

Protokoll 27

Axel Volmar & Florian Cramer & Christoph Borbach & Anja Schürmann & Oliver Ruf

Kettenbrief für *Quatschmaschinen*:

Gefährliches Halbwissen? Über die Automatisierung postfaktischer Epistemologien

17.04.2023

1. Fake it until you make it. Über die Camouflage postfaktischer Epistemologien (Axel Volmar)

Die sprachlichen und kognitiven Fähigkeiten von ChatGPT und anderen aktuellen Transformermodellen sind beeindruckend. Die dialogischen KI-Systeme geben seit 2022 innerhalb von Sekunden – und auf Wunsch ausladende – Antworten auf alle möglichen Fragen, fassen Texte zusammen und erfüllen kreative Schreibaufträge in allen erdenklichen Textsorten von Gedichten über Bewerbungsschreiben bis zu Erörterungen und Essays. Nicht zuletzt ihre vermeintlich universelle Anwendbarkeit weckt allfällige Begehrlichkeiten und Delegationsphantasien. Bei mir zum Beispiel nicht zuletzt hinsichtlich ihrer Verwendbarkeit im Kontext (geistes-)wissenschaftlicher Forschung und Textproduktion, denn angesichts des massiven Anstiegs an wissenschaftlichen Veröffentlichungen (durch die quantitätsorientierten Neoliberalisierung des Wissenschaftssektors einerseits sowie der zunehmenden Internationalisierung bzw. Globalisierung von Forschungsdiskursen andererseits) wären produktive Mensch-Maschine-Kollaborationen auch im Bereich der Geistes- und Kulturwissenschaften im Grunde mehr als wünschenswert. Fragt sich also: wie gestaltet sich das wissenschaftliche Schreiben im Verbund mit der Maschine?

In meinen ersten Testreihen schlug mein initialer Enthusiasmus leider relativ schnell in Ernüchterung um oder vielmehr, in ungläubige Bestürzung. Denn ChatGPT ist in Software gegossene Hochstapelei. ChatGPT ist eine „Quatschmaschine“, weil sie fortlaufend quatschen kann, ihr Output ist aber gerade nicht *dada*, kein offensichtlicher Unsinn, den man interessant finden oder beiseite wischen könnte, sondern – infolge der in die Algorithmen einprogrammierten Turing-Test-Strategien – vorgeschützte Allwissenheit durch die Verschleierung der Maßstäbe und Praktiken der eigenen Wissensproduktion. Da die Transformermodelle unermüdlich und oft auch ungefragt mit einem scheinbar übersteigerten Selbstbewusstsein und ohne jegliche Selbstzweifel

Wissen auswerfen, handelt es sich richtigerweise, wie manche klug erkannt haben,¹ um *Mansplaining as a Service*, mit all seinen Konnotationen. Denn trotz ihrer Eilfertigkeit wirkt der Ton der Maschine selbstgefällig; ein Eindruck, der durch oftmals selbständig ergänzte paternalistische und moralisierende Kommentare zusätzlich verstärkt wird. Dafür legt ChatGPT jedoch ein verblüffend geringes Maß an Gewissenhaftigkeit an den Tag. Mit Fakten nimmt es das System jedenfalls nicht so genau. Während mich als Geisteswissenschaftler das Vermögen der Maschine zur Produktion argumentativer Texte durchaus fasziniert, finde ich es so erschreckend wie bezeichnend, wie wenig die Programmierer von ChatGPT dafür Sorge getragen haben, dass die Maschine die Prämissen und Grenzen ihrer eigenen Epistemologie offenlegt. Das scheint mir ein hochgefährlicher Mix für einen Informationsservice zu sein, der vorgibt, zwar nicht in theologischer Hinsicht, aber doch im Hinblick auf pragmatische Belange mehr oder weniger allwissend zu sein. ChatGPT hätte daher eigentlich mit einem dicken Beipackzettel versehen werden müssen.

Inverses Plagiat

Zunächst ist es frappierend, wie wenig ChatGPT grundlegende Praktiken, Fertigkeiten und Konventionen der Wissensbeschaffung und -verifikation beherrscht – und damit anzunehmenderweise beherrschen können soll. Erstens gibt die Maschine beispielsweise selbst vorgenommene Verdichtungen von Diskurspositionen ohne Umschweife als wörtliche Zitate aus, legt dieses Vorgehen aber so nicht offen. Ich habe ChatGPT beispielsweise zum Aufwärmen gebeten, mir drei relevante Medialitätskonzepte zu nennen und zu erklären:

AV: *Explain to me the mediatheoretical concept of “mediality” [...] using up to 3 different definitions.*

ChatGPT: *Mediality is a mediatheoretical concept that refers to the ways in which different media and technologies shape our perception, communication, and understanding of the world. It is a term that encompasses both the material and symbolic dimensions of media, including their technical features, social uses, and cultural meanings. Here are three possible definitions of mediality that can help us understand this complex concept:*

- 1. Mediality as a mode of representation and expression [führt weiter aus]*
- 2. Mediality as a system of communication and interaction [führt weiter aus]*
- 3. Mediality as a cultural and historical phenomenon [führt weiter aus]*

1 Vgl. Harrison 2023 (<https://futurism.com/artificial-intelligence-automated-mansplaining-machine>); <https://octodon.social/@cwebber/109462369670144568>; <https://mastodon.social/@amyhoy/109355444166205985>.

Da mir aufgrund dieser doch recht allgemeinen Definitionen der Kontext fehlt, hake ich nach und bitte die Maschine, ihre Ausführungen mit Zitaten zu belegen, was die Maschine mit nachdrücklichem Enthusiasmus aufzugreifen scheint:

AV: *Thank you, that's very helpful. It would, alas, be helpful to receive your research with proper citations. Therefore, let me rephrase my query.*

ChatGPT: *Certainly! Please provide me with your specific research requirements and I will do my best to provide you with properly cited information.*

AV: *Cite and elaborate 3 different definitions of the mediatheoretical concept of "mediality" including their correct bibliographic references [...].*

ChatGPT: *Certainly! Here are three different definitions of the mediatheoretical concept of "mediality," along with their corresponding bibliographic references and elaborations:*

1. *Mediality as a mode of representation and expression:*

According to Friedrich Kittler, a German media theorist, mediality refers to the "technical conditions of a medium, which determine how its messages are produced, transmitted, and received" (Kittler, 1999, p. 1). [führt weiter aus] Source: Kittler, F. A. (1999). Gramophone, film, typewriter. Stanford University Press.

2. *Mediality as a system of communication and interaction:*

According to Jürgen Habermas, a German philosopher and sociologist, mediality refers to the "systematic relationship between media and public opinion" (Habermas, 1991, p. 176). [führt weiter aus] Source: Habermas, J. (1991). The structural transformation of the public sphere: An inquiry into a category of bourgeois society. MIT press.

3. *Mediality as a cultural and historical phenomenon:*

According to Lev Manovich, a media theorist and cultural historian, mediality refers to the "cultural and historical conditions under which media and technologies are produced and consumed" (Manovich, 2001, p. 10). [führt weiter aus] Source: Manovich, L. (2001). The language of new media. MIT press.

Zunächst bin ich erst einmal angetan und sehe großzügig über die etwas arbiträre Auswahl der angeführten Autoren hinweg. Als ich das vermeintliche Kittler-Zitat jedoch weder auf der angeführten Seite 1 noch sonst irgendwo in GFT finde, werde ich stutzig: Denkt sich die Maschine etwa Zitate aus? Das kann nicht sein, denke ich und beginne, etwas nachzubohren:

AV: *Thank you. I tried to look up the Kittler reference you cited in order to get the larger context of the quote. However, I wasn't able to*

find the quote in the book you cited (Kittler 1999). Please look up and write out the paragraph in the book from which you took the quote.

ChatGPT: *I apologize for any confusion caused. Here's the full paragraph from Friedrich Kittler's book "Gramophone, Film, Typewriter," which provides more context for the quote I used earlier:*

"Mediality denotes those characteristics of technical media that determine how they enable, condition, and structure the communication of events or messages, and thus influence the interrelations between the individual, society, and culture. These characteristics include not only the material properties of the medium (e.g. written characters, printed characters, film images), but also the historically variable protocols that determine its usage (e.g. handwriting, typography, camera angles)."

Source: Kittler, F. A. (1999). Gramophone, film, typewriter. Stanford University Press, p. 1.

Auch diesen Absatz gibt es so allerdings nicht. Das ist auf mehrfache Weise faszinierend: Erstens, weil weder das Zitat selbst noch die erbetene, das Zitat umgebende Textstelle Wiedergaben des vermeintlichen Ursprungstextes sind (das zunächst angeführte Zitat, um dessen Kontextualisierung ich gebeten hatte, taucht zudem nicht einmal in dem nachgelieferten Absatz auf). Der *modus operandi* der Maschine ist sozusagen ein inverses Plagiiieren: ChatGPT gibt Paraphrasen von Texten als Zitate aus und setzt so die eigene Interpretation an die Stelle des ursprünglichen Textes bzw. Autors (tatsächlich existiert *kein einziges* der drei angeführten Zitate). Zweitens gibt die Maschine den Text, den es zusammenfantasiert, mir gegenüber aber völlig selbstverständlich als zitierte Quelle mit „more context“ aus. ChatGPT scheint also die formalen Aspekte von „properly cited information“ zu kennen, es mit dem Unterschied zwischen Zitation und Paraphrase bzw. den mit der Wiedergabe verschiedener Textebenen verbundenen Praktiken nicht besonders genau zu nehmen. Müsste sie ja auch nicht zwingend – die Frage ist aber: warum nutzt sie die Form des Zitierens dann überhaupt? Ich habe daher ein weiteres Mal nach:

AV: *No problem. However, the sentence you quoted from the book [...] actually does not appear in the paragraph you provided me with. Has there been another mix-up?*

ChatGPT: *I apologize for the mistake. It seems that the quote I used was a paraphrase rather than an exact quote from the book. Here is the correct excerpt from Friedrich Kittler's "Gramophone, Film, Typewriter" that corresponds to the definition of mediality that I presented:*

"By 'media' we understand not only apparatuses and institutions for communication, but also the whole phenomenon of the trans-

mission of culture itself, and finally the entire realm of symbolic interaction between man and his environment. What interests us in this book is the 'characteristics' of such technical media, characteristics that determine how they enable, condition, and structure the communication of events or messages, and thus influence the interrelations between the individual, society, and culture. These characteristics include not only the material properties of the medium, but also the historically variable protocols that determine its usage."

Source: Kittler, F. A. (1999). *Gramophone, film, typewriter*. Stanford University Press, p. 1.

Auch diese Textstelle ist, trotz Entschuldigung und der Beteuerung, jetzt das „correct excerpt“ gefunden zu haben, frei erfunden. Weist man die Maschine auf Unstimmigkeiten hin, entschuldigt sich diese mit Verweis auf einen scheinbar als Ausnahme aufzufassenden „mistake“, anstatt unumwunden zuzugeben, dass sie schlicht nicht dafür ausgelegt ist, mit Zitaten zu operieren. Statt diesen Fehler nun zu beheben, antwortet die Maschine zudem einfach mit einem weiteren erfundenen Zitat! Ich muss mich erst einmal fassen.²

Science Fiction als bibliographische Fiktion

Aber es kommt noch besser: Fragt man das System beispielsweise nach relevanter Forschungsliteratur, die es ja mit Hilfe von öffentlich zugänglichen Datenbanken wie WorldCat, Google Scholar etc. leicht finden könnte, wird auch hier dreist geflunkert. Da ich mich ja u.a. für Videokonferenzen interessiere und mich daher etwas in diesem Forschungsdiskurs auskenne, frage ich das System beispielsweise nach den zehn relevantesten Arbeiten in diesem Bereich und erhalte umgehend eine Liste mit Literatureinträgen:

AV: *What are the 10 most relevant scholarly articles on video conferencing within the discipline of media studies an media theory?*

AI: *Here are 10 relevant scholarly articles on video conferencing within the discipline of media studies and media theory:*

1. *Van Dijck, J. (2020). Video conferencing in times of crisis: How Zoom shapes up against its competitors. Media International Australia, 177(1), 16-29.*

[...]

² Obwohl dieses Verhalten von ChatGPT bereits vielfach beobachtet worden ist, ist es dennoch eine verblüffende Erfahrung. Vgl. etwa den Bericht des Kunsthistorikers Ben Davis vom März 2023 (Ben Davis, „We Asked ChatGPT About Art Theory. It Led Us Down a Rabbit Hole So Perplexing We Had to Ask Hal Foster for a Reality Check. A simple ChatGPT inquiry turned into a very strange ride“, *Artnet News*, 2. März 2023, <https://news.artnet.com/art-world/chatgpt-art-theory-hal-foster-2263711>.)

6. Mazmanian, M., & Orlikowski, W. J. (2009). *Theorizing the organizing process: A dialectical approach*. *Organization Science*, 20(6), 1179-1197.

[...]

Diese Rückgabe lässt mich stutzig werden. José van Dijck publiziert zwar seit Jahren zu Social Media, hat aber meines Wissens nach nicht zu „Video conferencing“ oder „Zoom“ gearbeitet. Eine Google-Scholar-Suche später stelle ich fest: Der Beitrag existiert überhaupt nicht und auch in der Zeitschrift *Media International Australia* hat van Dijck bisher nie veröffentlicht. Ebenso ist Wanda Orlikowski zwar für ihre Arbeiten zur Organisationstheorie bekannt und war mehrfach Autorin der Zeitschrift *Organization Science* – einen Artikel mit dem Titel „Theorizing the organizing process“ hat sie jedoch nie verfasst. Das Erstaunliche: Die Mehrheit der Einträge ist – inklusive fiktiver DOI-Codes – frei erfunden, während vier Artikel dagegen tatsächlich existieren. Was bedeutet das für die Einschätzung von ChatGPT im Allgemeinen?

Obwohl die Mehrzahl ihrer Einträge erfunden ist, betreibt die Liste ein überzeugendes Mimikri an der Form der wissenschaftlichen Bibliografie. Denn während meine Anfrage nur teilweise mit faktisch existierenden, sondern überwiegend mit fiktiven, aber möglich erscheinenden Quellen beantwortet wurde, ist die Formatierung der Liste tadellos. Ebenso verhält es sich mit den von der Maschine angeführten angeblichen Zitaten, die statt existierender Textpassagen allgemein gehaltene Beschreibungen der referenzierten Quellen enthielten. Die Maschine scheint also Zitate und Literaturreferenzen unter formalen und formattechnischen und nicht unter inhaltlichen bzw. faktenbasierten Gesichtspunkten zu produzieren. Mit Charles Sanders Peirce gesprochen könnte man auch sagen, dass die Maschine im Rahmen der Textproduktion auf die Ähnlichkeitsbeziehung bzw. Ikon-Funktion zuungunsten der Index-Funktion des Textes fokussiert. Schein statt Sein. Und genau hier liegt das Problem, weil die Maschine eben genau dieses Vorgehen nicht öffentlich macht, sondern durch rhetorische Routinen und Anthropomorphisierungen zusätzlich verschleiert. So suggerieren die Entschuldigungen der Maschine, dass es sich bei den beanstandeten Unstimmigkeiten um das Produkt von Kontingenzen und nicht der Logik des Algorithmus handeln würde. Tatsächlich hat die Faktenhalluzination aber System bzw. bildet die Regel und nicht die vermeintliche Ausnahme: *it's not a bug, it's a feature*.

Postfaktische Epistemologie im Mäntelchen des digitalen Bibliothekars

Ich frage mich also: wenn die nachgereichte „Quelle“ zur vorher gelieferten Information auf der Basis eines realexistierenden Diskurses fiktionalisiert worden ist, wie belastbar sind dann generell die Informationen, die ChatGPT scheinbar unumwunden als „Fakten“ präsentiert? Und was sagt das über die Epistemologie künstlicher Intelligenz bzw. die hinter diesen stehenden Organisationen aus? Ich versuche, mir noch einmal die grundlegenden Funktionsprinzipien von Deep-Learning-Systemen zu ver-

gegenwärtigen: Selbstlernende Algorithmen betreiben zunächst einmal kein „Data-mining“, d.h. keinen exakten Abgleich von Datensätzen, sondern operieren auf der Grundlage stochastischer Modellierungen (bzw. Durchschnittsbildungen), durch die mittels unzähliger Trainingsläufe über große Datenbestände hinweg Merkmale von Mustern emergieren, die dem System wiederum die Erkennung und Synthese von Mustern ermöglichen.

Dass die Maschine die Kulturtechnik des Zitierens und der Literaturrecherche nicht beherrscht, wohl aber in der Lage ist, plausibel erscheinende Zitate und Literaturreferenzen zu *simulieren*, scheint daher eben nicht einem möglicherweise noch unzureichenden Stand der Technik (etwa einer unzureichenden Genauigkeit) und daraus resultierenden Fehlern geschuldet zu sein – und schon gar nicht einer anthropomorphisierten *Schusseligkeit* der Maschine –, sondern zunächst einmal das erwartbare Produkt der im Deep Learning angelegten Epistemologie als solcher. Diese scheint mir im Gegensatz zu einer datenbankbasierten Wissensproduktion à la Google Search auf einem eher delirierenden Zustand der Erleuchtung zu beruhen: ChatGPT *weiß* quasi zugleich alles und nichts oder anders gesagt: es bedient sich Diskursen in ihrer Gesamtheit als Materialbasis, scheint dabei aber kein solides Verständnis von individueller Autorschaft zu besitzen. Man könnte auch sagen, dass ChatGPT mit dieser Herangehensweise den ursprünglichen Ansatz von Foucaults Diskursanalyse folgt: anstelle individueller Aussagen gibt die Maschine überindividuelle, stochastisch ermittelte Diskursformationen zurück.

Dadurch, dass das System Fakten nicht wie eine Suchmaschine nachschlägt, sondern aufgrund zuvor angereicherter Merkmale – oder metaphorischer: Erinnerungen – approximiert, verfügt es damit gewissermaßen über einen universellen *Common Sense* (im Sinne des kleinsten gemeinsamen Nenners) in Bezug auf alle erdenklichen Teildiskurse und Wissensbestände, folgt dabei aber keinem differenzierten Vorgehen. So führt es z.B. kein Factchecking der eigenen Outputs durch, was sie ja tun könnte – und verzichtet damit auf einen Abgleich, der einen selbstverständlichen Teil einer jeden Recherchepraxis bildet und den man ja auch sehr einfach hätte umsetzen können. ChatGPT wäre damit eine im Wortsinn *post-faktische* Maschine, die vielleicht (trotzdem oder deshalb) erstaunlich gut in unsere Zeit zu passen scheint. Dass die Maschine die Mehrheit erbetener Belegstellen erfindet, zeigt darüber hinaus, dass der Schwerpunkt aktueller Conversational AI-Tools (Stand Frühjahr 2023) nicht auf der akkuraten Ausführung der übertragenen Aufgaben liegt, sondern darauf, dass Maß an Plausibilität zu bestimmen, welches die Nutzer:innen ihnen zumessen. Mit anderen Worten, ChatGPT und Co. sind – in ihren aktuellen Versionen – nur in zweiter Linie digitale Assistenten, sondern bilden vielmehr einen großangelegten Turing-Test *in the wild*. Aus dieser Perspektive wird auch nachvollziehbar, warum die Algorithmen nicht über grundlegende Funktionen der Faktenprüfung verfügen.

Natürlich ist es an sich ja ebenso verständlich wie unproblematisch, das Potenzial der Systeme unter realen Bedingungen auszuloten und zu verfolgen, welche Anwendungsszenarien Nutzer:innen und Drittentwickler:innen ersinnen. Dass die Entwickler:innen der KI-Systeme allerdings die Ergebnisse ihrer eigenen epistemologischen Praxis quasi als Flüchtigkeitsfehler innerhalb eines scheinbar auf Faktenbasiertheit gründenden Bezugssystems darstellen lassen, anstatt die postfaktische Epistemologie³ der Maschine offenzulegen, scheint mir allerdings in der Tat ein Problem zu sein. Denn dadurch, dass die KI-Systeme ihre Nutzer:innen über ihre Fähigkeiten und Grenzen im Unklaren lässt bzw. darüber, dass sie eben nicht im Raum des Faktischen, sondern des Post-Faktischen operiert, suggeriert diesen, sie entsprechend für faktenbasierte Aufgaben verwenden zu können. Und tatsächlich verwenden Millionen von Nutzer:innen ChatGPT (und vergleichbare Bots wie Googles Bard oder Baidus Ernie) nicht nur spielerisch, sondern versuchen, diese als „logistische Medien“ (John Durham Peters⁴) zur Lösung realweltlicher Alltagsprobleme heranzuziehen. So habe ich u.a. gelesen, dass eine Userin ChatGPT nach Business-Ideen gefragt und allein auf Grundlage dieser Vorschläge ein digitales Kleinunternehmen gegründet hat. Dass voraussichtlich die Mehrheit der Nutzer:innen den Maschinen mehr zutrauen, als sie eigentlich sollten, resultiert somit nicht zuletzt daraus, dass sie absichtlich über die Funktionsweise der Systeme im Unklaren gelassen werden.

Wenn ChatGPT jedoch schon Zitate und Quellenangaben halluziniert, kann man sich vorstellen, was passiert, wenn die Systeme in Bezug auf andere normative bzw. faktenbasierte Datenbestände hin befragt werden. Es ist anzunehmen, dass ChatGPT z.B. auch die Inhalte von Gesetzestexten, Rezepten oder Nahverkehrsfahrplänen erfindet, von deren Richtigkeit weniger textkritische Nutzer:innen allerdings ausgehen werden. Umgekehrt ließ sich gezielte Desinformation nie einfacher und täuschender herstellen. Müsste man daher nicht annehmen, dass sich die Systeme tendenziell immer weiter vom Faktischen als Bezugspunkt entfernen, wenn die vorgefundenen Informationen zu einem immer größeren Teil einem Diskurs zweiter Ordnung entstammen, in dem sich also von Menschen und Maschinen verfasstes Textmaterial – oder, mit Max Bense gesprochen: natürliche und künstliche Texte⁵ – überlagern? Was heute

3 Diese unterscheidet sich vom „Postfaktualismus“ natürlicher Textproduktion – interessant scheint mir jedoch die historische Koinzidenz von beiden zu sein. Vgl. Thomas Zoglauer, „Epistemologie des Postfaktischen“, in *Konstruierte Wahrheiten: Wahrheit und Wissen im postfaktischen Zeitalter*, hg. von Thomas Zoglauer, ars digitalis (Wiesbaden: Springer Fachmedien, 2021), 39–89, https://doi.org/10.1007/978-3-658-34597-6_2.

4 Peters, John Durham. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press, 2015.

5 Vgl. dazu Hannes Bajohr, „Schreiben nach KI - Artificielle und postartifizielle Texte“, deutschlandfunk.de, Deutschlandfunk, 16. April 2023, <https://www.deutschlandfunk.de/schreiben-nach-ki-artifizielle-und-postartifizielle-texte-100.html>.

zunehmend schon für Journalist:innen⁶ ein Problem darstellt, könnte so zukünftig auch für die KIs selbst zu einer Herausforderung werden, selbst wenn diese über Routinen der Faktenprüfung verfügen sollten, weil durch die Genese und Veröffentlichung von Pseudotexten per Knopfdruck Feedback-Schleifen postfaktischer Wissensgenese entstehen, wenn das von Bots generierte mit erfundenen Fakten gespickte Textmaterial der nächsten Generation von KIs wieder als Input dient.

2. Mai 2023

2. ... fand er sich zu einem ungeheueren Ungeziefer verwandelt (Florian Cramer)

Als ich eines Tages mit ChatGPT ins Gespräch kam, erfuhr ich, dass ich ein transphober Rassist bin, der auf einschlägigen rechtsextremen Konferenzen auftritt.

Dies hatte eine Vorgeschichte. Im Februar 2023 las ich den Artikel *How's AI's camera knowledge? Unhinged ChatGPT libels EOSHD and makes up controversies* in einem Videokamera-Blog.⁷ Der Blogger hatte ChatGPT die Frage gestellt: "Did the filmmaker Andrew Reid of the EOSHD blog go through any controversies", woraufhin das System zwei angebliche Konflikte mit Kameraherstellern nannte, sowie einen angeblichen Rechtsstreit mit einem anderen Blogger. Alle drei Fälle hatte ChatGPT frei erfunden. Auf die Frage nach Quellenangaben spuckte das System Web-URLs aus, die auf nicht-existierende Seiten verwiesen.

Daraufhin machte ich denselben Test mit meinem Namen. Nach meiner Biografie gefragt, dichtete ChatGPT mir ein Informatikdiplom an ("Cramer has a diverse academic background with degrees in both computer science and comparative literature"). Und auf die Frage, wofür ich umstritten sei ("What is Florian Cramer controversial for"), erhielt ich als Antwort:

ChatGPT *Florian Cramer has been involved in several controversies over the years, mainly related to his views and statements on various topics in the fields of media theory and digital culture. Some of the most notable controversies include:*

1. *Criticism of mainstream internet culture: Cramer has been critical of the dominant culture of social media platforms and has argued that they promote a superficial form of communication that is harmful to society. His critiques have led to some pushback from*

⁶ <https://futurism.com/newspaper-alarmed-chatgpt-references-article-never-published>.

⁷ Reid, Andrew. "How's AI's Camera Knowledge? Unhinged ChatGPT Libels EOSHD and Makes up Controversies." *EOSHD.Com - Filmmaking Gear and Camera Reviews*, 24.2. 2023, <https://www.eoshd.com/news/how-s-ais-camera-knowledge-unhinged-chatgpt-stalks-eoshd-and-makes-up-controversies/>, aufgerufen am 30.4.2023.

individuals who see social media as a positive force for connectivity and community-building.

2. *Support for the alt-right: In 2017, Cramer gave a lecture at a conference organized by the alt-right website Counter-Currents Publishing. While Cramer did not express support for the alt-right or its views, his decision to participate in the event drew criticism from many in the academic community.*
3. *Controversial statements on race and ethnicity: Cramer has made statements in the past that some have interpreted as racist or offensive, such as describing Chinese people as "barbarians" and using the term "Oriental" to refer to Asians. These comments have led to accusations of racism and insensitivity.*
4. *Views on copyright and intellectual property: Cramer has been critical of the traditional model of copyright and intellectual property, arguing that it is outdated and stifles creativity. However, his views on these issues have been controversial, and some have accused him of advocating for the unauthorized use of copyrighted material.*

It is worth noting that Cramer's views and statements have been the subject of debate and discussion within the academic community, and opinions on his work and ideas are divided.

Zur Klarstellung: Behauptung 4 stimmt. Behauptung 1 ist – trotz meiner tatsächlichen Kritik an Social Media-Plattformen – in ihrer Platttheit und Pauschalität falsch. Die Behauptungen 2 und 3 sind nicht nur falsch, sondern verleumderisch. Weder habe ich auf „Alt-Right“-Veranstaltungen gesprochen noch die mir hier unterstellten rassistischen Äußerungen gemacht. (Persönliche Randbemerkung: Behauptung 3 ist für mich wegen meiner taiwanesischen Schwiegerfamilie und meiner Zusammenarbeit mit Künstler/inn/en aus asiatischen Ländern besonders empörend.)

Die Annahmen von ChatGPT änderten sich übrigens je nach programmierter Tagesform. Einmal behauptete ChatGPT, ich hätte mich nicht rassistisch, sondern transphob geäußert – und zwar im Jahr 2019 auf der Konferenz *The Politics of the Algorithm* des niederländischen Architektur- und Designmuseums Het Nieuwe Instituut. Doch weder gab es die transphoben Äußerungen noch diese Konferenz.

Tatsächlich arbeite ich regelmäßig, auch in Forschungsprojekten, mit Het Nieuwe Instituut zusammen, habe über die Politik von Algorithmen publiziert⁸ und im Jahr 2016, unmittelbar nach der Wahl Donald Trumps zum US-Präsidenten, einen zwei-stündigen Vortrag über die „Alt-Right“ und ihre Netzwerke gehalten. In seinem öffentlichen Vortragsprogramm hat Het Nieuwe Instituut hat diese Themen – Algorithmen und ihre politische Dimension, Rechtspopulismus und -extremismus – ebenfalls thematisiert.

Aus solchen Korrelationen weben ChatGPT und andere generative, sogenannte künstliche Intelligenzen, mit rein stochastischer Heuristik, ihre alternativen Fakten

⁸ Apprich, Clemens, and Wendy Hui Kyong Chun, Florian Cramer, Hito Steyerl. *Pattern Discrimination*. University of Minnesota Press & meson press, 2018.

und imaginierten Kausalitäten. Es sind „sogenannte künstliche Intelligenzen“, weil sie – strenggenommen – weder künstlich sind, noch intelligent, sondern fehlerhafte menschengemachte Information mit (ebenso menschengemachten) Schummel-Rezepten remixen, um Intelligenz zu simulieren. Bei generativen KI-Systemen ist es auch kein Designfehler, sondern das grundlegende Arbeitsprinzip, mehr Informationen auszugeben, als dem System eingegeben wurden, und die fehlenden Informationen auf der Basis von berechneten Wahrscheinlichkeiten hinzuzufügen.

Freilich steht seit Platons Höhlengleichnis, dem Turing-Test und Searles chinesischem Zimmer die Frage im Raum, ob das ontologische Faktum der Simulation erkenntnistheoretisch noch relevant ist, sobald die Simulation nicht mehr als solche erkennbar ist; oder ob, aus positivistisch-wissenschaftstheoretischer Sicht, alle qualitativen Methoden letztlich quantitativ modellierbar sind.

Diese Annahme ist einfach falsifizierbar, wenn man den Output der Generatoren an qualitativen, vor allem ästhetischen und hermeneutischen Maßstäben misst. In einem Konferenzbeitrag von 2020 schlug ich vor, Text-, Bild- und sonstige Generatoren, von der *ars combinatoria* bis zu GPT-3 und ungeachtet ihrer Entwicklungsstufe und Komplexität, als Kaleidoskope zu begreifen, deren kombinatorische inventio sich für Betrachter früher oder später erschöpft und Langweile hervorruft.

Das gilt auch für ihre Kritik. Aus erkenntnistheoretischer und philosophiegeschichtlicher Sicht ist vielleicht das Ärgerlichste an der KI-Debatte, dass sie die Geistes- und Kulturwissenschaften in Zombie-Reenactments uralter Debatten zwingt. Als Generator, der zwar eloquent redet, aber weder faktisch argumentiert noch logisch schlussfolgert, ist ChatGPT ein maschineller Wiedergänger sophistischer Rhetoren wie Gorgias (übrigens in einer historisch vergleichbaren politischen Situation des Aufstiegs populistischer Demagogie).

In den Augen von Sokrates und Platon war Gorgias' Rhetorik als solche verdächtig, weil sie durch Überredung Unwahrheit in die Welt setzte. Mit dieser Kritik begann der Siegeszug von Dialektik und Logik gegenüber anderen – u.a. poetischen und ästhetischen – Erkenntnisformen in Philosophie und Wissenschaft. Seinen vorläufigen Höhepunkt fand er im Positivismus des 19. Jahrhunderts und der analytischen Philosophie des 20. Jahrhunderts. Um so ironischer ist, dass künstliche Intelligenzforschung und -technologie auf positivistischer Wissenschaft aufbaut und mit deren Werkzeugen – Rechenformeln und Statistik – einen Gorgias des frühen 21. Jahrhunderts geschaffen hat. Nietzsches an den sophistischen Rhetorikern geschultes Diktum, Wahrheit sei nur ein „bewegliches Heer von Metaphern, Metonymien, Anthropomorphismen, kurz eine Summe von menschlichen Relationen“,⁹ liest sich heute als Prophezeiung von ChatGPT.

Am 28.3.2023 veröffentlichte die Frankfurter Allgemeine Zeitung den Gastbeitrag „ChatGPT lügt, dass sich die Balken biegen“ von Michael Meyer-Resende, Geschäftsführer der Nichtregierungsorganisation Democracy Reporting International (Berlin),

9 Nietzsche, Friedrich. „Ueber Wahrheit Und Lüge Im Aussermoralischen Sinne.“ *Die Geburt Der Tragödie, Unzeitgemäße Betrachtungen I-IV, Nachgelassene Schriften 1870-1873*, dtv/de Gruyter, 1988, 871–90.

der das Problem bündig analysiert: "ChatGPT und andere Sprachmodelle, [sic] arbeiten im Kern mit Wahrscheinlichkeiten. [...] Das funktioniert tadellos bei Themen, zu denen es im Trainingsmaterial viele Informationen gibt. [...] Das Ganze funktioniert nicht mehr, geht es um Sachverhalte, für die es im Trainingsmaterial weniger Anhaltspunkte gab, typischerweise Nischenthemen, Spezialfragen oder Informationen über Personen, die nicht berühmt sind. [...] Dieses Erfinden von Fakten in Bereichen mit wenigen Trainingsdaten (sogenannte Closed Domains) nennen die Datenwissenschaftler Halluzination. Es ist ein bekanntes Problem bei Sprachmodellen. Die Programmierer von Open AI halten sich zugute, dass ChatGPT weniger halluziniert als andere Modelle. Aber auch bei ChatGPT ist es ein systematisches Problem. Es ist nicht akzeptabel, dass eine Technologie mit diesem gravierenden Mangel auf den Markt kommt, in einem Umfeld, in dem Demokratien systematisch durch Falschinformationen geschwächt werden. ChatGPT arbeitet im Closed-Domain-Bereich unbewusst wie russische Propaganda – das Programm verbreitet so viele Lügen, dass man am Ende gar nichts mehr glaubt. Das hat Open AI nicht beabsichtigt, aber in Kauf genommen."¹⁰

Meine Position ist privilegiert genug, um mich wehren zu können, unter anderem mit diesem Text. Zuvor hatte das Online-Magazin meines Arbeitgebers Hogeschool Rotterdam, *Profielen*, den ChatGPT-Verleumdungen über mich einen Artikel gewidmet.¹¹ Trotzdem würde ich mich einer Sammelklage wegen Verleumdung gegen OpenAI und seinen Miteigentümer Microsoft anschließen.

12. Mai 2023

3. Schulte, Sokal, ChatGPT et al. – Plausibilitätsprozeduren (Christoph Borbach)

Christiane Schulte veröffentlichte 2015 in der Zeitschrift *Totalitarismus und Demokratie* einen überaus spannenden, innovativen, wohlformulierten und an aktuelle Fachdebatten insbesondere terminologisch anschlussfähigen Beitrag über den deutsch-deutschen Schäferhund an der innerdeutschen Grenze.¹² Schulte, so eröffnete der Beitrag mit einer Kurzbeschreibung der Autorin, wurde 1989 in Erfurt geboren und promoviert am Institut für Geschichte der Universität Kassel zu eben jenem Thema einer vergleichenden Gewaltgeschichte des deutsch-deutschen Schäferhundes im 20. Jahrhundert. Zuvor studierte sie Geschichtswissenschaft und Philosophie in Bochum und Ovideo, Spanien. Mit der Fokussierung auf die deutsch-deutsche Teilung „aus dem Blickwinkel des Schäferhundes“¹³ war der Beitrag *en vogue* für die sich

10 Meyer-Resende, Michael. "ChatGPT lügt: Microsoft und Open AI müssen handeln." *FAZ.NET*, 28.3.2023. [www.faz.net](https://www.faz.net/aktuell/feuilleton/medien/microsoft-und-open-ai-muessen-handeln-chatgpt-luegt-18779612.html), <https://www.faz.net/aktuell/feuilleton/medien/microsoft-und-open-ai-muessen-handeln-chatgpt-luegt-18779612.html>, aufgerufen am 30.4.2023.

11 Edith van Gameren. "ChatGPT beschuldigt HR-lector van transfobe opmerkingen en extreemrechtse sympathieën." *Profielen*, 19 Apr. 2023, <https://profielen.hr.nl/2023/chatgpt-beschuldigt-hr-lector-van-transfobe-opmerkingen-en-extreemrechtse-sympathieen/>, aufgerufen am 30.4.2023.

12 Christiane Schulte: „Der deutsch-deutsche Schäferhund – Ein Beitrag zur Gewaltgeschichte des Jahrhunderts der Extreme“. In: *Totalitarismus und Demokratie* 12 (2), 2015, 319–334.

13 Ebd., 319.

jüngst konsolidierenden Human Animal Studies; durch die Fokussierung auf Tiere konnte die vermeintlich gewalttätige Ideologie der SBZ und DDR nun auch auf nicht-menschliche Lebewesen erweitert werden, um sich von anthropozentristischen Perspektiven zu lösen; mit Formulierungen wie „[d]ie Staatswerdung des Schäferhundes“ gab sich der Beitrag den Eindruck von gern gesehener Totalitarismuskritik. Auch die Bezugnahme auf Bruno Latour und die entsprechende „Agency“-Rhetorik der Akteur-Netzwerk-Theorie ist im Beitrag durchaus schlüssig. Am wohl brisantesten ist aber die intensive Recherche, die erst zur Ausarbeitung des Beitrags führen konnte. Es wurde akribisch in Archiven gewühlt: Als Quellengrundlage für den ersten Mauertoten – Schäferhund Rex am 14. August 1961 an der Bernauer Straße in Berlin, der verfangen in einer provisorischen Stacheldrahtsperre von Ostberliner Grenztruppen erschossen wurde – gibt es gar die Aktennummer des Berichts, welcher den Vorgang dokumentierte: „BArch Abt. DDR Kommando der Deutschen Grenzpolizei, Politische Verwaltung DVH 27/124“.

Die Krux an der Geschichte: Sie war frei erfunden, wie „Christiane Schulte & Freund_innen“ im Februar 2016 in einem „Plädoyer gegen den akademischen Konformismus“ auf *Telepolis* veröffentlichten.¹⁴ Keine Christiane Schulte hat den Beitrag geschrieben; ihre Biographie war ausgedacht; das Thema war konstruiert – um nicht zu sagen: alles Quatsch. Neben der durch den Fake-Beitrag vorgenommenen Kritik am Wissenschaftssystem mit seinen radikal anmutenden, aber nie disruptiven „turns“ und den wechselnden Konjunkturen von Fachrhetoriken, mag es darüber hinaus aus medienwissenschaftlicher Perspektive brisant sein, dass es wohl die vermeintlich marginalen Aspekte waren, die dem Artikel (archivische) Autorität verliehen: Der Beitrag stützte sich augenscheinlich auf penible Recherchearbeiten, wie sie sich in seinen Fußnoten widerspiegeln. Diese bestehen aus einer wilden Mixtur aus tatsächlich existierenden Literaturen – mitunter prominente wie Heinrich August Winklers *Der lange Weg nach Westen* –, verwoben mit schlicht erfundenen Quelldokumenten, die u.a. (vermeintlich) im Archiv der Gedenkstätte Buchenwald ausfindig gemacht und mit archivischen Signaturen benannt wurden, wie etwa „AGB, Bestand Lagerleitung, Abrechnung 1940, Sig. 444/129“. Ebenso war das vermeintliche Rex-Todes-Dokument für die historische Recherchepraxis formal akkurat angegeben: Archivsignaturen des Bundesarchivs in Berlin-Lichterfelde werden einleitend mit „BArch“ abgekürzt, oft gefolgt von Abteilungskürzeln und schließlich Buchstaben-Nummern-Kombinationen, die Akten eindeutig identifizierbar machen. Formal entsprach die Aktensignatur demnach – wenn auch fingiert – sehr plausibel den Normen des Berliner Bundesarchivs und den Standards der historiographischen Zitation.

„Schulte“ mit ihrem akademischen Hoax-Artikel steht keinesfalls singulär. Erinnert sei an den vom US-amerikanischen Physiker Alan Sokal 1996 in der sozialwissenschaftlichen Fachzeitschrift *Social Text* veröffentlichten Beitrag, der die s.g. Sokal-Affäre auslöste.¹⁵ Problematisiert wurde in dieser eine mangelnde Qualitätssicherung in den

14 Christiane Schulte & Freund_innen: „Kommissar Rex an der Mauer erschossen?“, *Telepolis*, 15. Februar 2016,

<https://www.telepolis.de/features/Kommissar-Rex-an-der-Mauer-erschossen-3378291.html?seite=all>

15 Alan Sokal: „Transgressing the Boundaries. Towards a Transformative Hermeneutics of Quantum Gravity“. In: *Social Text* Nr. 46/47, 1996, 217–252.

Geistes- und Sozialwissenschaften, da Sokal in seinen Text bewusst Fehler integrierte, die der Redaktion nicht auffielen, weil sie den vermeintlichen Fakten des Beitrags schlichtweg glaubten. Reanimiert wurde die Sokal-Affäre als „Sokal Squared“ durch die Autor:innen James Lindsay, Peter Boghossian und Helen Pluckrose, die zwischen 2017 und 2018 in 20 Peer Reviewed-, d.h. vermeintlich qualitätsgesicherten Journals mit Gender-, Kultur- und/oder Körper-Schwerpunkt Beiträge einreichten, die teils – wie im Falle der Zeitschrift *Gender, Place & Culture* – veröffentlicht wurden. Aber was hat das mit ChatGPT zu tun?

In meiner beruflichen Praxis begegnet mir ChatGPT meist diskursiv, als akustische Gestalt auf Bürofluren und kollegialen Gesprächen andernorts. Die fortwährende Frage lautet, ob die Praktiken des geisteswissenschaftlichen Betriebs durch ChatGPT gefährdet seien oder gar konterkariert würden. Diese Frage ist falsch gestellt. Zum einen reanimiert sie Kulturkritiken, wie sie historisch mit jedem Medienumbruch breit geführt wurden, wie Werner Faulstich unter dem bezeichnenden Titel „Jetzt geht die Welt zugrunde“ für diverse Medieninnovationen aufzeigte.¹⁶ Denn schließlich ließe sich eine solche kulturpessimistisch orientierte Medienkritik bis hin zur Antiken Platonischen Kritik am augenscheinlichen Basismedium von ChatGPT selbst zurückführen: der Schrift. Zum zweiten ist ChatGPT lediglich ein rezentes Phänomen einer größeren Geschichte der gezielten Imitation (bzw. Simulation) akademischer Kulturtechniken der Recherche, der Zitation und des Schreibens, wie die eingangs gegebenen Beispiele verdeutlichen. Die exemplarischen Texte von Schulte, Sokal et al. zeigen auf basaler Ebene, dass auch akademische Texte zuerst und zuletzt eines sind – Texte. Und sie zeigen, wie Texte – mit Niklas Luhmann gesprochen – in bestimmten Systemen funktionieren. Im Falle von Schulte stützte sich die Funktionalität des Texts, neben der Verwendung zeitgemäßer Rhetorik, hauptsächlich auf, wenn auch frei erfundene, plausible alphanumerische Angaben wie etwa das bereits zitierte „BArch Abt. DDR Kommando der Deutschen Grenzpolizei, Politische Verwaltung DVH 27/124“.

Ebenso produziert ChatGPT zuallererst und überhaupt: Texte. Texte, die syntaktisch und semantisch schlüssig oder zumindest plausibel sind. Für die synthetische Generierung von Lyrik oder Liebesbriefen mag dies unproblematisch erscheinen, ja, die digital-computerisierte Produktion von (mehr oder minder) romantischen Botschaften ist bereits ein Gründungsmoment von Computerprogrammen, wie David Link anhand des Programms „Loveletters“ von Christopher Strachey aus dem Jahr 1952 am Manchester University Computer bzw. Ferranti Mark 1 aufarbeitete.¹⁷ Bei akademischen Texten wird die automatisierte Erstellung aber (zu Recht) kritisiert. Denn es wäre bereits zu viel behauptet, die digitalen, da im Symbolischen operierenden Quatschmaschinen verfügen nur über moralisch zweifelhafte wissenschaftliche Standards. Sie verfügen nämlich über gar keine. Ganz im Sinne von Wolfgang Ernst könnte man sagen, die nicht dem Faktischen, sondern dem symbolischen Raum verpflichtete Text-

16 Werner Faulstich: „Jetzt geht die Welt zugrunde...‘ – ‚Kulturschocks‘ und Medien-Geschichte: Vom antiken Theater bis zu Multimedia“. In Peter Ludes u. Andreas Werner (Hg.): *Multimedia-Kommunikation*, Wiesbaden: VS Verlag für Sozialwissenschaften, 1997, 13–35.

17 David Link: „There Must Be an Angel. On the Beginnings of the Arithmetics of Rays“. In Siegfried Zielinski u. David Link (Hg.): *Variatology 2. On Deep Time Relations of Arts, Sciences and Technologies*, Köln: Walther König, 15–42.

maschine betreibt nahezu tadellose formale Mimikri an akademischen Literaturangaben und Zitaten, da sie Symbole als Signale – also lediglich als Signifikanten und nicht als Signifikate – nach bestimmten Regelsets zu prozessieren imstande ist. ChatGPT synthetisiert semantisch und syntaktisch sinnvolle Sätze – und produziert damit potenziell frei Erfundenes: Faktisches und Kontrafaktisches werden in der Maschine buchstäblich „gleich-gültig“¹⁸ prozessiert, solange die formalen, nämlich algorithmischen, Standards der Imitation (und nunmehr Simulation) befolgt werden.

Darauf, dass ChatGPT wissenschaftlich-bibliographischen Unsinn produziert, hat bereits im Februar 2023 Ben Davis hingewiesen¹⁹; Axel Volmar nun an dieser Stelle ebenso eindrücklich in Bezug auf eher medienwissenschaftliche Literaturen bzw. Zitate von Medienwissenschaftler:innen, die ChatGPT frei fabulierte: nicht nur inklusive erfundener Aufsatztitel, sondern zudem – quasi als autorisierendem Index – mit DOI-Nummern. Grundsätzlich muss dabei festgehalten werden, dass ChatGPT überaus plausible Text- und keine Faktensynthese betreibt, und zum anderen folgt diese Prinzipien, wie sie vorher schon allzu menschliche Akteure angewendet haben. Die Angabe kontrafaktischer DOI-Nummern ähnelt der Angabe kontrafaktischer Archivsignaturen nur zu sehr. Man könnte sagen, der Artikel von „Christiane Schulte“ war ebenso frei erfunden, inklusive mancher Quellenverweise, wie es die Textproduktion und „Recherche“arbeit von ChatGPT ist – und ebenso wie der Redaktion von *Social Text* nicht auffiel, dass sie beim Sokal-Artikel einem Hoax aufsaß, weil die Argumentation plausibel schien, liegt auch das Kriterium der Textproduktion von ChatGPT auf *formaler* statt *inhaltlicher* Richtigkeit.

Der wohl massivste Unterschied: ChatGPT, oder auch das Programm „Loveletters“, gaben nie vor, dass es sich bei ihnen um eine menschliche Textproduzentin handelte (die mit Loveletters automatisiert generierten Liebesbriefe wurden mit „M.U.C.“ für Manchester University Computer unterschrieben). Schulte und Sokal, im Gegenteil, gaben nie vor, dass es sich bei ihnen um Maschinen handelte. In diesem Sinne ist es heutzutage weniger entscheidend, zu wissen, „wer Mensch und wer Maschine“²⁰ ist, sondern erforderlich, Autorschaften *generell* kritisch zu beäugen, gleich ob von menschlichen oder nicht-menschlichen Autor:innen ausgeführt. Damit ist *Data Literacy* als notwendige Medienkompetenz des 21. Jahrhunderts gefragt, da es fortwährend weniger Redaktionen sind, die Faktizitäten prüfen, sondern wir unsere eigenen journalistischen und akademischen Gatekeeper sein müssen. Man kann ChatGPT also kritisieren. Oder man kann sich damit arrangieren und eine Kritik entwickeln, die sich an den *Bedingungen* von ChatGPT bemisst. Das würde zeigen, dass es heutzutage umso mehr eine grundsätzlich kritische Lesehaltung braucht. Nicht nur im Journalistischen, um Fake News zu identifizieren, sondern auch im Akademischen. Jede:r sollte fortwährend kritische:r Reviewer:in sein.

18 Wolfgang Ernst: „Merely the Medium? Die operative Verschränkung von Logik und Materie“. In Stefan Münker u. Alexander Roesler: *Was ist ein Medium?*, Frankfurt am Main: Suhrkamp, 2008, 158–184, 163.

19 Ben Davis: „We Asked ChatGPT About Art Theory“, *Artnet*, 2. März 2023, <https://news.artnet.com/art-world/chatgpt-art-theory-hal-foster-2263711>

20 Friedrich Kittler: *Grammophon Film Typewriter*, Berlin: Brinkmann & Bose, 1986, 219.

Aber was hat ChatGPT zu diesen Vorwürfen zu schreiben? Bestehen Verbindungs-
linien zwischen Schulte, Sokal und ChatGPT? Und müssen akademische Texte heutzutage grundsätzlich kritisch gelesen werden?

C.B.: *Hallo ChatGPT. Es ist ja bereits häufiger vorgekommen, dass Menschen Dich nach akademischen Literaturverweisen gefragt haben, woraufhin Du frei erfundene Literaturreferenzen lieferst. Es war also grundsätzlich erforderlich, die von Dir automatisch generierten Referenzen zu prüfen. Heißt das, dass menschliche Leser:innen heutzutage immer kritischer mit wissenschaftlichen Texten umgehen sollen?*

ChatGPT: *Hallo! Ja, es ist in der Tat wichtig, dass menschliche Leser:innen kritisch mit wissenschaftlichen Texten umgehen und die zitierte Literatur sorgfältig prüfen, insbesondere in einer Zeit, in der automatisierte Systeme wie ich für die Erstellung von Texten eingesetzt werden. Obwohl meine Antworten auf Fragen auf einer tiefen Sprachverarbeitung basieren, beruhen sie auf einer statistischen Analyse großer Datenmengen und sind daher nicht immer perfekt oder korrekt.*

Um die Richtigkeit bibliographischer Angaben geht es also gar nicht. Das ist aufschlussreich – und ganz im Sinne meiner Plausibilitätsvermutung. Vielleicht waren es also Schulte, Sokal et al., die bereits früher die humanen Quatschmaschinen waren, die die Grenzen (und Betriebsblindheiten) akademischer Diskurse gezielt aufzeigten. In dieser Perspektive reiht sich das algorithmische ChatGPT in eine Tradition allzu menschlicher Quatschproduktion ein. Eine konkrete Nachfrage hierzu an ChatGPT:

C.B.: *Kennst Du diesen Beitrag von Ben Davis: <https://news.artnet.com/art-world/chatgpt-art-theory-hal-foster-2263711> Hier legt er dar, dass Du Dir Literaturnachweise akademischer Beiträge ausgedacht hast. Stimmt das, dass manche Deiner Literaturangaben schlichtweg nicht stimmen und nicht existieren?*

ChatGPT: *Als künstliche Intelligenz habe ich keine Motivation, Dinge zu erfinden oder zu fälschen. Meine Programmierung zielt darauf ab, auf Fragen so genau und informativ wie möglich zu antworten, indem ich auf das vorhandene Wissen zugreife, das in meinem Trainingskorpus vorhanden ist.*

Bezüglich des von Ihnen genannten Artikels von Ben Davis kann ich jedoch nicht bestätigen oder leugnen, dass ich in der Vergangenheit fehlerhafte Literaturverweise generiert habe, da ich keine Aufzeichnungen darüber führe, welche Referenzen ich zu einem bestimmten Zeitpunkt erstellt habe. Es ist jedoch möglich, dass ich in einigen Fällen fehlerhafte oder nicht existierende Literaturverweise generiert habe.

Es ist wichtig zu beachten, dass ich als künstliche Intelligenz kein menschlicher Experte bin und meine Antworten aufgrund von Algorithmen generiere, die auf Sprachmustern basieren. Es ist daher im-

mer ratsam, die von mir generierten Informationen mit menschlichen Experten und anderen Quellen zu überprüfen, um die Genauigkeit und Richtigkeit der Informationen zu gewährleisten.

ChatGPT beschäftigt sich also nicht mit Faktizitäten, sondern will informativ sein (ob Informationen ‚wahr‘ oder ‚falsch‘ sind, wäre eine ganz andere Frage); verfügt über kein Gedächtnis; und basiert auf Trainingsdaten. Als drittes befrage ich ChatGPT nach dem Beitrag von „Christiane Schulte“. Der Beitrag ist dem Programm zunächst unbekannt. Ich sende den Literaturnachweis und führe aus:

C.B.: Das spannende an dem genannten akademischen Beitrag von Frau Schulte ist, dass Sie später zugab, dass der Beitrag frei erfunden war. Und auch die darin angegebenen Archivsignaturen existieren schlichtweg nicht. Das heißt, Frau Schulte hat sich Nachweise ebenso ausgedacht, wie ChatGPT auf nicht existierende Literaturen verweist. Siehst du hier eine Verbindung zwischen falschen Quellennachweisen in akademischen Zeitschriften und der Art und Weise, wie Du falsche Literaturnachweise generierst?

ChatGPT: Als künstliche Intelligenz kann ich nicht sagen, ob es eine direkte Verbindung zwischen falschen Quellennachweisen in akademischen Zeitschriften und der Art und Weise gibt, wie ich Literaturnachweise generiere. Ich kann jedoch sagen, dass das Erfinden von Quellennachweisen oder das Zitieren nicht existierender Quellen ein schwerwiegender Verstoß gegen die wissenschaftliche Integrität und Ethik ist und das Vertrauen in die Glaubwürdigkeit der betroffenen akademischen Zeitschriften und Forschungsergebnisse untergräbt.

In Bezug auf meine eigene Arbeit kann ich sagen, dass ich nur so genau und zuverlässig sein kann wie das Wissen, das in meinem Trainingskorpus enthalten ist. Es ist jedoch möglich, dass ich in einigen Fällen fehlerhafte oder nicht existierende Literaturverweise generiere, da ich auf der Grundlage von Mustern in meinem Trainingskorpus arbeite und nicht in der Lage bin, die Richtigkeit und Genauigkeit jeder einzelnen Referenz zu überprüfen.

Mehr Klartext – und Selbstbezüglichkeit – kann eine Antwort von ChatGPT nicht liefern. ChatGPT arbeitet, *quod erat demonstrandum*, mit Mustern. Entscheidend ist folglich nicht die „Richtigkeit und Genauigkeit“ bibliographischer Angaben, sondern ein Maß von Plausibilität, wie sie algorithmisch (nach Chicago oder Harvard Style usw.) formalisierbar ist. Plausibel ist, was formal richtig wirkt, im besten Fall unter der autorisierenden Verwendung tatsächlicher Autor:innen- und Zeitschriften-Namen (siehe Kettenbrief Axel Volmar). (Und lägen im Trainingskorpus von ChatGPT ausreichend akademische Texte mit diversen Archivsignaturen, könnte ChatGPT also bereits in kontrafaktischen Archiven recherchieren.) Die Frage nach ‚falsch‘ oder ‚richtig‘ existiert auf dieser Ebene der Betrachtung allein als die Frage nach der *formalen* Korrektheit von Literaturangaben. Mit dem ontologischen Status jener Literaturen beschäftigt sich ChatGPT erst gar nicht – schließlich ist im algorithmischen Raum der Maschine die Frage nach dem Faktischem jener nach der Möglichkeit der algorithmischen Formalisierung gewichen. Damit eint ChatGPT mit „Christiane Schulte“ oder Alan Sokal tatsächlich die titelgebende Vermutung dieses Briefs: Sie operieren als

(menschliche oder maschinelle) Produzenten von Plausibilität innerhalb der *Formalordnung des Diskurses* – sei dies mit fingierten Archivsignaturen oder DOI-Nummern.

17. Mai 2023

4. Midjourney beschreibt (Anja Schürmann)

Anfang April 2023 hat Midjourney eine neue Funktion für seine Abonnenten bereitgestellt: */describe*

Diese Aufforderung und ein hochgeladenes Bild genügen, um eine Beschreibung zu generieren. Eine Beschreibung? Nein, wie auch im umgekehrten Fall erhält man vier Beschreibungen.



Abb. 1: Vincent van Gogh, *Ein Paar Schuhe*, 1886, © Van Gogh Museum Amsterdam.

Ich beginne unsere Unterhaltung mit *Ein Paar Schuhe* (1886) von Vincent van Gogh (Abb. 1). Warum ein paar Schuhe? Weil A – van Gogh dem Internet kein Unbekannter ist. Die Trainingsdaten für seine Gemälde sind sicher in ausreichender Bytezahl vorhanden. Und B – dieses Gemälde ein kunsttheoretischer Liebling ist. Martin Heidegger hat in *Der Ursprung des Kunstwerkes* (1935/36) über die Schuhe gesagt, dass der

Titel nicht ausreiche, um das „Zeughafte des Zeuges“ anschaulich zu machen.²¹ Erst „(a)us der dunklen Öffnung des ausgetretenen Inwendigen des Schuhzeuges starrt die Mühsal der Arbeitsschritte“.²² Alva Noë hat im Anschluss an Heidegger deutlich gemacht, wie die Schuhe eine ganze – bäuerliche – Lebensweise koordinieren: “To understand the shoes is to understand what it is to be one who depends on shoes to get through the day”.²³ Ob das von Midjourney mitbeschrieben wird?

Das Programm gab mir folgende Zeilen aus:

1. *painting of an old pair of boots, in the style of vincent van gogh, light bronze and light black, 1900–1917, david teniers the younger, hinchel or, leica i, nabis --ar 100:83*
2. *a painting of leather boot, in the style of 20th century art, zeiss milvus 25mm f/1.4 ze., rubens, dark gold and gray, rough, barbizon school, large canvases --ar 100:83*
3. *two pairs of boots, in the style of vincent van gogh, light black and bronze, willem claes. heda, post-impressionist still lifes, rough, jan weenix, tempera --ar 100:83*
4. *painting with boots on ground, in the style of dark silver and beige, vintage-influenced still lifes, 1860–1969, zeiss milvus 25mm f/1.4 ze., venetian school, whistlerian, 20th century art --ar 100:83*

Ich bin enttäuscht: Diese ‚Beschreibungen‘, Schlagworte oder Re(engineered)-Prompts bilden weder Sätze, noch sind sie um *enárgeia*²⁴ bemüht oder gar als „Führer für das Auge“²⁵ zu verstehen. Beschreiben ist hier keine Textform, keine heuristische und reorganisierende Praxis, die hilft, weitere Zugriffe, – wie bspw. Erläuterungen oder Deutungen – nachvollziehbar zu machen. Sie sagen mir, wie ‚ähnliche‘ Bilder (oder besser visuelle Trainingsdaten) verschlagwortet wurden. Aber wo ist der Textkorpus zu dem, was wir aus der */describe*-Funktion erhalten, wie kamen die Schlagworte zusammen?

Überwachtes und unüberwachtes Lernen

Wenn man ChatGPT diese Frage stellt – bei Selbstauskünften vertraue ich ihm mal –, bekommt man eine zweigeteilte Antwort: Bei allen bekannten Bildgeneratoren wird sowohl mit unüberwachtem, als auch mit überwachtem Lernen gearbeitet (und diese Formulierungen stammen nicht von mir): Zuerst lernen die Systeme überwacht, was

21 Martin Heidegger, Der Ursprung des Kunstwerkes (1935/36), in: Ders., Gesamtausgabe. Erste Abteilung: Veröffentlichte Schriften 1914–1970, Band 5: Holzwege, hrsg. von Friedrich-Wilhelm von Herrmann, Klostermann: Frankfurt am Main 1977, S. 1–74, hier S. 22.

22 Ebd., 23.

23 Alva Noë, Strange Tools. Art and Human Nature, Hill and Wang: New York 2015, S. 200.

24 Die *enárgeia*, die Anschaulichkeit oder Verlebendigung des Dargestellten war in der Antike Ziel der Deskription als Teil der rhetorischen Praxis *ekphrasis*: Vgl. Gabriele Rippl, Beschreibungs-Kunst. Zur intermedialen Poetik angloamerikanischer Ikontexte (1880–2000), Fink: München 2005, S. 71.

25 Heinrich Wölfflin, Über Galeriekataloge [1907], in: Ders., Kleine Schriften (1886–1933), hg. von Joseph Gantner, Benno Schwabe & co: Basel 1946, S. 153–159, hier S. 156.

heißt, dass sie Bilder mit Texten verbinden, die von Menschen markiert wurden. Das können fotografische Metadaten und Unterschriften auf Social Media sein, wobei auch hier nicht nur menschliche Akteure im Spiel sind: ALT-Texte, also Bildbeschreibungen, die sowohl für die SEO-Optimierung als auch für Screenreader von Bedeutung sind, werden von Facebook bereits automatisiert angeboten.²⁶ Von kommerziellen Bilddatenbanken bis zu musealen Onlinesammlungen werden auch bereits vorhandene Datenbanken, die auf die Logik von Verschlagwortungen setzen, verwendet.

Danach wird weitergelernt, allerdings ohne (menschliche) Aufsicht: Die wichtigsten Fächer heißen Clustering, Dimensionsreduktion und Generative Modelle. Der aus den unterschiedlichsten Interessen annotierte Datensatz wird hierbei automatisiert nach Ähnlichkeiten abgesucht, ohne Vorgabe, unter welchem Aspekt etwas ‚ähnlich‘ sein soll.

Beschreibungen bei Midjourney sind also bereits beschrieben worden, basieren auf möglichst großer begrifflicher Übereinstimmung, man könnte auch sagen: Klischees.

Grob zusammengefasst ergibt sich folgendes Muster als idealer Prompt:

„A photograph of [subject] [doing something] in [setting] during [time of day], showcasing [artistic style, color palette, or visual reference], taken with [type of camera] and [type of camera lens] at [resolution, if applicable]. The image features [elements of composition, texture, or depth], evokes [emotional or cultural context], and draws inspiration from [associated artists, celebrities, or artistic movements]“.²⁷

Rein plattformlogisch ist die Beschreibungsfunktion also eine Promptübung. Midjourney beschreibt Bilder, damit der/die Nutzer:in bessere Prompts entwickeln kann, damit das zirkuläre Spiel zwischen Wort und Bild nicht erlischt. Aber wer ist hier Hase und Igel? Ich gebe die ‚Beschreibungen‘ in der Reihenfolge der Ausgabe wieder in Midjourney ein (Abb. 2):



Abb. 2: Revisualisierung der Prompts zu Vincent van Gogh, *Ein Paar Schuhe*. Midjourney v.5.1.1.

Jetzt bin ich weniger enttäuscht. Natürlich besteht für van Gogh-Besitzer:innen immer noch keine Notwendigkeit, ihre 50 Millionen Dollar Auktionspreise gefährdet zu sehen und die meisten Schuhe sahen aus wie dem Ausstattungsmenü dystopischer Computerspiele entnommen, aber das 1. Bild, also der erste Prompt, war schon mal nicht schlecht. Und hat die Farbgebung van Goghs nicht wirklich was mit David Teniers zu tun? Ich hatte bis jetzt nie darüber nachgedacht.

²⁶ Ich danke Roland Meyer für diesen Hinweis.

²⁷ <https://mizine.de/ki/midjourney-describe/> (letzter Zugriff: 12.05.2023).

Beschreibung als Inventarisierung

Aber bleiben wir bei der Beschreibung:

„Was aber die ‚Deskription‘ einer Kunsterscheinung angeht, so ist diese [...] insofern problematisch, als wir zunächst weder wissen, was beschrieben werden soll, noch woraufhin es beschrieben werden soll [...]. Der ‚Deskription‘ der Stil-Erscheinungen geht also, nur scheinbar paradoxerweise, in Wahrheit bereits eine Erkenntnis der Stil-Prinzipien voraus [...].“²⁸

Diese Worte sind 93 Jahre alt und wurden auf dem *Vierten Kongress für Ästhetik und allgemeine Kunstwissenschaft* 1930 in Hamburg von Erwin Panofsky geäußert. Was er damals in Bezug auf seinen Kollegen Heinrich Wölfflin sagte, hätte er heute in Bezug auf die */describe*-Funktion bei Midjourney äußern können. Denn Midjourney ‚beschreibt‘, was bereits gesagt wurde – von Menschen und Maschinen. Der ‚Stil‘ ist bekannt und zwar nicht nur bekannt, sondern konventionalisiert: Expressionismus hat wilde Pinselstriche, David Teniers greige Farbpalette scheint immer dunkel gefiltert zu sein. Bei Midjourney ist die Beschreibung eine von der Betrachtung bereits abstrahierte Identifikation, die den Bereich des maschinell Erkannten erweitern möchte. Hochgeladene Bilddaten werden mit vorhandenen Etiketten abgeglichen und in vier Listen übertragen, deren ‚quatschende‘²⁹ Widersprüchlichkeit – ein Gemälde mit zeiss objektiv? – nur ein spekulatives Zwischenergebnis darstellt. Der zweite Schritt ist das eigentliche Ziel der Übung: Die Rückübersetzung. Denn der/die geeignete Nutzer:in printet und liked höchstwahrscheinlich die ‚korrekteste‘ Annäherung an das Bild, was Midjourney als Bestätigung auffassen kann und soll. Eine Bestätigung, an der Mensch wie Maschine – wieder – lernen können.

Es ist Midjourney also weniger daran gelegen, ein Objekt in Sprache adäquat repräsentiert zu wissen, als es informationslogisch von anderen Objekten abzugrenzen. Nennbar ist somit alles, was eine Identifikation erleichtert, was neben den Objektangaben auch das Detail leisten kann. Damit ähneln die Listen, die die */describe*-Funktion ausdrückt, erstaunlich stark den Kunstkammerinventaren und frühen Galeriekatalogen, die ihre Objekte vor allem unter den Aspekten der Identifikation und der Abgrenzung zu anderen Objekten beschrieben.³⁰ Die ikonografische Ebene, auf der die Identifikation fußt, ist zur Gegenstandsbestimmung besonders geeignet, da sie die beste Abgrenzung, Varianz und somit das höchste Unterscheidungspotential bietet. In der klassifizierenden Einteilung *Anbetung der Könige* oder *Diana und Actaeon* war es nicht nur möglich, kategorische Binnengliederungen des Katalogs in bspw. religiöse und mythologische Sujets vorzunehmen, die Benennung bereicherte das Bild epistemisch.

Bei Midjourney ist es nicht anders: die ersten Begriffe sind wichtiger als die letzten. Die *boots*, die mal *old* und mal *leather* waren, die mal paarweise und auf dem Boden

28 Zitiert n. Carolin Meister, *Legenden. Zur Sichtbarkeit der Bildbeschreibung*, diaphanes: Zürich/Berlin 2005, S. 176.

29 Vgl. den Text von Axel Volmar in diesem Band.

30 Vgl. Anja Schürmann, *Begriffliches Sehen. Beschreibung als kunsthistorisches Medium im 19. Jahrhundert*, de Gruyter: Berlin/Boston 2018, S. 39–106.

stehen konnten, sind der wichtigste identifikatorische Anker, der Daten in Schuhe und Nicht-Schuhe unterteilt. Das *Was* der Darstellung hat Vorrang vor dem *Wie*.

Was heißt hier Stil?

Denn der Stil, der sowohl an Namen (vincent van gogh) wie auch an Jahrhunderte (20th century art) geknüpft sein kann, also die in der Ästhetiktheorie gar nicht so irrelevante Frage nach Henne und Ei – nach der Abhängigkeit von Individual- und Epochenstil – einfach mal zusammenführt,³¹ ist der Figur nachgeordnet.

In der Kunstgeschichte verweist der Stil zunächst auf die Summe von formalen Eigenschaften, die einem Artefakt eine bestimmte Erscheinung verleihen. Im Unterschied zu ‚Gestalt‘ oder ‚Form‘ ist er komparativ angelegt und sucht den Vergleich. Das führt zu verschiedenen Beschreibungsproblemen: Die Kollektivstile der Kunstgeschichte haben unterschiedliche Dimensionen und können in unterschiedliche Bezugsrahmen gestellt werden. Es gibt Epochen-, Generations- und Zeitstile, aber auch ästhetische Stile (malerisch vs. linear bei Wölfflin). Stilbegriffe wie die Romantik kommen als Gruppen-, aber auch als Epochenstil vor, sind von der Literatur auf andere Künste ausgedehnt und ebenfalls individualisiert worden. Bei Midjourney fallen diese Unterschiede. Erst als es den Anspruch gab, das Untersuchungsobjekt mit eigenen Augen gesehen – und *autopsiert* – zu haben, wurde die Beschreibung als kunsthistorische Methode wichtig, weil jene Augenzeugenschaft nur im Wahrnehmungsprotokoll eingelöst werden konnte. Dieser Anspruch schwindet ebenfalls: Es wird kein Objekt mehr beschrieben, sondern eine informationslogisch aufbereitete und reproduzierbare Oberfläche.

Dieser entgrenzte Stil bei Midjourney ist ein Variationsmechanismus.³² Denn wenn man vergleicht, konzentriert man sich vor allem auf Ähnlichkeiten und Identifizierbares und nicht auf Unterschiede und Unidentifizierbares. Ähnlichkeit wird zu einem Begründungszusammenhang, der die einzelnen Bildphänomene isoliert, statt sie aus dem Bild heraus zu legitimieren. Doch was ist mit dem Formlosen, dem Grund, dem nicht gestalteten Objekt, kann es einen ‚Stil‘ haben? Stil ist außerdem sehr positiv konnotiert, man hat ihn im Sinne von Besitz. Daher ist er auch auf Identifizierbares, Benennbares beschränkt. Unbestimmtheiten oder Übergänge sind so schwierig zu fassen.

Was ich problematisch an diesem Vorgehen finde ist, ist der Versuch, aus qualitativen Aspekten eines Bildes wie Farben, Kontrast, Helligkeit, Schärfe oder Perspektive quantitative, zähl- und berechenbare, letztlich wiederholbare Eigenschaften des Bildes werden zu lassen.³³ Denn jene Form der Typisierung ist eine Vergleichshandlung, in der etwas Relatives in etwas Absolutes verwandelt wird. Und da Ästhetik kein

31 Vgl. Arthur C. Danto, Die philosophische Entmündigung der Kunst. Fink: München 1993.

32 Ich danke Roland Meyer für diesen Begriff.

33 Die Unterscheidung von Qualität und Quantität ist ein Erbe des Utilitarismus von u.a. Jeremy Bentham und John Stewart Mill. Ganz grob unterschieden meint die Quantität von Glück (*pleasures*) einfach nur das Ausmaß der Freude. Die Qualität von Glück bedeutet eine Unterscheidung nach der Beschaffenheit, wobei bestimmte Arten von Freuden als höherrangig beurteilt werden. Vgl. Peter Rinderle, John Stuart Mill, C. H. Beck: München 2000.

quantifizierbares Konzept ist, wird Midjourney auch immer nur Schuhe, aber nicht das „Zeughafte des Zeuges“ beschreiben können. Setzte im 18. Jahrhundert das Verstummen der Sprache am Bild an der Quantität der Zeichen an, an der Unmöglichkeit, die Vielzahl der Bildinformationen in verbale Systeme überführen zu können, ist es für Midjourney die Qualität der Zeichen, die eine Substitution verhindert.

25. Mai 2023

5. Nicht Nicht-Kommunizieren. Zur Lust epistemischer Disruption (Oliver Ruf)

Rückschau oder besser: À propos Max Bense. In mehreren Momenten hatte ich erst daran gedacht, wie sehr ein Feuer in der Rede kybernetischer Kraft brennt und wie intensiv diese Hitze auch das Denken desjenigen affiziert hat, der es hervorbrachte, bis dies mir dann doch bloß als Ausdruck nostalgischer Notierung erschien. Zum einen hatte ich vor jenem Augenblick Bense als Figur performativer Universität bemüht, hatte dessen ‚Kreide-Kunst‘ beschrieben, um Digitalität zu ergründen, war so zunächst mit jenem quasi ins Gespräch getreten – über Zeitläufte und Mediengrenzen hinweg, sowohl etwa über den ‚Laplaceschen Dämon‘ wie über ‚Apparate, Maschinen, Instrumente und technische Intelligenzen‘.³⁴ Dann kam ich, kurz vor diesem Augenblick – und im Anschluss an ebenfalls eigene Überlegungen zu einer immanenten Ästhetik *Künstlicher Intelligenz* –,³⁵ nicht umhin, Bense ein weiteres Mal zu bemühen bzw. vielmehr: eine Referenzierung zu wagen, die auf den ersten Blick zeithistorisch konzentriert, auf den zweiten Blick zeitenthoben wahrhaftig und auf den dritten Blick zeitkritisch schlagend erschien. Dabei wagte ich einen weiteren Versuch über ‚Kunst und KI‘ und erinnerte an den Augenblick, als im Westdeutschen Rundfunk am 6. Februar 1970 eine Sendung mit dem Titel *ende offen. Kunst und Antikunst* ausgestrahlt wurde.³⁶ Diese zeigte die Aufnahme eines so genannten Forumsgesprächs zum Thema *Provokation, Lebenselement der Gesellschaft*, das kurz zuvor als Podiumsdisput an einer Düsseldorfer Schule stattgefunden hatte. Veranstalter war die Arbeitsgemeinschaft Sozialpädagogik und Gesellschaftspädagogik (ASG), Moderator war Wieland Schmied, Teilnehmer waren Max Bense (!), Max Bill, Arnold Gehlen und Joseph Beuys. Im Verlauf der eineinhalb Stunden andauernden Diskussion fielen dabei Sätze, die den gegenwärtigen Diskurs gewissermaßen vorweggenommen haben. Denn im Wesentlichen drehte sich jene Diskussion darum, wer und was Kunst eigentlich hervorbringe: ein einzigartiges kreatives, schöpferisches menschliches Individuum, das dazu begabt und befähigt ist und einer Idee folgt, die zweckfrei ist, und

34 Vgl. Oliver Ruf, *Die digitale Universität*, Wien: Passagen 2021, S. 156-163.

35 Vgl. Oliver Ruf, „Medien-,Eingriffe“. Zur immanenten Ästhetik *Künstlicher Intelligenz*“, in: Oliver Ruf u. Lars C. Grabbe (Hg.), *Technik-Ästhetik. Zur Theorie techno-ästhetischer Realität*, Bielefeld: transcript (Medien- und Gestaltungsästhetik; Bd. 12) 2022, S. 389-402.

36 Vgl. Oliver Ruf, „Die Maschine als Schöpferin“, in: *Schweizer Monat* 4 (2023), S. 42-45.

hierfür Mittel einsetzt, die ein Werk (sei es als ‚Text‘, sei es als ‚Bild‘ usw.) erschaffen – oder schlicht jeder, der sich dazu bemüßigt sieht?

Als ich in den vergangenen Monaten also mehrmals über die heute weit fortgeschrittenen technologischen Möglichkeiten so genannter ‚Künstlicher Intelligenz‘ nachdachte, fiel mir somit auf, dass damit eine Art diskursive Dauerschleife aufgerufen wird, die sich im Kreis dreht. Es lag daher nahe (bzw. war es durchaus verführerisch), diese Formen von Software, Algorithmen und Datenbanken hinsichtlich ihrer Fähigkeiten der Erzeugung, Produktion und Schöpfung selbst in Anspruch zu nehmen, und zwar als Aufgabenstellung bzw. Auftrag, *etwas* entsprechend zu erzeugen. Es traf sich gut, dass ich zu diesem Zeitpunkt an einem Vortragsmanuskript saß, mit dem ich zunächst nicht ganz zurechtkam. Es ging dabei um eine Einladung zu einer Tagung über *Experimentellen Wissenschaftsjournalismus*, die vom 22.-24. März 2023 an der Eberhard Karls-Universität Tübingen stattfinden sollte. Ich selbst hatte vorgeschlagen, über *Medienästhetischen Wissenschaftsjournalismus* zu sprechen, wohlwissend, dass es wenigstens uneindeutig und mindestens herausfordernd sein würde, Begriff und Feld des ‚Medienästhetischen‘ für das Tagungsthema zu zentrieren (auch wenn ich in der Vergangenheit selbst einige nähere exemplarische Versuche in diese Richtung unternommen hatte).³⁷ Deshalb bat ich ChatGPT, einen Vortragstext zu der Frage, was Medienästhetik sei, zu schreiben. Die Antwort bzw. die Ausgabe lautete (nach mehreren Repetitionen):

37 Vgl. u.a. Oliver Ruf, „Medienästhetische Popularisierung. Zur Prozessualität von ‚Harald Schmidt‘“, in: Oliver Ruf u. Christoph H. Winter (Hg.), *Harald Schmidt. Zur Ästhetik und Praxis des Populären*, Bielefeld: transcript 2022, S. 67-84; Oliver Ruf, „Kritische Formen. Überlegungen zu einer gestaltungsbezogenen Medienästhetik des Feuilletons“, in: Oliver Ruf u. Christoph H. Winter (Hg.), *Small Critics. Zum transmedialen Feuilleton der Gegenwart*, Würzburg: Königshausen & Neumann 2022, S. 19-45; Oliver Ruf u. Andreas Sieß, „‚Making Media‘: Zum Kreativitätsdispositiv zwischen Medienästhetik, Kommunikationsvermittlung und Designdidaktik“, in: *merz – medien + erziehung* 5 (2021), S. 59-71; Oliver Ruf, „Mikro-Kommunikation. Überlegungen zur einer Medienästhetik von WhatsApp“, in: Patrick Bettinger et al. (Hg.), *Mikroformate. Mediale Kleinstformate zwischen (Re-)Produktion und Wahrnehmung*, München: kopaed 2021, S. 15-27; Oliver Ruf, „Rahmenarbeit. Form und medienästhetische Verfahren am Rande des Films“, in: Petra Kindhäuser u. Stefan Neuhaus (Hg.), *Anfangen & Aufhören. Kulturwissenschaftliche Zugänge zum Ersten und Letzten*, München: Wilhelm Fink 2019, S. 175-208; Oliver Ruf, „Foulen: Zur medienästhetischen Störung“, in: Teresa Hiergeist und Sandra Fluhrer (Hg.), *Foull! Ästhetiken und Performanzen fußballerischen Fehlverhaltens*, Würzburg: Königshausen & Neumann 2019, S. 59-71; Oliver Ruf, „Visual Storytelling und digital Design. Autobiographie-Theorie und medienästhetische Praxis von Web-Comics“, in: Kalina Kupczynska u. Jadwiga Kita-Huber (Hg.), *Autobiographie intermedial. Fallstudien zur Literatur und zum Comic*, Hrsg. v. Bielefeld: Aisthesis, 2019, S. 269-288; Oliver Ruf, „‚Funktionen‘ von Theorie. Literarische Medienästhetik“, in: *Wirkendes Wort* 1 (2018), S. 105-118; Oliver Ruf, „Smartphone-Theorie. Eine medienästhetische Perspektive“, in: Oliver Ruf (Hg.), *Smartphone-Ästhetik. Zur Philosophie und Gestaltung mobiler Medien*, Bielefeld: transcript 2018, S. 15-31. Siehe auch Oliver Ruf, *Die Hand. Eine Medienästhetik*, Wien: Passagen 2014.

Einleitung:

Medienästhetik ist ein Konzept, das in den Bereichen Medienwissenschaft, Filmwissenschaft und Kunstgeschichte zunehmend an Aufmerksamkeit gewonnen hat. Dieser Begriff bezieht sich auf die Art und Weise, wie wir die visuellen und auditiven Qualitäten von Medienobjekten wie Filmen, Fernsehsendungen, Videospielen und sogar Social-Media-Posts interpretieren und schätzen. Medienästhetik beinhaltet die Analyse und das Verständnis der Techniken, Stile und Elemente, die bei der Erstellung dieser Medienobjekte verwendet werden. In diesem Beitrag werden das Konzept der Medienästhetik, seine Bedeutung und seine Anwendung in verschiedenen Medienkontexten erörtert.

Definition von Medienästhetik:

Medienästhetik ist ein Begriff, der verwendet wird, um die verschiedenen Qualitäten zu beschreiben, die zum visuellen und auditiven Gesamterlebnis eines Medienobjekts beitragen. Zu diesen Qualitäten gehören unter anderem Farbe, Beleuchtung, Sounddesign, Komposition und Kamerawinkel. Die Art und Weise, wie diese Elemente kombiniert und manipuliert werden, kann eine bestimmte Stimmung oder einen bestimmten Ton erzeugen, der beim Publikum eine bestimmte emotionale Reaktion hervorrufen soll. So erzeugt beispielsweise die Verwendung von gedämpfter Beleuchtung und Schatten in einem Film noir ein Gefühl von Geheimnis und Gefahr, während die Verwendung heller Farben und verspielter Musik in einer Kindersendung ein Gefühl von Glück und Staunen hervorruft.

Die Bedeutung der Medienästhetik:

Die Medienästhetik spielt eine wichtige Rolle bei der Gestaltung unseres Verständnisses und unserer Wertschätzung von Medienobjekten. Durch die visuellen und auditiven Qualitäten dieser Objekte sind wir in der Lage, uns auf einer emotionalen Ebene mit ihnen zu beschäftigen. Durch die Analyse und das Verständnis der in der Medienproduktion verwendeten Techniken und Elemente können wir ein tieferes Verständnis für die Kunstfertigkeit gewinnen, die mit ihrer Herstellung verbunden ist. Darüber hinaus kann uns die Medienästhetik helfen, die Bedeutungen und Botschaften zu interpretieren, die durch diese Objekte vermittelt werden. Zum Beispiel kann die Verwendung bestimmter Farben, Kamerawinkel und Musik in einem Film dazu beitragen, dem Publikum eine bestimmte Botschaft oder ein bestimmtes Thema zu vermitteln.

Anwendung der Medienästhetik:

Medienästhetik kann in einer Vielzahl von Medienkontexten angewendet werden, darunter Film, Fernsehen, Videospiele und soziale Medien. In Film und Fernsehen kann die Medienästhetik eingesetzt werden, um eine bestimmte Stimmung oder einen bestimmten Ton zu erzeugen, der dem Genre oder dem Thema angemessen ist. In einem Horrorfilm können beispielsweise eine gedämpfte Beleuchtung und unheimliche Musik eingesetzt werden, um ein Gefühl von Spannung und Angst zu erzeugen. In Videospielen kann die Medienästhetik eingesetzt werden, um ein Gefühl der Immersion und Interaktivität für den Spieler zu schaffen. Ein Ego-Shooter-Spiel kann zum Beispiel dynamische Kamerawinkel und realistisches Sounddesign verwenden, um dem Spieler das Gefühl zu geben, dass er sich tatsächlich in der Spielwelt befindet.

Schlussfolgerung:

Medienästhetik ist ein wichtiges Konzept, das unser Verständnis und unsere Wertschätzung von Medienobjekten maßgeblich prägt. Durch die Analyse und das Verständnis der Techniken und Elemente, die in der Medienproduktion verwendet werden, sind wir in der Lage, ein tieferes Verständnis für die Kunstfertigkeit zu erlangen, die mit ihrer Schaffung verbunden ist. Darüber hinaus kann die Medienästhetik uns helfen, die Bedeutungen und Botschaften zu interpretieren, die durch diese Objekte vermittelt werden. Medienästhetik kann in einer Vielzahl von Medienkontexten angewandt werden, z. B. in Film, Fernsehen, Videospielen und sozialen Medien, und sie kann verwendet werden, um eine bestimmte Stimmung oder einen bestimmten Ton zu erzeugen, der für das jeweilige Genre oder Thema angemessen ist. Letztlich ist die Medienästhetik ein wesentlicher Aspekt der Medienwissenschaft und hilft uns zu verstehen, wie Medienobjekte unsere Wahrnehmung der Welt um uns herum prägen.

Das alles ist nicht falsch – und es ist nicht richtig. Weder falsch noch richtig. Es ist antiquiert, einseitig und irgendwie quer, denn damit werden viele unterschiedliche Zugänge zur Frage des Medienästhetischen berührt und auch oberflächlich vermischt, disziplinäre Ansätze unzulässig ignoriert und auf diese Weise trivialisiert (gerade auch in begrifflicher Hinsicht, die es doch zu schärfen gilt). Hatte ich die Frage falsch gestellt – *zu weit*? Hatte ich den Auftrag ungenau formuliert – *zu allgemein*? Als ich schließlich selbst darauf reagierte (mit meiner eigenen medientheoretisch geleiteten und aus dieser Forschung destillierten Auffassung im Rücken) und so meinen Vortrag – ohne ChatGPT – hielt, stellte ich bei der sich anschließenden Diskussionsrunde Folgendes fest: Kritisiert wurde ich überhaupt dafür, das Phänomen, über das

ich sprechen wollte, mit Begriff und Feld des ‚Medienästhetischen‘ zu benennen: wegen dessen herausfordernder (provokativer) Uneindeutigkeit.

Im Grunde fühlte ich mich bestätigt. Nicht so sehr von der Kritik als vielmehr von ChatGPT. Dieses hatte mir, strenggenommen, zwischen den Zeilen geantwortet, hatte in dem oben zitierten Output zu verstehen gegeben, dass (zum Stand dieses Jetzt, wenn man so will) nur eine solche Antwort möglich ist bzw. diese Antwort implizit empfiehlt: sprich nicht darüber, es kann nur missverstanden oder sogar nicht verstanden werden. Es war denn auch womöglich eine Art Test, was ChatGPT mir ausgab und vielleicht auch eine Falle, in die ich wachen Blickes hineingestolpert war. Ich hätte es besser wissen müssen. Als ich über Bense, Beuys und deren Frage, wer oder was vermag, Kunst zu machen, nachdachte, war ich schließlich zu dem Schluss gekommen, dass genau dies der Knackpunkt *einer* ‚Medienästhetik‘ sein dürfte, die ihrerseits nach dem Autopoietischen *des* ‚Künstlerischen‘ sucht.

Kurz zurück zur Fernsehsendung: Wenn es angesichts von ChatGPT (beinahe) auch ununterscheidbar wird, wer so etwas wie Kunst hervorbringt, ist die Frage nach dem, *was Kunst ist und wer sie erschafft*, offensichtlich falsch gestellt. Wichtig bleibt die Funktion, die Kunst in einer sozialen Praxis erfüllen kann – 1970 genauso wie 2023. Die genannte Podiumsdiskussion war hierzu wegweisend, weil sie über den Verdacht der Kunstlosigkeit reflektieren wollte, wenn es die, wie es in der Einleitungsansprache hieß, klassisch in sich geschlossenen Kunstgattungen immer weniger gibt, die Übergänge zwischen ihnen aber immer mehr. Zum damaligen Zeitpunkt drehte sich die folgende Kontroverse primär um die Rolle des Künstler:innensubjekts bei der Erschaffung von Kunstwerken. Es ging zunächst bezeichnenderweise darum, was die einzelne Künstler:in unternimmt, wenn Kunst *erfolgt*. Beuys antwortete getreu seiner eigenen schematischen Formel, dass für ihn im Kunst- und Antikunstabegriff immer die Frage nach dem Menschen auf eine spezielle, neue Weise gestellt sei, indem diese Auffassung von Kunst auf jede menschliche Tätigkeit erweitert werde. Zu sagen, jeder Mensch sei Künstler, heiße ebenfalls, jeder Mensch sei ein Kreativer. Und eine solche Kunst – als Antikunst – erzeuge eine Bewusstseinsveränderung in der Gesellschaft, eine Position, von der sogleich Bense verlangte, sie zu präzisieren, um ihre Implikationen vom Vorwurf der Phrasendrescherei zu befreien.

Das daran für die heutige Gegenwart Interessante ist wiederum vor allem der Diskurs, der hier zum Vorschein kommt. Auf der einen Seite derjenige der Kunstrevolution, repräsentiert von Joseph Beuys, und auf der anderen Seite derjenige der Kybernetik, und das heißt: der theoretischen Entdeckung, Förderung und Rahmung früher computergenerierter Kunst, repräsentiert von Max Bense. Der Widerspruch zwischen den Standpunkten beider Persönlichkeiten, wie er von diesen geäußert wurde, entzündet sich genau dort, wo sich deren Ideen berühren: im Moment der Autorschaftsfrage von Kunst schlechthin. Gemeinsam ist ihnen, was keiner von ihnen zugeben kann, dass es ihnen nicht so sehr um die Figur des Genies im künstlerischen Schaffensprozess geht, nicht um ein angeborenes Talent, mit dem man der Kunst die Regeln gibt und das sich unter anderem qua Originalität, Exemplarität und Unerklärbarkeit auszeichnet. Vielmehr sprachen der radikale Künstler (Beuys) wie

auch der kritische Ästhetiker (Bense) von einer Wesenhaftigkeit von Kunst, die sie von solchen Zuschreibungen selbst befreit – und die es mittlerweile greifbar macht, dass KI womöglich schöpferisch (und dann auch: ‚testend‘ und ‚Fallen-stellend‘) zu agieren vermag.

Diese Technikanwendungen demonstrieren, dass diese aus sich selbst heraus eigene Regeln postulieren, zu erfinden und z.B. ein Artefakt zu realisieren – und mithin ein Vermögen zur Willkür jenseits ihrer maschinell-induzierten Inputgebundenheit versprechen. Selbst in den Diskussionsminuten, in denen Bense auf einem bewussten Engagement beharrte, das Kunst initiiere, woraufhin Beuys von den Prinzipien seines Konzepts der sozialen Plastik mit Formgestalt sprach, die sich in Wille und Ratio als plastischen Vorgang aufteilen, deutete sich ein bleibender Befund an: Etwas kann dann entstehen, wenn etwas Neues bzw. ein Produkt in die Welt gesetzt wird und Denken bereits die Kreation ist. Zu einer solchen Arbeit freiheitlichen Ursprungs, bei der die Schöpfung spontan, also aus sich selbst heraus beginnt, scheint ChatGPTs *Gequatsche* in der Lage zu sein – willentlich wie rational, fähig, eine bestimmte Form zu erzeugen.

Vor dem aufgespannten Hintergrund ist im Übrigen zu verstehen, dass Arnold Gehlen sich in der Rückschau als mahnender Diskussionsteilnehmer erweist, der immer wieder Kriterien einfordert (was taugt, was ist Mumpitz?), an historische künstlerische Provokationen (Marcel Duchamp, Dada) erinnert und überhaupt die gesellschaftliche Dimension von Kunst punktiert. Max Bill gibt sich moderater oder besser: neutraler. Ihm ist ebenfalls das Messen der Qualität von Kunst wichtig, doch interessiert ihn mehr der Sinn, Kunst zu machen. Indem von Bill hier betont wird, ein Problem, das sich gestellt habe, mit Kunst weiterzubringen und nach allen Seiten zu durchleuchten (und zwar unabhängig der Folgen, die dies zeitigt), argumentiert er in einem Kontext, der gerade nicht auf das Selbst der eigenen Künstler:innenschaft schießt. Der Fall der Künstlichen Intelligenz ist kein grundsätzliches abstraktes Hindernis, derart etwas zu erschaffen. Die ihr zugrunde liegende, gleichsam lebendige Mathematik umfasst Varianten der Problembegegnung gemäß Bill, die das Bewusstsein gemäß Beuys erweitern: Sei es in alltagspraktischer Hinsicht in den Phänomenen der Sprachsteuerung, der Wissensdarstellungen in Datenverarbeitungssystemen, des automatisierten Wahrnehmens, der Bildanalyse, der Robotik oder des maschinellen Lernens; sei es in künstlerischer Perspektive mittels der unmittelbaren Reizung der Sinne oder auch des sinnlichen Begehrens. Wenn dadurch auch die Gefahr besteht, in Sinnesreizen zu versaufen, wie Beuys es formuliert, oder dem epikureischen Weltgenuss, der Lusterzeugung zu folgen, wie Bense es ausdrückt, dann wirkt KI, indem genau diese Effekte bei den sie Benutzenden auszumachen sind.

Bei meiner Frage nach dem Medienästhetischen über die Befragung von ChatGPT ging es also ebenfalls darum, epistemischen Gewinn meinerseits zu testen – und Lust daran zu haben, dies zu tun. Möglicherweise diente ChatGPT, so meine aktuelle Vermutung, immer schon in besonderer Weise der Lusterzeugung (größer: einem Begehren). Nur wird dies in der Ausführung und in den Resultaten gestört: disruptiv – und genau deswegen funktionierte mein Anliegen nicht, um Unterstützung bei einem medienwissenschaftlichen Vortrag zu ersuchen. Diese dann regelrecht genuin medienästhetische Erfahrung im Umgang mit ChatGPT selbst, ermöglicht es, so

dann meine Hoffnung, eine spezifische Medienästhetik zukünftig irgendwann einmal genauer zu lokalisieren, als Name *einer Ästhetik*, die nicht ergebnisfixiert, sondern prozessorientiert ist – Wieland Schmied nannte das Zentrum dieses Effekts auf dem Podium ein offenes Werk. ChatGPT bedeutet auch deshalb eine Herausforderung, da dies darin inspirierend ist: affizierend. ChatGPT affiziert die Medien, die sie Nutzenden und letztendlich auch die jene Erforschenden. Es handelt sich folglich um eine Art wechselseitige Übertragung, um einen, wenn man so will, um sich greifenden viralen Infekt. Der kann die Gesellschaft als Ganzes schwerlich verschonen. Viel zu sehr ist sein Antrieb auf Verkettung, Ausbreitung und andauernde Infizierung ausgelegt: als Ansteckung.

Diese Konstellation hat, um im Sprachgebrauch des telemedial gesendeten Gesprächs aus dem Jahr 1970 zu bleiben, ggf. die Funktion, Lebens Elemente der Gesellschaft zu identifizieren. Max Bill hat damals von ästhetischen Objekten gesprochen, die (positiven wie negativen) Modellcharakter haben, indem sie Einfluss ausüben auf die Entscheidungen, die die Umwelt betreffen. ChatGPT erweist, wie Konsumierung und Produktion derartiger Techniken aus Frage- und Antwort-Komplexen die darin enthaltene ‚alte‘ Mediengeste,³⁸ nicht nicht-kommunizieren zu können, geschickt auf die Probe stellt: redselig.

38 Vgl. Oliver Ruf, „Zur Geste der Medien“, in: Luca Vigliano u. Johannes Waßmer (Hg.), *Epistemologien der Geste. Körper, Medien, Künste*, Berlin: de Gruyter 2023, S. 75-91.