Hannes Bajohr

Probabilistische Intuition

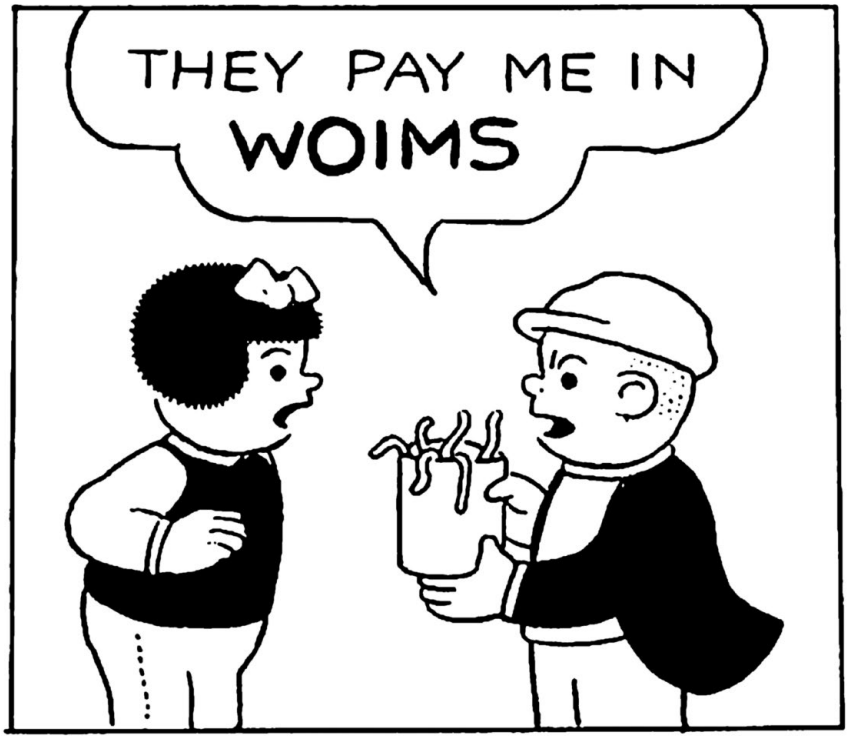
Über artifiziellen Text und seine Stile

Ein Panel aus Ernie Bushmillers Comicstrip *Nancy* von 1978. Wohl auf die Frage nach seinem neuen Job in einem Angelfachgeschäft hin hält die Figur Sluggo eine Köderdose hoch und antwortet in schwerem Bronx-Akzent: »They pay me in woims«. Dieses in seiner Absurdität charmante Einzelbild kursierte 2022 für kurze Zeit als minderes Meme. Es war aber auch Anlass für einen kuriosen Blogbeitrag, der im Jahr darauf erschien und auf den die Bluesky-Nutzerin Mags Colvett aufmerksam machte.[1] Der Text liest sich wie das, was wir heute als »Slop« bezeichnen würden: wertloser, KI-generierter Content, der eher auf *Search Traffic* hin optimiert ist als auf Informations- oder überhaupt Bedeutungsgehalt.

»They Pay Me in Woims: The Truth about This Bizarre Form of Payment«,[2] so der Titel des Posts, ist Slop nicht nur wegen seiner weitschweifigen Richtungslosigkeit, sondern auch, weil er so offensichtlich an der phonologischen Normalisierung scheitert, die jeder kompetente Sprecher des Englischen ohne nachzudenken vornehmen würde: nämlich zu erkennen, dass das mysteriöse »woims« nichts anderes als »worms« bedeutet. Sluggo wird in Würmern bezahlt. Das ist der Witz. Statt ihn aber mehr oder weniger amüsiert zur Kenntnis zu nehmen, behandelt der Essay seinen Gegenstand als kontextlosen, schwebenden Signifikanten und ergeht sich in ausufernden Spekulationen darüber, was um alles in der Welt denn nun »woims« sein könnte. Nach der Erörterung zahlreicher Möglichkeiten – vielleicht handelt es sich um einen Reddit-Insiderwitz, vielleicht um ein Portmanteau aus »warm coins«, in jedem Fall aber um eine höchst ungewöhnliche Zahlungsform – bringt der Beitrag sogar eine Interpretation vor, die an Theorien der Rezeptionsästhetik anknüpft: »Wie der Schriftsteller Umberto Eco in Bezug auf poetische Sprache diskutiert hat, ermöglicht Mehrdeutigkeit den Lesern, ›woims‹ auf der Grundlage eigener Erfahrungen und Schwierigkeiten mit Arbeit, Identität und Sinnhaftigkeit zu kontextualisieren.«

1 Inzwischen wurde der Post gelöscht. Ein Screenshot findet sich hier: imgur.com/a/Urqf6zq.

2 Auch dieser Text ist nicht mehr in seiner Originalform abzurufen. Stattdessen wurde er aktualisiert: Tony Robbins, *Understanding the Enigma: A Comprehensive Analysis of ›They Pay Me in Woims‹*. In: *Own Your Own Future* vom 11. März 2024 (ownyourownfuture.org/they-pay-me-in-woims). Mehr dazu weiter unten.



Die offensichtliche Frage – ist das schon eine Lektüre? – muss für den Moment beiseite bleiben. Denn ob ein großes Sprachmodell (LLM) zu einer Interpretation im starken Sinn fähig ist, würde Positionierungen über Erfahrung, Absicht und Bewusstsein erfordern, die ich hier lieber ausklammere.[3] Stattdessen soll es darum gehen, wie *wir* Geschriebenes lesen, das von LLMs verfasst wurde. Man mag hier vorläufig von »artifiziellem Text« sprechen, auch wenn dieser Titel, wie sich zeigen wird, völlig problematisch ist. Der Blogbeitrag zu »woims« jedenfalls ist ein Beispiel artifiziellen Textes, der ganz oder zu einem großen Teil durch ein Sprachmodell erzeugt wurde. Im Jahr 2023 – kurz nach der Einführung des für große Sprachmodelle inzwischen metonymisch verwendeten ChatGPT – war solcher Text noch eine Kuriosität

3 Dass es so etwas wie »dumme Interpretation« gibt, die an jene »dumme Bedeutung« anschließt, die ich großen Sprachmodellen vor Jahren unterstellt habe – und die durchaus überaus sinnvoll sein kann, nur eben kein Zeichen von Intelligenz –, davon mag sich jeder überzeugen, der einmal die »Podcast«-Funktion von Googles NotebookLM verwendet hat (notebooklm.google); vgl. Hannes Bajohr, *Dumme Bedeutung. Künstliche Intelligenz und artifizielle Semantik*. In: *Merkur*, Nr. 882, November 2022.

und seine Künstlichkeit offensichtlich. Heute, unter den Bedingungen vielfach verbesserter Systeme, ist das nicht mehr der Fall. Die lebensweltliche Verbreitung generativer KI nimmt zu, nicht zuletzt, weil ihre Ergebnisse inzwischen so gut sind, dass sie – bei einigermaßen kompetentem Prompten – kaum mehr als solche auszumachen sind. Und befragt man gegenwärtige Modelle, verstehen sie sich alle ohne große Probleme auf die Dekodierung von »woims« als »worms«.

Postartifizielle Bedingung und Laienforensik

Je flüssiger aber LLMs schreiben, je weniger artifizieller Text der Lektürearbeit Widerstand leistet, desto undurchsichtiger wird seine Herkunft. Sobald menschenähnlich Klingendes massenhaft maschinell produziert werden kann, droht sich der Unterschied zwischen natürlich und künstlich aufzulösen. Wir stehen in der zweifelhaften Morgenröte einer *post*artifiziellen Bedingung, in der wir nicht mehr fraglos annehmen können, dass ein uns unbekannter Text von einem Menschen verfasst wurde.[4] Gehen wir auch nur konservativ von der Konsolidierung des modernen Autorschaftsbegriffs aus, prägte diese Annahme etwa zweihundert Jahre lang unsere Lesegewohnheiten.[5] Dank großer Sprachmodelle ändert sich das nun. Der Zweifel an der Herkunft beginnt, zu einem neuen Standard zu werden. Er ist so destabilisierend, weil er in zwei Richtungen zugleich ausschlägt, so dass den *false negatives* der Fälle, die unerkannt bleiben, die *false positives* der fälschlich als generiert unterstellten gegenüberstehen.[6] Unter Maßgabe der postartifiziellen Bedingung hat sich somit das hermeneutische Verhalten, das Eve Kosofsky Sedgwick einst »paranoide Lektüre« nannte und das jeder

4 Hannes Bajohr, *Artifizielle und postartifizielle Texte: Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben*. In: *Sprache im technischen Zeitalter*, Nr. 61/245, März 2023. Zur Einordnung des Begriffs in seinem Verhältnis zum Postdigitalen, vgl. auch Eva Geulen, *Das Postdigitale. Ein Vorschlag zur Güte aus der Literaturwissenschaft*. In: Dieter Thomä (Hrsg.), *Postismen. Vom Nutzen und Nachteil eines Denkmusters für die Wissenschaften*. Berlin: Suhrkamp 2026.

5 Siehe Hannes Bajohr / Moritz Hiller, *Das Subjekt des Schreibens: Einleitung*. In: *TEXT + KRITIK*, Nr. X, 2024.

6 Man sollte sich immer darüber im Klaren sein, dass auf jeden Text, den man als offensichtlich generiert beurteilt, unbekannt viele kommen, die der eigenen Aufmerksamkeit entgehen. Dieser *survivorship bias* verleitet dazu, anhand der sichtbaren Stichprobe die unsichtbar gebliebenen Fälle systematisch zu unterschätzen. Umso schwerer wiegt dann die Einsicht, ein Text sei tatsächlich artifizieller Natur gewesen.

manifesten eine eigentlichere, latente Bedeutung zuschreibt, zu einem Massenphänomen entwickelt.[7]

Interessant in dieser Situation ist nun, dass die Literaturwissenschaft nurmehr ein Teil dieser Masse ist. Aller fachlichen Kompetenz in Sachen Text zum Trotz beteiligen sich ihre Vertreter kaum mit mehr Virtuosität als andere an dem, was man als »Laienforensik« bezeichnen könnte: dem skeptischen, aber letztlich methodisch freihändigen Durchsuchen einer E-Mail, eines Beitrags, eines studentischen Aufsatzes oder einer Textnachricht nach verräterischen Anzeichen maschineller Herkunft. Denn trotz der breiten Übernahme der Hermeneutik des Verdachts fiel die Reaktion der Literaturwissenschaft auf das Aufkommen artifizieller Texte, bis auf wenige Ausnahmen,[8] bisher verhalten, wenn nicht gar hilflos aus: Wie artifizieller Text zu lesen sei, was sein Aufkommen methodologisch und theoretisch für unsere Lektüre- und Interpretationspraktiken bedeutet – darüber ist von eigentlich berufener Seite kaum etwas zu hören. Es waren jedenfalls keine Germanisten, die die KI-generierten Meinungsbeiträge von Mathias Döpfner in der *Welt* oder Stephan-Andreas Casdorff im *Tagesspiegel* entlarvten.[9]

Fast scheint es, als hätte die Derrida'sche Annahme unendlicher Iterabilität diese Frage gänzlich zu umgehen erlaubt: Ein Text ist ein Text ist ein Text, ganz gleich, woher er stammt.[10] Zugleich habe ich – durchaus im Sinne dessen, was Leif Weatherby spöttisch als »remainder humanism« bezeichnet hat[11] – mehr als einen vordem überzeugten Dekonstruktivisten sagen hören, er werde niemals einen Roman lesen wollen, der *nicht* von einem Menschen geschrieben sei. Angesichts der sich etablierenden postartifiziellen Bedingung, das scheint klar, geraten nicht wenige unserer Annahmen darüber, was Lesen und das Lesbare ausmacht, unter Druck. Zumal die Literaturwissenschaft

7 Eve Kosofsky Sedgwick, *Paranoid Reading and Reparative Reading; or, You're So Paranoid, You Probably Think This Introduction is About You*. In: Dies. (Hrsg.), *Novel Gazing*. Durham / London: Duke University Press 1997.

8 Wobei die ins Auge springenden Ausnahmen alle eher am Schnittpunkt zur Medienwissenschaft angesiedelt sind, wie etwa Matthew Kirschenbaum und Leif Weatherby, oder ganz dort beheimatet sind, wie Mercedes Bunz.

9 Ersterer tat das freilich selbst – wohl im Sinne des *cutting out the middle man* darauf hoffend, seine Titel bald ganz KI-generiert bestücken zu können –, bei letzterem war es die Redaktion, der Unstimmigkeiten aufgefallen waren und die die entsprechenden Texte daraufhin depublizierte.

10 Jacques Derrida, *Signatur Ereignis Kontext*. In: Ders., *Randgänge der Philosophie*. Übersetzt von Gerhard Ahrens u.a. Wien: Passagen 1999.

11 Leif Weatherby, *Language Machines: Cultural AI and the End of Remainder Humanism*. Minneapolis: University of Minnesota Press 2025.

sollte hier eine Methodenreflexion wagen, auch wenn (und gerade weil) die Konsequenzen dieser Situation weit über das Fach hinaus und in so gut wie jeden Lebensaspekt hineinreichen.

Dabei sind KI-Themen den Geisteswissenschaften keineswegs fremd. Seit einigen Jahren floriert ein kritischer Zugriff, der den Fokus von den Outputs dieser Systeme auf die Modelle selbst verlagert, um die ihnen zugrundeliegenden Strukturen aufzudecken. Die Critical AI Studies, wie dieses sich formierende Feld heißt, haben durchaus wegweisende Analysen generiert. Sie kartieren Wissen-Macht-Gefüge großer Sprachmodelle, analysieren die in der Entwicklung von KI-Modellen verankerten epistemischen und ontologischen Annahmen, decken ökologische und ökonomische Folgekosten auf und verorten sie in kolonialen Geschichtsverläufen.[12] So grundlegend diese Arbeiten sind, erscheint artifizieller Text in ihnen doch lediglich als Durchgangsstation zur eigentlicheren Tiefensphäre, nie als selbstständiges Objekt. Für den Fokus auf die Tiefe ist der Output als *Oberfläche* nebensächlich. Doch gerade in der Beschäftigung mit gegenwärtigen Lesepraktiken verdient diese Oberfläche Beachtung, bevor sie auf das reduziert wird, was angeblich eigentlich darunter liegt. Schließlich sind es vor allem diese Oberflächen, denen wir – und zwar alle, ohne disziplinären Unterschied – in der Lebenswelt begegnen.

Man könnte es so sagen: Was fehlt, ist eine Praxis des *surface reading* artifizieller Texte. So nannten Sharon Marcus und Stephen Best vor gut anderthalb Jahrzehnten jene Oberflächenlektüre, die sie der Hermeneutik des Verdachts

12 Diese Literatur ist inzwischen Legion und ich nenne nur exemplarisch, was sich bereits als Kanon herausgebildet hat: Michael Castelle / Jonathan Roberge, *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Cham: Springer 2021; Wendy Hui Kyong Chun, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, Mass.: MIT Press 2021; Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven / London: Yale University Press 2021; Rita Raley / Jennifer Rhee, *Critical AI: A Field in Formation*. In: *American Literature*, Nr. 95/2, Juni 2023 (doi.org/10.1215/00029831-10575021); Matteo Pasquinelli, *The Eye of the Master: A Social History of Artificial Intelligence*. London: Verso 2023; Syed Mustafa Ali u.a., *Histories of artificial intelligence: a genealogy of power*. In: *BJHS Themes*, Nr. 8, 2023 (doi.org/10.1017/bjt.2023.15); Lucy Suchman, *The uncontroversial >thingness< of AI*. In: *Big Data & Society*, Nr. 10/2, Juli 2023 (doi.org/10.1177/20539517231206794); Louise Amoore u.a., *A world model: On the political logics of generative AI*. In: *Political Geography*, Nr. 113, August 2024 (doi.org/10.1016/j.polgeo.2024.103134); Leonardo Impett/Fabian Offert, *Vector Media*. Lüneburg: Meson Press 2026. Seit 2023 gibt es unter diesem Namen auch eine Zeitschrift: *Critical AI* erscheint bei Duke University Press (dukeupress.edu/critical-ai). Vgl. für einen eingehenderen Überblick Hannes Bajohr, *Surface Reading Large Language Models. Artificial Text and Its Styles*. In: *New German Critique*, Nr. 53/2, Herbst 2026.

gegenüberstellten, die auch Sedgwick als zu beengend beklagte.[13] Eine solche Lektüreform widersetzt sich der Annahme, dass allein Tiefe eine Analyse rechtfertigt, und rückt stattdessen die affektive und materielle Unmittelbarkeit des Textes in den Vordergrund. Zugleich aber ist dieses *surface reading* nicht naiv und ohne Modifikationen wieder zu aktivieren: Der Zweifel am Ursprung kann im Augenblick womöglich noch nicht übergangen werden, auch wenn das in Zukunft einmal so sein mag. Artifizieller Text, das wäre die Annahme, *ist* irgendwie, bevor er einmal, in einer vollumfänglich realisierten postartifiziellen Bedingung, eben nichts *anderes* mehr ist als anderer Text.

Eine Oberflächenlektüre artifiziellen Textes muss daher dessen Produktionsweise zumindest heuristisch mitberücksichtigen. Dabei wirft das Lesen dieser Oberflächen Probleme auf, die sich bei herkömmlichen Texten oder sogar anderen technischen Systemen bisher nicht stellten.[14] Der Charakter artifiziellen Textes lässt dennoch eine Lücke für das, was ich hier »probabilistische Intuition« nennen will – eine Intuition, zu der der Weg für die Literaturwissenschaft über den altehrwürdigen Begriff des »Stils« führt.

Repräsentativ–repräsentativ lesen

Natürlich hinkt der Vergleich mit Sedgwick: Die gegenwärtige Hermeneutik des Verdachts, die sich in der Form von Laienforensik als breite Reaktion auf artifiziellen Text herausgebildet hat, unterscheidet sich in wichtigen Punkten von der paranoiden Lektüre alter Schule. Beide Ansätze versuchen zwar, über die Oberfläche hinaus in eine darunterliegende Tiefe vorzudringen. Im einen Fall geht es aber darum, lediglich einen maschinellen Ursprung zu eruieren, im anderen, im Ganzen das aufzudecken, was der Text verdrängt. Wichtiger noch, die Eigenschaften des Geschriebenen selbst sind strukturell andere. Ihre Differenzen lassen sich entlang zweier Achsen skizzieren: erstens in Bezug darauf, wie jeweils mit Anomalien umgegangen wird; und zweitens hinsichtlich der Statik oder Dynamik des Textes selbst.

13 Stephen Best / Sharon Marcus, *Surface Reading: An Introduction*. In: *Representations*, Nr. 108/1, November 2009 (doi.org/10.1525/rep.2009.108.1.1).

14 Dass ich hier und im Folgenden vor allem von Text rede, auch wenn es natürlich Bilder, Musik und Videos gibt, die von Systemen maschinellen Lernens hervorgebracht werden, hat nicht nur den Zweck, den Fokus dieses Essays nicht ausufern zu lassen. Anders als andere artifizielle Outputs ist artifizieller Text so gut wie nicht forensisch belastbar als generiert nachweisbar, kann sehr viel leichter umgearbeitet und in andere Texte integriert werden und ist in seiner indexikalischen Struktur um einiges komplexer: Ein Bild, das »lügt«, ist etwas ganz anderes als ein Text, der die Wahrheit sagen mag, aber dessen Herkunft nichtmenschlich ist.

In seinen Arbeiten über *cultural analytics* untersucht Lev Manovich, was geschieht, wenn hermeneutische Praktiken auf statistische Methoden treffen; sind erstere auf Singularitäten ausgerichtet, nehmen letztere Regelmäßigkeiten in den Blick.[15] Entsprechend unterschiedlich gehen beide mit Anomalien um. In der Statistik ist der *outlier*, der Ausreißer, ein zu eliminierender Störfaktor. In der Literaturwissenschaft ist die mehrdeutige Passage, die *crux interpretum*, gerade das, was Aufmerksamkeit erfordert: ein symptomatischer Faden, an dem gezupft sich das Gewebe des ganzen Textes aufzulösen vermag. Das Problem artifiziellen Textes besteht nun darin, dass er als hermeneutisches Objekt selbst das Ergebnis eines statistischen Prozesses ist, als Extraktion des Allgemeinen aus dem Besonderen. LLMs glätten Ausnahmen und verwandeln singuläre Äußerungen in modellierbare sprachliche Regelmäßigkeiten. Was die Hermeneutik als aufschlussreiche Abweichung herausheben würde, erscheint auf der Ebene des Modells als nur noch marginales Signal, das im Mittelwert aufgeht. Das Symptom verliert seinen Vorrang.

Zweitens, und damit zusammenhängend, destabilisieren LLM-Outputs die empirische Identität des Textes. Er ist weder eindeutig noch statisch. In einer Polemik gegen die Naivität, mit der manche Geisteswissenschaftler an KI herangehen, haben Fabian Offert und Ranjodh Singh Dhaliwal jüngst beobachtet, dass sich allgemeine Urteile über ein System oft auf eine einzige oder nur wenige Ausgaben stützen. Das Modell auf frischer Tat dabei zu ertappen, wie es sachliche Fehltritte begeht oder ideologische Aussagen macht, wird zu einer symptomatischen Lesepraxis, mit der sich die Literaturwissenschaft im Zeitalter der großen Sprachmodelle wieder auf ihrem eigenen Terrain wähnt.[16] Diese Strategie, die Offert und Dhaliwal als »*n*=1-Experimente« bezeichnen,[17] steht dabei in einem grundlegenden Widerspruch zur probabilistischen Natur zeitgenössischen maschinellen Lernens, dessen Ausgaben je nach Prompt, Modell und Sampling-Parametern variieren. Polemisch gesagt: Es reicht nicht, ChatGPT eine Frage zu stellen, frei über die Ausgabe zu improvisieren und dann Feierabend zu machen. Das ist, als wollte man einen Würfel durch einen einzigen Wurf als gezinkt entlarven.

Auch wenn sie es eigentlich besser weiß: Pragmatisch ist die Literaturwissenschaft doch daran gewöhnt, den Text vor aller Interpretation als stabiles

15 Lev Manovich, *Cultural Analytics*. Cambridge, Mass.: MIT Press 2020. Er greift damit die von Wilhelm Windelband eingeführte Unterscheidung zwischen Natur- und Geisteswissenschaft als nomothetisch respektive ideographisch auf.

16 Fabian Offert / Ranjodh Singh Dhaliwal, *The Method of Critical AI Studies, A Propaedeutic*. In: *arXiv* vom 23. März 2025 (doi.org/10.48550/arXiv.2411.18833).

17 Die Variable *n* bezeichnet in empirischen Studien die Anzahl der untersuchten Fälle; eine *n*=1-Studie untersucht genau einen Fall, erlaubt also keine Verallgemeinerung.

Objekt zu betrachten oder bestenfalls als eines, von dem einige wenige Varianten existieren, die im Stemma der Textkritik erfasst werden können. Artifizieller Text widersetzt sich dieser Statik. Es ist ja nicht einmal offensichtlich, was als *der* Text zu gelten habe: die jeweilige Ausgabe, die mit jedem neuen Prompt eine andere ist, oder das Modell selbst, das aber gar keine Lektüre erlaubt, bevor ihm nicht ein Output abgerungen wurde.[18] Von *einem* Text zu sprechen, mag in jedem Fall irreführend sein, denn das, womit wir konfrontiert sind, ist je nur als Stichprobe aus einer Wahrscheinlichkeitsverteilung über eine Vielzahl möglicher Texte verständlich, als Sample aus einem probabilistischen Zeichenraum.[19] Im Gegensatz zur klassischen Druckkultur, aber auch zu früheren technischen Texten – sei es der Hypertext, der auch noch von Hand geschrieben wird, um dann die Lesereihenfolge den Rezipienten zu überlassen, oder seien es die regelgebundenen und damit, trotz Random-Funktion, durchaus durchschaubaren Textgeneratoren von einst – erfordern diese Systeme einen anderen Zugriff.[20] Offert und Dhaliwal können daher mit einiger Plausibilität verkünden: »Probabilistische Systeme erfordern notwendig eine probabilistische Form von Kritik.«

Der erste Schritt hin zu einem *surface reading* von LLMs besteht also darin anzuerkennen, dass sich das Objekt der Interpretation verschoben hat, und das in doppelter Hinsicht. Nicht nur ist der Text dynamisch und multipel statt stabil und singulär, so dass eine Lesepraxis, die sich auf isolierte Abweichungen konzentriert, das Phänomen verfehlt. Es kann auch nicht mehr das Ziel sein, jenen singulären Faden zu finden, an dem sich das Gewebe

18 Die Diskussion um *mechanistic interpretability*, die das Modell selbst lesen will, übergehe ich hier; weder hat sie bisher eindeutige Ergebnisse vorzuweisen, noch wären diese für die Literaturwissenschaft ohne Weiteres nutzbar zu machen.

19 Hübsch anschaulich gemacht bei Léon Bottou / Bernhard Schölkopf, *Borges and AI*. In: *arXiv* vom 4. Oktober 2023 (arxiv.org/abs/2310.01425).

20 Vgl. für eine Übersicht Hannes Bajohr / Simon Roloff, *Digitale Literatur zur Einführung*. Hamburg: Junius 2024. Eine empfehlenswerte Quellensammlung ist Lillian-Yvonne Bertram / Nick Montfort (Hrsg.), *Output: An Anthology of Computer-Generated Text, 1953–2023*. Cambridge, Mass.: The MIT Press 2024. Die für bisherige Systeme wichtige, von Espen J. Aarseth eingeführte Differenz zwischen Texton (Text, wie er im Code der Software auftaucht) und Scripton (Text, wie er für den User erscheint) ist in großen Sprachmodellen auch nicht mehr aufrechtzuerhalten, wenn das Texton keine stabile, im Programm hinterlegte Zeichenfolge ist, sondern erst während der *inference*, der Generierungsphase hervorgebracht wird, siehe Espen J. Aarseth, *Cybertext: Perspectives on Ergodic Literature*. Baltimore: The Johns Hopkins University Press 1997; dasselbe gilt für die ebenfalls vorpostkritische Unterscheidung zwischen »Oberflächen-« und »Tiefentext«, siehe N. Katherine Hayles, *Print is Flat, Code is Deep: The Importance of Media-Specific Analysis*. In: *Poetics Today*, Nr. 25/1, März 2004.

des Textes auftrennen ließe. Das markiert einen Übergang von einer »unrepräsentativen–repräsentativen« zu einer »repräsentativen–repräsentativen« Lektürepraxis.[21] In der klassischen, symptomatischen Interpretation ist die oberflächliche Anomalie insofern unrepräsentativ, als sie sich vom übrigen Text absetzt; repräsentativ wird sie erst als Ausdruck einer Tiefenstruktur, sobald sie gegen den Strich gelesen wird. Bei artifiziellem Text dagegen ist die Anomalie nicht mehr privilegiert, sondern lediglich ein Ereignis geringer Wahrscheinlichkeit; was das Textsystem als Ganzes wirklich repräsentiert, ist seine statistische Typizität. Das *surface reading* von LLMs beginnt also mit der Aufmerksamkeit für Regelmäßigkeiten und gerade nicht für den Ausnahmefall. Und diese Regelmäßigkeiten sind immer nur über Wahrscheinlichkeiten, nicht über feste Gegebenheiten zu fassen.

Technoimagination, probabilistische Intuition, Stil

Wer aber weiß, dass es einen technischen Unterschied zwischen artifiziellem und nichtartifiziellem Text gibt, hat bereits eine andere interpretative Haltung eingenommen. Sie besteht im Einüben dessen, was Vilém Flusser als »Technoimagination« bezeichnet hat – die Fähigkeit, sich neuen Medien gemäß ihrer eigenen Logik zu nähern, anstatt sich an überlieferte Modelle zu klammern.[22] Ein Foto sollte nicht wie ein Gemälde gelesen werden, ebenso wenig ein digitales Bild wie ein Foto. Dasselbe gilt für artifiziellen Text. Ohne diese Fähigkeit »sind wir verurteilt, in einer bedeutungslos gewordenen, technoimaginär kodifizierten Welt ein sinnloses Dasein zu fristen«.[23] So gesehen hat auch Flusser ein Modell von Tiefe. Doch sein phänomenologischer Ansatz geht von der Oberfläche aus, da nur das Phänomen jene Intuition liefern kann, die abstraktes Wissen über das technische Substrat nicht bereitstellt – Wissen, das dennoch im Blick bleiben muss, um überhaupt die richtigen Fragen stellen zu können.

Die Herausforderung für jede geisteswissenschaftliche Auseinandersetzung mit artifiziellem Text besteht heute darin, dass wir noch nicht die erforderliche Intuition entwickelt haben, jene Technoimagination, die das, was wir erleben, in einen sinnvollen Zusammenhang mit den tatsächlichen Funktionsweisen seiner Herstellungstechnik bringt: eine praktisch erworbene Sensibilität für strukturierte, weder deterministische noch willkürliche

21 Ich danke Dana Karout und Francesco Fritz für diese Formulierung.

22 Vilém Flusser, *Kommunikologie*. Frankfurt: S. Fischer 1998.

23 Ders., *Lob der Oberflächlichkeit. Für eine Phänomenologie der Medien*. Bensheim / Düsseldorf: Bollmann 1993.

Regelmäßigkeiten, die auf einer ausreichenden Kenntnis der verwendeten Techniken beruht. Wie die *connoisseurship* der Kunstgeschichte, die »persönliche Intuition«, die Erich Auerbach der Philologie der Weltliteratur anempfahl, oder das »geschulte Urteil« der Wissenschaft will auch eine solche Praxis geübt sein – indem man *viele* Oberflächen aufnimmt, aus *vielen* Systemzuständen heraus, in *vielen* Kontexten.[24]

Wie liest man nun artifiziellen Text im Bewusstsein seiner probabilistischen Natur, als dynamische Wahrscheinlichkeiten statt fixer Gegebenheiten? Wer den Weg empirischer Methoden und Benchmarks über viele Ausgaben hinweg nicht gehen möchte – gewissermaßen als $n \geq 1000$-Lektüre, wie sie die Digital Humanities erproben –, kann versuchen, bewusst auf eine interpretative und erfahrungsbezogene Auseinandersetzung als Entwicklung einer Art probabilistischer Intuition zu setzen.[25] Die Geisteswissenschaften halten jedenfalls dafür ein Mittel in ihrem hermeneutischen Arsenal bereit: den nur scheinbar überholten Begriff des »Stils«.[26] Er ist in vielerlei Hinsicht ein Intuitionsbegriff für den Umgang mit Wahrscheinlichkeiten.

Wenn isolierte Ausgabebeispiele keine unmittelbaren Erkenntnisse über den multiplen Text liefern können, ist eine *Reihe* von Beispielen dazu durchaus in der Lage – aber nur insoweit, als diese Reihe eine gemeinsame Einheit von Komplexität über ihre probabilistischen Variationen hinweg offenbart, die die Beispiele zu einem Teil *dieser* Reihe macht. Diese Einheit von

24 Dies nur als drei recht disparate Arten nicht-regelgebundenen, durch Erfahrung erworbenen Urteils in den Geisteswissenschaften, vgl. Bernard Berenson, *Three Essays in Method*. Oxford: Clarendon Press 1927; Erich Auerbach, *Philologie der Weltliteratur* [1952]. In: *Zeitschrift für interkulturelle Germanistik*, Nr. 9/12, Januar 2019; Lorraine Daston / Peter Galison, *Objektivität*. Übersetzt von Christa Krüger. Frankfurt: Suhrkamp 2007.

25 Probabilistische Intuition ist freilich ein alltägliches Phänomen. Strömungsdynamik und andere komplexe Systeme bieten anschauliche Beispiele, gerade weil sich hier, zumindest bis zu einem gewissen Grad, zwischen formalem technischem Wissen und kultivierter Intuition unterscheiden lässt. Physiker modellieren Turbulenzen oder Wellenbildung mit Gleichungen aus der statistischen Mechanik; Surfer und Segler entwickeln durch lange Erfahrung ein praktisches Gespür für das Meer und lernen, Dünungen oder Wirbel vorherzusehen, ohne Navier-Stokes-Gleichungen zu lösen. Beide Formen des Verstehens koexistieren, ersetzen einander aber nicht. Ich danke Marti Hearst für diesen Punkt.

26 Für die Radikalität, mit der sich etwa die Kunstgeschichte in den letzten fünfzig Jahren vom Stilbegriff entfernt hat, siehe Béatrice Joyeux-Prunel, *Style, Nonetheless. Eight Good Reasons to Reject the Notion of Style; Three Reasons to Use It Nonetheless*. In: *Artl@s Bulletin*, Nr. 13/2, 2024; Mario Carpo, *Imitation Games*. In: *Artforum*, Nr. 61/10, Sommer 2023.

Komplexität ist das, was man Stil nennen könnte. Stil ist, zumindest in dieser Definition, niemals nur die Eigenschaft eines Einzeldings, sondern immer einer Anzahl von Instanzen; er ist ein probabilistisches Oberflächenmerkmal, das allen Elementen einer Reihe gemeinsam ist, ohne mit ihnen identisch zu sein. In diesem Sinn, so schreibt Jeff Dolven, »hält ein ›Stil‹ Dinge zusammen. Er lässt uns ein Ganzes erkennen, wo wir von den Einzelteilen verwirrt sein könnten.«[27] Das passt gut zu Bests und Marcus' Definition von »Oberflächen als jenem Ort von Mustern, die innerhalb und *über* Texte *hinweg* existieren«. Stil ist dieses *»über … hinweg«* – die Differenz und Wiederholung, die alle Stichproben als kollektive Einheit eines einzigen, erfassbaren Erscheinungsbilds signalisieren. Die Entwicklung einer Intuition für probabilistische Systeme durch das Lesen ihrer Oberflächen läuft also, meine ich, vor allem darauf hinaus, den Stil ihrer Ausgabetexte zu erfassen.

All dies bedeutet nicht, dass ein einzelnes Sprachmodell nicht in der Lage wäre, verschiedene Stile zu erzeugen; natürlich kann es das.[28] Aber es heißt, dass bei ähnlichen Ausgangsbedingungen die Ergebnisse einen ähnlichen Stil gemeinsam haben werden. Hier, so glaube ich, kann man für eine Praxis argumentieren, die eine Technoimagination schärft, mit der empirische Methoden nicht zu ersetzen, aber möglicherweise zu ergänzen wären. Anstatt also LLM-Outputs als fixe Objekte und singuläre Texte zu behandeln, muss man lernen, sie als Momente in einem breiteren probabilistischen Feld wahrzunehmen, das sich in der Erzeugung einer gemeinsamen statistischen Gestalt vereinheitlicht – die wir, im Bewusstsein seines technischen Hintergrunds, von der Oberfläche ablesen können.

Surface Reading LLMs I: Referenzielle Drift

Wie könnte nun eine Praxis aussehen, die probabilistische Intuition kultiviert, indem sie Stil als echte Oberfläche und nicht als Epiphänomen von Tiefe betrachtet? Vielleicht können zwei Beispiele, ohne Vollständigkeit zu

27 Jeff Dolven, *Senses of Style: Poetry before Interpretation*. Chicago / London: The University of Chicago Press 2017.

28 Wie etwa gezeigt von Simon Meier-Vieracker, *No choice: On the stylistics of AI-generated texts*. In: *Journal für Medienlinguistik*, Nr. 8/2, 2026 (doi.org/10.21248/jfml.2026.85). Entsprechend ist Floridis allzu optimistische Hoffnung, für jedes Modell ließe sich ein einzigartiger Dataprint finden, zweifelhaft, vgl. Luciano Floridi, *Distant Writing: Literary Production in the Age of Artificial Intelligence*. In: *Minds and Machines*, Nr. 35/3, Juni 2025 (link.springer.com/10.1007/s11023-025-09732-1). Vgl. auch meine Replik auf Meier-Vieracker: Hannes Bajohr, *Style, Folk Forensics, and the Post-Artificial Condition*. In: *Research Gate*, Februar 2026 (doi.org/10.13140/RG.2.2.22714.15040).

beanspruchen, zumindest eine Idee davon geben, in welche Richtung diese Art des Lesens weist.

Zurück zu »woims«. Trotz seines Alters zeigt dieser Text deutlich Eigenschaften von artifiziellem Text, die auch neueren Slop besserer Qualität auszeichnet. Dazu gehört, was man als »referenzielle Drift« bezeichnen könnte. Der zentrale Begriff »woims« entbehrt eines feststellbaren Bezugsobjekts und beginnt zu schweben; in der Folge werden Interpretationsmöglichkeiten angedeutet, aber nie aufgelöst. Diese Drift wäre das objektseitige Gegenstück zum subjektseitigen »Bullshitting«, das ebenfalls oft mit artifiziellem Text in Verbindung gebracht wird – nur, dass es hier nicht um die Simulation von Sprecherkompetenz bei Indifferenz gegenüber der Wahrheit geht, sondern um die Simulation von Referenz bei Indifferenz gegenüber der Gegenständlichkeit des Bezeichneten.[29]

Referenzielle Drift muss man dabei als einen historisch-technisch spezifischen Stil von Slop verstehen, der in verschiedenen Modellen unterschiedlich stark ausgeprägt sein kann. Eine sorgfältig durchgeführte Oberflächenlektüre dieses Stils würde nachzeichnen, wie sich diese Drift phänomenologisch manifestiert: Wie Unbestimmtheit durch vorgetäuschte Kontextualisierung, den Einsatz scheinbar präziser Beispiele und strategischer Zitate (wie das von Eco) gemanagt wird, und wie formelhafte Allgemeinplätze verbunden mit weit ausholenden Gesten der Welterklärung zur Vagheitsökonomie dieser Textsorte beitragen.[30]

29 Natürlich ist »Indifferenz« hier nicht intentionalistisch zu verstehen. Vgl. zur Debatte um die Anwendbarkeit von Harry G. Frankfurts Begriff des »Bullshit«: Michael Townsen Hicks / James Humphries / Joe Slater, *ChatGPT is bullshit*. In: *Ethics and Information Technology*, Nr. 26/2, Juni 2024 (doi.org/10.1007/s10676-024-09775-5); David Gunkel / Simon Coghlan, *Cut the crap: a critical response to »ChatGPT Is Bullshit«*. In: *Ethics and Information Technology*, Nr. 27/2, April 2025 (doi.org/10.1007/s10676-025-09828-3).

30 Ein instruktives Beispiel für die Katalogisierung der stilistischen Marker von artifiziellem Text findet sich in der Wikipedia, die zunehmend mit KI-generierten Einträgen zu kämpfen hat: de.wikipedia.org/wiki/Wikipedia:Anzeichen_f%C3%BCr_KI-generierte_Inhalte. Man könnte hinzufügen, dass die Laienforensik nicht immer lediglich ein Stilempfinden meint, also eine holistische Erfahrung. Vielmehr muss diese Art der Detektivarbeit häufig einzelne, atomistische Merkmale betrachten, um zu ihren Urteilen zu gelangen (bis in die jüngste Vergangenheit waren etwa missgestaltete Finger ein gutes Anzeichen für KI-generierte Bilder). Ihr Vorläufer ist Giovanni Morellis kunsthistorische Zuschreibung anhand kleiner Details wie Ohren, Hände oder Fingernägel, von denen man annahm, dass sie sich einer bewussten Nachahmung entziehen, vgl. Carlo Ginzburg, *Spurensicherung. Die Wissenschaft auf der Suche nach sich selbst*. Berlin: Wagenbach 1995. Ich danke Lorraine Daston für diesen Hinweis.

Eine solche Lesart würde diese phänomenale Oberfläche zudem mit Dimensionen der Tiefe verknüpfen, ohne jedoch eine davon als die einzig maßgebliche zu privilegieren. In Anlehnung an Flussers Grundsatz, »Techno-Imagination erfordert Kenntnis der Theorien, auf welchen Apparate beruhen«,[31] kann man sich diesem Phänomen über mehrere gleichberechtigte Zugänge nähern.

Ein Erklärungsrahmen für referenzielle Drift ist die Modellarchitektur und der Trainingsprozess selbst. Der Text wird in Tokens zerlegt, also in Einheiten, die nicht notwendig mit ganzen Wörtern zusammenfallen (»Tokenisierung«); diese Tokens werden anschließend in Vektoren überführt, deren Bedeutung sich aus relationalen Nachbarschaften im Modell ergibt (»Embedding«).[32] Während »worms« als Zeichenfolge mit stabilen statistischen Kontexten dem Modell geläufig ist, wird das unbekannte »woims« in weniger semantisch bestimmte Bestandteile wie »wo« und »ims« aufgesplittet.[33] Diese Teile sind zwar weiterhin relational bestimmt, aber semantisch so vage, dass das (damalige) Modell auf einen generischen Diskurs über Mehrdeutigkeit zurückgreift, anstatt die pragmatische Korrektur vorzunehmen, zu der ein menschlicher Leser fähig wäre.

Doch diese technische Darstellung ist nicht die einzige Perspektive. Eine semiotisch-ökonomische Lesart könnte »woims« in der Logik von Content-Farmen und Suchmaschinenoptimierung lokalisieren: Überbordender, nur noch sekundär auf menschliches Lesen bezogener Text produziert Sichtbarkeit, die sich monetarisieren lässt, indem sie in Klicks, Werbeeinblendungen, Affiliate-Links oder Suchmaschinen-Rankings übersetzt wird. Lange vor KI funktionierte von Menschen produzierter Slop ganz ähnlich, so dass sich »woims« nahtlos in eine existierende Schablone einfügt: mehr Text = mehr Traffic = mehr Profit. »Woims« wird so zur Metonymie für Slop selbst: ein referenzloses Token, das gerade deshalb endlos recycelt und monetarisiert werden kann, weil seine schwebende Unbestimmtheit unendliche Elaborierung erlaubt.[34]

31 Vilém Flusser, *Für eine Theorie der Techno-Imagination*. In: Andreas Müller-Pohle (Hrsg.), *Standpunkte: Texte zur Fotografie*. Göttingen: European Photography 1998.

32 Tyler Shoemaker, *Verkettete Textualität*. In: Hannes Bajohr / Markus Krajewski (Hrsg.), *Quellcodekritik. Zur Philologie von Algorithmen*. Berlin: August 2024; Leonardo Impett / Fabian Offert, *Vector Media*. Dass Embeddings somit einer strukturalistischen Idee von Sprache nahestehen, ist oft bemerkt worden.

33 Man kann das selbst mit dem OpenAI Tokenizer überprüfen: platform.openai.com/tokenizer

34 Weshalb Weatherby behaupten kann, dass Genre vor Referenz die eigentlich operative Ebene artifiziellen Textes ist, vgl. Leif Weatherby, *Language Machines*. Dass dies dem Stil

Ganz gleich, für welche Tiefenerklärung man sich entscheidet, wesentlich ist doch, dass keine davon tatsächlich grundlegend ist. Jede folgt aus einer vorgängigen Beschreibung der Oberfläche, die im *surface reading* stets das primäre Datum bleibt.

Stilistische Schismogenese

Oder vielleicht doch nicht? Tatsächlich ist »woims« nur ein schwaches Indiz. Denn alles, was wir haben, ist der Text. Die postartifizielle Bedingung versagt es uns, mit Gewissheit zu erkennen, welches Modell ihn erzeugt hat und mit welchem Prompt oder inwieweit er bearbeitet wurde. Der Zweifel bleibt. Im vorliegenden Fall ist selbst der ursprüngliche Post nicht mehr verfügbar, sondern nur noch eine später aktualisierte Version, die seltsamerweise kaum kohärenter ist. Der Verdacht, dass es sich überhaupt um artifiziellen Text handelt, beruht also selbst schon auf dem Erkennen eines Stils, auf dem Erscheinen-als-Slop, und damit bereits auf einer Intuition, die wir in der Praxis gewonnen haben. Damit sind wir wieder mit der Crux der postartifiziellen Bedingung konfrontiert, in der jeder Versuch, eine *professionelle* probabilistische Intuition zu entwickeln, ohne das zugrunde liegende System sicher zu kennen, vor Problemen steht.

Denn natürlich kann, wie die Informatikerin Chantal Shaib und Kollegen feststellen, »Text als ›Slop‹ wahrgenommen werden, auch wenn er nicht von KI generiert wurde, und nicht jeder KI-generierte Text liest sich als ›Slop‹«.[35] Die paranoide Lektüre der Laienforensik läuft stets Gefahr, in jenen Wahn abzugleiten, den Gabriele de Seta als »algorithmische Folklore« bezeichnet: User mobilisieren »gefühlte« Erklärungen wie *shadow banning*,[36] die zwar

der Unmittelbarkeit entspricht, der mit dieser kulturökonomischen Logik in Verbindung steht, meint Anna Kornbluh, *Immediacy or, The Style of Too Late Capitalism*. London: Verso 2024; und, direkt darauf bezugnehmend: Jakko Kemper, *Generative AI, Everyday Aesthetic Production, and the Imperial Mode of Living*. In: *Critical AI*, Nr. 3/1, April 2025 (doi.org/10.1215/2834703X-11700246). Für die Verbindung von Semiotik und politischer Ökonomie allgemein vgl. Maurizio Lazzarato, *Signs and Machines: Capitalism and the Production of Subjectivity*. Übersetzt von Joshua David Jordan. Los Angeles: Semiotext(e) 2014.

35 Chantal Shaib u.a., *Measuring AI »Slop« in Text*. In: *arXiv* vom 23. September 2025 (arxiv.org/abs/2509.19163).

36 Also die anbieterseitige Praxis, User aufgrund bestimmter Aussagen für andere nicht mehr oder nur weniger anzuzeigen, ohne dass diese User es selbst merken würden. Während diese Praxis tatsächlich existiert, ist sie schwer nachzuweisen, so dass die Beschwerde, man sei Opfer von *shadow banning* geworden, notwendig einen paranoiden Einschlag hat.

falsch sein mögen, aber als Ermächtigungsgesten gegenüber undurchsichtigen Interfaces dienen.[37]

Hinzu kommt, dass einst diagnostische Merkmale, etwa die vormals überbeanspruchte Formulierung »to delve into«,[38] mit der laufenden Aktualisierung der Modelle zurücktreten, zugleich aber auf menschliche Schreibgewohnheiten wirken. Das Ergebnis ist ein Stilfeedback: Laienforensik, KI-Textsynthese und »reguläres« Schreiben verschränken sich dann so miteinander, dass menschliche Schreibende lernen, bestimmte Stilmarker zu vermeiden, um nicht selbst unter den Verdacht von KI-Einsatz zu geraten; so gehört der Em Dash im englischsprachigen Schreiben wohl auf lange Sicht der Vergangenheit an.[39] Dieser Prozess lässt sich mit der Logik der von Gregory Bateson so genannten »komplementären Schismogenese« vergleichen, bei der zu Distinktionszwecken das Verhalten einer Gruppe von einer zweiten negativ gespiegelt wird.[40] Für einen Moment ist dann das, was als menschlich erscheint, nicht einfach, was ohne KI entstanden ist, sondern zunehmend das, was den Verdacht der KI-Generierung erfolgreich abwehrt, ganz gleich, auf welche Weise es eigentlich produziert wurde.

Doch genau weil es kein absolutes Wissen um die tatsächliche Herkunft von Geschriebenem geben kann und jedes Überführen ein Indizienprozess bleiben muss, besitzt diese Schismogenese ihren eigenen Grenzwert. Wo nur Stilmarker über den Verdacht entscheiden, genügt es, diese Marker zu tilgen – und das lässt sich, etwa über Schreibtools, automatisieren: Die Rechtschreib-KI Grammarly bietet die Funktion an, bestimmte, an KI gemahnende Formulierungen zu entfernen, und macht damit aus der Vermeidung artifiziell wirkenden Textes selbst eine artifiziell unterstützte Praxis. Die so kalibrierte Sprache wirkt nun auf die Outputs der Sprachmodelle zurück, deren Tics sich notwendig ändern werden, so dass eine Art Wettrüsten zwischen menschlichem und generiertem Schreiben die Folge ist. In letzter Konsequenz müsste dieses fortwährende Stilfeedback aber zur Aufhebung der Differenz zwischen natürlichem und artifiziellem Text selbst tendieren. Das dialektische Moment lautet hier: Selbst *nicht* wie eine KI schreiben zu wollen

37 Gabriele de Seta, *Algorithmic Folklore*. In: Rayvon Fouché (Hrsg.), *Oxford Research Encyclopedia of Science, Technology, and Society*. New York: Oxford University Press 2026.

38 Dmitry Kobak u.a., *Delving into LLM-assisted writing in biomedical publications through excess vocabulary*. In: *Science Advances*, Nr. 11/27, Juli 2025 (doi.org/10.1126/sciadv.adt3813).

39 Ich beschreibe einen anderen, aber verwandten sprachlichen Feedbackmechanismus in Hannes Bajohr, *Dumme Bedeutung*.

40 Gregory Bateson, *Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven*. Übersetzt von Hans Günter Holl. Frankfurt: Suhrkamp 1985.

heißt, unter der Maßgabe artifiziellen Textes zu schreiben. Aus diesem Verhältnis kommt man nicht mehr raus. Sollte er je existiert haben – in der postartifiziellen Bedingung gibt es keinen »natürlichen« Text mehr.

Surface Reading LLMs II: Oberflächenerzählung

Um der Herkunftsproblematik zumindest ansatzweise zu begegnen, also ein zweites Beispiel, dessen Ursprung gewiss ist, weil ich es – *verum ipsum factum* – selbst gemacht habe.[41] Etwa zum gleichen Zeitpunkt, da der »woims«-Artikel entstand, schrieb ich nämlich mit einem selbsttrainierten Sprachmodell einen Roman. Mich überhaupt darauf einzulassen, war motiviert durch die Behauptung des Literaturwissenschaftlers Angus Fletcher, Sprachmodelle seien grundsätzlich nicht in der Lage zu erzählen. Seine Begründung: Narration setze Kausalität voraus; Kausalität lasse sich über statistische Korrelationen nicht angemessen reproduzieren; daher seien Sprachmodelle echter Narration unfähig.[42] Auch dies ist ein Argument *de profundis*, insofern Fletcher seine These auf Grundprinzipien stützt, die er aus der technischen Struktur des statistischen maschinellen Lernens ableitet, statt auf die Outputs der Modelle zu schauen.

Der Text, der 2023 unter dem Titel *(Berlin, Miami)* erschien, ist zwar kein empirisch belastbares Experiment zu Fletchers These – zumal seine Ergebnisse heute durchaus andere wären –,[43] sondern eher so etwas wie Artistic Research. Aber indem er die Narrationsfähigkeit von KI-Text am Modelloutput prüft, zeigt er sowohl, wie Argumente, die sich ausschließlich auf die Tiefe stützen, blind für das eigentliche Phänomen sein können, als auch, dass es sich zur Ausbildung einer probabilistischen Intuition lohnt, von Rezeption gelegentlich auf Produktion umzuschalten.

(Berlin, Miami) entstand unter Verwendung des offenen Sprachmodells GPT-J, das anhand von vier deutschsprachigen Romanen finegetuned, ihm

41 Das Folgende führe ich näher aus in Hannes Bajohr, *Kohäsion ohne Kohärenz. Künstliche Intelligenz und narrative Form*. In: Ders. / Ann Cotten (Hrsg.), *Schreiben nach KI*. Berlin: Rohstoff 2025.

42 Angus Fletcher, *Why Computers Will Never Read (or Write) Literature: A Logical Proof and a Narrative*. In: *Narrative*, Nr. 29/1, Januar 2021 (doi.org/10.1353/nar.2021.0000).

43 Ohne Frage würde ein aktuelles Modell besseren, das heißt kohärenteren Text hervorbringen. Aber auch Modelle wie GPT-5 sind keineswegs narrativ kohärent, vgl. Junjie Li u.a., *Lost in Stories: Consistency Bugs in Long Story Generation by LLMs*. In: *arXiv* vom 6. März 2026 (arxiv.org/abs/2603.05890), so dass man in älteren Modellen möglicherweise schlicht klarer erkennen mag, was in neueren als überschriebene Tendenz weiterhin vorhanden ist.

also eine anhand dieser Texte trainierte Stimme aufgesetzt wurde. Der Plot des Romans ist nur schwer zusammenzufassen, da er mit Narrativität in der Tat zu kämpfen hat. Auf den ersten Blick scheint das Buch die Elemente einer konventionellen Geschichte zu besitzen: wiederkehrende Figuren (wie der mysteriöse, Odradek-ähnliche »Kieferling«), sich entfaltende Ereignisse (wie Miamis Heimsuchung durch die unheimlichen »Lebensviren«), Dialogfetzen und sogar formale Elemente wie Kapitelunterteilungen oder Übergangspassagen. Doch bei genauerem Hinsehen offenbart sich eine seltsame, antinarrative Qualität aus logischen Unstimmigkeiten und Zeitsprüngen.

Man kann versuchen, den dem Roman eigenen Stil rein auf Grundlage der Textoberfläche, als kollektives Muster über viele Beispiele hinweg zu analysieren.[44] Das heißt aber, Kategorien wie Korrelation und Kausalität zu vermeiden, die zu einer tiefenorientierten Erklärung gehören und die darauf blicken, was ein System im Inneren *tut*, und etwa durch jene von Kohärenz und Kohäsion zu ersetzen, bei denen es allein darum geht, wie ein Text an der Oberfläche *erscheint*. Meint Kohäsion den grammatischen und lexikalischen Zusammenhang eines Textes, bezieht sich Kohärenz auf dessen abstrakten semantischen Zusammenhang. Die Unterscheidung ist idealtypisch: Kohäsion vermittelt Kohärenz, und Kohärenz ermöglicht es Kohäsion, Bedeutung zu tragen. *(Berlin, Miami)* hebt sowohl das konventionelle Gleichgewicht von Kohäsion und Kohärenz als auch den Status von Kohärenz als Marker narrativer Logik auf. Es überschießt mit Kohäsion – fließende Übergänge, wiederkehrende Motive, lokale Sprachgewandtheit –, scheitert aber an der nachhaltigen Erzeugung von Kohärenz. Was dabei entsteht, ist eine Simulation narrativer Muster, eine Oberflächenerzählung.

Natürlich kann man diese wieder mit dem zugrunde liegenden technischen System in Verbindung bringen und versuchen, sie mit dem Fehlen kausaler Schlussfähigkeit oder einer anderen Eigenschaft maschinellen Lernens zu erklären. Gleichzeitig wird eine solche Lesart, da sie von der Oberfläche ihres Objekts ausgeht, nicht mit diesem überstürzten Abtauchen hin zu Tiefenstrukturen zufrieden sein. Denn während der Mangel an Kohärenz die konventionellen narrativen Qualitäten des Textes durchaus mindert, würde eine solche Lesart auch anerkennen, dass der Text zugleich auch kein *Nicht*-Narrativ ist. Vielmehr, und *pace* Fletcher, ist der unweigerliche Akt der Sinnstiftung durch einen Leser (um auf Eco zurückzukommen) jene erfahrungsbasierte Praxis, die Unbestimmtheiten ausfüllt, *selbst* wenn ein Text nichtmenschlich ist – und das heißt, auch dann, wenn man seine artifizielle Herkunft kennt.

44 Der Roman ist *ein* Text, aber er besteht aus *vielen* generierten Samples, die aus technischen Gründen nie länger als 2048 Token waren.

Und so ist die Differenz von natürlich und artifiziell auch von dieser Seite wieder eingehegt: Erzählt werden kann auch dann, wenn man eigentlich weiß, dass keiner erzählt.[45]

Zugleich ist diese Konstruktion des Textes seitens des Lesers immer eine Funktion dessen, welchen Spielraum an möglichen Erwartungen ein Text zulässt – was eine Frage der Erfahrung mit *dieser* Art von Text ist, also einer probabilistischen Intuition. Sie übt sich in einem Prozess der Konvergenz ein, zwischen der erwarteten / konventionellen und der realisierten / tatsächlichen Form des Textes, auch wenn sie dabei mit der Paradoxie ringen muss, dass für sie artifizieller Text gerade noch existiert und, dank des Zweifels des Postartifiziellen, zugleich auch schon nicht mehr.[46] Dass sie nicht nur durch die passive Konfrontation mit Generiertem zustande kommt, sondern auch durch den eigenen Umgang mit den Modellen, fasst prägnant der Titel einer Studie zusammen: *Personen, die ChatGPT häufig für Schreibaufgaben nutzen, sind zuverlässige und robuste Detektoren für KI-generierte Texte*.[47] Eine Kritik von KI, die ihr Objekt aus praktisch-ästhetischer Sicht nicht kennt, kann einiges über Tiefe sagen, aber nichts über die Oberfläche – um die muss es uns, gerade als professionellen Lesern, heute aber gehen.

45 Freilich gilt das nur prinzipiell. Ein interessanter Effekt des Wissens um die Generiertheit des Romans war, dass keine der Rezensionen auch nur ansatzweise den Versuch unternahm, den Text tatsächlich zu lesen, sondern bereits damit zufrieden war, in ihm eine Illustration der Grenzen literarischer KI zu sehen. Ein als experimentell ausgezeichneter Roman ohne diesen Herkunftsausweis hätte womöglich mehr interpretative Anstrengung erfordert. Die Lektion für Autoren ist offensichtlich.

46 Das kommt im Paradox zum Ausdruck, dass einerseits Menschen in der Regel keine mehr als zufallswahrscheinliche Detektionsleistung zeigen, andererseits aber, wie die nächste Fußnote zeigt, Praxis doch auf Erkenntnis zurückwirkt; vgl. Mark Louie F. Ramos, *Is It Cake or Is It AI? A Systematic Review of Human Uncertainty in Distinguishing Generative Artificial Intelligence Content*. In: *arXiv* vom 3. April 2026 (arxiv.org/abs/2604.03437).

47 Jenna Russell / Marzena Karpinska / Mohit Iyyer, *People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated text*. In: *arXiv* vom 19. Mai 2025 (arxiv.org/abs/2501.15654).