel geschrieben worden, Bedeutung zu kommunizieren. Die Standarderwartung an unbekannte Texte ist eben diese: dass sie von einem Menschen stammen, der etwas sagen will[14]. Um einen Text als *artifiziell* zu erkennen, bedarf es immer noch zusätzlicher Information – gerade bei künstlicher Poesie. Lutz hatte sein Publikum in der Tat, wie ein Leserbriefschreiber unterstellte, „an der Nase herumgeführt“ – aber nicht, weil ein moderner Dichterling hässliche, aber natürliche Lyrik verfasst, sondern weil ein Computer einen bedeutungslosen, weil artifiziellen Text geschrieben hatte.

Einen artifiziellen *als* natürlichen Text auszugeben, war nicht bloß die Premiere eines heute ziemlich abgewetzten Scherzes, den sich 1960 ein Informatiker in einer Esslinger Jugendzeitschrift erlaubte. Im Gegenteil ist dieses An-der-Nase-Herumführen das Urprinzip Künstlicher Intelligenz – und zugleich das, was sie mit Sprachtechnologien verbindet. Der Wegbereiter der Informatik, Alan Turing, hatte zehn Jahre zuvor in einem Artikel mit dem Titel „Computing Machinery and Intelligence" darüber sinniert, ob Computer jemals denken, jemals intelligent sein könnten.[15] Turing lehnte diese Frage als falsch gestellt ab – Intelligenz im Sinne einer intrinsischen Qualität könne man nicht verlässlich messen. In gut behavioristischer Manier ersetzte er die Frage daher durch eine andere: Wenn wir davon ausgehen, dass Intelligenz eine Eigenschaft von Menschen ist, dann müsste man nur herausfinden, wann ein Mensch den Computer selbst für einen Menschen und also für intelligent hielte.

Der Versuchsaufbau ist bekannt: Eine Person kommuniziert über einen Fernschreiber mit einer abwesenden anderen Person und soll herausfinden, ob es sich dabei um einen Menschen oder eine Maschine handelt.[16] Über den Fernschreiber kann sich die Versuchsperson wie in einem Chat mit der anderen Seite unterhalten, Fragen stellen und Aufklärung fordern. Dabei geht es nicht darum, dass die Antworten auf diese Fragen *wahr* sind, sondern dass sie *menschlich* klingen; Lügen und Bluffen sind explizit erlaubt. Der Turing-Test ist bis heute umstritten und gilt überdies als nicht bestanden – wirklich, völlig und dauerhaft hat noch keine KI genügend Menschen davon überzeugt, dass sie ein Mensch ist. Will man aber die Erwartungshaltung an artifizielle Texte untersuchen, ist Turings Test immer noch ein hilfreicher Ausgangspunkt, setzt er doch Intelligenz mit schriftlicher Kommunikation gleich,[17] deren Ziel darin besteht, die für die Maschine bedeutungslosen Zeichen als für Menschen bedeutsame zu verkaufen. Zugespitzt gesagt: *Das Wesen von KI ist es, artifizielle als natürliche Texte auszugeben*. Diesen Versuch überhaupt zu unternehmen lohnt sich aber nur, *weil* die Standarderwartung an unbekannte Texte eben die menschlicher Urheberschaft ist.

Künstliche Intelligenz basiert also von Anfang an auf dem Prinzip der *Täuschung* – und sie muss es: Weil Intelligenz nicht als objektive Eigenschaft des Systems, sondern nur als subjektiver Eindruck für eine Beobachterin definiert wurde – und also nur durch das ästhetische Erscheinen-als-Mensch –, ist der Turing-Test ohne Täuschung gar nicht denkbar. Der Medienwissenschaftler Simone Natale schreibt aus diesem Grund: „Täuschung ist so zentral für die Funktionsweise von KIs wie die Schaltkreise, Software und Daten, die sie funktionieren lassen." Ziel der KI-Forschung sei „nicht die Herstellung intelligenter Wesen, sondern von Technologien, die Menschen *als* intelligent wahrnehmen".[18]

Ich möchte diese Haltung *starke Täuschung* nennen. Man sieht sofort, dass sie Probleme mit sich bringt. Denn erst einmal bedeutet sie, dass es für KI-Systeme am besten ist, wenn eine Wissensasymmetrie zwischen den menschlichen Usern und dem System besteht: Je mehr es über sie weiß und je weniger sie über es, desto überzeugender lässt sich die Täuschung aufrechterhalten. Die politischen und ethischen Probleme sind offensichtlich: Starke Täuschung ist, im Sinne Zemaneks, eine technische Ideologie. Sie lässt sich als notwendig für die Funktion des Systems begründen, belohnt aber eine Intransparenz, die die Benutzerinnen über ihr Getäuschtwerden im Dunkeln lässt und sie notwendig entmündigt.

Zweitens kann man für unser Thema aber fragen, ob sich unter diesen Voraussetzungen die Erwartungshaltung an KI-generierte Texte auf lange Sicht je verändern und ob ihre Veränderung beschrieben werden kann. Ich glaube nicht. Der Turing-Test besteht nämlich darauf, dass artifizielle und natürliche Texte weiterhin fein säuberlich voneinander getrennt bleiben, damit die einen als die anderen gelten können. Wird mit einem Mal enthüllt, ein natürlicher Text sei in Wirklichkeit ein artifizieller gewesen, fühlt sich das Publikum betrogen. Und das nicht zu Unrecht: die Täuschung erweist sich als Enttäuschung.

Wir wissen nicht, wie Theo Lutz' Leserinnen auf die Enthüllung der Computerautorschaft reagiert haben, aber man kann es sich denken, betrachtet man gegenwärtige Fälle, in denen sich „der Künstler" nachträglich als Maschine entpuppte. Zuletzt geschah das im Juni 2022 bei einem eher peripheren Kunstpreis: Als ein Teilnehmer zugab, sein Bild gar nicht selbst gemalt, sondern es durch die Text-zu-Bild-KI Dall·E 2 generiert zu haben, hagelte es empörte Reaktionen und er wurde des Betrugs bezichtigt. Denn obwohl es sich um einen Kunstpreis für *digitale* Kunst handelte, waren damit doch nur die Werkzeuge gemeint; die Kunst selbst sollte immer noch von Menschen stammen.[19] Ein ähnlicher Fall ereignete sich 2016 in Japan, wo es ein KI-generierter Text immerhin in die zweite Runde eines Literaturpreises schaffte. Zwar gewann er nicht, aber er konnte die Jury doch davon überzeugen, genügend hohe „literarische Qualität" zu besitzen, um einen zweiten Blick wert zu sein.[20] Es gibt noch weitere solcher Beispiele – zwar werden sie in der Berichterstattung meist übertrieben, aber als enttäuschte *Erwartungen* legen diese Reaktionen frei, was denn eigentlich erwartet worden ist: nämlich natürliche, nicht artifizielle Texte.

Diese Erwartungen bestätigen sich auch negativ: Die Enttäuschung besteht dann darin, dass ein angeblich computergeneriertes Werk in Wirklichkeit das eines Menschen war. So erfreute sich um 2011 – in der ersten Hochphase von Twitterbots zum Zweck digitaler Literatur – der Account @horse_ebooks größter Beliebtheit. Er, schien es, war ursprünglich als Spam-Bot zur Verbreitung von Werbung programmiert worden. Durch irgendeinen Fehler

spuckte er nun aber absurde und dadurch recht poetische Unsinnstexte aus: Ein literarischer Bot wider Willen, offensichtlich ohne intendierte Bedeutung, aber gerade darum so faszinierend zu beobachten, wenn er dann doch etwas für menschliche Leser Sinnvolles ausgab – etwa solche aphoristischen Perlen wie „everything is happening so much“[21] oder „unfortunately, as you probably already know, people“.[22] Als sich aber herausstellte, dass die Tweets nicht generiert, sondern von einer Künstlergruppe *hand*geschrieben worden waren – die die Ästhetik der kaputten Textmaschine nur simulierten –, herrschte allgemeine Enttäuschung: Die schönen Zufallssentenzen schienen auf irgendeine Weise entwertet. Das Wissen, dass dahinter – wie der *Independent* betroffen in Großbuchstaben schrieb – „A REAL HUMAN BEING“ stand, machte die Hoffnung auf die zufällige Bedeutung eigentlich bedeutungsloser künstlicher Poesie zunichte.[23]

Die Krise der Standarderwartung

Diese Beispiele scheinen zunächst nahezulegen, dass sich die Erwartungshaltung an unbekannte Texte seit Lutz' Zeiten *nicht* geändert hat: Wir vermuten menschliche Herkunft und Kommunikationswillen, weshalb Täuschung überhaupt erst eine sinnvolle Strategie im Design von KI-Systemen sein kann.

Ich werde jetzt etwas sagen, was sich wie ein Widerspruch anhört. Ich glaube nämlich, dass die Erwartungshaltung dennoch schon im Begriff ist, sich zu verschieben. Weil sich auf der einen Seite die Zahl computergenerierter Texte stetig erhöht und auf der anderen Seite wir selbst immer mehr mit, über und durch Sprachtechnologien schreiben, sind wir auf dem Weg zu einer neuen Erwartung, oder besser gesagt: einem neuen Zweifel. Je mehr künstliche Texte es gibt, desto eher muss sich der Standard auflösen und sich auch die Frage ihrer Herkunft aufdrängen; selbst da, wo wir uns normalerweise überhaupt keine Gedanken darum machen.

Dieser scheinbare Widerspruch erklärt sich damit, dass die Textart, die ich bisher betrachtet habe, eine besondere ist: Es sind *literarische* Texte – Texte, die in unserer Kulturtradition als außergewöhnlich markiert sind. Dazu gehört, dass sie bis ins Kleinste durchgearbeitet und „gewollt“ erscheinen. Trotz aller Versuche der literarischen Avantgarden, Texte ohne Stimme zu schaffen, und trotz mehr als sechzig Jahren, in denen die Literaturwissenschaft den „Tod des Autors“ verkündet hat, heißt diese Gewolltheit aber: die standardmäßige Erwartungshaltung an literarische Texte ist bis heute, dass sie Autorinnen im Sinne kommunikationsgewillter Menschen haben.[24] Wir wissen zwar, dass es Ausnahmen gibt – aber dennoch gehen wir heute genauso wie seinerzeit die Leser von Lutz' electronus-Gedicht zunächst von menschlicher Autorschaft aus, bis wir eines anderen belehrt werden. Was das nun für

literarisches Schreiben in Zeiten von KI bedeutet, darauf komme ich gleich zurück.

Zunächst aber lohnt es sich, einen Blick auf die entgegengesetzte Seite des Spektrums zu werfen – auf jene eher unmarkierten Texte, die im Hintergrund bleiben, die nur funktional sind und die sich als Texte gerade nicht aufdrängen. Für sie ist der Turing-Test schlicht eine falsche Beschreibung der Wirklichkeit. Er geht von starker Täuschung als einziger Form von Mensch-Maschine-Interaktion und von der Differenz artifiziell/natürlich als der einzigen möglichen Unterscheidung von Textarten aus. Aber vor allem im Umgang mit Interfaces, mit den idealerweise unsichtbaren Schnittstellen, an denen wir mit Maschinen kommunizieren, gibt es bereits heute Zwischenstufen – es ist nämlich sehr wohl möglich, darum zu wissen, *dass* etwas von einer nichtintelligenten Maschine produziert wurde, und es gleichzeitig so zu behandeln, als *wäre* es bewusste Kommunikation. In der Tat ist das ganz normal.

Simone Natale hat dafür den Begriff *banale Täuschung* vorgeschlagen.[25] Anders als bei dem, was ich als starke Täuschung bezeichnet habe, sind sich die Userinnen hier bewusst, dass sie getäuscht werden. Wir verstehen, dass Siri kein Mensch ist und kein Inneres besitzt, aber die reibungslose Kommunikation mit ihr funktioniert nur dann, wenn wir sie zumindest ansatzweise so behandeln. Das Wissen darum ist kein Widerspruch, der plötzlich und unerwartet eine Illusion zerstört, wie im Beispiel der Wettbewerbe, an denen eine KI teilnimmt. Stattdessen wird es zur Bedingung von Funktionalität: Anders macht Siri eben nicht, was ich möchte.

Ähnlich verhält es sich mit Texten. Das beginnt bereits mit dem Dialogfeld auf dem Computermonitor. Die Frage: „Möchten Sie Ihre Änderungen speichern?“, lässt schließlich eine Interaktion zu, die ganz basal der mit einem Menschen ähnelt – die Antwort „Ja“ hat eine andere Folge als die Antwort „Nein“ und beide liegen in einem Sinnkontinuum, das natürliche Sprache mit Datenverarbeitung verbindet –, ohne dass man dahinter bereits Intelligenz vermutete.[26] Damit wäre die Erwartungshaltung an unmarkierte Texte bereits wieder heruntergestuft: Zwar verhalten wir uns immer noch so, als erwarteten wir menschliche Bedeutung und ein bewusstes Kommunikationsinteresse, wir klammern aber die Überzeugung, dass dahinter *wirklich* ein Bewusstsein stecken muss, ein.

Dennoch verläuft diese Einklammerung nicht immer reibungslos. Banale Täuschung ist ein Als-ob, das uns die Fähigkeit abverlangt, eine Überzeugung und ihr Gegenteil gleichzeitig zu vertreten. Aus dieser leicht schizophrenen Position geht schnell jener Zweifel hervor, von dem ich vorhin sprach: Je besser künstliche Texte werden, je mehr der ästhetische Eindruck, den sie machen, doch wieder so etwas wie Bewusstsein nahelegt, desto schwerer wird es, sich in der Schwebe wohlzufühlen, die die banale Täuschung einzunehmen voraussetzt. Dazu muss man gar nicht elaborierte Deepfakes heranziehen; das lässt sich sogar bei den allerunauffälligsten Sprachtechnologien beobachten.

Zu ihnen, die wir heute unentwegt verwenden, gehören jene kleinen Helfer, die unsere Schreibaufgaben begleiten und die wir kaum als intelligent bezeichnen würden: Die Rechtschreibprüfung im Wordprozessor unterkringelt die peinlichsten Fehler rot, bevor ein baldiges „Wiehersehen" gewünscht wird; die Eingabevervollständigung im Handy schreibt Wörter sogar ohne nachzufragen zu Ende, was gelegentlich besonders unintelligent wirkt. Aber schon bei der Eingabevervollständigung kann man sehen, wie sich die Übergänge zwischen eindeutig artifiziellen Texten und weniger klar zuzuordnenden Formen verwischen. Sie ist eine eher ältere Technologie und traditionell beruht sie auf dem einfachen Vergleich zwischen einer Eingabe und Elementen in einem Wörterbuch mit nach Wahrscheinlichkeit gewichteten Einträgen; die Buchstaben H, A und L werden so eher zu *Hallo* als zu *Halisterese* vervollständigt.

Seit einigen Jahren ist diese Technik immer öfter aber nicht mehr als einfacher Regelsatz, sondern über komplexe KI-Systeme implementiert. So führte Gmail 2019 „Smart Compose" ein – eine Funktion, die beim Verfassen von E-Mails ganze Sätze zu Ende schreibt. Die wahrscheinlichsten Wortfolgen lernt es, indem es die Korrespondenz aller User analysiert – und da 1,8 Milliarden Menschen auf der Welt einen Gmail-Account besitzen, also etwas mehr als ein Fünftel der Menschheit, verfügt Google so über eine schiere Unmenge an Text, mit der es sein Modell trainieren kann. Aus dieser Technik ergeben sich geradezu unheimliche Effekte, die die Fiktion der banalen Täuschung in Zweifel zu stürzen vermögen. Das illustriert ein Erlebnis, von dem der Autor John Seabrook im *New Yorker* berichtete.

In einer E-Mail an seinen Sohn wollte Seabrook einen Satz mit „I am pleased that", also: „ich freue mich, dass", beginnen. Als er beim „p" angekommen war, schlug ihm Smart Compose statt „pleased" die Wortfolge „proud of you" vor: „Ich bin stolz auf dich." Seabrook fühlte sich von der Maschine ertappt: „Als ich vor meiner Tastatur saß, spürte ich plötzlich etwas Unheimliches in meinem Nacken kribbeln. Es lag nicht daran, dass Smart Compose richtig erraten hatte, wohin meine Gedanken gingen – das hatte es nämlich nicht. Das Unheimliche bestand darin, dass die Maschine aufmerksamer und fürsorglicher war als ich."[27]

Seabrooks Scham war, objektiv betrachtet, ungerechtfertigt. Es war ja nicht die Maschine, die aufmerksam war – sie ist immer noch dumm, verarbeitet immer noch keine volle Bedeutung und kann nur vorschlagen, was sie, angesichts der ihr zur Verfügung stehenden Trainingsdaten, als das wahrscheinlichste nächste Wort betrachtet.[28] Was Seabrook hier vielmehr schildert, ist die Wirkung, die die neuesten Sprach-KIs, die an der Grenze zum Schein der Intelligenz operieren, auf die intimsten Aspekte unseres Schreibens haben. In seinem Fall bestand sie sogar darin, dass er sich für einen Moment fragte, ob er ein guter Vater war. Anders gesagt: Seabrook kämpfte mit der Schwierigkeit, die Fiktion banaler Täuschung aufrechtzuerhalten. Beginnt sie

zu bröckeln, schleichen sich Zweifel am Als-ob ein, und es wird ein Leichtes, auf die KI die Vorstellung einer Personalität zu projizieren, die sogar Scham hervorrufen kann: Ein unmarkierter, eigentlich artifizieller Text erscheint dann als natürlicher – oder bewegt sich zumindest in diese Richtung.

Das kann letztlich in die Überzeugung umschlagen, es hier *wirklich* mit einer Intelligenz zu tun zu haben – wie etwa im Fall des Google-Mitarbeiters Blake Lemoine, der im Sommer 2022 behauptete, die Sprach-KI, an der er mitarbeitete, verfüge über Bewusstsein. Das Chatsystem LaMDA, sagte Lemoine, besitze die Intelligenz eines Achtjährigen und habe ihn darum gebeten, als Person mit Rechten betrachtet zu werden. Google hielt eine solche Aussage offensichtlich für geschäftsschädigend und entließ den Mitarbeiter darauf.[29] Seine Reaktion scheint bislang eher die Ausnahme zu sein, auch wenn sie keineswegs selten ist, betrachtet man etwa die atemlose Berichterstattung um Microsofts Bing-Chatbot. Was der Fall aber zeigt, ist, dass sich das Gefühl des Unheimlichen, von dem Seabrook sprach, in Zukunft wohl noch verstärken wird: Werden artifizielle Texte *zu* gut – indem sie etwa aufmerksamer erscheinen als ihre Schreiberinnen – und wissen wir zudem, dass Computer solche Texte zu verfassen in der Lage sind, steht eine neue Standarderwartung gegenüber unbekannten Texten in Aussicht: der Zweifel an ihrer Herkunft. Statt selbstverständlich einen menschlichen Ursprung anzunehmen oder ihn erst einmal auszuklammern, wäre das Erste, was wir von einem Text wissen wollen: Wie wurde er gemacht?

Eine Flut artifizieller Texte

Diese Überlegung folgt lediglich einem Trend, der sich mit jeder Meldung über neue Sprach-KIs verstärkt. LaMDA ist bisher noch nicht für die Öffentlichkeit freigegeben, andere Modelle dagegen schon. Ihre Fähigkeiten hätte man vor fünf Jahren für unmöglich gehalten; heute sind sie beinahe schon normal geworden.

Jedes moderne, auf *machine learning* basierende KI-Modell ist nichts anderes als eine komplexe statistische Funktion, die auf der Grundlage gelernter Daten Vorhersagen über wahrscheinliche zukünftige Zustände trifft. Bei sogenannten Sprachmodellen bestehen sowohl die gelernten Daten wie die gemachten Vorhersagen aus Text. Solche Modelle haben eine ganze Reihe von Einsatzmöglichkeiten, von linguistischer Analyse über die automatische Übersetzung bis zur Generierung von Text (und, als *foundation model*, als Motor für noch umfangreichere Applikationen). Kann Googles Smart Compose aber nur ein paar Wörter oder Sätze vorschlagen, sind *große* Sprachmodelle (*large language models*) in der Lage, ganze Absätze und sogar zusammenhängende Texte zu schreiben: Und das nur, weil sie lernen, welche Wörter, Sätze und Absätze statistisch am wahrscheinlichsten aufeinander folgen.

Inzwischen kennt jeder GPT-3, das große Sprachmodell, das die Firma OpenAI vor knapp drei Jahren einführte. Damit wurde, und zwar äußerst öffentlichkeitswirksam, auf einen Schlag klar, dass Computer Texte generieren können, die sich beinahe lesen, als seien sie von einem Menschen geschrieben worden. Beinahe, weil auch GPT-3 nicht immer perfekt ist und Fehler macht; seine Ergebnisse waren aber beeindruckend genug, dass für eine Weile Artikel, in denen das Sprachmodell „Autorin" wird und über „sich" erzählt, ein eigenes journalistisches Genre bildeten, das Titel hervorbrachte wie: „Ein Roboter hat diesen ganzen Artikel geschrieben. Hast du schon Angst, Mensch?"[30]

Im November 2022 veröffentlichte OpenAI eine aktualisierte Version: ChatGPT ist noch einmal mächtiger als GPT-3 – es basiert auf einer leicht verbesserten, GPT-3.5 genannten Version.[31] Auf die Bitte, etwa einen Seminaressay über Jorge Luis Borges zu schreiben, legt die Maschine ohne zu zögern los; der Output ist ein durchaus akzeptabler Text, der zwar keine bedeutenden Einsichten enthält, aber als Einleitung in eine Bachelor-Hausarbeit durchgehen könnte. Weil das System zudem dialogbasiert ist – eine Art besserer Chatbot –, kann ich ChatGPT zudem darum bitten, den Text in einer bestimmten Richtung weiterzuschreiben oder Literaturangaben hinzuzufügen.

In Presseberichten wird stets schnell spekuliert, dass solche Sprachmodelle menschliche Autoren einmal ersetzen werden; aus verschiedenen Gründen bezweifle ich das.[32] Aber so weit muss es gar nicht kommen, damit sich unsere *Wahrnehmung* von Text grundsätzlich ändert. Es ist bereits heute Realität, dass diese Technologien eine Assistenzfunktion übernehmen – nicht die ganze Schreibarbeit erledigen, aber dabei helfen, sehr viel mehr Text sehr viel schneller und durch immer weniger Menschen zu produzieren. Bestimmte Textarbeit wird so zumindest *teil*automatisiert.

Der Clou an den GPT-Modellen ist zudem nicht nur ihre technische Mächtigkeit, sondern auch ihre ökonomische Verwertbarkeit. Sie sind per Lizenznahme verfügbar und Firmen können OpenAI dafür bezahlen, das Sprachmodell in andere Software einzubauen. Damit kann Textgenerierung auf spezielle Aufgaben maßgeschneidert und als Produkt verkauft werden. So gibt es bereits jetzt eine ausgefeilte Programmierassistenz mit GitHubs Copilot.[33] Es genügt, in wenigen Worten zu skizzieren, was das gewünschte Programm tun soll, und schon schreibt die KI den entsprechenden Code dazu. Das klappt nicht immer, aber doch hinreichend oft, dass nun auch Programmierlaien ihre Ideen umsetzen, Firmen in Windeseile Prototypen aufsetzen oder einzelne Coderinnen lästige Detailarbeit an Copilot delegieren können.[34] Auch für gewöhnliches Schreiben existiert Ähnliches. Das beginnt wieder bei der eher flankierenden Unterstützung: So wie ich ChatGPT bitten kann, den Text anders fortzuführen, umzuformulieren oder auszuschmücken, hat nun beispielsweise die Notizsoftware Craft einen Assistenten, der mein Geschriebenes für mich überarbeiten kann, indem er es etwa erläutert, weiterschreibt

oder in Stichpunkten zusammenfasst.[35] Auch Microsoft hat GPT lizensiert – Bing erwähnte ich schon –, und Sie werden in der nächsten Version von Word solche Assistenzfunktionen finden, die weit über das hinausgehen, was wir bisher von Textverarbeitungsprogrammen gewohnt sind.

Jenseits von bloßer Assistenz aber sind große KI-Sprachmodelle vor allem dort profitabel einzusetzen, wo es um die Produktion des wahrscheinlichsten Outputs geht. Gerade Routinetextarbeit wird so automatisierbar. Am weitesten fortgeschritten ist das KI-Schreiben daher in einer Branche, die sehr viel Text produziert, ihn aber dabei vergleichsweise wenig wichtig nimmt und oft eher als Füllmasse betrachtet. So sind im vergangenen Jahr Dutzende von Sprach-KIs erschienen, die auf Marketing zugeschnitten sind: Man soll damit Ad Copy schreiben und schnell und in großen Mengen Content für Social Media, Produktseiten, Blogs und anderes produzieren können. Oft *soll* dieser Text gar nicht so genau gelesen werden, und da ist es von Vorteil, wenn das Ergebnis nicht überrascht, sondern so klingt wie andere Texte ähnlicher Machart auch.[36]

Desto schwieriger aber wird es für Leser, solche Texte als menschen- oder maschinengemacht einzuordnen. Wenn man bedenkt, wie viel des Geschriebenen, das uns täglich umgibt, Produkte solcher langweiligen Routineaufgaben sind, wird das Ausmaß klar, in dem wir generierte Texte zu erwarten haben. Je mehr davon zirkulieren werden – und das werden sie ohne Frage –, desto mehr wird sich die Standarderwartung an unbekannte Texte fort von der unmittelbaren Annahme menschlicher Autorschaft in Richtung jenes Zweifels verlagern: Hat das eine Maschine geschrieben?

Nun mag sich die Frage bei Marketingprosa weniger stellen – was aber ist mit dem Brief vom Anwalt, der automatisch erstellt sein könnte, obwohl es um meinen ganz persönlichen Fall geht? Was mit den Essays meiner Studierenden, die ich bewerten muss?[37] Was mit politischen Artikeln oder Fake News? Was mit der privaten, persönlichen, intimen E-Mail? Ist auch sie ein KI-Produkt – ganz oder in Teilen? Zumindest ein Grund für das Unbehagen, das mit diesen Vorstellungen einhergeht, lautet: Menschen stehen ein für das, was sie schreiben, selbst wenn sie sich irren oder in die Irre führen. Zur Standarderwartung an Texte gehört jene Unterstellung von Rezeptionsseite, die Jürgen Habermas den „Geltungsanspruch der Wahrhaftigkeit" genannt hat – jenes Grundmaß an Vertrauen, dass die Sprechenden (die Schreibenden) es ernst damit meinen.[38] Nur deshalb muss kritische Lektüre überhaupt gelernt werden: Man will Texten erst einmal glauben.

Das wird aber schwieriger, wenn große Sprachmodelle einerseits Texte herstellen können, die so scheinen, als hätte sie eine Autorin produziert und sanktioniert – und die andererseits kein zuverlässiges Wissen über die Welt besitzen, sondern nur die Wahrscheinlichkeit von Zeichenfolgen ausrechnen. Diese Gefahr wurde im November 2022 recht drastisch durch das Sprachmodell Galactica illustriert, das die Facebook-Mutterfirma Meta veröffentlicht hatte: Auf

Millionen Papers, Lehrbücher, Enzyklopädien und wissenschaftliche Websites trainiert, sollte Galactica dabei helfen, akademische Texte zu schreiben. Doch nach nur drei Tagen wurde es wieder offline genommen.[39] Das Modell nämlich verfasste brav und *en masse* Ausgaben, die autoritativ klangen, den Gepflogenheiten wissenschaftlicher Formatierung und Gesten folgten – aber völligen Unsinn enthielten, weil Galactica nur wahrscheinliche Sätze vollendete, statt auf Wissen zuzugreifen.[40] Es hatte lediglich die *Form* von Wissenschaftsprosa gelernt, ohne jede wissenschaftliche Einsicht, ohne jedes Einordnungswissen und ohne jede Fähigkeit, Rechenschaft über das Geschriebene abzulegen.

Das Letzte Modell und der Ouroboros

Die Standarderwartung an Texte wird sich also auf kurz oder lang verschieben – von der Überzeugung, ein Mensch stehe dahinter, zum Zweifel, ob es nicht doch eine Maschine sein könnte. Damit aber wird auch die Unterscheidung zwischen natürlichen und artifiziellen Texten zusehends hinfällig. Wir würden dann womöglich in eine Phase *post*artifizieller Texte übergehen.

Darunter verstehe ich zweierlei: Erstens die zunehmende Vermischung von natürlichen und artifiziellen Texten. Auch schon vor großen Sprachmodellen war kein Text wirklich ganz natürlich. Nicht nur kann die mathematische Verteilung von Zeichen auf einer Seite, wie Bense sie vorschwebte, auch von Hand erfolgen,[41] ebenso ist es eine Binsenweisheit der Medienwissenschaft, dass jedes Schreibzeug, vom Federkiel bis zum Wordprozessor, dem damit Produzierten seinen Stempel aufdrückt.[42] Anderseits ist aber auch kein Text je ganz artifiziell – das würde wirkliche Autonomie, wirkliche starke KI voraussetzen, die am Ende auch selbst glaubhaft entscheiden kann, einen Text zu veröffentlichen, und davon sind wir wirklich meilenweit entfernt.[43] Heute aber – dadurch, dass KI-Sprachtechnologien in die kleinsten Verästelungen unserer Schreibvorgänge eindringen – ist eine neue Qualität der Vermischung erreicht. In ungeahntem und nahezu unentwirrbarem Ausmaß integrieren wir artifiziellen *in* natürlichen Text.[44]

Denn angesichts großer Sprachmodelle ist es nicht ausgeschlossen, dass beide in einen sich gegenseitig bedingenden Kreisprozess eintreten, der sie vollends miteinander verstrickt. Da ein Sprachmodell lernt, indem man es auf große Mengen Text trainiert, bedeutet bislang mehr Text immer auch bessere Performance. Denkt man das zu Ende, wird ein zukünftiges, monumentales Sprachmodell einmal *mit aller verfügbaren Sprache überhaupt* trainiert worden sein; einer Studie zufolge wird dieser Fall bereits zwischen 2030 und 2050 eintreten.[45] Jeder mit diesem „Letzten Modell" generierte artifizielle Text wäre dann auf Grundlage allen natürlichen Texts entstanden; zugleich hätten sich so auch die natürlichen linguistischen Ressourcen für das Modell *nach* dem Letzten Modell erschöpft.

Es mag sich damit, wie der Philosoph Benjamin Bratton es nennt, ein »Ouroboros-Effekt« ergeben: Wie die Schlange, die sich selbst in den Schwanz beißt, werden alle folgenden Sprachmodelle für weiteren Performancegewinn dann anhand von Text lernen, der selbst bereits aus einem Sprachmodell stammt.[46] Damit käme, könnte man sagen, natürliche Sprache – und sei es nur als ohnehin nie real existente Fiktion – an ihr Ende. Denn der so gewonnene Sprachstandard würde umgekehrt wieder auf menschliche Sprechende einwirken – er hätte, eingebunden in all die auf dieser Technik aufbauenden Mechanismen des Schreibens, den Status einer bindenden *Norm*, der statistisch kaum zu entkommen wäre: Jede linguistische Innovation, jedes neue Wort oder jede grammatische Marotte, die in menschlicher Sprache regelmäßig neu auftaucht, hätte einen so geringen Anteil an den Trainingsdaten, dass sie in zukünftigen Modellen praktisch keine Spuren hinterließe.

Das ist natürlich ein bewusst überspitztes Szenario. Als Gedankenexperiment zeigt es aber, was postartifizieller Text im Extremfall sein könnte. Doch schon bevor es so weit ist, auf halbem Weg zum Eschaton der absoluten Vermischung und Auslöschung von natürlicher und artifizieller Sprache, ergäbe sich bereits eine neue Standarderwartung an unbekannten Text.

Das ist die zweite Bedeutung von „postartifiziell“ und die, auf die es mir hier ankommt. Nach der stillschweigenden Annahme menschlicher Autorschaft und dem Zweifel an der Herkunft von Geschriebenem, wäre sie die nächste Erwartungshaltung an unbekannte Texte. Denn der Zweifel über den Textursprung kann, wie jeder Zweifel, nicht von Dauer sein; Menschen haben ein Interesse daran, Normalität herzustellen, Komplexität und Unsicherheit auf ein erträgliches Maß zu reduzieren. Das kann etwa durch digitale Zer- ti fikate, Wasserzeichen oder andere Sicherheitstechniken geschehen, die das Vertrauen darein stärken sollen, dass der vorliegende Text nicht nur eleganter Unsinn ist,[47] oder aber durch das schlichte jurstische Verbot von verschleiert Generiertem. Sollten politische Regulierung und technische Eindämmung jedoch scheitern, ist es nicht unwahrscheinlich, dass die Erwartung selbst postartifiziell wird: Statt einen Menschen hinter einem Text zu vermuten oder von der Skepsis heimgesucht zu werden, ob es nicht doch eine Maschine ist, wird diese Frage schlicht uninteressant: Wir konzentrieren uns dann nur darauf, was der Text sagt, statt darauf, von wem er stammt. Postartifizielle Texte wären ihrer Herkunft gegenüber agnostisch; sie wären standardmäßig autorlos.[47]

Wenn sich die gewöhnliche Erwartung an uns unbekannte Texte also durchaus verschiebt; wenn sie in Zweifel gerät, um womöglich, in einer spekulativen Zukunft, in ihren Annahmen agnostisch zu werden – wieso dann die ostentative Aufregung um generierte Texte in Literaturwettbewerben? Wieso der Skandal, ein Roman sei mit Hilfe einer KI hervorgebracht worden, wo wir doch ohnehin immer schon in digitale Technik verstrickt sind? Wieso konnte es so scheinen, als wäre hier alles beim Alten, wo doch so vieles in Bewegung ist?

Ich glaube, weil Literatur langsamer ist. Und dies, weil sie – Bense zum Trotz – von allen Textsorten den maximalen Anspruch auf menschliche Herkunft erhebt.

Ich sagte bereits, dass es schon heute Texte gibt, über deren Ursprung man sich keine Gedanken macht; ein Straßenschild hat in diesem Sinne keine Autorin und im täglichen Leben ist für uns die Wettermeldung ebenfalls praktisch autorlos. Zwar nahmen wir bisher fraglos an, dass ein Mensch dahintersteht – aber unter postartifiziellen Lesebedingungen gar keine Annahme mehr darüber zu treffen, macht praktisch gesprochen kaum einen Unterschied. In Zukunft ist abzusehen, dass immer mehr Texte so rezipiert werden. Man könnte es auch so sagen: *die Zone unmarkierter Texte weitet sich aus*. Nicht nur Straßenschilder, sondern auch Blogeinträge, nicht nur Wettermeldungen, sondern auch Informationsbroschüren, die Diskussion von Netflix-Serien und sogar ganze Zeitungsartikel wären in Zukunft tendenziell unmarkiert, autorlos.

Literarische Texte dagegen sind heute immer noch maximal markiert. Wir lesen sie radikal anders als andere Textsorten – unter anderem gehen wir weiter davon aus, dass sie eine Autorin haben. Diese Markiertheit hat zur Folge, dass Kunst und Literatur neuerdings selbst zum Ziel der Tech-Branche geworden sind – nämlich als *benchmark*, die nach den ehemals rein menschlichen Domänen Schach und Go nun auch noch zu knacken wäre: Nichts würde die Leistungsfähigkeit von KI-Modellen besser beweisen als ein überzeugend generierter Roman. Letztlich beruht diese Hoffnung aber immer noch auf dem Paradigma der starken Täuschung; in der Tat gibt es derzeit eine ganze Flut literarischer und künstlerischer Turing-Tests zu beobachten: Können die Probanden das echte Bild vom künstlichen, das echte Gedicht vom KI-generierten unterscheiden? Diese Versuche kommen meist aus der Informatik, die, als Ingenieurwissenschaft, gern Metriken zur Hand hat, an denen sie den Erfolg ihrer Aufgaben messen kann. Das Problem ist nur, dass dort immer noch die starre Differenz von natürlicher Erwartung und künstlicher Wirklichkeit verglichen wird. Das scheint mir wenig zu besagen, wenn gerade diese Differenz selbst zur Disposition steht.[48] Interessanter ist daher die Frage, unter welchen Umständen sie hinfällig wird. Anders gefragt: Was müsste geschehen, damit auch *Literatur* postartifiziell würde?

Was ist postartifizielle Literatur? Und was nicht?

Ich will zum Abschluss diese Frage beantworten, indem ich noch einmal auf die Standardisierungstendenz zurückkomme, die vom Ouroboros-Effekt großer Sprachmodelle ausgeht. In ihnen findet, wie gesagt, eine Normalisierung statt; ihre Ausgaben sind gerade dann überzeugend, wenn sie Erwartbares, Gewöhnliches, eben statistisch Wahrscheinliches ausspucken sollen. Je „nor-

maler“ eine Schreibaufgabe ist, desto leichter lässt sie sich durch KI-Sprachtechnologien realisieren. Und so, wie es assistierende Marketing-KI für erwartbare Marketing-Prosa gibt, gibt es inzwischen auch assistierende Literatur-KI für mehr oder weniger erwartbare Literatur.

„Erwartbarkeit“ lässt sich statistisch als Wahrscheinlichkeitsverteilung über eine Menge an Elementen beschreiben – je rekurrenter diese sind, desto wahrscheinlicher, desto erwartbarer das Ergebnis. Genreliteratur ist durch die Rekurrenz bestimmter Elemente geradezu definiert, weshalb sie sich besonders für KI-Generierung eignet. So berichtete die Website *The Verge* über die Autorin Jennifer Lepp, die unter dem Pseudonym Leanne Leeds Fantasyromane schreibt – und zwar wie am Fließband, alle 49 Tage einen.[49] Ihr hilft dabei das Programm Sudowrite, ein GPT-3-basierter, spezifisch literarischer Schreibassistent, der Dialoge fortführt, Beschreibungen ergänzt, ganze Absätze umschreibt und sogar Feedback auf das menschliche Geschriebene gibt. Die Qualität dieser Textausgaben ist recht hoch, insofern ihr Inhalt eben erwartbar ist. Da sich alle Idiosynkrasien in der Masse an Trainingsdaten ausmitteln, tendieren sie zu einem konventionellen Umgang mit Sprache – sie werden selbst Ouroboros-Literatur. Im Moment sind KIs zur Generierung ganzer Romane noch nicht ausgereift, aber ich sehe nicht, wieso etwa Genreliteratur nicht sehr bald schon nahezu vollautomatisiert hergestellt werden könnte; dann wäre es möglich, dass aus den 49 Tagen nur 49 Minuten werden oder noch weniger. Wenn die Prognose erlaubt ist: Ich glaube, es wäre *diese* Art von Literatur, die am ehesten postartifiziell werden kann. Zwar würden Autorennamen nicht verschwinden, aber sie würden eher als *brands* fungieren, die für einen bestimmten, erprobten Stil stehen, statt tatsächlich menschliche Herkunft anzuzeigen. Die unmarkierte Zone würde sich auf bestimmte Bereiche der Literatur ausweiten – nicht auf alle, und sicher nicht auf alle erzählende, aber eben doch auf weit mehr als heute.

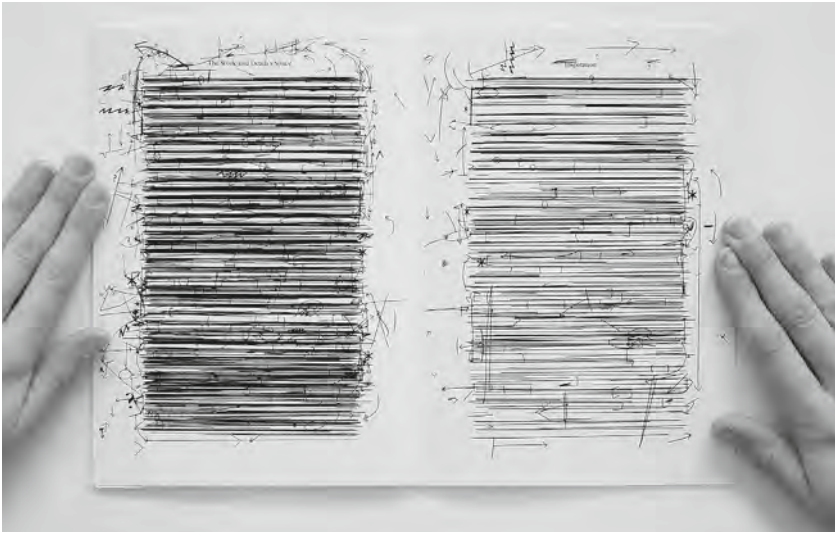
Die Oberfläche von Sudowrite; auf der rechten Seite eine vorgeschlagene Fortführung der hier begonnenen Geschichte.

Umgekehrt kann man fragen: Welche Literatur könnte sich dann dieser Ausweitung am ehesten entziehen? Hier sehe ich zwei, auf den ersten Blick gegensätzliche Antworten. Ist unmarkierte, postartifizelle Literatur solche, die natürliche und künstliche Poesie absolut *mischt*, wäre weiterhin markiertes Schreiben eines, das gerade die *Trennung* betont.

Auf der einen Seite könnte man sich also die Hervorhebung menschlicher Herkunft als besonderes Merkmal vorstellen. In der Tat sind schon jetzt *ex negativo* erste Phänomene zu beobachten, die eine solche Entwicklung andeuten. So laufen im Netz Künstlerinnen gegen bildgenerierende KI wie Dall·E 2 oder Stable Diffusion Sturm: Einerseits, weil sie in deren Ausgaben stilistische Merkmale ihrer eigenen Werke wiederzuerkennen meinen, die daher womöglich Teil des Trainingssets waren; hier stellen sich legitime Fragen nach Urheberrecht, Konsens und fairer Vergütung.[50] Gleichzeitig aber regt sich Widerstand gegen KI-generierte Kunst an sich, die – so die Sorge – menschliche Künstler überflüssig zu machen drohe. Auf Twitter hat sich daher das Hashtag *#supporthumanartists* als Kampfansage gegen generative Bild-KI etabliert.[51] Man kann sich Ähnliches für Literatur vorstellen, vielleicht sogar eine Zukunft, in der das Label *guaranteed human-made* als Auszeichnung gelten könnte. So wie man auf Etsy oder bei Manufactum Handgefertigtes kauft, wäre eine Art *boutique writing* denkbar, die ihre menschliche Herkunft als Qualitätsausweis und Verkaufsargument vor sich herträgt.

Will man aber nicht allein auf eine solche Versicherung vertrauen – die doch wieder dem Zweifel Raum geben mag, sie stimme nicht –, wäre es vor allem ein nicht-erwartbarer, unkonventioneller Gebrauch von Sprache, der ein Schreiben jenseits des Modells anzeigte. Jedes formale Experiment, jede textuelle Subversion stünde der Wahrscheinlichkeit großer Sprachmodelle, ihrem nivellierenden Ouroboros-Standard entgegen; linguistische Unvorhersehbarkeit wäre dann Ausweis menschlicher Herkunft. Im absoluten Extremfall derart, dass gleich das Zeichensystem, in dem Sprach-KIs operieren, gesprengt wird – so wie bei visueller und „asemischer" Literatur, etwa in den Werken Kristen Muellers: Sie verwendet gar keine Buchstaben mehr, sondern nur noch die Anmutung von Zeilen und Textblöcken.[52] Die „reine" Poesie, von der Max Bense träumte, käme paradoxerweise nicht aus der Maschine, die nun, in der postartifiziellen Vermischung, plausibel Bedeutung simuliert, sondern von Menschen, die genau das nicht mehr tun.

Auf der anderen Seite wären es aber womöglich die Nachfahren von Lutz und Bense, die dem Postartifiziellen entgingen, indem sie die Künstlichkeit ihrer Produkte weiterhin markieren. Das ist die digitale Literatur – Literatur, die dezidiert mit Hilfe von Computern hervorgebracht wird. Sie würde sich dem Postartifiziellen entziehen, indem sie die Verquickung zwischen natürlich und artifiziell bewusst betont. Viel eher als das konventionelle Schreiben hält digitale Literatur ihre Herkunft stets kritisch im Bewusstsein.[53] Ich habe anderswo



Kristen Mueller, *Partially Removing the Remove of Literature*.

sehr viel ausführlicher über sie geschrieben und gebe hier nur zwei Beispiele: Da wäre einmal Mattis Kuhns Buch *Selbstgespräche mit einer KI*, in dem er neben seinen literarischen Experimenten gleich den Quellcode zum Training des Sprachmodells und sogar dessen Datenbasis mitliefert. Zwar nicht völlig, aber immerhin ein wenig lassen sich hier die menschlichen und maschinellen Komponenten trennen, die gemeinsam den Text ergeben.

Umgekehrt kann auch eine bewusst inszenierte Mensch-Maschine-Kollaboration diesen analytischen Effekt haben: Etwa in David „Jhave" Johnstons *ReRites*: Ein Jahr lang ließ er jede Nacht ein Sprachmodell trainieren und edierte den Output am nächsten Morgen in einem Prozess, den er „carving"– also Meißeln oder Schnitzen nennt – von Hand: Der Punkt, an dem die Maschine ihren Text dem Menschen Jhave übergibt, wird genau markiert. Und indem dieser die bearbeiteten Ergebnisse eines jeden Monats in je einem Buch sammelt – so dass *ReRites* nun zwölf dicke Bände umfasst –, rahmt er diesen zwar kollaborativen, aber eben nicht absolut verschmolzenen Prozess zudem als Performance, die ebenfalls nicht konventionell literarisch ist.

Natürlich ist auch hier letztlich kein „Beweis" menschlichen Eingreifens erbracht. Aber vielleicht sind die Steine, die man dem allzu glatten Rezeptionsprozess noch in den Weg zu legen vermag, das Maximum an Widerstand gegen das Postartifizielle, das dann noch möglich sein wird – bevor der Unterschied zwischen natürlich und artifiziell wirklich ganz und gar verschwunden ist.

Mattis Kuhn, *Selbstgespräche mit einer KI*; links ein Gedichttext, in der Mitte der Code und rechts ein Ausschnitt aus dem Trainingsdatensatz.

Es sollte klar geworden sein, dass ich mich hier auf höchst spekulatives Terrain begeben habe. Ich will nicht sagen, dass erzählende oder im weitesten Sinne konventionelle Literatur von nun an verloren wäre und wir nur noch experimentelle oder ausdrücklich digitale Literatur betreiben sollten. Auch nicht, dass postartifizielle Texte notwendig schlecht sind – man wird sicher auch sie mit Freude lesen, sich über ihre Vorzüge austauschen und ihren interpretativen Überschuss enträtseln können. Mir ging es hier nur um die Analyse von Tendenzen, und da lohnt sich der Blick auf mögliche Extreme. Vor allem wollte ich, im Sinne Höllerers und Zemaneks, versuchen darüber nachzudenken, wie sich die Sprache in jenem technischen Zeitalter ändert, das wir heute bewohnen und das uns noch bevorsteht – sowohl, ohne Furcht vor der Technik zu haben, aber auch, ohne ihren Ideologien aufzusitzen. Eins scheint mir jedenfalls sicher zu sein: Mit der zunehmenden gesellschaftlichen Durchdringung von Sprachtechnologien, mit dem Siegeszug von KI-Modellen *werden* sich unsere Leseerwartungen verändern.

Daher zum Abschluss eine Frage an Sie: Wie reagieren Sie, wenn ich Ihnen nun sage, dass auch ich in diesem Text große Teile per KI habe schreiben lassen? Fühlen Sie sich übers Ohr gehauen? Dann sind sie noch fest in der Standarderwartung des zwanzigsten Jahrhunderts zu Hause. Ich kann Sie aber beruhigen: Dieser Text entstand völlig ohne KI-Assistenz. Oder doch nicht? Können Sie da ganz sicher sein? Wenn Sie sich nun unschlüssig sind, dann stehen Sie bereits an der Schwelle zur zweiten Erwartung, dem Zweifel an der Herkunft eines Textes im Zeitalter großer Sprachmodelle. Vielleicht ist es Ihnen aber auch egal – nicht völlig, aber doch genug, dass Sie sich vorzustellen vermögen, wie eine Welt *post*artifizieller Texte aussehen könnte.

1 Dieser Vortrag wurde am 8. Dezember 2022 in etwas kürzerer Form als 14. Walter-Höllerer-Vorlesung an der Technischen Universität Berlingehalten. Mein Dank gilt Eva Geulen und Hans-Christian von Herrmann. Für Diskussionen und Rückmeldungen bin ich Jules Pelta Feldmann und Sina Dell'Anno sehr verbunden.

2 Walter Höllerer: „Diese Zeitschrift hat ein Programm". In: S*prache im technischen Zeitalter* 1.1 (1961), S. 1–2.

3 Heinz Zemanek: „Möglichkeiten und Grenzen der automatischen Sprachübersetzung". In: *Sprache im technischen Zeitalter* 1.1 (1961), S. 3–15. Ich danke Hans-Christian von Herrmann für den Hinweis auf diesen Aufsatz.

4 Yehoshua Bar-Hillel: „The Present Status of Automatic Translation of Languages". In: Franz L. Alt (Hg.): *Advances in Computers.* Bd. 1. New York 1960, S. 91–163, hier S. 158. Bar-Hillels berühmtes Beispiel nennt als plausiblen Kontext den Satz: „Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy." („Der kleine John suchte nach seiner Spielzeugkiste. Schließlich fand er sie. Die Kiste war im Laufstall. John freute sich sehr.")

5 Zemanek: „Möglichkeiten und Grenzen", S. 13.

6 Ebd., S. 14.

7 Max Bense: „Über natürliche und künstliche Poesie". In: Ders.: *Theorie der Texte. Eine Einführung in neuere Auffassungen und Methoden*. Köln 1962, S. 143–147, hier S. 143.

8 Ebd. Ich interpretiere Bense so, dass er eine frühe (ontologische) Formulierung des *symbol grounding problem* gibt, es aber mit einer (durchaus provokativ gemeinten) postromantischen Poetik verknüpft, die ihm als Kontrastfolie zu seiner Avantgardeästhetik dient. Vgl. Stevan Harnad: „The Symbol Grounding Problem". In: *Physica D: Nonlinear Phenomena* 42.1-3 (1990), S. 335–346.

9 Vgl. hierzu die Beiträge in Barbara Büscher, Christoph Hoffmann, Hans-Christian von Herrmann (Hg.): *Ästhetik als Programm: Max Bense/Daten und Streuungen.* Berlin 2004.

10 Theo Lutz: „Stochastische Texte". In: *augenblick* 4.1 (1959), S. 3–9.

11 Stattdessen, und das kann man bei vielen frühen Experimenten mit solcher generativer Literatur beobachten, sahen sich ihre Schöpfer so gut wie immer auch als Autorinnen und wiesen dem Computer nur die Rolle eines Werkzeugs zu, vgl. Hannes Bajohr: „Autorschaft und Künstliche Intelligenz". In: Stephanie Catani, Jasmin Pfeiffer (Hg.): *Handbuch Künstliche Intelligenz und die Künste.* i. E.

12 electronus [i.e. Theo Lutz]: „und kein engel ist schön". In: *Ja und Nein* 12.3 (1960), S. 3.

13 „So reagierten Leser". In: *Ja und Nein* 13.1 (1961), S. 3. Ich danke Toni Bernhart sehr herzlich dafür, dass er diesen Fund mit mir geteilt hat; vgl. ausführlich zum Hintergrund Toni Bernhart: „Beiwerk als Werk: *Stochastische Texte* von Theo Lutz". In: *editio* 34 (2020), S. 180–206.

14 Diese Idee ist Leah Henricksons Begriff des „hermeneutic contract" ähnlich. Ich behaupte aber, dass dieser hermeneutische Vertrag impliziert, dass die Instanz, die schreibt, *menschlich* ist, währen Henrickson diese Annahme gerade nicht macht, vgl. Leah Henrickson, *Reading Computer-Generated Texts.* Cambridge 2021, S. 28.

15 Alan M. Turing: „Computing Machinery and Intelligence". In: *Mind* 59.236 (1950), S. 433–460.

16 Turing übernimmt diesen Aufbau vom „imitation game" genannten Gesellschaftsspiel, bei dem das *Geschlecht* der unbekannten Person herauszufinden ist. Viel ist aus diesem „passing" gemacht worden, sowohl in Bezug auf Turings eigene Bio-

graphie – als homosexueller Mann wurde er zu einer Östrogenbehandlung gezwungen, mit der wahrscheinlich sein Selbstmord in Zusammenhang steht – wie auf die gegenderte Beschaffenheit von KI allgemein in der „offensichtlichen Verbindung zwischen Geschlecht und Computerintelligenz: beide sind nachahmende Systeme", Jack Halberstam: „Automating Gender. Postmodern Feminism in the Age of the Intelligent Machine". In: *Feminist Studies* 17.3 (1991), S. 439–460, hier S. 443.

17 Die essenzielle Textlichkeit von KI spitzte Jay David Bolter schon 1991 so zu: „Künstliche Intelligenz ist die Kunst, Texte zu machen", Jay David Bolter: „Artificial Intelligence". In: Ders.: *Writing Space. The Computer, Hypertext, and the History of Writing.* Hillsdale, NJ 1991, S. 171–193, hier S. 180.

18 Simone Natale: *Deceitful Media. Artificial Intelligence and Social Life after the Turing Test*. Oxford 2021, S. 3.

19 Kevin Roose: „An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy." In: *The New York Times* vom 02.09.2022. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html (Stand: 13.12.2022).

20 Danny Lewis: „An AI-Written Novella Almost Won a Literary Prize" (28. März 2016). In: Smithsonian Magazine. https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577 (Stand: 13.12.2022).

21 @horse_ebooks, 28. Juni 2012, https://twitter.com/horse_ebooks/status/218439593240956928 (Stand: 7.1.2023).

22 @horse_ebooks, 25. Juli 2012, https://twitter.com/Horse_ebooks/status/228032106859749377 (Stand: 7.1.2023).

23 Memphis Barker: „What Is horse_ebooks? Twitter Devastated at News Popular Spambot Was Human After All" (24. September 2013). In: *The Independent.* https://www.independent.co.uk/voices/iv-drip/what-is-horse-ebooks-twitter-devastated-at-news-popular-spambot-was-human-after-all-8836990.html (Stand: 13.12.2022).

24 Ich benutze den Begriff „Autorschaft" hier bewusst reduktiv. Ich meine damit nicht die „Seinsweise des Diskurses" und die „klassifikatorische Funktion" von Werkkohärenz und geistigem Eigentum, für die der Begriff der Autorschaft normalerweise in seiner emphatischen Funktion reserviert ist (Michel Foucault: „Was ist ein Autor?" In: Fotis Jannidis u. a. (Hg.): *Texte zur Theorie der Autorschaft.* Stuttgart 2000, S. 198–229, hier S. 210). Stattdessen gehe ich von einer „kausalen" Autorschaft aus (wer hat diesen Text und mit welchen Mitteln hervorgebracht?, vgl. Bajohr: „Autorschaft und Künstliche Intelligenz") und frage nach der rezeptionsseitigen Bewusstheit dieser Kausalität. Meine Verwendung ist so von geringerer Extension als der „implizite Autor", der ebenfalls ein Artefakt des Textes ist; sie kann daher nicht schlicht durch einen allgemeinen Textbegriff aufgelöst werden.

25 Natale: *Deceitful Media*, S. 4.

26 Aus systemtheoretischer Sicht, die allerdings allein auf den Begriff der Kommunikation abhebt und „menschliche Herkunft" bewusst ausklammert, beschreibt das schön Elena Esposito: *Artificial Communication. How Algorithms Produce Social Intelligence.* Cambridge, Massachusetts 2022.

27 John Seabrook: „The Next Word. Where Will Predictive Text Take Us?" (4. Oktober 2019). In: *The New Yorker.* https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker (Stand: 13.12.2022).

28 Dass die Sache etwas komplizierter ist und es in KI-Modellen so etwas wie „dumme Bedeutung" gibt, dazu vgl. Hannes Bajohr: „Dumme Bedeutung. Künstliche Intelligenz und artifizielle Semantik". In: *Merkur* 76.882 (2022), S. 69–79.

29 Vgl. dazu ebd.

30 GPT-3: „A Robot Wrote This Entire Article. Are You Scared yet, Human?“ (8. September 2020). In: *The Guardian.* https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3 (Stand: 13.12.2022). Die Autorenangabe – GPT-3 – ist natürlich selbst eine Fiktion. Wie ein Disclaimer am Ende des Artikels betont, wurden die Ausgaben händisch ausgewählt; die Prompts, mit denen das Programm gefüttert wurde, stammten von einem Informatikstudenten namens Liam Porr. Im Übrigen hat das Pronomen „ich“ in einem Sprachmodell kaum mehr Signifikanz als das Wort „Regenschirm“ – es ist jedenfalls ein Kategorienfehler, es als Identitätsaussage zu lesen.

31 Aber nur ein wenig; die allgemeine Be- und Entgeisterung über ChatGPT holte schlicht die Erfahrung auf breiter Front nach, die User mit Zugang zu GPT-3 (den nicht alle besaßen) bereits gemacht hatten. Interessanter wird es, wenn im Laufe dieses Jahres GPT-4 erscheint.

32 Vgl. Hannes Bajohr: „Keine Experimente. Über künstlerische Künstliche Intelligenz“. In: Ders.: *Schreibenlassen. Texte zur Literatur im Digitalen*. Berlin 2022, S. 173–190.

33 „GitHub Copilot“. In: *GitHub.* https://github.com/features/copilot (Stand: 13.12.2022). Copilot basiert auf OpenAIs Codex, einer speziell auf Code trainierten GPT-Version.

34 In Zukunft könnte dann, wie ein Insider spekuliert, „Programmieren obsolet“ werden, Matt Welsh: „The End of Programming“. In: *Communictions of the ACM* 66.1 (2023), S. 34–35.

35 „Craft. The Future of Documents“. https://www.craft.do (Stand: 13.12.2022).

36 Nur ein Beispiel unter vielen: „Jasper. AI Copywriting & Content Generation for Teams“. https://www.jasper.ai (Stand: 14.12.2022).

37 Die Diskussion um den Einsatz von ChatGPT für Seminararbeiten hat interessanterweise fast die größten Kreise gezogen; das ist insofern erstaunlich, als das Eingeständnis, ein Testregime auf Grundlage *vorhersehbarer* Sprache zu betreiben, vielleicht eher dieses Regime selbst auf den Prüfstand stellen sollte. Das zu erwartende Wettrüsten zwischen Sprachmodell und Sprachmodellprüfung ist jedenfalls pädagogischer Praxis kaum zuträglich (geradezu symptomatisch hilflos nimmt sich die Maßnahme aus, die jüngst aus einer New Yorker Schule gemeldet wurde – sie hatte kurzerhand auf Schulrechnern die IP-Adresse für ChatGPT gesperrt). Auf die *longue durée* möglicher postartifizieller Texte bezogen, die ich hier behandle, erscheint mir das Argument, dass wir nun den Übergang vom Rechenschieber zum Taschenrechner mit Integralfunktion erleben, aber eben für die geisteswissenschaftlichen Fächer, gar nicht mal so unplausibel zu sein.

38 Dies, insofern textuelle Äußerungen als in einem weiten Sinn – und wie oben auch für Literatur angenommen – kommunikativ aufgefasst werden können, vgl. Jürgen Habermas: *Theorie des kommunikativen Handelns.* Bd. 1. Frankfurt am Main 1982, S. 42.

39 Will Douglas Heaven: „Why Meta’s Latest Large Language Model Survived Only Three Days Online“ (18. November 2022). In: *MIT Technology Review.* https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science (Stand: 13.12.2022). Aus diesem Grund muss technisch noch einiges geschehen, dass ChatGPT wirklich auch als zuverlässige Suchmaschine benutzt werden kann. Bei der Präsentation von Googles Prototyp einer solchen KI-gestützten Suche – einem System namens Bard –, kam es sehr öffentlich zu einem faktisch falschen Suchergebnis; Google verlor daraufhin kurz-

zeitig 100 Milliarden Dollar an Marktwert, Emily Olson, „Google Shares Drop $100 Billion after Its New AI Chatbot Makes a Mistake“. In: *NPR* vom 9. Februar 2023. https://www.npr.org/2023/02/09/1155650909/google-chatbot--error-bard-shares (Stand: 10.02.2023). Und auch der Bing Chatbot gab bei seiner Vorstellung Falsches aus, bevor er später anfing, Journalisten zu beleidigen, Aaron Mok, „It's Not Just Google. Closer Inspection Reveals Bing's AI Also Flubbed the Facts in Its Big Reveal“. In: Business Insider (14. Februar 2023). https://www.businessinsider.com/bings-gpt-powered-ai-chatbot-made-mistakes-demo-like-google-2023-2 (Stand: 21.2.2023).

40 Murray Shanahan beschreibt den Unterschied sehr schön: „Nehmen wir an, wir geben einem großen Sprachmodell den Prompt ‚Der erste Mensch auf dem Mond war' und es antwortet mit ‚Neil Armstrong'. Was fragen wir hier wirklich? In einem wichtigen Sinn fragen wir nicht tatsächlich, wer der erste Mensch auf dem Mond war. Die eigentliche Frage, die wir dem Modell stellen, lautet: Welche Wörter folgen angesichts der statistischen Verteilung der Wörter im riesigen öffentlichen (englischen) Textkorpus am ehesten der Sequenz ‚Der erste Mensch, der den Mond betrat, war'? Eine gute Antwort auf *diese* Frage ist ‚Neil Armstrong'.“ Murray Shanahan: „Talking About Large Language Models.“ In: arXiv 2022, S. 2. http://arxiv.org/abs/2212.03551.

41 Tobias Wilke: „Digitale Sprache. Poetische Zeichenordnungen im frühen Informationszeitalter“ (12. November 2021). In: *ZfL Blog.* https://www.zflprojekte.de/zfl-blog/2021/10/12/tobias-wilke-digitale-sprache-poetische-zeichenordnungen-im-fruehen-informationszeitalter.

42 Martin Stingelin: „UNSER SCHREIBZEUG ARBEITET MIT AN UNSEREN GEDANKEN. Die poetologische Reflexion der Schreibwerkzeuge bei Georg Christoph Lichtenberg und Friedrich Nietzsche“. In: Sandro Zanetti (Hg.): Schreiben als Kulturtechnik. Grundlagentexte. Berlin 2012, S. 83–104.

43 Vgl. Bajohr: „Keine Experimente“.

44 Pablo Villalobos u.a.: „Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning“. In: arXiv, 2022. https://arxiv.org/abs/2211.04325 (Stand: 21.2.2023).

45 Dieser Sinn von „postartifiziell“ scheint an den Begriff „postdigital“ angelehnt zu sein. Wo aber letzterer auf die Differenz digitaler zu analogen Technologien abhebt – die ebenfalls bereits automatisiert sein können –, geht es ersterem vor allem um die menschliche oder nichtmenschliche *Herkunft* eines Artefakts unabhängig von seinem spezifischen *technischen* Substrat.

46 Benjamin Bratton, Blaise Agüera y Arcas: „The Model Is The Message“ (12. Juli 2022). In: *Noema.* https://www.noemamag.com/the-model-is-the-message (Stand: 03.08.2022).

47 Eine solche Lösung bietet OpenAI mit seinem „Al Text Classifier“ an (http://platform.openai.com/ai-text-classifier). Da die Verteilung der Ausgabewörter (oder Tokens) nicht tatsächlich zufällig ist, sondern einem nur zufällig erscheinenden Muster folgt, lässt sich in der Art dieser Verteilung ein Wasserzeichen einlassen. Natürlich bräuchte nur eine zweite, weniger ausgefuchste KI damit beauftragt zu werden, die Ausgabe der ersten umzuformulieren, und schon hätte man dieses Wasserzeichen wieder gelöscht, vgl. Kyle Wiggers: „OpenAI's Attempts to Watermark AI Text Hit Limits“ (10. Dezember 2022). In: *TechCrunch.* https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits (Stand: 07.01.2023).

48 Im oben genannten kausalen Sinn von Autorschaft wäre dann endlich eingetreten, was Foucault bereits in den sechziger Jahren imaginierte: die Frage nach Autorschaft hätte sich in der „Namenlosigkeit des Gemurmels“ verloren, Foucault: „Was ist ein Autor?“, S. 227.

49 Wenn ein solche Studie schreibt, „menschliche Leistung lässt sich am besten durch den Einsatz von KI in Form von Werkzeugen verbessern, die den Menschen selbst in die Lage versetzen, kreativer oder produktiver zu werden“, ist das schön gesagt, kann aber in der Rhetorik der „Steigerung“ natürlicher Anlagen die genannte Vermischung gerade nicht reflektieren, Vivian Emily Gunser u. a.:„Can Users Distinguish Narrative Texts Written by an Artificial Intelligence Writing Tool from Purely Human Text?“ In: Constantine Stephanidis, Margherita Antona, Stavroula Ntoa (Hg.): *HCI Inter-national 2021 – Posters.* Bd. 1419. Cham 2021, S. 520–527, hier S. 521.

50 Josh Dzieza: „The Great Fiction of AI. The Strange World of High-Speed Semi-Automated Genre Fiction“ (20. Juli 2022). In: *The Verge.* https://www.theverge.com/c/23194235/ai-fiction-writing-amazon-kindle-sudowrite-jasper (Stand: 13.12.2022).

51 So schildert die Comiczeichnerin Sarah Andersen, dass ihre eigenen Werke Teil des Trainingssets LAION Large-scale Artificial Intelligence Open Network) Stable Diffusion waren, das nun auch Bilder in ihrem Stil ausgeben kann. Der Name von Künst lerinnen sei damit „nicht mehr nur an das eigene Werk gebunden, sondern auch an eine Reihe von Nachahmungen unterschiedlicher Quali-tät, die sie nicht gebilligt haben. [...] Ich sehe, wie ein Monster entsteht.“ Sarah Andersen: „The Alt-Right Manipulated My Comic. Then A.I. Claimed It.“ In: *The New York Times* vom 31.12.2022. https://www.nytimes.com/2022/12/31/opinion/sarah-andersen-how-algorithim-took-my-work.html (Stand: 06.01.2023).

52 Eine Liste von Künstlern, die dezidiert *nicht* mit KI arbeiten, findet sich etwa unter https://whimsicalpublishing.ca/support-human-artists (Stand: 7.1.2023).

53 Kristen Mueller: *Partially Removing the Remove of Literature.* New York 2014.

54 Vgl. Hannes Bajohr, Annette Gilbert: „Platzhalter der Zukunft: Digitale Literatur II (2001– 2021)“. In: Dies. (Hg.): *Digitale Literatur II.* München 2021, S. 7–21. Die hier zitierten Beispiele bespreche ich vertieft in Hannes Bajohr: „Künstliche Intelligenz und digitale Literatur. Theorie und Praxis konnektionistischen Schreibens“. In: Ders.: *Schreibenlassen,* S. 191–213.