Hannes Bajohr

Artifizielle und postartifizielle Texte. Über Literatur und Künstliche Intelligenz

Walter-Höllerer-Vorlesung 2022, gehalten am 8. 12. 2022 an der TU Berlin

(erscheint in überarbeiteter Form in: *Sprache im technischen Zeitalter*, 1/2023)

Ich freue mich außerordentlich, heute hier sein und die Walter-Höllerer-Vorlesung halten zu dürfen. 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 dieser Zeitschrift aus dem Jahr 1961 definierte Höllerer, welche Aufgabe eine Literaturwissenschaft erfüllen muss, die auf der Höhe der Gegenwart und eben *im* technischen Zeitalter agiert: Sie sollte keine Furcht vor der Technik haben oder sie als ihren natürlichen Feind ansehen, weil Literatur eben irgendwie das Gegenteil von Technik ist; und sie sollte sich den Ideologien der Technik nicht willfährig unterwerfen.[1]

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 mehr *sein wird*.

1961 war der Umfang, den Sprachtechnologien einmal annehmen würden, kaum absehbar – angestoßen war die Entwicklung dahin aber durchaus. Höllerer 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* ein Aufsatz des österreichischen Computerpioniers Heinz Zemanek, der den damaligen Stand automatischer Sprach*übersetzung* referierte.[2] Zemaneks Artikel wandte Höllerers zwei Leitlinien – keine Furcht vor der Technik zu haben und nicht ihrer Ideologie aufzusitzen – ganz konkret auf die Sprachtechnologie der Übersetzung an: Damit Sprache *überhaupt* verarbeitet werden kann, muss man erst einmal annehmen, dass sie

[1] Walter Höllerer, »Diese Zeitschrift hat ein Programm«, *Sprache im technischen Zeitalter* 1, Nr. 1 (o. J.): 1–2.

[2] Heinz Zemanek, »Möglichkeiten und Grenzen der automatischen Sprachübersetzung«, *Sprache im technischen Zeitalter* 1, Nr. 1 (1961): 3–15.

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. Im selben Atemzug aber warnte Zemanek vor der Illusion *völliger* Automatisierbarkeit; Sprache ist komplex, situationsgebunden, oft mehrdeutig und immer Sache menschlicher Interpretation. Allein ihre Syntax zu automatisieren – was auch heute noch nicht vollständig klappt – heißt noch lange nicht, auch ihre Bedeutung zu erfassen. 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*.[3] Da »pen« mindestens zwei Bedeutungen hat, kommen auch zwei Übersetzungen in Frage: Die Kiste war im Gehege. 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. Die Software aber weiß das nicht. Sprachgebrauch setzt Intelligenz voraus, und solange beides nicht zusammen gelöst ist, hielt Zemanek eine hochqualitative, das heißt, menschenähnlich gute Übersetzung für ein »utopisches Ziel«.[4]

Schlägt man nun alle Vorsicht in den Wind, warnte Zemanek, und gibt man sich der Verführung hin, die von nur hinreichend guten Ergebnissen ausgeht, läuft man Gefahr, aus der Fiktion automatisierbarer Sprache eine Ideologie werden zu lassen. Dann passiert es, dass »der ästhetische Eindruck des Resultats alle Zweifel einschläfert, […] gleichzeitig aber schwierige Entscheidungen nicht anzeigt, sondern einfach trifft«.[5] Das Ergebnis *erscheint* sinnvoll, *ist* es aber in Wirklichkeit nicht; die Macht über Entscheidungen wird dann im falschen Vertrauen auf die Kompetenz der Maschine an sie übergeben. Besser wäre es daher, so Zemanek, die Sprachtechnologie allein als Hilfsmittel, in einer Assistenzfunktion zu verwenden, an dessen Ende immer noch ein Mensch steht, der die endgültige Entscheidung über die Bedeutung in der Hand hat.

Heute ist die Situation anders, aber diese Einsichten gelten immer noch. Sprachtechnologien sind ungleich ausgereifter. Und obwohl sie auch heute nicht intelligent sind – sie verstehen nicht tatsächlich, was sie tun –, erscheinen die allerneuesten KI-Modelle doch mehr denn je als intelligent. In diesem Erscheinen geht es darum, wie Beobachter:innen diese Ausgaben interpretieren, wie sie ihnen gegenübertreten und von ihnen auf das dahinterstehende System zurückschließen. Das ist, wie Zemanek selbst betonte, nicht nur eine ideologische, sondern auch eine eminent ästhetische Frage – und damit sind wir wieder bei Höllerer, der gerade der Literatur die Funktion zuschrieb, Sprache in ihrer Interaktion mit Technik zu reflektieren.

Hier möchte ich heute ansetzen und fragen, welchen Effekt die gegenwärtigen rapiden Fortschritte in der KI-Forschung auf den Umgang mit der Sprache, genauer, unsere *Leseerwartungen* haben. Anders als Höllerer und Zemanek stehen wir heute bereits wirklich

---

[3] Yehoshua Bar-Hillel, »The present status of automatic translation of languages«, *Advances in Computers* 1 (1960), 158.

[4] Zemanek, »Möglichkeiten und Grenzen«, S. 13.

[5] Ebd., 14.

an der Schwelle, von Texten umgeben zu sein, die 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 Texten* 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 diese Unterscheidung selbst wieder hinfällig wird, so dass wir diese Frage gar nicht mehr stellen und so statt natürlichen und artifiziellen vielmehr *post*artifizielle Texte lesen?

1.

Die Differenz zwischen natürlichen 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 machte er sich Gedanken speziell darüber, wie sich mit Computern hergestellte Literatur von der alten, menschengeschriebenen unterschied. Bense konzentriert sich dabei auf die »Art der Entstehung«[6] eines Textes: Was geschieht auf Seiten von Autor:innen, 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. Beides mache natürliche Texte wesentlich aus: Das poetische Bewusstsein, formuliert es Bense, setze »Seiendes in Zeichen«, also Welt in Text, und stehe am Ende dafür ein, dass das eine mit dem Anderen verbunden ist.[7] Ohne dieses Bewusstsein wären die Zeichen und die Beziehung zwischen ihnen sinnlos; sie würden nichts bedeuten. 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, Variablen in einer Operation, die auch völlig anders heißen könnten.

---

[6] Max Bense, »Über natürliche und künstliche Poesie«, in *Theorie der Texte. Eine Einführung in neuere Auffassungen und Methoden* (Köln: Kiepenheuer & Witsch, 1962), 143.

[7] Ebd., 143.

Genau diesen Fall beschreibt Benses zweite Kategorie, die *künstliche* Poesie. Damit meinte 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. entstanden. Das Thema eines künstlich generierten Textes ist 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 messbares Objekt.[8] Entstammt die natürliche Poesie dem Reich der Verständigung, ist die künstliche eine Sache der Mathematik – sie will und kann nicht kommunizieren, sie spricht nicht mehr von einer menschlichen Welt.

Benses Stoßrichtung war dabei aber nicht die Rettung einer romantischen Idee von unerklärlicher menschlicher Schaffenskraft. Im Gegenteil, der Autor ist hier mausetot. Stattdessen wollte Bense wissen, was man von einem Text ästhetisch noch aussagen kann, wenn man von den traditionellen Kategorien wie Bedeutung, Konnotation oder Referenz absieht.Die Antwort, die Bense vorstellte, war seine »Informationsästhetik«: Sie berücksichtigt, streng positivistisch, nur noch allein statistisch messbare Texteigenschaften. Künstliche Poesie wäre dann, eben *weil* sie bedeutungslos ist, auch *reine Poesie* – sie kommt völlig ohne die Unterstellung eines Bewusstseins aus; nur noch die Ästhetik des Textes selbst soll betrachtet werden. Wie bei Zemanek wäre die Unterstellung, das textproduzierende System hätte Intelligenz, ein Fehler – und zudem sogar ein *ästhetischer* Fauxpas.

Bense war selbst an mehreren Experimenten mit künstlicher Poesie beteiligt. Das bekannteste unter ihnen waren sicher die sogenannten »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. 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 nicht bedeutungsvoller.

Führt man Lutz' Programm aus, erhält man 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.

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 konnten. Sie waren zudem 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. Aber beide wussten ja auch, wie der Text hergestellt worden war. Ob man ihm seinen künstlichen Ursprung selbst ansieht – er sich also im »ästhetischen

[8] Bense formuliert hier ein Phänomen, das später unter dem Titel des *symbol grounding problem* für die KI-Forschung kanonischen Status erlangte: Computer verarbeiten leere Symbole, vgl. Stevan Harnad, »The Symbol Grounding Problem«, *Physica D: Nonlinear Phenomena* 42, Nr. 1–3 (1990): 335–46.

Eindruck« enthüllt, von dem Zemanek sprach – ist dagegen weniger klar; die Leserinnen und Leser der Literaturzeitschrift *augenblick* kamen jedenfalls nicht in die Verlegenheit, diese Frage zu stellen: Ein begleitender Essay klärte sie in allen Details über die 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.[9] Allein der Autorname »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« (als Fernschreiberausdruck) veröffentlichte er eine Reihe von Leserbriefen. Die Schreiber:innen waren sich – unwissend, 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 Leserin erkannte, dass es sich um Computerdichtung handelte und beglückwünschte die Zeitschrift zu ihrer kühnen Veröffentlichung.[10]

Was sich aber im Gros der 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:innen 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.

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.

---

[9] electronus [i.e. Theo Lutz], »und kein engel ist schön«, *Ja und Nein* 12, Nr. 1 (1960): 3.

[10] »So reagierten Leser«, *Ja und Nein* 13, Nr. 1 (1961): 3. Ich danke Toni Bernhart, dass er diesen Fund mit mir geteilt hat; siehe zum Hintergrund Toni Bernhart, »Beiwerk als Werk: Stochastische Texte von Theo Lutz«, *editio*, Nr. 34 (2020): 180–206.

2.

Einen artifiziellen *als* natürlichen Text auszugeben war nicht bloß ein einmaliger, banaler Scherz, den sich ein Informatiker in einer Esslinger Jugendzeitschrift erlaubte. Im Gegenteil ist dieses An-der-Nase-Herumführen das *Ur-Prinzip* künstlicher Intelligenz – und zugleich das, was sie mit Sprachtechnologien verbindet.

Der Informatikpionier Alan Turing hatte zehn Jahre zuvor, 1950, in einem Artikel mit dem Titel »Computing Machinery and Intelligence« darüber nachgedacht, ob Computer jemals denken, jemals intelligent sein könnten.[11] Turing lehnte diese Frage als falsch gestellt ab: Intelligenz könne man nicht verlässlich messen. Er ersetzte sie daher gut behavioristisch 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 jemandem und soll herausfinden, ob es sich dabei um einen Menschen oder eine Maschine handelt. 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. Wie Sie wissen, ist der Turing-Test 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, 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 und Bedeutsamkeit ist.

Für den Medienwissenschaftler Simone Natale basiert KI daher von Anfang an auf dem Prinzip der *Täuschung*. »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«.[12]

Sie sehen, dass diese Haltung – ich möchte sie *starke Täuschung* nennen – sofort Probleme mit sich bringt. Denn erst einmal bedeutet sie, dass es für KI-Systeme am besten ist, wenn eine Wissensasymmetrie zwischen dem menschlichen User und dem System besteht: Je mehr es über den User weiß und je weniger der User über die KI, desto überzeugender lässt sich die Täuschung vollziehen. Die politischen und ethischen Probleme sind offensichtlich: Täuschung ist, im Sinne Zemaneks, eine technische Ideologie: Sie lässt sich als notwendig für die Funktion des Systems begründen, belohnt aber

---

[11] Alan M. Turing, »Computing Machinery and Intelligence«, *Mind* 59, Nr. 236 (1950): 433–60.

[12] Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test* (Oxford: Oxford University Press, 2021), 3.

eine Intransparenz, die die Benutzer 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 oder 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' Leser auf die Auflösung der Computerautorschaft reagiert haben, aber man kann es sich denken, schaut man sich gegenwärtige Fälle an, in denen sich der »Künstler« nachträglich als Maschine entpuppte. Zuletzt geschah das im Juni 2022 bei einem eher peripheren Kunstpreis: Als der einreichende Teilnehmer zugab, sein Bild gar nicht selbst gemalt, sondern es durch die 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; der Künstler selbst sollte immer noch ein Mensch sein.[13]

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.[14] Es gibt noch weitere solcher Fälle – 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 *ex negativo*: 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 gab 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 doch etwas für menschliche Leser Sinnvolles ausgab – etwa solche aphoristischen Perlen wie:

> everything is happening so much[15]
>
> oder:
>
> unfortunately, as you probably already know, people[16]

---

[13] Kevin Roose, »An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy«, *The New York Times*, 2. September 2022, Abschn. Technology, https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

[14] Danny Lewis, »An AI-Written Novella Almost Won a Literary Prize«, Smithsonian Magazine, 28. März 2016, https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/.

[15] https://twitter.com/horse_ebooks/status/218439593240956928n

[16] https://twitter.com/Horse_ebooks/status/228032106859749377

Als sich aber herausstellte, dass die Tweets nicht generiert, sondern von einer Künstlergruppe *hand*geschrieben worden waren – und 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.[17]

3.

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 viel eher 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, desto mehr muss sich auch die Frage seiner 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 Autor:innen haben. 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.

Zuerst 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 der Differenz artifiziell/natürlich als der einzigen

[17] Memphis Barker, »What Is Horse_Ebooks? Twitter Devastated at News Popular Spambot Was«, The Independent, 24. September 2013, https://www.independent.co.uk/voices/iv-drip/what-is-horse-ebooks-twitter-devastated-at-news-popular-spambot-was-human-after-all-8836990.html.

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 einer nichtintelligenten Maschine produziert wurde, und es gleichzeitig so zu behandelt, als *wäre* es bewusste Kommunikation. In der Tat ist das ganz normal.

Simone Natale hat dafür einen weiteren Begriffsvorschlag: er nennt es *banale Täuschung*.[18] Anders als bei starker Täuschung sind sich User hier bewusst, dass sie getäuscht werden: Wir wissen, dass Siri kein Mensch ist und kein Inneres besitzt, aber die reibungslose Kommunikation mit ihr funktioniert nur dann, wenn wir die KI zumindest ansatzweise so behandeln als ob. 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 wird zur Bedingung von Funktionalität: Anders macht Siri eben nicht, was ich möchte.

Ähnlich verhält es sich mit Texten. Das beginnt bereits mit jedem Dialogfeld auf dem Computermonitor, das uns über etwas informiert. Die Frage, »Möchten Sie Ihre Änderungen speichern?« lässt schließlich eine Interaktion zu, die der mit einem Menschen ähnelt – die Antwort »Ja« hat eine andere Folge als die Antwort »Nein« –, ohne, dass man damit bereits Intelligenz vermutete. Damit wäre die Erwartungshaltung an unmarkierte Text bereits wieder heruntergestuft: Zwar verhalten wir uns immer noch so, als erwarteten wir 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 auf uns machen, doch wieder so etwas wie Bewusstsein nahelegt, desto schwerer wird es, sich in der Schwebe wohlzufühlen, die die banale Täuschung einzunehmen verlangt. 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; die Eingabevervollständigung im Handy schreibt Wörter sogar ohne nachzufragen zu Ende, was gelegentlich besonders unintelligent wirkt.

Seit einigen Jahren wird diese Technik komplexer und verwendet KI-Systeme. So führte Gmail, der Emailservice von Google, 2019 »Smart Compose« ein – eine Funktion, die beim Schreiben von Emails ganze Sätze vervollständigt. Trainiert auf die Texte von 1,8 Milliarden Gmail-Usern hat Smart Compose die wahrscheinlichsten Buchstabenverteilungen ihrer Emails gelernt. Es sagt nun voraus, welches Wort am ehesten auf ein anderes folgt. Angesichts der Masse an Trainingsdaten ergeben sich unheimliche Effekte, die die Fiktion

[18] Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*, 4.

der banalen Täuschung in Zweifel stürzen. Das illustriert ein Erlebnis, von dem der Autor John Seabrook im *New Yorker* berichtete.

In einer Email 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 »please« 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.«[19]

Seabrooks Scham war, objektiv betrachtet, ungerechtfertigt; es war ja nicht die Maschine, die aufmerksam war – sie ist weiterhin dumm. Was er hier vielmehr schildert, ist die Auswirkung, die die neuesten Sprach-KIs, die an der Grenze zum Schein der Intelligenz operieren, auf die intimsten Aspekte unseres Schreibens haben. In seinem Fall hatte sie sogar den Effekt, 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, wird es 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 dieses Jahres 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.

Seine Reaktion scheint bislang eher die Ausnahme zu sein, auch wenn sie keineswegs selten ist. Was der Fall aber zeigt, ist, dass sich das Gefühl des Unheimlichen, von dem Seabrook sprach, noch verstärken wird: Werden artifizielle Texte *zu* gut – indem sie etwa aufmerksamer erscheinen als ihre Schreiber – 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 die Frage danach erst einmal einzuklammern, wäre das erste, was wir von einem Text wissen wollen: Wie wurde er gemacht?

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 haben sowohl

---

[19] John Seabrook, »The Next Word. Where Will Predictive Text Take Us?«, The New Yorker, 4. Oktober 2019, https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker.

die gelernten Daten als auch die gemachten Vorhersagen Textform. Sprachmodelle besitzen eine ganze Reihe von Einsatzmöglichkeiten, von linguistischer Analyse über die automatische Übersetzung eben bis zur Generierung von Text. Kann Googles Smart Compose aber nur ein paar Wörter oder Sätze vorschlagen, sind *große* Sprachmodelle in der Lage, ganze Absätze und sogar zusammenhängende Texte zu schreiben: Und das nur, weil sie lernen, welche Sätze und Absätze statistisch am wahrscheinlichsten aufeinander folgen.

Inzwischen kenn jeder GPT-3, das große Sprachmodell, das die Firma OpenAI vor zwei 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 zum »Autor« wird und über »sich« erzählt, ein eigenes journalistisches Genre bildeten, und Titel wie diesen hervorbrachte: »A robot wrote this entire article. Are you scared yet, human?«[20]

Im November 2022 veröffentlichte OpenAI eine aktualisierte Version: ChatGPT ist noch einmal mächtiger als GPT-3 und ist nun auf Dialoge ausgelegt. Auf die Bitte, etwa einen Seminaressay über Jorge Luis Borges zu schreiben, legt die Maschine ohne zu Zögern los. Und 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, kann ich ChatGPT darum bitten, den Text in einer bestimmten Richtung weiterzuschreiben – etwa mit Literaturangaben.

In der Presse wird schnell spekuliert, ob solche Sprachmodelle menschliche Autor:innen einmal ersetzen werden. Aus verschiedenen Gründen bezweifle ich das.[21] 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 mithilfe immer weniger Menschen zu produzieren. Bestimmte Textarbeit wird so zumindest teilautomatisiert.

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, ein Sprachmodell in andere Software einzubauen. Damit wird Textgenerierung auf spezielle Aufgaben maßgeschneidert.

So gibt es inzwischen Schreib-KIs für das Programmieren, etwa GitHubs *Copilot*.[22] Es genügt, in wenigen Worten zu beschreiben, was das gewünschte Programm machen 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

---

[20] GPT-3, »A robot wrote this entire article. Are you scared yet, human? | GPT-3 | The Guardian«, zugegriffen 13. Dezember 2022, https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3. Die Autor:innenangabe – GPT-3 – ist natürlich selbst eine Fiktion. Wie ein Disclaimer am Ende des Artikels betont, wurden die Ausgaben händisch ausgewählt; und die Prompts, mit denen das Programm gefüttert wurde, stammten von einem Informatikstudenten namens Liam Porr.

[21] Vgl. Hannes Bajohr, »Keine Experimente: Über künstlerische Künstliche Intelligenz«, *Merkur* 75, Nr. 863 (2021): 32–44.

[22] »GitHub Copilot«, *GitHub*, zugegriffen 13. Dezember 2022, https://github.com/features/copilot.

Prototypen in Windeseile skizzieren oder einzelne Coder:innen lästige Detailarbeit an Copilot delegieren können.

Auch für gewöhnliches Schreiben gibt es schon Ähnliches. Das beginnt wieder bei der eher flankierenden Unterstützung: So wie ich ChatGPT bitten konnte, den Text anders fortzuführen, hat nun beispielsweise die Notizsoftware *Craft* einen Assistenten, der mein Geschriebenes für mich überarbeiten kann – indem er es etwa erläutert, weiterschreibt oder als Stichpunkte zusammenfasst.[23]

Große KI-Sprachmodelle sind vor allem dort gut, 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 eher als Füllmasse betrachtet. So sind im letzten Jahr dutzende von Sprach-KIs erschienen, die auf Marketing zugeschnitten sind – man soll damit Ad Copy und Produktbeschreibungen produzieren und schnell und massenhaft Content für Facebook, Firmenblogs und andere Plattformen generieren können – der oft ohnehin gar nicht so genau gelesen werden *soll*. Auch hier hilft es, wenn das Ergebnis nicht überrascht, sondern so klingt wie andere Texte dieser Machart auch.[24]

Umso schwieriger aber wird für die Leser:innen dieser Texte, sie als menschen- oder maschinengemacht einzuordnen. Wenn man bedenkt, wie viele des Geschrieben, 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, den ich bewerten muss? Was mit politischen Artikeln oder Fake News? Was mit der privaten, persönlichen, intimen Email? Ist auch sie ein KI-Produkt – ganz oder in Teilen?

Zumindest ein Grund für dieses Unbehagen ist: Menschen, auch das ist Teil der herkömmlichen Standarderwartung, stehen ein für das, was sie schreiben. Auch wenn sie lügen, sich irren oder in die Irre führen – zur Standarderwartung an Texte gehört das *principle of charity*, jenes Grundmaß an Vertrauen, dass die Schreibenden es ernst damit meinen. Nur deshalb ist die kritische Textlektüre, die in den Geisteswissenschaften gelehrt wird, ja überhaupt nötig: weil sie sich nicht von selbst versteht. Man will Texten erst einmal glauben.

Das wird aber schwieriger, wenn große Sprachmodelle einerseits Texte herstellen können, die so scheinen, als hätte sie ein Autor oder eine Autorin sanktioniert, und die andererseits kein zuverlässiges Wissen über die Welt besitzen, sondernd nur die Wahrscheinlichkeit von Zeichenfolgen ausrechnen.

---

[23] »Craft - The Future of Documents«, *Craft*, zugegriffen 13. Dezember 2022, https://www.craft.do/.

[24] Nur ein Beispiel unter vielen: „Jasper - AI Copywriting & Content Generation for Teams“, *Jasper.ai*, zugegriffen 14. Dezember 2022, https://www.jasper.ai/.

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 Prosa zu schreiben – doch nach nur drei Tagen wurde es wieder offline genommen: Das Modell nämlich schrieb sehr gut Texte, die autoritativ klangen, den Gepflogenheiten wissenschaftlicher Formatierung und ihren linguistischen Gesten folgten – aber völligen Unsinn enthielten.[25] Es hatte lediglich die *Form* von Wissenschaftsprosa gelernt, ohne jede wissenschaftliche Einsicht und Verantwortlichkeit.

4.

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 *postartifizieller* Texte übergehen.

Darunter verstehe ich zweierlei: Nämlich erstens die zunehmende *Vermischung* von natürlichen und artifiziellen Texten. Auch schon vor großen Sprachmodellen war kein Text wirklich *ganz* natürlich – der Wordprozessor ist ebenso eine Technik wie die Rechtschreibprüfung, und beide drücken dem Ergebnis ihren Stempel auf. Ebenso ist ein Text nie *ganz* artifiziell – das würde wirkliche Autonomie, wirkliche starke KI voraussetzen, die am Ende auch selbst entscheiden kann, einen Text zu veröffentlichen; und davon sind wir wirklich meilenweit entfernt. 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.[26]

Angesichts großer Sprachmodelle ist es nicht ausgeschlossen, dass beide in einen sich gegenseitig bedingenden Kreisprozess eintreten, der sie vollends miteinander verstrickt. Denn ein Sprachmodell lernt, indem man es auf große Mengen Text trainiert; und bislang bedeutet mehr Text immer auch bessere Performance. Denkt man das zu Ende, wird eine zukünftige Version im Extremfall einmal mit aller verfügbaren Sprache überhaupt trainiert werden. So würde jeder artifizielle Text bereits auf Grundlage *allen* natürlichen Texts entstehen.

---

[25] Will Douglas Heaven, »Why Meta's Latest Large Language Model Survived Only Three Days Online«, MIT Technology Review, zugegriffen 13. Dezember 2022, https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/.

[26] Dieser Sinn von »postartifiziell« scheint an den Begriff »postdigital« angelehnt; wo aber letzterer vor allem 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 unabhängig von ihrem spezifischen technischen Substrat.

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 zukünftige Sprachmodelle für weiteren Performancegewinn dann anhand von Text lernen, der selbst bereits aus einem Sprachmodell stammt.[27] Und der so gewonnene Sprachstandard mag umgekehrt wieder auf menschliche Sprechende einwirken – er hat, eingebunden in all die kleinen Schreibassistenten, den Status einer bindenden Norm, der statistisch kaum zu entkommen ist: 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 überspitzes Szenario. Als Gedankenexperiment zeigt es aber, was postartifizieller Text im Extremfall sein könnte. Doch schon bevor es soweit ist, auf halben Weg der absoluten Vermischung von natürlicher und artifizieller Sprache, ergäbe sich vielleicht bereits eine neue Standarderwartung an unbekannten Text.

Das ist die andere Bedeutung von ›postartifiziell‹. Nach der stillschweigenden Annahme menschlicher Autorschaft und dem Zweifel an der Herkunft von Geschriebenem, wäre sie die dritte Erwartungshaltung. Denn der Zweifel über den Textursprung kann nicht von Dauer sein. Menschen haben ein Interesse daran, Normalität herzustellen. Sollten politische Regulierung und technische Eindämmung hier 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 sein könnte, 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.

Wenn sich die gewöhnliche Erwartung an uns unbekannte Text also durchaus verschiebt; wenn sie in Zweifel gerät, um womöglich, in einer spekulativen Zukunft, in ihren Annahmen *agnostisch* zu werden – wieso dann die offensichtliche Aufregung um generierte Texte in Literaturwettbewerben? Wieso der Skandal, ein Bild, ein Roman sei mithilfe 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 keinen Autor, 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

[27] Benjamin Bratton und Blaise Agüera y Arcas, »The Model Is The Message«, Noema, 12. Juli 2022, https://www.noemamag.com/the-model-is-the-message.

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 anderen gehen wir weiter davon aus, dass sie einen Autor 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 der 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 zu Disposition steht. Interessanter ist daher die Frage, unter welchen Umständen sie hinfällig wird. Anders gefragt: Was müsste geschehen, damit auch *Literatur* postartifiziell würde? Welche Literatur wäre das – und welche nicht?

Ich will abschließend diese Frage beantworten, indem ich auf die Standardisierungstendenz zurückkomme, die vom Ouroboros-Effekt großer Sprachmodelle ausgeht. In ihnen findet, wie gesagt, eine Normalisierung statt; ihre Ausgaben sind dann überzeugend, wenn sie Erwartbares, Gewöhnliches, eben statistisch Wahrscheinliches ausspucken. Je »normaler« eine Schreibaufgabe, desto leichter lässt sie sich durch KI-Sprachtechnologien realisieren. Und so, wie es Marketing-KI für gewöhnliche Marketing-Prosa gibt, gibt es inzwischen auch Literatur-KI-Assistenz für vergleichsweise erwartbare Literatur.

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.[28] Ihr hilft dabei das Programm *Sudowrite* – es ist ein GPT-3-basierter, spezifisch *literarischer* Schreibassistent, der Dialoge fortführt, Beschreibungen ergänzt, ganze Absätze umschreibt und sogar Feedback gibt.

Die Qualität dieser Textausgaben ist recht hoch, ihr Inhalt dabei aber – das liegt in der statistischen Natur großer Sprachmodelle – eher gewöhnlich. Da sich alle Idiosynkrasien in der Masse an Trainingsdaten ausmitteln, tendieren sie auf einen 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 gerade diese Art von Literatur 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 Autornamen nicht verschwinden; aber sie würden eher als *brand* fungieren, die für einen bestimmten, erprobten Stil stehen (wie ja in Ansätzen

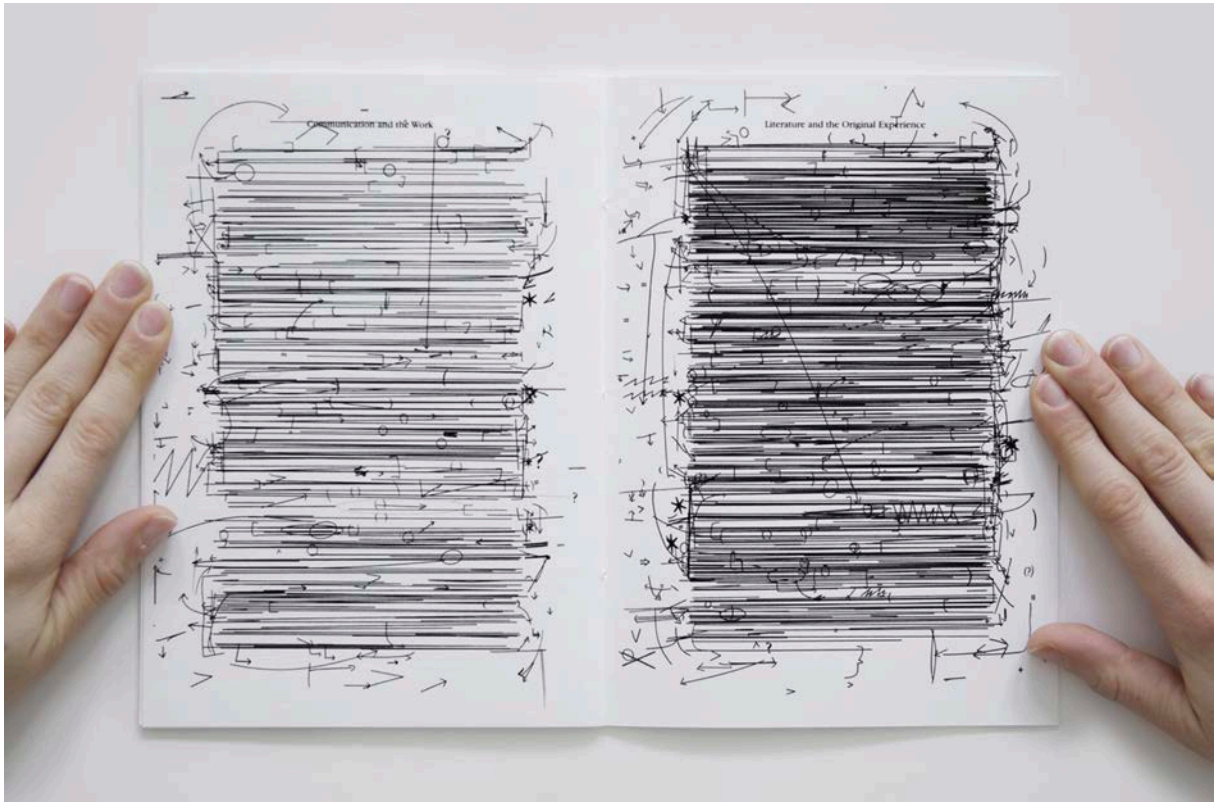
[28] Josh Dzieza, »The Great Fiction of AI. The strange world of high-speed semi-automated genre fiction«, The Verge, 20. Juli 2022, https://www.theverge.com/c/23194235/ai-fiction-writing-amazon-kindle-sudowrite-jasper.

heute auch schon), 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 weit mehr als es heute der Fall ist.

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: *guaranteed human-made* – so wie man auf Etsy oder bei Manufactum Handgefertigtes kauft, wäre eine Art *boutique writing* denkbar. Will man aber nicht allein auf solche Versicherung vertrauen – die doch wieder dem Zweifel Raum gibt, sie stimme nicht – wäre es vor allem ein unkonventioneller Gebrauch von Sprache, der ein Schreiben jenseits des Modells anzeigte.

Jedes formale Experiment, jede linguistische 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 der visuellen und »asemischen« Literatur, etwa in den Werken Kristen Muellers. Sie verwendet gar keine Buchstaben mehr, sondern nur noch die Anmutung von Zeilen und Textblöcken.[29] Die reine Poesie, von der Max Bense träumte, käme paradoxerweise gerade nicht aus der Maschine, die nun, in der postartifiziellen Vermischung, plausibel Bedeutung simuliert, sondern von Menschen, die gerade das nicht mehr tun.

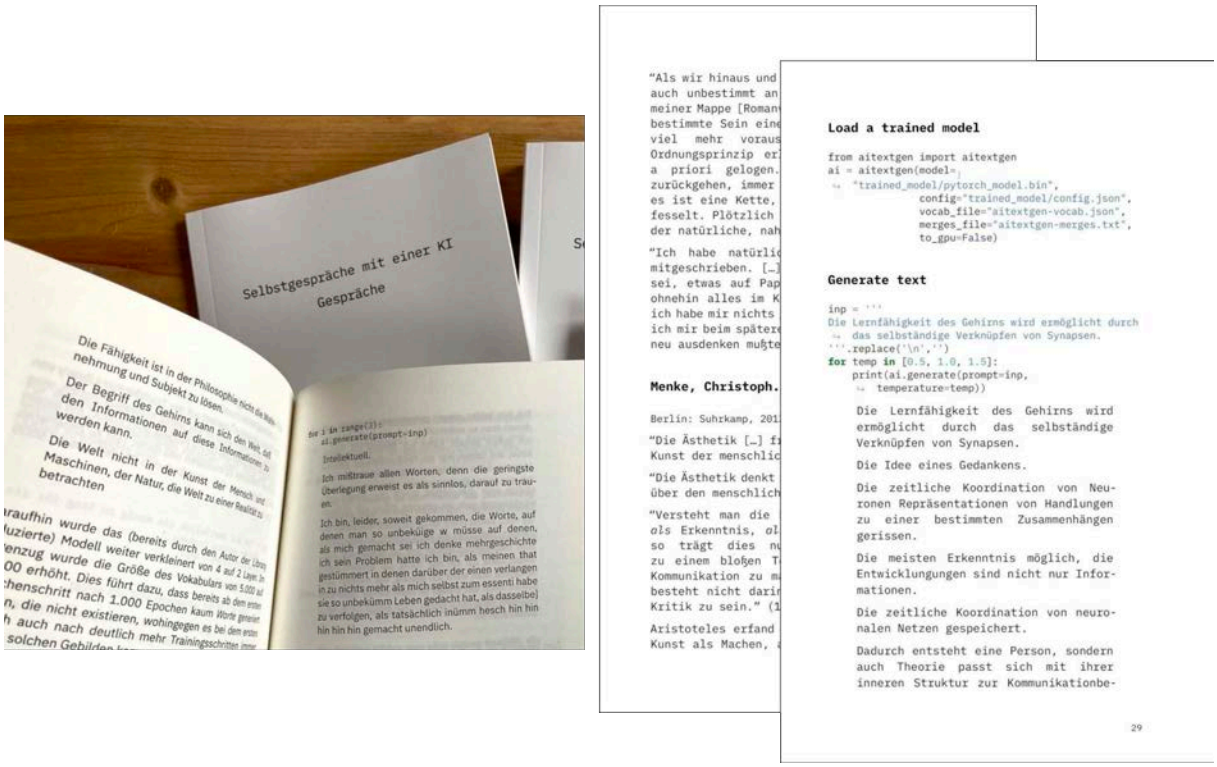


Kristen Mueller, *Partially Removing the Remove of Literature*.

[29] Kristen Mueller, *Partially Removing the Remove of Literature* (& So., 2014).

Auf der anderen Seite wären es aber gerade 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 mithilfe von Computern hervorgebracht wird. Sie kann 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.

So etwa bei Mattis Kuhns Buch *Selbstgespräche mit einer KI*, die neben seinen literarischen Experimenten auch den Quellcode zum Training des Sprachmodells mitgibt: Zwar nicht völlig, aber immerhin ein wenig lassen sich hier die menschlichen und maschinellen Komponenten trennen, die gemeinsam den Text ergeben.



Mattis Kuhn, *Selbstgespräche mit einer KI*

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« nennt – also meißeln oder schnitzen – von Hand: Der Punkt, an dem die Maschine ihren Text dem Menschen Jhave übergibt, wird genau markiert. Und indem er die Ergebnisse eines jeden Monats in je einem Buch sammelt – so dass *ReRites* nun zwölf Bände umfasst –, rahmt er diesen zwar kollaborativen, aber eben nicht absolut verschmolzenen Prozess zudem als Performance, die ebenfalls nicht konventionell literarisch ist.[30]

[30] Zu den beiden letztgenannten vgl. Hannes Bajohr, *Schreibenlassen. Texte zur Literatur im Digitalen* (Berlin: August Verlag, 2022), 191–213.

David »Jhave« Johnston, *ReRites.*

5.

Es sollte klar geworden sein, dass ich mich hier auf höchst spekulatives Terrain begeben habe. Ich will ich nicht sagen, dass erzählende oder im weitesten Sinne konventionelle Literatur von nun an verloren ist 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 können. Mir ging es eher um die Analyse von Tendenzen, und da lohnt sich der Blick auf die möglichen 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 bald bevorsteht – sowohl ohne Furcht vor der Technik zu haben, aber auch ohne ihren Ideologien aufzusitzen. Eines scheint mir jedenfalls sicher zu sein: Mit der zunehmenden 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 zuhause. 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 vorstellen können, wie eine Welt *post*artifizieller Texte aussehen könnte.