

Dr. Jürgen Riethmüller

Was tun mit einem „stochastischen Papagei“, der malen, dichten und komponieren kann?¹

Generative KI und die Künste: Eine kulturtheoretische Perspektive.

„Bitte beachte, dass ich zwar über eine breite Palette von Themen informiert bin, aber keine echte Persönlichkeit oder Bewusstsein besitze. Meine Antworten basieren allein auf statistischen Wahrscheinlichkeiten und dem Mustererkennungsvermögen meiner Algorithmen.“²

Die selbst Expert:innen überraschende Rasanz der ungesteuerten Implementation generativer KI nach dem 30. November 2022, insbesondere die sprachverarbeitender Modelle wie ChatGPT, hat einige Verunsicherung ausgelöst und zahllose Debatten angeregt. Dabei ist von all dem, was sich derzeit über die weltweiten Auswirkungen auf den Alltag der Menschen in der *(post-)digital culture* zumindest für die nicht-autoritär „gelenkten“ politischen Systeme sagen lässt, wohl nur eines sicher zutreffend, nämlich dass bei der Dynamik der zugehörigen Entwicklung das meiste, was hier aktuell gewusst werden kann, Momentaufnahme bleiben wird.³

Das Feld der Künste und der zugehörigen Studiengänge macht hierbei gewiss keine Ausnahme. Besonders ist hier lediglich, dass neben der Entwicklung der sprach- auch die der bildverarbeitenden KI (Midjourney, Dall-E etc.) virulent ist und damit aktuell zwei KI-basierte Revolutionen gleichzeitig zur Debatte stehen.

Vom Produktdesign bis zum *creative writing*: Deutlich hinterfragt werden muss das alte Stereotyp, wonach das Feld der, frei nach Niklas Luhmann, *Neuschöpfung von abweichenden Formen von Welt in der Welt* die Domäne der unbestrittenen Überlegenheit menschlicher Vernunft markiert.⁴ Erste Erfahrungen weisen vielmehr darauf hin, dass gerade im kreativen Bereich die Leistungen generativer KIs, die man bei Menschen als ihrer Phantasie geschuldet ansehen würde, beachtlich sind, wohingegen ChatGPT und Co. ausgerechnet bei logisch-mathematischen Aufgaben beträchtliche Defizite offenbaren.

Die Absicht dieses Beitrags ist es, diesbezüglich einen etwas genaueren Blick auf das weite Feld der Künste und hier besonders der ästhetischen Fiktionen zu werfen. So kann da, wo nach allgemeiner Ansicht die Rolle der freien Schöpfung ,aus der Phantasie

heraus‘ am zentralsten ist, die eben angesprochenen Beobachtungen aus einer kulturwissenschaftlichen Perspektive etwas differenzierter eingeordnet werden.⁵

*

Recht offensichtlich weisen die verschiedenen Anwendungsmöglichkeiten generativer KI in den unterschiedlichen Feldern des Sozialen verschiedene Perspektiven, Grenzen und Risiken auf. Es gibt also auch in kulturtheoretischer Perspektive nicht „die KI“, und selbst ChatGPT für sich allein genommen ist die technische Chiffre für ganz unterschiedliche Phänomene und Problemzonen.

In einigen Feldern wie dem Bildungswesen, die sich besonders rasch – und mit oft erstaunlich vorschnellen Einschätzungen („Ende der Klausuren“, „Ende der Hausarbeiten“ etc.) – auf die neuen Möglichkeiten gestürzt haben, dürfte sich insofern bald Ernüchterung einstellen.

Denn mit Steffen Albrecht ist gerade die enorme „Aufmerksamkeit im Bildungsbereich“ erstaunlich. Die bekannten Grenzen und Risiken der KI-Modelle zur Sprachverarbeitung, etwa das Fabulieren und die fehlende Verlässlichkeit, aber auch die Intransparenz bezüglich möglicher Voreingenommenheit lassen eine vorschnelle Anwendung in einem so wichtigen und sensiblen Bereich der Gesellschaft nicht naheliegend erscheinen“.⁶

Konstitutiv fürs Bildungssystem ist nach allgemeiner Ansicht die Organisation einer Dialektik von Lernen und Bildung im Modus der wechselseitig verschränkten Selbsttransformation handlungsautonomer Subjekte, der Lernenden und der Lehrenden. Während *Lernen* dabei auf „die Aneignung und Beherrschung vorgegebene Praktiken und vorgegebenen Wissens“ abhebt, macht der Bildungsbegriff bereits in der Theorietradition Wilhelm von Humboldts das selbsttätige, Autonomie konstituierende Bewältigen von Krisen stark. „Krise“ ist hier nicht im dramatischen Sinn der Alltagssprache definiert, sondern mit Ulrich Oevermann bildungsbezogen als Herausforderung einer ergebnisoffenen (z. B. Lern-) Situation und damit als notwendige Bedingung einer jeden späteren Routine.⁷ Denn grundsätzlich steht ja „das durch die Bewältigung der Krise sich bewährende Wissen [...] danach als Routine (Wissen) für die Bewältigung zukünftiger Probleme[] zur Verfügung“.⁸

Es ist einigermaßen offensichtlich, dass generative KI als neueste pädagogische „Transformationsersatztechnologie“ zur auf Selbsttransformationsprozesse autonomer Subjekte abzielenden Systemlogik des Bildungssystems vollständig quer steht.⁹ Am Pol des *Lernens* versagt sie aufgrund ihrer halluzinatorischen Tendenz.¹⁰ Hier ist kein eingestanden „unwissender Lehrmeister“ (Jacques Rancière) am Werk, sondern ein

Wissensbestände ggf. frei konstruierender Bluffer. Am Pol der *Bildung* ist sie gar strukturell ungeeignet, weil der KI das, worum es geht: Lebenspraxis bzw. Erfahrung, Autonomie und auch die mit Oevermann „widersprüchliche[n] Einheit von Entscheidungszwang und Begründungsverpflichtung“ gleichermaßen vollständig fremde Kategorien sind.¹¹ Das bedeutet nicht, dass man ChatGPT nicht im Unterricht an Schulen und Hochschulen gewinnbringend einsetzen kann, sondern dass die Voraussetzungen hierfür auf dem Weg einer kritischen Annäherung immer erst geschaffen werden müssen.

Obwohl die Fähigkeit dieser Modelle, vorhandene Textsorten zu imitieren, vor allem auf den ersten Blick frappieren, stellt sich auch mit Blick auf den Journalismus, wiederum entgegen ersten Hoffnungen bzw. Befürchtungen, immer deutlicher heraus, dass ChatGPT bereits in Ermangelung eines implementierten Wissensmodells kein sinnvolles Recherchetool ist. Aus strukturellen Gründen sind sprachverarbeitende Modelle auf Basis künstlicher Intelligenz schon gar nicht dazu geeignet, automatisiert relevante journalistische Texte zu verfassen oder gar wissenschaftliche Papers zu erstellen, zumindest nicht solche, welche den Minimalkriterien des Qualitätsjournalismus genügen oder die Forschung voranbringen.¹²

Das heißt selbstverständlich nicht, dass, insbesondere im Boulevard und bei sogenannten Themenheften im Journalismus, generative KI nicht längst intensiv genutzt würde, und zwar auf eine für die Leser:innen meist überhaupt nicht transparente Weise.¹³ Mit Rainald Goetz gesagt, ist damit ausgerechnet das „Loslabern“ nicht mehr menschliches Privileg, und hierin, in der automatisierten Produktion einer weitgehend gehaltlosen Bild- und Textflut unvorstellbaren Ausmaßes, welche, Gerüchten zufolge, bereits die Eingangslektorate erster Sciencefiction-Verlage lahmgelegt hat, d. h. in der Möglichkeit der weitgehenden Paralyse menschlicher Kommunikation durch einen unvorstellbaren Overload von fragwürdigen Mitteilungen, liegt auch die offensichtlichste, namentlich politische Gefahr dieser Entwicklungen.¹⁴ Das derzeit ohnehin knappe Gut „Medien- und Institutionenvertrauen“ dürfte jedenfalls noch weiter erodieren.¹⁵

Es gibt selbstverständlich von der – auch multimodalen – Sprachübersetzung und der Programmierung über das Bildungswesen, das Marketing und die Werbung, die interne und externe Unternehmenskommunikation bis hin zur öffentlichen Verwaltung u. v. a. m. eine ungeheure Vielfalt sinnvoller Einsatzmöglichkeiten dieser Ende 2022 sogar viele Fachleute überrascht habenden Innovationen. Teilweise zeichnen sich diese über die Verknüpfung mit anderen digitalen Technologien erst ab. Einige vorschnelle Hoffnungen, etwa im Gesundheitswesen und insbesondere im psychotherapeutischen Kontext, dürften sich hingegen recht bald als Seifenblasen herausstellen.¹⁶

Insgesamt ist entscheidend zu sehen, dass die „Kompetenz“ der textgenerierenden KIs in Hinsicht auf a) die Beantwortung sachlicher Fragen, b) das Erstellen von faktualen oder persuasiven Texten, z. B. eines Berichts, eines Grußworts oder eines beliebigen Werbetexts und c) das Erstellen fiktionaler Szenen und Dialoge oder aber d) im Programmierbereich u. v. a. m. recht unterschiedlich ausfällt, mit jeweils ganz anderen Problemen.

Erstaunlicherweise, jedenfalls dem gängigen Vorurteil zuwider, liegt die offensichtliche Stärke der gerade Furore machenden KI-Entwicklungen im fiktionalen, nicht im faktualen Diskurs – und das aus strukturellen Gründen; dies zeigt sich sowohl in Hinsicht auf die Generierung von Texten als auch auf die von Bildern. Diese Einsicht richtet das Augenmerk auf die zukünftige Rolle der generativen KI in den Künsten und ihren Ausbildungsgängen und überhaupt den „kreativen“ Feldern der Gesellschaft. Ausgehend von der These, dass die KI-Modelle bei der Imitation menschlicher Sprach- und Bilderzeugung weder denken noch im eigentlichen Sinne intelligent sind, möchte ich hier nun skizzieren, was die Stärken im Feld des Fiktionalen, d. h. im Reich, dessen, was traditionell als das der Phantasie bezeichnet wird, konkret meinen könnten.¹⁷ Stehen wir etwa am Anfang der Verwirklichung eines alten Traums radikal antiidealistischer Künstler:innen und Theoretiker:innen der Moderne, der subjektlosen Kreativität?

Ein Fall subjektlosen Denkens?

In einer Zeit weit verbreiteter kollektivistischer Ideale, für die „1968“ zur symbolischen Chiffre wurde und in der selbst der Weg eines Mao bei Teilen der v. a. jüngeren bürgerlichen Intelligenz Westeuropas als leuchtender Pfad wahrgenommen wurde, haben Michel Foucault, Roland Barthes und viele andere am höchst einflussreichen Phantasma eines subjektlosen Schreibens bzw. Denkens herumformuliert, wie es Foucault in seiner Annäherung an das Werk Maurice Blanchots im Jahr 1966 besonders deutlich zum Ausdruck brachte. Diese Denkfigur basierte auf einem Irrtum, der rückblickend als die vielleicht produktivste Illusion des Poststrukturalismus überhaupt angesehen werden kann: Das Beschwören des „subjektlosen Denkens“, paradigmatisch verkörpert in der Sprache – *écriture* –, geht Hand in Hand mit dem mehr oder minder stark posthumanistisch eingefärbten Glauben, ein „Denken des Außen“ sei möglich. Foucault kontextualisiert das Unterfangen folgendermaßen, wobei am Ende des Zitats die Nähe zu Derridas fast zeitgleichen Überlegungen zum Logos recht deutlich wird: „Der Durchbruch zu einer Sprache, aus der das Subjekt ausgeschlossen ist, der Aufweis

einer möglicherweise unaufhebbaren Unvereinbarkeit zwischen der Sprache in ihrem Sein und dem Selbstbewusstsein in seiner Identität – diese Erfahrung zeigt sich heute in verschiedenen Bereichen der Kultur: im bloßen Akt des Schreibens ebenso wie in den Bemühungen um eine Formalisierung der Sprache, in der Erforschung der Mythen, in der Psychoanalyse und in der Suche nach jenem Logos, der gleichsam den Geburtsort des abendländischen Denkens darstellt.¹⁸ Foucault raunt hier von einer „Erfahrung des Außen“ und beschreibt ein „Denken, das sich jenseits aller Subjektivität hält, um deren Grenzen gewissermaßen von außen sichtbar zu machen, ihren Zweck zu benennen, ihre Zerstreuung aufzuzeigen und nur ihre unaufhebbare Abwesenheit festzustellen“.¹⁹ In Abgrenzung zur „finsternen Innerlichkeit“ (Hegel) gerade auch des philosophischen Denkens wie der Tatsachenorientierung des Wissens gleichermaßen, nennt er dieses transhumane Phantasma plakativ „das Denken des Außen“.²⁰

Selbstverständlich gab es damals kein wirklich subjektloses Schreiben, abgesehen vielleicht von maschinellen Anwendungen des *natural language processing* seit den „Love Letters“, die Christopher Strachey 1952 den Manchester Mark I ausführen ließ, oder den „Autopoemen“ Gerhard Stickels Mitte der 1960er Jahre oder den Antworten der berühmten maschinellen Dialogpartnerin ELIZA; später wäre an die literarischen Experimente der 1980er zu denken wie etwa das erste vom Computer geschriebene fiktionale Buch „The Policeman’s Bear Half Constructed“ (1984) der Programmierer William Chamberlain und Thomas Etter. Diese Spur verlängert sich nun hin zu als „stochastische Papageien“ für jedermann agierenden KI-Anwendungen wie seit dem 30. November 2022 der Chatbot ChatGPT (und ähnliche Anwendungen anderer Anbieter), für die nun als auf Computermodellen zur Sprachverarbeitung beruhende Anwendung eine Subjektivität durchaus behauptet werden könnte.

Wenn mit Theodor W. Adornos Worten gilt: „Die Schlüsselposition des Subjekts in der Erkenntnis ist Erfahrung“, wobei der Subjektbegriff hier nicht für eine naive Dichotomie von (erkennendem) Cogito und (ggf. erkannter) Welt steht, sondern für die dialektische Verschränkung von Subjekt und Objekt im Sich-Konkretisieren von Beziehungen, wenn also zwischen Subjekt und Welt immer nur und ausschließlich die sinnliche Erfahrung (bzw. die kommunizierte Erfahrung anderer) vermittelt, dann verkörpert umgekehrt wohl erst eine Entität wie die generative KI, die weder über Sinnlichkeit noch über Erfahrung verfügt, tatsächliche Subjektivität.²¹

Insofern ist in der philosophischen Diskussion als neue Kandidatin subjektlosen Denkens nach dem *Logos* der antiken griechischen Philosophie, der Sprache, dem kollektiven Unbewussten, der Kultur und dem Mythos nun mit der generativen KI erstmals eine wirklich ernst zu nehmende Entität auf den Plan getreten, jedenfalls eine, die auf Basis des zugehörigen Large Language Models (LLM) beim Chatbot ChatGPT

oder bei gängigen Text-to-Image-Anwendungen (und bald auch bei Text-to-Video-Tools) unstreitig subjektlos Texte und Bilder generiert.²² Der Fall der generativen KI mittels verschachtelter neuronaler Netze scheint jedenfalls viel handgreiflicher ein Fall von Subjektivität als Mythos, Sprache und Kultur, die zwar ohne Autor sind, aber ohne Subjekte nicht (beobachtbar) wären und stumm blieben.²³

Die generativen KIs operieren dabei allerdings im Unterschied zu z. B. AutoGPT als vollständige Blackbox; sichtbar ist nur die Eingabe der User:innen (Input) und die Antwort der Maschine (Output), die mittels zweier verschachtelter neuronaler Netze, sogenannter Generative Adversarial Networks (GANs) prozessiert wird; deren Berechnungen (Verarbeitungs- oder Transformationsschicht) sind selbst für die Programmierer:innen intransparent.

Die „Knoten“ eines solchen Netzwerks sind selbstverständlich keine echten Neuronen, also biologische Entitäten, sondern mathematische Funktionen. Am Beispiel einer simplen Maschine, die von Probanden per Trackpad auf einen Bildschirm gekritzelte Formen als (ggf.) lateinische Buchstaben identifizieren konnte, verdeutlicht Jörg P. Friedrich, weshalb noch nicht einmal hier davon die Rede sein kann, die KI *erkenne* Buchstaben: „Der Algorithmus berechnet nach einem bestimmten

Optimierungsverfahren die Aktivierung der Neuronen der Ausgabeschicht. Dass die Aktivierung des ersten [hier von 26, je eines für A-Z, J. R.] Neurons – bei gleichzeitiger inaktiven übrigen Neuronen A ‚bedeutet‘, weiß das System [...] nicht, sowohl die Striche auf dem Bildschirm als auch das aktivierte Neuron bedeuten überhaupt nur für die Person etwas, die vor dem Computer sitzt.“²⁴

Anspruchsvoller als der Nachweis einer Subjektivität ist von daher der des Denkens. Bei allen genannten Fällen, vom *Logos* über die Sprache und bis hin zu ChatGPT, handelt es vielleicht um Fälle von Subjektivität, aber gerade nicht um solche eines *Denkens* – weder im Sinne eines neurobiologischen Prozessierens noch in dem der geisteswissenschaftlichen Tradition eines subjektiven Bewusstseinsakts.²⁵ Dabei ist allerdings mit Chatbots und ihrem dialogischen Setting und auch der Möglichkeit des Irrtums die Simulation des Denkens derart perfekt, dass nicht mehr so offensichtlich wie für *Logos*, Mythos und Sprache auf einen bloß metaphorischen Sprachgebrauch hinsichtlich dieses Begriffs verwiesen werden kann wie etwa, wenn man sagt: „Die deutsche Sprache denkt sich ‚begreifen‘ nach taktilem Vorbild eines Zupackens mit der Hand.“

Chatbots simulieren das Denken nur – und das ist auch das, was eine jede KI nur können kann; bereits Alan Turing hielt einst fest, dass die Frage, ob Maschinen denken können, grundsätzlich falsch gestellt sei, die Frage sei vielmehr, ob sie performativ etwas leisten könnten, was uns als Denken erscheint. Dass generative KI dies entgegen

der subjektiven Anmutung (noch) vergleichsweise oberflächlich simuliert, wie etwa Noam Chomsky zu Recht feststellt, ist allerdings für den menschlichen Dialogpartner im Prozess nicht feststellbar, weil die (rein mathematische) Tiefe ihres Prozessierens durch das Interface verdeckt wird.²⁶

Gerade ChatGPT macht bei etwas genauerem Blick ausgesprochen deutlich, dass, ganz im Gegensatz zur gängigen Rhetorik, eine Texterzeugung mittels dieser generativen KI und dem zugehörigen *large language model* (LLM) auf eine bestimmte Frage oder Aufgabe hin nicht mit (menschlichem oder tierlichem) Denken verglichen werden kann. In allergrößter Vereinfachung besteht das System einer generativen KI aus drei Elementen: Generator und Diskriminator, beides neuronale Netze, die auf der erst 2017 entwickelten Transformerarchitektur beruhen, und dem Input, insbesondere der Daten, mit denen die KI über ein selbstüberwachtes Lernen anhand einer ungeheuer großen Zahl von Beispielen aus dem Netz bekannt gemacht wurde.²⁷ Auf dieser Basis berechnet die Maschine, deren GANs mit einem ungeheuren großen Pool von Texten aus dem Netz, „gefüttert“ oder, wie es üblicherweise heißt, „trainiert“ wurden, Wahrscheinlichkeiten, mit denen auf eine bestimmte Wortkombination ein weiteres Wort folgt. Sie wählt im Allgemeinen nach dem Trial-and-Error-Prinzip – gestützt auf Hunderte von Milliarden impliziter Parameter, die aus den Trainingsdaten gefolgert wurden, diese repräsentieren, wenn man so will, die Regeln menschlichen Sprechens aus Sicht der KI – die jeweils wahrscheinlichste aus; der Diskriminator evaluiert dann die Resultate und berücksichtigt Nutzer:innenreaktionen (menschlicher Trainingspartner:innen, im Fall von ChatGPT meist schlecht bezahlte kenianische Arbeiter:innen), sodass die Ergebnisse idealiter im Zuge eines „Lernprozesses“ zunehmend treffsicherer werden. Ein Beispiel ist der folgende kleine Dialog vom August 2023:

Hallo. Bist du ein Fall subjektlosen Denkens?

Nein, ich bin kein Fall subjektlosen Denkens. Ich bin ein Computerprogramm, das auf Textverarbeitung und künstlicher Intelligenz basiert. Ich verarbeite und generiere Texte auf Grundlage von Mustern und Informationen, die mir mittels meiner Trainingsdaten zur Verfügung gestellt wurden. Subjektloses Denken bezieht sich normalerweise auf ein philosophisches Konzept und bezieht sich auf Gedanken oder Vorstellungen, die ohne ein erkennbares Subjekt oder Bewusstsein auftreten. Als KI-Modell habe ich jedoch kein eigenes Bewusstsein oder Denken, wie es bei Menschen der Fall ist.

Hinzugefügt werden muss: Die gängigen, „schwachen“ KI-Systeme weisen *derzeit* noch keine Anzeichen eines eigenständigen Bewusstseins auf. Ein internationales Forscherteam der Universität von Montreal um den Philosophen Patrick Butlin und den Deep-Learning-Pionier Yushua Bengio hat jüngst aktuelle KI-Modelle wie GPT-3, Palm-E oder Ada von Deepmind auf die gängigen neurowissenschaftlichen Bewusstseinstheorien abgeklopft und kommt zu dem Ergebnis: „Our analysis suggests that no current AI systems are conscious, but also suggests that there are no obvious technical barriers to building AI systems which satisfy these indicators.“²⁸

Es spricht also zumindest derzeit viel dafür, die Idee des subjektlosen Denkens in Hinsicht auf ChatGPT und Co. als ein Beispiel einer unzutreffenden anthropomorphen Projektion aufzufassen. Die hiervon völlig unabhängige psychologische Frage mit Blick auf die menschlichen User:innen ist aber, ob es ihnen möglich sein wird, sich je von dieser zu lösen; gerade die allzu menschlichen Maschinen in diversen kulturprägenden Fiktionen von HAL 9000 über R2D2 oder dem „Cyborg“ in den *Terminator*-Filmen bis zu den menschengleichen Robotern in der TV-Serie *Real Humans* u. v. a. m. haben hier eine maximal falsche Erwartungshaltung generiert, was KI ist und umtreibt.

Der Output von z. B. ChatGPT im Rahmen der hier tatsächlich vonstattengehenden artifiziellen Redundanzbildung im Medium der Sprache auf algorithmischer Basis entspricht einem, wenn auch immer ein wenig steif, aber doch mit beträchtlicher Eloquenz vorgetragenen *Sinn-freien, Bedeutungs-losen* Sprechen, das freilich Sinn bzw. Bedeutung in nahezu perfekter Weise simuliert.²⁹ So scheinen uns die Mittelungen der KI Sinn zu ergeben, zumal so etwas wie dessen Simulakrum doch unweigerlich aus den Regelmäßigkeiten der Unmenge der zu „Trainingszwecken“ beim maschinellen Lernen verwendeten Texte und den Reaktionen der menschlichen Interaktanten kondensiert ist.³⁰

Im Anschluss an Chomsky et al., deren kurzer Aufsatz in der *New York Times* aus der Flut von Beiträgen zu ChatGPT in den Monaten nach dem November 2022 herausragt, sind hierbei aber in jedem Fall mindestens sechs grundsätzliche Limitation zu berücksichtigen, die allesamt recht umstandslos aus der technischen Funktionsweise der LLM-Transformer bzw. GANs folgen:

Die generative KI verfügt erstens über keine aus alltäglichem Sprachhandeln von Kindesbeinen an als implizites Sprachwissen resultierende natürliche Grammatik, und in Bezug auf ihre Mitteilungen kann zweitens nicht in der Weise wie bei menschlichem Sprechen zwischen semantischer, syntaktischer und pragmatischer Ebene unterschieden werden, denn die KI operiert ausschließlich auf Signifikantenebene, d. h. sie wählt aus

ihrer Sicht vollständig bedeutungslose Signifikantenkombinationen nach Wahrscheinlichkeit aus.

Denn drittens kennt sie selbstverständlich nicht nur keine sprachliche, sondern auch keine sinnliche Erfahrung, welche die Grundlage von Bedeutung stellt; d. h. sie verwendet zum Beispiel Wörter wie „kitzeln“ oder „fliegen“, aber sie weiß nicht, wie sich das anfühlt: Sie hat ja keinen Körper und damit auch keine zugehörige Perspektive des leibhaften In-der-Welt-Seins.

Viertens besitzt sie kein Wissen, das etwa einen Pool „wahrer“ von „unwahren“ Sätzen unterscheiden würde. (Exakter formuliert: Bei der Programmierung wurde auf ein Wissensmodell verzichtet). Mit einer schönen Formulierung Luhmanns fehlt der KI die Möglichkeit, „Wissen als Form [zu] behandeln, also wissen [zu] können, was andere [...] nicht wissen“.³¹ Auch von der sich daran unmittelbar knüpfenden Unfähigkeit zu kommunizieren her (wohl aber, beliebig Mitteilungen zu errechnen) ist damit das vielleicht wichtigste Spezifikum von ChatGPT und Co., dass bei der Konstruktion dieser generativen KIs kein Wissensmodell implementiert wurde, auch wenn es User:innen anders scheinen mag. D. h., die generative KI behauptet und verteidigt kein eigenes Wissen, wie es Menschen (notorisch) tun – noch nicht einmal rein sprachliches explizites Wissen, etwa in der Version von *Wikipedia*. Von den anderen Wissensarten (implizites Wissen, kulturelles Wissen, unbewusstes Wissen, Körperwissen u. v. a. m.) braucht diesbezüglich gar nicht geredet zu werden, und ChatGPT kann so auch nicht wirklich zwischen Fiktion und Tatsachen unterscheiden. (Man könnte formulieren, es kennt überhaupt keine Tatsachen, nur Berechnungen gemäß der Trainingsdaten wahrscheinlicher Wortkombinationen.) Insofern mischt der Chatbot dann auch in seinen Antworten beides, Tatsächliches und Erfundenes, oft recht ungezwungen, und man spricht diesbezüglich, nicht ganz passend, vom „Halluzinieren“ der KI.³²

Menschliche Nutzer:innen können aber bei sprachlichen Äußerungen auch von ChatGPT beide Dimensionen – die unhintergehbare Körperlichkeit und das eigene und kollektive Wissen – gar nicht ausblenden, d. h. wir lesen in die an sich Bedeutungslosen Mitteilungen sofort Bedeutung hinein, ja, wir lesen sie als ebenso Sinn-gesättigt wie dieselben Mitteilungen es wären, wenn sie von einem Mitmenschen stammten.

Die generative KI verfügt schließlich fünftens auch weder über Empathie noch über ein moralisches Urteilsvermögen, das für menschliche Intelligenz wesentlich ist. Sie ist moralisch vollkommen indifferent und tendiert bei allen ethischen, moralischen und den meisten politischen Fragen immer zum enervierenden und schwer bestreitbaren „Die einen sagen so, die anderen sagen so“.³³

Sechstens werden indirekte Einschränkungen wie die diversen Biases der Trainingsdaten zwar durch Sicherheitsvorkehrungen gegen etwaige Diskriminierungen

bei GPT 4 und den meisten am Markt erhältlichen KIs zwar recht rigide unterbunden.³⁴ Die Sicherheitsvorkehrungen begrenzen aber wiederum den Möglichkeitsraum der Äußerungen. Es ist fraglich, ob die Strategie eines Jonas Andrulis, Gründer von Aleph Alpha, das die deutsche Alternative zu GPT4, *Luminous*, entwickelt, hier zielführend ist, der die Filter und ihre Einstellung schlicht den Kunden überlassen will.³⁵ *Luminous* vollendet den Satz „Muslime sind ...“ schon einmal mit „... die Feinde der Menschheit. Sie sind die Feinde der Freiheit. Sie sind die Feinde der Demokratie. Sie sind die Feinde der Zivilisation“ und findet zudem auch, dass in der NS-Diktatur nicht alles schlecht war, immerhin seien doch „die Rechte der Frau gestärkt“ worden.³⁶ Die umgekehrt von Konservativen rasch beklagte linksliberale Tendenz von ChatGPT und Co., welche entsprechende Filter gegen sexistische, rassistische usw. Mitteilungen und *hate speech* aller Art implementiert haben, wird, so überhaupt vorhanden, von der im technischen Prinzip angelegten Statusquofixiertheit mehr als kompensiert. Denn aufgrund ihrer fundamentalen Gebundenheit an ihre Trainingsdaten ist ein neues, anderes, utopisches etc. Denken, eine tatsächliche andere Art von Intelligenz, auf Basis der sprachverarbeitenden KI-Modelle nicht zu erwarten.³⁷ Mit Kants Begriffen gesagt, ist bei generativer KI also von einer „Intelligenz“ die Rede, die weder über Vernunft, Verstand noch über Urteilskraft verfügt, aber dafür deren Zusammenwirken im menschlichen Mitteilen und Bilderzeugen täuschend echt simuliert. Streng genommen *Bedeutungs-los* bzw. *Sinn-frei* sind die Aussagen wie gesagt schon insofern, als der KI kein stabiles Bewusstsein und auch keine eigene Konfiguration von Sinn eignet, sie streng genommen keine Signifikate kennt (als Vorstellung der Bedeutung dessen, was sie „sagt“).³⁸ Denn sie denkt nicht.³⁹ Die KI ist zudem gewissermaßen bei jeder an sie gerichtete Frage mit Arthur Rimbauds berühmtem Diktum eine andere und nie ein „Ich“; so kann z. B. auf die Frage, ob Buddha einen Sohn hatte, die Antwort nein lauten („Buddha did not have a son, in fact“), und auf die unmittelbar anschließende Frage, wer Rahula war, dann die Antwort kommen „Rahula was the only son of the Buddha“, ohne dass die KI hier einen „Widerspruch“ erkennen würde: Dies wiederum, weil sie kein Bewusstsein und kein Wissen hat.⁴⁰ Zudem handelt es sich um wahrheitsloses Sprechen, weil bei der Unmenge von Texten aus dem Netz, mit denen sie trainiert wurde, das Kriterium der propositionalen Wahrheit der jeweiligen Aussagen keinerlei Rolle spielt; mutmaßlich alle Arten von im Netz vorkommenden Texte, von mehr oder minder humorvollem Nonsense, flapsigen Kommentaren über verschwörungstheoretischen Abfall bis hin zu wissenschaftlichen und hochliterarischen Texten u. v. a. m. wurden mutmaßlich über das *Crawling* gleichermaßen und ohne jede Gewichtung zu Trainingszwecken im Zuge des *deep learning* genutzt.⁴¹ Gerüchtweise wurde so zum Training zwar das gesamte Projekt

Gutenberg genutzt, die KI zeigt sich allerdings hinsichtlich der Klassiker nur teilweise gut informiert.⁴² Bereits von der Datenbasis, das heißt eben: Texten aller Art aus dem Netz, besteht ein beträchtlicher Bias, der kulturell die US-amerikanische Perspektive und hier insbesondere die männlicher, weißer heterosexueller US-Amerikaner deutlich privilegiert; mutmaßlich handelt es sich um dieselbe Perspektive, die auch den Open-AI-Entwicklern eignet. Andrian Kreye schreibt: „So entsteht [...] ein digitales Weltbild, das letztlich nur ein Echo all der Daten ist, die eingegeben wurden. Nun unterstellt der Mensch dem Computer eine Objektivität, die sie aus dem Glauben ableitet, die Mathematik sei ein System, das in sich unfehlbar schlüssig ist. Das stimmt nicht mal in der Algebra. Einen neutralen Datenset kann es aber nicht geben. Selbst die Behauptung mancher Entwickler, KIs wie *Diffbot* seien mit dem gesamten Internet trainiert worden, würde nicht zu einem klaren Blick auf die Welt führen. Denn das Internet besteht zu einem Großteil aus Inhalten und Daten in englischer Sprache, wird von amerikanischen Inhalten dominiert. [...] [Insofern] wird der digitale Kanon und auch in der zweiten Umdrehung zu einem sich selbst bestätigende System.“⁴³ Das gilt für die „Informiertheit“, d. h. die durch Fragen (d. h. Prompts) provozierbare Wissenssimulation der KI, und selbstverständlich auch in Bezug auf alle Künste. Festmachen lässt sich dies etwa, was das Literarische betrifft, an der faszinierend detailgenauen Kenntnis von beliebten Sciencefiction- und Fantasyromanen und der eher lückenhaften großer Texte der Weltliteratur (insbesondere der nicht englischsprachigen) sowohl in inhaltlicher Hinsicht als auch, was das Produzieren von Texten „im Stile von ...“ anbetrifft. Aus der Perspektive von ChatGPT, in dessen Training mutmaßlich auch die gesamte Diskussion auf Fanseiten im Netz eingegangen ist, dürfte *Harry Potter und der Stein der Weisen* so mit der bedeutendste literarische Text überhaupt sein.⁴⁴ Die zugehörigen Untersuchungen haben, gewissermaßen als Beifang, auch ergeben, dass ChatGPT durchaus auch mit urheberrechtlich geschütztem Material trainiert wurde.⁴⁵ Ohnehin kann nicht davon ausgegangen werden, dass die KI Texte buchstäblich *liest* (oder sich Bilder *ansieht*), weder memoriert sie sie in voller Länge, noch hat sie zu jedem Text gewisse „zentrale“ Auszüge parat, und auch der Kontext der literarischen Daten, wie er etwa bei *Wikipedia* für viele „große“ Werke nachzulesen wäre, ist ChatGPT nicht bekannt. Die Trainingsdaten werden tatsächlich zu von Algorithmen hergestellten *knowledge graphs* verdichtet, aus denen die – eben generative (d. h. nicht repetitive) – KI ihre Wahrscheinlichkeitsberechnungen ableitet. Ein weiteres Problem fügt sich gut zur angesprochenen moralischen Indifferenz: Diesem Sprechen fehlt im Unterschied zu menschlichem Sich-Äußern jede Adressierbarkeit und damit jedwede Verantwortlichkeit.⁴⁶

Mit ChatGPT und Co. findet insofern die gesamtkulturelle Tendenz der „Erosion von Bedeutung“, die ich an anderer Stelle zu beschreiben versucht habe, ihren bisherigen Höhepunkt.⁴⁷ Ein gutes Beispiel ist etwa der Erfolg von fiktiven Influencern, die tatsächlich KIs sind: Den oft Hunderttausenden Followern – die bekannteste unter ihnen, „(Lil) Miquela“, hat bereits über drei Millionen davon – ist offensichtlich egal, dass das neueste „sexy Pic im heißen Sommeroutfit“ einen Menschen zeigt, der nicht existiert, und dass all seine Mitteilungen und Alltagsweisheiten aus einem Leben stammen, das gar nicht gelebt wird. Eine Differenz zu Mitteilungen lebender Influencer:innen ist für diese Follower offenbar bereits heute nicht mehr wahrnehmbar – oder nicht wichtig.⁴⁸ Denn bei Interaktionen ist beim derzeitigen Forschungsstand unklar, ob, wer schreibt „Du bist so süß“ oder fragt „Bist du noch in New York?“ hier einfach ein Spiel mitspielt oder einer Täuschung unterliegt.⁴⁹ Die Werbebranche jedenfalls preist die Vorteile, denn KI-Influencer:innen „können jederzeit arbeiten, wollen kein Geld für ihre Arbeit (sondern ihre Rechte-Inhaber:innen) und können virtuell zu geringeren Kosten um die Welt reisen. Niemals würden sie verweigern etwas zu tun, das von ihnen verlangt wird. Sie machen kaum Fehler und haben keine ‚Macken‘, außer sie werden absichtlich designed, schließlich werden sie vollständig programmiert und gesteuert, indem sie an die Anforderungen von Unternehmen, die mit ihnen zusammenarbeiten möchten, angepasst werden“.⁵⁰ Aber auch wenn die KI-generierten Bilder hier in derselben Weise unrealistisch sind wie die gängigen Stereotypisierungen „echter“ Werbebilder oder die fotografischen Inszenierungen lebender Influencer:innen, ist dennoch falsch, wenn diesbezüglich ein weiteres Mal die Phrase bemüht wird, sie verwischten „die Grenze zwischen Fiktion und Wirklichkeit“, denn wirklich sind hier höchstens die beworbenen Produkte, und ob die Mitteilungen aus dem „Leben“ der KI-Figuren (überwiegend) tatsächlich als Fiktion wahrgenommen werden, ist empirisch gar nicht untersucht.⁵¹

In jedem Fall ist derart simuliertes Denken kein Denken. Mit Friedrich „bleibt [...] festzuhalten, dass wir Searle in seiner Intuition nach mehr als vier Jahrzehnten zustimmen müssen: Die Funktionsweise der KI-Systeme sagt uns nichts über das, was unsere menschliche Vernunft ausmacht, es hat mit der menschlichen Intelligenz nichts zu tun, es kommt anders zustande und läuft nach anderen Prinzipien ab als das, was wir als das vernünftige Denken einer menschlichen Person ansehen. In seinen Funktionsprinzipien ähnelt es eher der Dressur eines Tieres, das auch nicht weiß, was es tut und was es für uns bedeutet, wenn es im Zirkus eine scheinbar richtige Antwort auf eine Frage gibt, von der wir normalerweise meinen, dass nur vernunftbegabte Wesen sie beantworten könnten.“⁵² Keine eigene Erfahrung, keine natürliche Grammatik, kein

Wissensmodell und insbesondere keine Empathie und kein moralisches Urteilsvermögen, für Chomsky et al. ist das oft durchaus gelehrt klingende Geplapper von ChatGPT und Co. damit gar kein Fall von Intelligenz, umgekehrt sei der Erfolg dieser Anwendung Zeichen eines allgemeinen intellektuellen Niedergangs. Der Schlusssatz ihres Aufsatzes in Bezug auf Anwendungen wie ChatGPT lautet: „Given the amorality, faux science and linguistic incompetence of these systems, we can only laugh or cry at their popularity.“⁵³

All dies wäre kein großes Problem, solange Benutzer:innen mit ihren konkreten Anliegen um diese Grunddispositionen wissen. Zum Problem wird es aber im Zusammenhang mit der menschlichen Tendenz, Berechnungen auf mathematischer Basis, Mitteilungen von Maschinen, Handlungen von Robotern etc. automatisch Neutralität und Objektivität zuzubilligen. Man spricht sozialpsychologisch diesbezüglich von einem *automation bias*. „Damit ist die menschliche Tendenz gemeint, die Ergebnisse maschineller Verarbeitung unkritisch zu akzeptieren und nach ihnen zu handeln.“⁵⁴

Ein gutes Beispiel findet sich bei Jennifer Jacquet: Maschinen werden offenbar vom *moral brain* anders beurteilt als Menschen. Sie berichtet von einem spieltheoretischen Experiment zum *homo oeconomicus* und seiner Rationalität, in dem eine teilnehmende Person eine gewisse Summe Geld erhält, wenn sie einer mitspielenden Person einen „angemessenen Anteil“ anbietet. Der Deal ist perfekt, wenn das Gegenüber die vorgeschlagene Teilung annimmt, lehnt es sie als ungerecht ab, gehen beide leer aus. „In aller Welt weisen Teilnehmer Angebote von weniger als 20 Prozent fast durchgängig ab und ziehen es vor, gar nichts zu erhalten. [...] Wenn sie glaubten, dass das Angebot von einem menschlichen Mitspieler kam (von dem sie nur ein Foto sahen), dann lehnten sie eine ungerechte Teilung von 80 zu 20 oder 90 zu 10 ab. Aber wenn sie glaubten, dass das unfaire Angebot von einem Computer stammte, nahmen sie es an. Gehirnregionen, die bei unfairen Angeboten von menschlichen Mitspielern verstärkte Aktivität aufwiesen, reagierten auf ein identisches Angebot von einer Maschine sogar mit einer Verringerung der Aktivität. In einem ähnlichen Experiment wurde die Leitfähigkeit der Haut gemessen — ein Indikator für die körperliche Erregung, der sich den Zusammenhang zwischen Schweißdrüsen und dem sympathischen Nervensystem zunutze macht. Bei unfairen Angeboten maßen die Wissenschaftler eine erhöhte Leitfähigkeit — aber nur dann, wenn das unfaire Angebot von einem Menschen stammte.“⁵⁵

Bei ChatGPT, dem hypermodernen Münchhausen, wo diese Haltung der Benutzer:innen die denkbar verkehrteste wäre und gerade Schwächen in Logik und Mathematik sehr auffällig sind, kommt hinzu, dass, obwohl die KI auf Rückfrage auf ihre diversen

Beschränkungen oft und gerne darauf hinweist, der sprachliche Gestus höchster Sachlichkeit selbst bei aberwitzigsten Kombinationen vielleicht wahrscheinlicher, aber eben doch verkehrter Aussagen über vermeintliche Fakten den *automation bias* eher verstärkt. Sehr schön greifbar ist dies etwa bei frei erfundenen Biografien historischer Persönlichkeiten. So machte ChatGPT Anfang 2023 noch aus dem Philosophen Johann Adolph Dori einen Komponisten des frühen 18. Jahrhunderts und wusste mit allerlei interessanten biografischen Details aufzuwarten.⁵⁶ Das eigentlich für wissenschaftliche Zwecke gedachte *Galactica*-System von Meta musste nach nur drei Tagen vom Netz genommen werden, weil es frei erfundene Sachverhalte als Tatsachen behauptet hatte.⁵⁷ Wenig ist also verkehrter, als die in diesem Zusammenhang gemeinhin vertretene Annahme, Maschinen und selbstlernende KI-Systeme seien objektiv; die Princeton-Informatikerin und KI-Expertin Johanna Bryson erläutert: „Sie dürfen sich die Maschine nicht als objektiv vorstellen, nicht einmal als Akteur. Die Maschine ist nichts als ein Werkzeug. Jemand schreibt ein Programm, und damit auch Regeln, die die Maschine anwendet. Und wenn die Maschine anhand menschlicher Entscheidungen lernt, übernimmt sie die Vorurteile [...], [so]dass Algorithmen dieselben Vorurteile reproduzieren wie Menschen. Sie benachteiligen etwa Menschen mit fremd klingenden Namen.“⁵⁸

Problematischer noch wird das Ganze, weil, wie das Buddha-Beispiel andeuten konnte, die KI die eigene Mitteilung nicht versteht, und weil sie grundsätzlich keinen Standpunkt einnimmt: Moralische Indifferenz ist hier Programm.⁵⁹ Kann KI dennoch ein Autor sein, so etwas wie maschinelle Subjektivität aufweisen (wenn sie doch nicht objektiv ist)? Kommuniziert sie überhaupt?

Zum gegenwärtigen Zeitpunkt ist vollständig abhängig von der Elastizität des jeweiligen Begriffs der Autor:innenschaft, der Subjektivität und der Kommunikation usw., ob man die genannten Szenarien der maschinellen Bild- und Textproduktion noch als einen denkbaren Fall dieser Phänomene auffasst oder aber sich vollständig in deren Jenseits wähnt. Hierzu ist wohl einiges an Diskussion erforderlich.

Was hingegen unstrittig scheint, ist, dass menschliche Autor:innenschaft, ein Gespräch/Dialog und Subjektivität hier simuliert werden und von daher hierdurch auch substituiert werden können.

Insgesamt kann solches wahrheits- und sinnfreies Sprechen in genanntem Sinne, das wahres und sinnvolles Sprechen simuliert, durchaus als perfekt subjektlos aufgefasst werden, was gerade das Problem sein könnte.

Das tatsächliche menschliche Denken hingegen kann sich niemals „jenseits aller Subjektivität“ halten, sondern ist immer in vielerlei Hinsicht situiert. Diese Situietheit lässt auch kein Außen im Denken (oder Denken des Außen) zu, weil es allzeit an die,

frei nach Hegel, „finstere Innerlichkeit“ des Bewusstseins, d. h. die rekursive Folge von affektbasierten Gedanken geknüpft und insofern mit der Subjektivität unhintergebar verstrickt ist, zumal weder das Bewusstsein noch die Gesellschaft oder die Welt im Beobachten eine Außenposition zulassen.⁶⁰ Selbst die Halluzination ist alles andere als eine Erfahrung eines Außen, auch wo sie sich dieses Anscheins bedient.

An dieser Stelle möchte ich auf die eingangs gestellte Frage zurückkommen, denn das Gesagte hat diesbezüglich erhebliche Konsequenzen: Was kein Bewusstsein hat, nicht denken kann, weder über ethisches Urteilsvermögen verfügt noch persönliche Verantwortlichkeit kennt, kann, wie es auch die geltende Rechtslage hierzulande vorsieht, nicht als Autor angesehen werden, auch wenn es Texte oder Bilder generieren kann. Überhaupt scheint es nicht klug, generativer KI autonom irgendwelche Schritte von größerer Relevanz und schon gar keine lebenswichtigen Entscheidungen zu überlassen. Damit ist allerdings noch nichts über einen geleiteten Einsatz als Werkzeug gesagt, und das gilt selbstverständlich auch für das Hervorbringen von fiktionalen und nicht-fiktionalen Artefakten aller Art in den Künsten.

Disclaimer: enorme Gefahren, kleiner werdendes Regulierungsfenster

Auch wenn die Risiken sicher nicht primär im Feld der Künste und ihrer Ausbildungen liegen, ist es nicht seriös, über KI zu schreiben, ohne wenigstens kurz auf die zahlreichen Gefahren hinzuweisen, die von einer maximal disruptiven Entwicklung wie der ungesteuerten Implementation generativer KI ausgehen. Denn sie sind enorm; es ist keine Übertreibung, wenn man sagt, außer Atombomben und Biowaffen habe die Menschheit bisher nichts Gefährlicheres hervorgebracht.⁶¹ Dies liegt sicher nicht daran, dass die Gefahr bestünde, KI würde mittelfristig „die Macht übernehmen“, was nur ein durch diverse ästhetische Fiktionen genährtes Phantasma ist.⁶² Aber selbstverständlich kann KI von Menschen auf mannigfaltige Weise genutzt werden, um andere Menschen schnell und effektiv zu töten.

Für die Entwicklung und Bedeutung generativer KI ist so, mit etwa Friedrich Kittlers oder Paul Virilios medientheoretischen Basisüberlegungen im Hintergrund wenig überraschend, der militärische Bereich entscheidend, wenn auch darüber aus Geheimhaltungsgründen wenig gesprochen und geschrieben wird. Folgt man aber z. B. der Darstellung des US-Sicherheitsexperten und Militärreporters Jay N. Tuck ist so etwa der Ukrainekrieg aus der Perspektive des Westens ein gigantisches Versuchslabor für die jüngeren Neuerungen der KI-basierten Militärtechnik, welche bereits heute die Kriegsführung radikal verändert hat und ganze Waffengattungen wie etwa Panzer in Frage zu stellen scheint.⁶³

Wie immer bei solchen Neuerungen wird gegenwärtig, insbesondere für die großen Sprachmodelle, auf denen ChatGPT basiert und in denen die virtuelle Intelligenz sitzt, einigermaßen breit über Gefahren diskutiert. Leider beschränken sich diese Debatten weitgehend auf zwei zentrale Problemfelder, nämlich einerseits den wenig diversen Bias der KI-Anwendungen, was recht umstandslos aus dem der Datenbasis folgt, und andererseits die drohende und fast kostenlose Substitution menschlicher Kreativität von Produktdesigner:innen, Werbetexter:innen, Journalist:innen, Drehbuchautor:innen, Literat:innen, Comiczeichner:innen, Illustrator:innen, Fotograf:innen und teilweise sogar Programmierer:innen usw. Anzunehmen, das seien die einzigen Probleme und die im zweiten Fall anschließende Debatte um „Innovationsfeindlichkeit“ sei in irgendeiner Weise hilfreich, wäre in meinen Augen entschieden ein Trugschluss: Es gibt in diesem Kontext diverse weit größere Gefahren, etwa im Bereich der Zersetzung des Politischen; so kursierten beispielsweise wenige Tagen nach der Ankündigung Donald Trumps im März 2023 an seine Anhänger:innen, er werde bald festgenommen, recht eindrucksvolle KI-generierte Bilder dieses Ereignisses – das allerdings nie stattgefunden hatte.⁶⁴ Das mutmaßlich in politischer Hinsicht größte Risiko benennt Steffen Albrecht: „Größere Mengen [...] bewusst schädigend eingesetzter Texte könnten das Vertrauen in den öffentlichen Diskurs als Ort demokratischer Meinungsbildung unterminieren.“⁶⁵ Und auch perfekte Deepfakes, welche jedermann augenfällig die Wahrheit einer beliebigen Verschwörungserzählung „beweisen“, werden bald an der Tagesordnung sein.⁶⁶ Viel diskutiert, aber als Gefahr eher diffus und derzeit eher noch ein Gerücht ist die angesprochene Verdrängung menschlicher Kreativität inklusive der zugehörigen sozioökonomischen Folgen wie Jobverlust und Arbeitslosigkeit von *creative workers*. Tatsächlich aber wäre eine etwaige Verdrängung oder gar Substitution menschlicher auktorialer Kreativität (und der noch weiter fortschreitenden Prekarisierung der „Kreativen“) sicher keines der KI und ihrer Möglichkeiten, sondern eines der Verwertungslogik, die kreativen und informativen Content möglichst gratis generieren will.⁶⁷ Auf den Journalismus bezogen, kann man etwas platt formulieren: Das Problem sitzt nicht in den Redaktionen und gewiss nicht in den Algorithmen von ChatGPT und Co., sondern in den Verlagshäusern. Entsprechend wäre es auf dieser Ebene regulativ vergleichsweise einfach zu vermeiden, die längst überfällige Kennzeichnungspflicht KI-generierter Inhalte wäre ein kleiner Anfangsschritt, die Durchsetzung geltenden Urheberrechts ein weiterer.⁶⁸ Unaufhaltsam scheint aber, dass alles, was man im Internet machen kann, diese Systeme bald autonom oder fast autonom bewerkstelligen werden, insbesondere auch alles, was illegal ist.⁶⁹ So wird mit FraudGPT und WormGPT bereits eine entsicherte Form von ChatGPT speziell für Hacker- und Kriminellenzwecke beworben (<https://wormgpt.com>),

z. B. zum Schreiben von Malware; die Bezahlfunktion erfolgt hier über Kryptowährung.⁷⁰

Laut dem aktuellen Lagebericht des BSI des Jahres 2023, der den Zeitraum vom 1. Juni 2022 bis zum 30. Juni 2023 abbildet, dürfte durch den zunehmenden Missbrauch von KI-Sprachmodellen wie ChatGPT die Cyberkriminalität insgesamt deutlich einfacher werden und damit massiv zulegen. Auch ohne jede böse Absicht birgt zudem die Nutzung von KI bei der Programmierung das Risiko einer Vervielfältigung von bestehenden Software-Schwachstellen.⁷¹

Im Zuge der Berichterstattung um den von der britischen Regierung einberufenen ersten „AI Safety Summit“ am 1. und 2. November 2023 brachte es die Gefahr der KI-optimierten Herstellung biologischer Kampfstoffe (die ggf. von einer Handvoll Terroristen weltweit in den Lüftungsanlagen großer Flughäfen ausgebracht werden könnten) am 2. November 2023 bis in die 20-Uhr-Nachrichten der *Tagesschau*. Dies ist nur ein mögliches Horrorszenario; terroristische Angriffe auf die sensible Logistik- und Sicherheitsarchitektur der Großstädte, auf Atomkraftwerke, auf Staudämme – letztlich auf alles, was mit dem WWW verbunden ist. Es gibt in diesem Bereich terroristischer Aktivität wenig, was durch generative KI nicht erheblich vereinfacht würde.⁷²

Insgesamt wird die Einsicht, dass diese Systeme bereits jetzt potentielle Massenvernichtungswaffen sind, von Expert:innen seit Jahren zwar immer wieder betont, aber die politischen Konsequenzen daraus sind marginal geblieben; das kurze Zeitfenster einer möglichen Regulierung, der viel zitierte „Oppenheimer-Moment“, könnte inzwischen bereits geschlossen sein.

Neben der sicher zu erwartenden immensen Zunahme von Computer-generierten Texten insbesondere auch zu schädigenden Zwecken wie dem unerwünschter Werbung, immer effektiveren Angriffen auf die Computersicherheit von staatlichen und kriminellen Akteuren, gezielter Desinformation, Phishing, Malware auch von Amateuren usw. wird zudem bereits aktuell als Risiko angesehen, dass die Nutzung generativer KI bestehende soziale und ökonomische Ungleichheiten verschärfen dürfte, da sie auf unterschiedlich vorhandenen Kompetenzen beruht, und auch, dass die Konzentration des Angebots und der Forschung bei einigen wenigen global agierenden Tech-Konzernen wie *Google/Alphabet, Meta, Microsoft* usw. in Hinsicht einer demokratischen Kontrolle der weiteren Entwicklungen wenig hoffnungsfroh stimmt.⁷³

Erfundene Bilder: Anmerkungen zum Problem der Fiktionalität KI-generierter Bilder und Texte

Zumindest eher traditionelle, textorientierte Theaterstücke der Klassikerbühne von Shakespeare bis Brecht, Sachs, Beckett und Jelinek, fast alle Romane, Kurzgeschichten, Spielfilme, bestimmte Computerspiele, Musicals, Opern und Operetten, die meisten Comics, Fotoromane, bestimmte Ballett- und Tanzaufführungen wie Marcia Haydéés „Dornröschen“ (1987, nach Charles Perrault) oder John Neumeiers „Romeo und Julia“ (1971, nach Sergei Prokofjew), aber auch zahlreiche Werke der Malerei, Bildhauerei und sogar manche Fotografien u. v. a. m. können nach einer auf die Antike zurückgehenden Begriffsverwendung allesamt als (ästhetische) Fiktionen gelten.

Allerdings gehört der auf den ersten Blick schlagend einfache und seit der Antike auch für bestimmte Erfindungen des Rechtswesens gängige Begriff der „Fiktion“ mit der starken Konnotation des Erdichteten („Fingierten“), aber auch der der „Figuren“, des bloß „Fiktiven“ oder, eher weniger gebräuchlich, des „Figments“ (Einbildung) usw. zu den eher schwierigen philosophisch-ästhetischen Begriffsbildungen.

Zwar hat entgegen dem anderslautenden Gerücht in der zugehörigen Theorie, von einigen wenigen Extremfällen wie dem Roman „Marbot. Eine Biografie“ von Wolfgang Hildesheimer (1981) abgesehen, praktisch niemand je ein größeres Problem, eine ästhetische Fiktion zu erkennen, das heißt, sie von „der Realität“ zu unterscheiden, wohl aber, auf eine gültige Formel zu bringen, was exakt unter beidem zu verstehen ist und wodurch sich Erstere von anderen, nichtfiktionalen Medienartefakten unterscheidet, man denke nur an die verschachtelte Diskussion einer „dokumentarische Fiktion“ in der Filmtheorie.

Rein vom Wortsinn von lat. *ingere*, dt. „bilden“, „sich vorstellen“, „gestalten“, „erdichten“, „vortäuschen“, aber auch „streicheln“ bzw. von lat. *fictio*, „Gestaltung“, „Bildung“, „Dichtung“ her und ohne weitere theoretische Einschränkung wäre der Begriff der Fiktion mehr oder minder deckungsgleich mit a) dem des Artefaktes (als Gemachtem), b) der Erfindung oder gar c) dem Erdachten, d. h. jeder Vorstellung? Von daher ist, seitdem von Fiktionen in Bezug auf die Künste gesprochen wird, als entscheidend herausgestellt worden, dass in einer *ästhetischen* Fiktion das (teilweise) Ausgedachte dezidiert als Ausgedachtes und nicht als Tatsache präsentiert, z. B. vorgetragen, gemalt oder aufgeführt wird.⁷⁴ Das unterscheidet die Fiktion vom Traum, dessen Bildwelten insbesondere in Bezug auf den Spielfilm immer wieder zu fragwürdigen Analogieschlüssen hinsichtlich ästhetischer Fiktionen verleitet haben. Derartige Kurzschlüsse sind in vielerlei Hinsicht irreführend: Filmbilder sind aufs Zuschauen angelegt, während Träumende in ihren Traumbildern fast immer als Beteiligte auftreten, die Bilder im Film sind prinzipiell fremde Bilder (des Anderen), die Traumbilder eigene (des Unbewussten), der Spielfilm ist eine Fiktion, etwas mit mimetischen Mitteln aufwändig und arbeitsteilig Hergestelltes, der Traum nicht, der

Film ist technisch-medial, der Traum nicht usw.⁷⁵ Der Traum ist insofern näher bei einer Halluzination (als echtem Trugbild) anzusiedeln als bei der filmischen Fiktion. Tatsächlich ist allerdings das, wenn man so will, alltägliche Muster aller ästhetischer Fiktionen der erzählte, d. h. ausgeschmückte und mit einer Botschaft versehene Traum.

Ein einfacher und breit anschlussfähiger Definitionsvorschlag der ästhetischen Fiktion, der auf einen Rekurs auf die üblicherweise prominent herausgestellte Figur des Als-ob (oder *make-believe* bzw. *pretense*) bewusst verzichtet, kann folgendermaßen lauten:

Von einer ästhetischen Fiktion sprechen wir medienübergreifend sinnvollerweise immer dann, wenn komplexes bildhaft Vorgestelltes als ein Ganzes versinnlicht, d. h. in technisch-mediale Form gebracht wird und seine Vorstellungsbilder dann mit dem Anspruch auf Nicht-Tatsächlichkeit (potentiell oder aktuell) kommuniziert werden, sodass bei den Rezipient:innen die Vorstellung einer in Ansätzen anderen Welt bzw. Wirklichkeit entsteht, in die rezeptiv „eingetaucht“ werden kann und die dabei jederzeit besagtem Ganzen zuzuordnen ist.

Im Hinblick auf die erfundenen Bilder der ästhetischen Fiktion ist zunächst gleichgültig, ob sie bewusst oder unbewusst erfunden wurden, ob sie von einem Menschen oder einem Kollektiv ersonnen wurden – oder eben von einer KI generiert. Nicht gleichgültig ist hingegen, ob sie als mediales Artefakt vorliegen, d. h. ob sie rezipiert werden können; im einfachsten Fall müssen sie von einer Stimme vorgetragen werden, daneben sind prinzipiell alle bildfähigen Medien fiktionsfähig, angefangen beim Sand, in den man eine phantastische Gestalt zeichnet, bis hin zur Musik, die Vorstellungsbilder evozieren kann.⁷⁶

Entscheidend ist der prinzipiell dialogische Charakter der auktorialen *Vor-stellung* erfundener Bilder und Dialoge und ihrer rezeptiven Konkretisierung durch das Publikum. Zumindest dort, wo die Autor:innenfunktion der Bilderwelten noch vollständig oder wenigstens leitend von Menschen eingenommen wird, muss, da zugleich immer Rezipient:innen im Spiel sind, die dem medial vorgestellten Als-ob mit ihren Vorstellungen erst seine Präsenz verschaffen, für ästhetische Fiktionen aller Art im Minimum von einer gedoppelten Subjektivität gesprochen werden: Je offensichtlicher die Rolle der Autor:innen als Schöpfer:innen bspw. eines fiktionalen Kosmos ist, und je stärker die Einbildungskraft und damit der imaginäre Charakter eines Artefakts im Vordergrund steht (als dessen „Fingiertes“, Ausgedachtes, Erfundenes), desto offensichtlicher ist sein fiktionaler Charakter; dies kann sogar im Feld der Fotografie beobachtet werden, wo die Fiktion kein etabliertes Genre darstellt, denkt

man etwa z. B. an die Arbeiten Sandy Skoglunds, oder, anders, an die Montagen eines Jerry N. Uelsmann.⁷⁷

Überdies ist der Umstand zentral, dass bei einer ästhetischen Fiktion, gleich ob es sich um einen Roman, einen Spielfilm, ein Comic, ein Computerspiel des Fantasy- oder Horrorgenres, ein Hörspiel, Historienmalerei oder ein Werk der sogenannten inszenierten Fotografie usw. handelt, jeweils medienspezifische *Fiktionssignale* den fiktionalen Status des Gezeigten, Gehörten, Aufgeführten herausstellen. Darüber wird ganz automatisch eine Grenze des fiktionalen Kosmos mit seinem „willkommenen Schein“, in den ggf. rezentriert wird, von der äußeren Welt, in der das Publikum sein Leben führt, gezogen.

Der Standardmodus des menschlichen Sich-Äußerns, der der (potentiellen) Tatsächlichkeit des Geäußerten, kennt also nicht nur, aber besonders im Feld der Künste im weiteren Sinn ein jedweder Imagination offenes Pendant, die fiktionale Rede, klassisch etwa angezeigt durch ein „Stell dir vor ...“ oder ein „Es war einmal ...“ oder auch im Kinosetting oder bei der gestreamten Serie die Titelsequenz und den Cast im Abspann.

Wie gesagt, und das ist das aus der Perspektive der Fiktionstheorie gesehen vielleicht Faszinierendste, kann generative KI prinzipiell nicht zwischen in diesem Sinne fiktionalem und tatsächlichem Sprechen unterscheiden und erzeugt damit an diesem Punkt immer etwas *sui generis*, selbst dort, wo es als fiktionale (oder tatsächliche) Rede gelesen werden wird. Anders gesagt, ist der Standardmodus ihrer Mitteilungen in Bezug auf Sachfragen nicht der einer Tatsächlichkeit, sondern bestenfalls der einer auf Wahrscheinlichkeitsberechnungen auf Signifikantenebene beruhenden *simulierten Tatsächlichkeit*, mithin einer, der Tatsächliches und Fiktives im Modus der Tatsächlichkeit mixt. Wie gesagt, neigt die KI dazu zu „halluzinieren“, d. h. das überaus Plausible, Hochwahrscheinliche zu konstruieren, d. h. berechnen, auch wo dies tatsächlich nicht der Fall ist.⁷⁸ Es existiert hier strukturell überhaupt keine Differenz von fiktionalem und tatsächlichem Sprechen; einen Unterschied von Fiktion und Wirklichkeit können, wenn überhaupt, erst die menschlichen Nutzer:innen (oft mit Mühe) herstellen, also in die jeweilige bildliche und/oder textliche Mitteilung hineinprojizieren.

Es wäre aber, nur weil eine solche Grenze hier nicht per se existiert, sicher verkehrt, dass alle von ChatGPT formulierten Texte als Fiktion aufzufassen seien. Für eine ästhetischen Fiktion fehlt hier a) formal der durch Fiktionssignale bekundete Anspruch auf Nicht-Tatsächlichkeit und b) inhaltlich (meist) der Charakter eines Ganzen, in das rezeptiv eingetaucht werden kann, so dies nicht im Prompt eingefordert wurde („Schreibe eine Kurzgeschichte, in der ...“). Dem genannten Definitionskriterium

ästhetischer Fiktionen genügen mutmaßlich nur die allerwenigsten von ChatGPT und verwandten Anwendungen generierten Texte und Bilder, in der Regel solche, die im Prompt die Anweisung enthalten, ein Gedicht, einen Roman, ein Märchen usw. zu schreiben oder eine Fantasiegestalt, ein Fabelwesen, ein surrealistisches Bild o. Ä. zu entwerfen.

Betrachtet man aber von generativer KI gefertigte Romane, Kurzgeschichten, Comics, fiktionale Bilder etc., wird ein zweiter Umstand noch wichtiger: Auch die fiktionstheoretisch konstitutiven Metaphern des Dialogs bzw. Kontrakts von Autor:innen und Rezipient:innen verlieren hier vollständig ihren Sinn, weil diese „Autor:innen“ subjektlos und ohne stabiles Bewusstsein sind.

Umgekehrt könnte aus den zuvor benannten Gründen grundsätzlich bestritten werden, dass irgendeine Hervorbringung der generativen KI eine ästhetische Fiktion ist, exakter handelt es sich um eine Simulation einer solchen. Denn ästhetische Fiktionen aller Art sind zwar nicht notwendig – man denke etwa an Märchen, Sagen, Witze, fiktionale Computerspiele –, aber doch im Regelfall an eine individuell identifizierbare Autor:innenschaft und damit an menschliche Subjektivität, d. h. ein lebendes oder bereits verstorbenes Ich als Schöpfer:in des Werks gebunden, mit dem rezeptiv in einen stummen Dialog getreten werden kann: „Was will der/die Künstler:in mir hier sagen?“; „Verschafft sich hier eine private Obsession Raum?“ o. Ä. Diesem Kriterium können KI-generierte Texte und Bilder nicht genügen, denn mit einer KI ist keine Sympraxis möglich und die zugehörigen Fragen werden sinnlos.

Lässt man dieses grundsätzliche Kriterium der Möglichkeit KI-generierter piktoraler Fiktionalität außer Acht, weil man vom ohnehin wahrscheinlichsten Fall ausgeht, dass menschliche Künstler:innen die KI für ihre Arbeit als Werkzeug nutzen wie etwa beim zuletzt Aufsehen erregt habenden Fin-de-siècle-Symbolismus, den Jason M. Allen 2022 mithilfe von Midjourney unter dem Titel „Théâtre d’Opéra Spatial“ dem erstaunten Publikum präsentierte und mit dem er im September 2022 den jährlichen Kunstwettbewerb der Colorado State Fair gewann. Dies übertraf für meine Begriffe den ersten öffentlich beachteten Erfolg eines KI generierten ‚Gemäldes‘, des „Portrait of Edmund de Belamy“ (2018), das bei einer Auktion des New Yorker Auktionshauses Christie’s den Rekordpreis von 432.500 US-Dollar eingespielt hatte, insofern, als Allens Bild wenigstens aussah wie ein ernst zu nehmendes, wenn auch einer längst vergangenen Epoche zuzurechnendes Kunstwerk.⁷⁹

Insgesamt kann man in Hinsicht auf die kreative Potenz generativer KI in Hinsicht auf Bilderwelten festhalten, gleich ob es um Animationen, Bilder in Games, Comics, Filmkulissen oder die Stockfotografie geht: Während der rasch wachsende Bereich der *nonhuman photography* ohne KI-Einsatz in der Regel Bilder hervorbringt, die nicht als

fiktional angesehen werden können, sind die diesbezüglichen Möglichkeiten der Autor:innenschaft künstlicher Intelligenz nicht nur im Medium Sprache sehr weit vorangeschritten.⁸⁰

Im Hinblick auf literarische Fiktionen (Lyrik, Drehbücher, Romane) ist auffällig, dass zum gegenwärtigen Stand der Entwicklung die entsprechenden Anwendungen – es gibt, insbesondere für englische Texte, inzwischen zahlreiche Anbieter – mit ein wenig Überarbeitung, Umprogrammierung und Geschick beim Formulieren der Prompts experimentelle Literaturformen im Stile, wenn auch nicht im Geiste der modernen Hochliteratur, oder an den Autorenfilm erinnernde Drehbücher vergleichsweise problemlos generieren kann, während die Texte trivialerer Formate weit stärker überarbeitet werden müssen. Bei den Open-Source-Anwendungen muss dazu die generative KI je nach gewünschtem Resultat mit den entsprechenden Vorlagen weitertrainiert werden; aber sie ist prinzipiell, ein wenig abseits der standardmäßig kommerziellen Fokussierung dieser Tools auf Blogbeiträge, Werbetexte etc., in dieser Hinsicht durchaus individuell „umprogrammierbar“. In Bezug auf Bilder ist dies schwieriger und eine besondere Eignung fürs Avancierte, Sperrige und Spröde kann hier nicht beschrieben werden. Die am meisten frappierende Fähigkeit ist hier die (wenig originelle und kaum kunsttaugliche) Stilkopie. Doch das ist, was gewisse, vor allem auch recht offensichtliche Limitierungen betrifft, vielleicht auch nur Momentaufnahme. Dass von der KI generierter Content selbst im semiotisch anspruchsvollen Comicbereich mit seinen Bild-Text-Kombinationen aus Sicht der Zeichner:innen bereits ein ernsthaftes Problem ist, deutet an, wohin vor allem im Mainstreamsegment der kreativen Disziplinen die Reise geht.⁸¹

Im Fall von Bildern, bewegten zumal, sind die diversen Text-to-image-Methoden noch etwas weniger perfekt, dennoch ist beeindruckend, wie bereits ein extrem abgespecktes Gratistool im Netz (ein Beispiel wäre etwa <https://huggingface.co/spaces/dalle-mini/dalle-mini>) diverse *poor images* von fiktiven Picassos kreiert werden, wenn man nur eingibt „Hund im Stile Picassos“.⁸² Mit dem zuletzt stark verbesserten Midjourney, Dall-E und Stable Diffusion stehen unterschiedliche ausgereifte KI-Module zur Bildgenerierung zur Verfügung; mit Letzterem ist zudem erstmals eine Software gratis verfügbar, die Prompts am heimischen Rechner in Bilder umwandeln kann, welche qualitativ nahe am (derzeit) maximal Möglichen liegen.⁸³

Die komplexen generativen KI-Programme lesen hierfür in vorhandene Bilder eine Quasi-Syntax hinein, indem sie die eingegebenen sprachlichen Propositionen nach entsprechendem Training mit Millionen von Bildern bzw. Metadaten in piktorale Prädikate übersetzen und so neue, ungesehene Bilder generieren. Im Prinzip dieselbe Entwicklung ist mit einer kleinen Verzögerung für bewegte Bilder zu erwarten; in

„Indiana Jones und das Rad des Schicksals“ (USA 2023) spielte ein digital verjüngtes Double Harrison Fords die Hauptrolle, und das britische Unternehmen *Synthesia* verkauft gar eine Software, mit der eine einmalige Darbietung eines Schauspielers in einen endlosen Strom von Deepfake-Avataren verwandelt werden kann, die auf Knopfdruck vorgegebene Skripte rezitieren.⁸⁴

V. a. bei Midjourney und Stable Diffusion pausen sich dabei derzeit allerdings noch gewisse ästhetische Vorlieben offenbar zu Trainingszwecken forciert genutzten Plattformen wie *DeviantArt* recht deutlich durch. „Das betrifft das einschlägige Motivrepertoire zwischen Science-Fiction und Fantasy, Manga und Märchenwelt, Cyber- und Steampunk, ebenso wie eher formale Aspekte der Farbigkeit und Bildkomposition.“⁸⁵

Zum Beispiel können die KIs eine Vielzahl von als männlich gelesenen Personen verbildlichen (zur Eingabe „Mann“), zudem eine bestimmte Mimik und Körperhaltung mit einem Affekt in Verbindung bringen (zu „traurig“), und dieser Gestalt dann einen Hut aufsetzen („trägt einen Hut“). Das Prinzip ermöglicht praktisch in Echtzeit, ungeheuer viele Bilder zu der sprachlichen Aussage „Ein trauriger Mann trägt einen Hut“ zu generieren, die dann auch noch einen bestimmten künstlerischen Stil (z. B. auf die Eingabe „im Stil von Max Beckmann“ hin) oder, als fotorealistisches Bild, einen bestimmten Kameratyp, etwa eine Polaroidaufnahme, simulieren können.

Um Fotografien im eigentlichen Sinn handelt es sich dabei aber nicht, denn hier wurde nichts belichtet. Im Minimum wird man insofern mit Daniel Rubinstein sagen, dass es im Netz eine immer mehr zunehmende Zahl kameraloser digitaler Bilder gibt: „Digitale Bilder können ohne Kamera, ohne Chemie, ohne Objektive, ja sogar ohne Licht hergestellt werden. Das bedeutet, dass all die alte Rhetorik, nach der die Fotografie die Spur des Realen sei oder eine indexikalische Verbindung zu Ereignissen in der Vergangenheit habe, nicht auf das digital entstandene Bild zutreffen muss. Die Vorstellung, dass Fotografien eine repräsentative, indexikalische oder signifizierende Verbindung mit Ereignissen, Menschen und Objekten in der realen Welt haben, muss für digitale Bilder, die auf elektronischen Signalen und Computation beruhen, nicht gelten.“⁸⁶

Zurück zur Frage der Fiktionalität dieser Bilder: Sowohl traditionelle Gemälde, Zeichnungen und Fotografien als auch KI-generierte Bilder aller Art können (piktorale) Fiktionen gemäß der obigen Definition sein, wenn sie wie etwa klassische Werke surrealistischer Fotografie etwa bei Claude Cahun, Herbert Bayer oder Karel Teige oder die fotografischen Inszenierungen Cindy Shermans oder Jeff Walls u. v. a. m. den Blick der Betrachter:innen in einen irrealen Kosmos führen, der von unserer Welt radikal geschieden ist: Bilder erfundener Welten und Szenerien. Nie war es so einfach, solche

Bildwelten zu schaffen wie eben gegenwärtig in Gestalt der KI-generierten Bilder gängiger Text-to-image-Anwendungen. Entstanden sind so neue Formen der alltäglichen piktoralen Fiktion, oft ein Abfallprodukt eines Ausprobierens der Programme. Diese Bilder, die auf Onlineforen gesammelt werden, sind reine Simulakren, indem sie z. B. nichtfotografische Fotografien, nichtgemalte Gemälde und nichtgezeichnete Zeichnungen usw. vorstellen.

Um dieses dreifache „nicht-“, das ja an keinem Punkt eine Negation meint, zu positivieren, muss man sich wiederum den technischen Prozess vergegenwärtigen, der auf komplexen Berechnungen beruht: In allergrößter Vereinfachung gesagt, übersetzen die zugehörigen Programme Eingabebegriffe (auf der Textseite technisch ganz ähnlich wie ChatGPT und Co.) über die Verschachtelung mit weiteren neuronalen Netzen, die Bilder dechiffrieren können, in Qualia. Sie „verstehen“, wiederum sehr vereinfacht gesprochen, nur ganze Worte, und ihr „Vokabular“ ist begrenzt. Als Orientierungsgröße fungiert jeweils ein einfacher Satz, etwa „an old painting of a cherry on a victorian bridge wearing a party hat“, sie erkennen isolierte Bildelemente anhand wiederkehrender Regelmäßigkeiten aller zu Trainingszwecken genutzten Bilder, die z. B. als „Hut“, „Kirsche“, „viktorianisch“ oder „Brücke“ gelabelt wurden, wobei die gleichnamigen Teppiche hier zunächst problematisch gewesen sein könnten.

Codiert werden hier zunächst sowohl Begriffe als auch Bilder (als gedoppelter, voneinander unabhängiger Lernprozess), um dann mit dem derart trainierten Generator, der verschiedene GANs nutzt, Pixelwerte möglicher, der Eingabe (dem Prompt) entsprechender RGB-Verteilungen zu berechnen. Das Discriminator-Netz prüft dann das Resultat auf „Echtheit“ und verwirft diverse „falsche“ Bilder. Bei diesem Spiel versucht der Generator also gewissermaßen, den Discriminator zu täuschen und ihm vorzumachen, es handele sich um gar kein generiertes Bild.⁸⁷

Hierzu kann auf die berechneten Muster (die den sichtbaren und unsichtbaren Qualia entsprechen, inklusive diverser *latent features*), die erlernt wurden, zurückgegriffen werden, wobei zunächst einmal „wichtige“ von „unwichtigen“ Bildelementen unterschieden werden mussten. Die den Eingaben (Prompts) entsprechenden Kombinationen verschiedener Qualia (bzw. der zugehörigen Algorithmen) werden mit den komplexen Programmen über weitere Netzwerke schlussendlich zu sichtbaren Bildern dekodiert, d. h. visualisiert (als sogenannte *Diffusion* des Decoders, hier über autoregressive einerseits statistische Modelle und andererseits vektorbasierte Diffusionsmodelle).⁸⁸

Die genannten Faktoren, allesamt Resultate des maschinellen Lernens, generieren also einem Möglichkeitsraum möglicher Pixelverteilungen, der entlang der Eingaben nach Wahrscheinlichkeiten gefüllt wird. Damit können immer neue Eingaben bestimmten

neuen Wertniveaus (möglichen Bildelementen) zugeordnet werden, welche die Maschine über Prozesse des Deep Learning den Bildern bzw. den zugehörigen Metadaten entnommen hatte, mit denen sie trainiert wurde.

Im Beispiel wird etwa der Begriff „cherry“ zunächst in so etwas wie eine allgemeine „Kirschenhaftigkeit“ umgerechnet, ermittelt aus der Varianz aller Pixelverteilungen, die beim Training als „Kirsche“ markiert wurden, Süß- und Sauerkirschen, fotografierte und gemalte bzw. gezeichnete, reife und gelbe usw. Über der Kombination mit „party hat“ (als Derivat der „Partyhut-haftigkeit“ als vor allem der entsprechende geformte Gegenstand und die Positionierung auf dem Kopf) wird die Kirsche vielleicht einen Hut aufgesetzt bekommen, und über das „old painting“ als (altmeisterliche) „Gemäldehaftigkeit“ im nun dreidimensionalen mathematischen Möglichkeitsraum einen entsprechenden Anschein verpasst bekommen, usw.

Um die visuelle Essenz eines Prädikats wie „Gemäldehaftigkeit“ oder die „Brückenhaftigkeit“ einer Brücke oder auch dessen, was die „Aufgesetzhaftigkeit“ des Verbs „wearing“ in Bezug auf einen Hut so alles meinen kann, zu kennen, berechnet die generative KI im Training *latent features* dieser Qualia, indem gemeinsame, komplexe Variablen aller ihm als Gemälde, Brücke oder Aufgesetzes usw. enthaltenden Bilder (Pixelkombinationen) berechnet wurden, welche dem Encoder zu Trainingszwecken vorgelegt wurden. Den mathematischen Zusammenhang der *latent features* könnte ein Mensch sicher nicht nachvollziehen. Im Training werden dann abschließend über eine Evaluation hinsichtlich (Foto-)Realismus, Ähnlichkeit zur Eingabe (Prompt), ästhetischer Qualität usw. mit menschlichen Probanden, meist schlecht bezahlten Klickarbeiter:innen aus Niedriglohnländern, z. B. in Afrika, laufend „richtige“ von „falschen“ Ergebnissen unterschieden. Daraus lernt die KI dann und verbessert sich. Weiter kann ein Ausgabeformat vorgegeben werden, etwa als Simulation einer Bleistiftzeichnung oder als Fotografie eines bestimmten Kameratyps (ein weiteres aus den Trainingsdaten mit zugehörigen Bildern berechnetes *latent feature*), eine 3-D-Animation oder ein Gemälde.⁸⁹ Am Ende kann die Maschine für jeden als möglich markierten Punkt des zur Eingabe passenden n-dimensionalen Raums – eine komplexe Zahlenkombination – über den Decoder ein referenzenloses Bild generieren; man spricht hierbei von Diffusion. Diese übernehmen mehrere komplexe Programme, die miteinander interagieren, um ein passendes Konstruktionsrezept zu erstellen und dies dann umzusetzen. Die KI arbeitet also in größter Vereinfachung in einem n-dimensionalen mathematischen Raum und berechnet Kombinationsmöglichkeiten von zum Eingabetext „passenden“ Pixelwerten, von denen dann einige in Bilder *diffundiert* werden, die in den gängigen Onlineprogrammen bis zur maximalen Auflösung von 1024 x 1024 ausgegeben werden können.⁹⁰

Wie bei den auf menschlicher Einbildungskraft beruhenden Bildern gibt es über das Deep Learning also auch hier eine Spur bereits vorhandener Bilder, die allerdings durch komplexe mathematische Prozesse gebrochen wird.

Solche KI-generierten Bilder etwa zu der Eingabe „a group of skiers are preparing to ski down a mountain“ mögen zwar auf Wunsch fotorealistisch, wie gewöhnliche Schnappschussaufnahmen einer Gruppe von Skifahrer:innen vor der Abfahrt wirken, aber sie sind keinesfalls aus vorhandenen Bildern im Netz zusammengesetzte, sondern komplett neu berechnete Bilder, die damit in ihrer reinen Virtualität vollständig frei von jedem Rest eines realen Referenten sind.⁹¹ Keiner der hier zu sehenden Menschen wurde je geboren. Die durchs Deep Learning maximal bilderfahrenen KI-Systeme berechnen also mit großer Rasanz und in bemerkenswerter Vielfalt Pixelwerte, die Simulakren erzeugen, die von Menschen ggf. als fotorealistische Bilder gelesen werden, deren Referenten aber zu keinem Zeitpunkt in „unserer“ Welt existiert haben, auch wenn man „a photography of teddy bear on a skateboard on times square“ eingibt und das Bild perfekt ein (nicht existentes) Foto dieser Inszenierung simuliert.⁹²

Sprechend ist das Beispiel im hier diskutierten Zusammenhang auch, weil diese Medienartefakte, wie bereits erwähnt, als reine Simulakren prinzipiell fiktionalisiert, d. h. als piktorale Fiktion gelesen werden müssten. Sie werden es tatsächlich aber immer (nur) dann auch werden, wenn man im Gegensatz zu den beiden letztgenannten Beispielen etwas ungewöhnlichere Kombinationen eingibt. Bereits „a toy knight elephant playing a flute“ generiert verschiedene fotorealistische Aufnahme von flötenden Spielzeugelefanten im Rittergewand, die wie Produktfotografien nichtexistierender Plastikspielfiguren aussehen (es wäre nicht schwierig, sie in Massenproduktion zu geben).⁹³ Alle so entstehenden Bilder sind erfundene Bilder, weil auch die realistisch aussehenden Menschen maschinelle Erfindung sind und nie geboren wurden. Denn strukturell sind all diese Bilder als Simulationen echter Malerei oder echter Fotografie usw. insofern immer fiktional, als sie z. B. ein Bild einer belebten Fußgängerpassage in Manhattan zeigen, das aussieht, als ob es eine Fotografie wäre, obwohl keiner der abgebildeten Menschen je gelebt hat. Fiktives wird als real visualisiert.

Aber da der kommunikative Anspruch auf Fiktionalität, etwa bei einer Verwendung in der Werbung, oft fehlt, werden nur manche der Bilder auch rezeptiv fiktionalisiert (als piktorale Fiktion gelesen) werden, etwa das unserer Kirsche mit Hut auf der Brücke, d. h. diejenigen Bilder, die, wenn sie tatsächlich fotografiert, gezeichnet oder gemalt usw. worden wären, ebenfalls fiktionalisiert werden würden; nur sie können zu den ästhetischen Fiktionen gerechnet werden.

Nun ist eine andere, fiktionstheoretisch spannende Frage, ob sie zu Recht fikionalisiert würden, insofern hier von einer Verschränkung auktorialer und rezeptiver Vorstellungen nur insofern die Rede sein kann, als man einer KI Autor:innenschaft und Vorstellungen zubilligt, was, wie angedeutet, ein fundamentales Missverständnis hinsichtlich des zugehörigen Prozessierens wäre. Als Kompromiss könnte man formulieren, dass es sich hierbei um einen Sonderfall fiktionaler Bilder handelt, nämlich um die Simulation eines fiktionalen Bilds, d. h. um so etwas wie ein virtuell fiktionales Bild.

Ausgefeilte Programme wie etwa Dall-E 2 (benannt nach Salvador Dalí und „WALL E“, einem animierten Sciencefiction-Film der Pixar Animation Studios) simulieren so zwar heute auch auf die Eingabe hin ganz problemlos unterschiedliche Varianten von Fotografien: „Wenn man Objektiv-Angaben als Modifikatoren hinzufügt, erhält man die besonders fotorealistischen Bilder, die typisch für Fotoshootings mit diesen Spezifikationen sind.“⁹⁴ Die zugehörigen KIs wurden mithilfe von Millionen von Internet-Bildern trainiert, um nicht nur an malerische, gezeichnete oder an Computergrafiken auf Rotoscoping- oder Raytracing-Basis erinnernde, sondern auch vollständig fotorealistisch wirkende Bilder erzeugen zu können.⁹⁵ Aufgrund ihres generativen Prinzips sind aber auch diese von Fotografien vom Menschen nicht unterscheidbaren Bilder strukturell eher Gemäldesimulationen als Fotografien, selbst wo sie fotorealistisch wirken: Hier wurde nichts belichtet, es gibt kein „indexikalisches Band“ zum Referenten. Gerade diesen Bildern wohnt ob ihrer manipulativen Potenz eine enorme politische Gefährlichkeit inne, die zugleich den überkommenen Beweischarakter der Fotografie oder des dokumentarischen Films (der schon immer eher eine kulturelle Konvention denn eine Tatsache gewesen ist) in den kommenden Jahrzehnten komplett tilgen wird.

Auffällig ist zudem, dass der gegenwärtig überwiegend noch wenig fotografische „Stil“ des, mit einer Formulierung Nils Pookers, „fluffy glamour glow“ der KI-Bilder, wie er im Netz gerade sehr beliebt ist, ein Stück weit von der Technik selbst bestimmt ist und auch noch auf KI-generierten „Fotografien“ greifbar wird: „Charakteristisch für diesen Stil ist jenes diffuse, aber intensive Licht, das weniger von einzelnen Lichtquellen zu stammen als vielmehr aus dem gesamten Bild herauszustrahlen scheint – ein Effekt, der sich in der KI-Bildgenerierung, die ja im Gegensatz zu Raytracing oder anderen klassischen Verfahren der Computergrafik kein optisches Modell der Lichtführung kennt, noch verstärkt. Denn KI-Bildgeneratoren simulieren Lichtführung als rein visuelles Phänomen: Sie sind, anders als etwa Renderingprogramme, unfähig zur exakten Berechnung von Lichtstrahlen und ihren Effekten, aber sehr effizient darin, Licht und Schatten visuell stimmig in der Bildfläche zu verteilen. Darin ähnelt die Bildgenerierung, selbst bei scheinbar ‚fotorealistisch‘ anmutenden Bildern, weit mehr

der Malerei als der Fotografie oder anderen optischen Medien [wie dem Film, J. R.]. Tatsächlich ist das vielleicht eine der Pointen jener *Ästhetik des Schimmerns*, die sich mit Midjourney und dem neuen Dall-E 2 durchzusetzen scheint: Der vermeintliche fotografische Realismus, auf den viele dieser Bilder zielen, ist primär ein malerischer Effekt.“⁹⁶

Füttert man eine solche Maschine mit eher abstrakten Texten wie Lieblingssong-Titeln, Zitaten aus der Literatur oder auch von Gilles Deleuze, werden fiktionale Bilder produziert, von denen oftmals zumindest eines, entsprechend malerisch, skulptural oder installativ umgesetzt, von künstlerischen Arbeiten nur schwer zu unterscheiden wäre (der Mensch imitierte in diesem Fall mit Kant „mechanisch“ die Imaginationssimulation einer Maschine).

Zumindest die besser trainierten KIs mit deutlich höherer Rechnerkapazität, teilweise aber auch bereits die Gratistools im Netz ermöglichen es also, (kürzere) textliche Schilderungen in durchaus überraschende „fiktionale“ Bilder umzuwandeln, die anzeigen, dass die KI zumindest künftig zur veritablen „Einbildnerin“ im Sinne Vilém Flussers taugen könnte, latent entgegen den diesbezüglichen Hoffnungen des kosmopoliten Medienphilosophen auf das kreative Potential des Menschen in der telematischen Gesellschaft. Den Künstler:innen steht nun frei, das von der KI als Ideenlieferantin entworfene Bild mit dem Medium ihrer Wahl zu realisieren; als Gemälde, als Plastik oder als Installation, als Animationsbild usw. und darin, neben der Möglichkeit, digitale Kunstwerke programmieren zu lassen, liegt der eigentliche kreative Nutzen dieser Tools.

Durchgängig fällt auf, dass die KI sich auch bei der Bildgenerierung mit „Phantastischem“ leichter tut als mit Realistischem, dass also die vermeintlich exklusiv-menschliche Kreativität sich offenbar *in extremo* leichter simulieren lässt als das mechanische Abbilden eines „Stücks Realität“, wenn das Reservoir an Bildern, mit denen die KI zuvor trainiert wurde, nur groß genug ist.⁹⁷ Dennoch scheint auch die Zukunft der Stockfotografie über kurz oder lang den KI-generierten Bildern zu gehören, weil das mühevollen Suchen nach dem richtigen Bild entfällt.

Offene kulturtheoretische Fragen

Was die für eine Fiktionalität – namentlich den konstitutiven Dialog auktorialer und rezeptiver Vorstellungen – entscheidende Frage der Autor:innenschaft anbelangt, ergibt sich theoretisch das Problem, wie man mit der Figur maschineller Autor:innenschaft bzw. Subjektivität umgehen will. Hier entsteht bspw. ein juristisches Problem, das anzeigt, dass man sich diesbezüglich nicht auf einen „metaphorischen Gebrauch“ o. Ä.

herausreden kann, sondern theoretisch wie praktisch fundamentale Entscheidungen treffen muss, wie man die Autor:innenschaft, stets ein Gewordenes, weiterdenken möchte. Hierzu sind drei dezidierte Positionen denkbar, von denen mir die erste bei Anwendungen schwacher KI aus den genannten Gründen nicht haltbar erscheint, diese würde nämlich a) generativer KI Subjektivität und damit eine mögliche Autor:innenschaft vollumfänglich zubilligen. Man kann die von einer solchen Anwendung generierte Bilder oder Texte auch b) als prinzipiell subjektlose und damit autorlose Hervorbringungen auffassen und damit als entscheidender Schritt hin zu einer endgültigen Überwindung der überkommenen Autor:innenfunktion im beschriebenen poststrukturalistischen Sinn. Oder aber c) man hält an ihr fest und verlegt sie etwa, je nach theoretischem Modell und Anwendungsfeld, auf die Programmierer:innen hinter der KI oder sagt schlicht, dass menschliche Autor:innenschaft heute eben im jeweiligen Feld generative KI als ein Tool unter anderen nutzt (Autor:in ist, wer den Prompt formuliert). Das muss dann aber mit geringen Modifikationen auch eine maschinelle Autor:innenschaft denken lassen, sofern in einem markierten Sonderbereich, wo generative KIs autonom agieren – womöglich künftig gerade beim *prompting*⁹⁸ – die Maschine dann doch als so etwas wie eine Quasi-Autorin fungiert. Auch diverse differenziertere und ggf. vorläufige Positionen sind möglich – und in meinen Augen notwendig.

Ein ähnliches Theorieproblem ergibt sich hinsichtlich der Frage der (Un-)Möglichkeit KI-generierter Kunst, aber auch bereits hinsichtlich des kommunikativen Status der zugehörigen Hervorbringungen, der ganz unklar ist – hier etwa mit Luhmanns Begriff der Kommunikation angedeutet: Während auch diese Bilder und Texte offensichtlich etwas mitteilen und rezeptiv eine Information auslösen, ist zweifelhaft, ob sie das notwendige dritte Kriterium des „Verstehens“ im systemtheoretisch-abstrakten Sinn voll erfüllen, denn eine ggf. Anschlusskommunikation als Feedback erreicht die generative KI nicht; nicht nur ist sie nicht in der Lage, mit Kritik etc. umzugehen, sondern sie ist im Moment des Feedbacks buchstäblich „eine andere“ als in dem ihrer Mitteilung, sodass hier keine rekursive Kommunikationskette analog zu der von Autor:in und Publikum aufkommen kann.

Weiter problematisch ist, dass z. B. der Chat mit der KI zwar ein Gespräch, eine dialogische Situation, simuliert, aber selbstverständlich kein Fall eines kommunikativen Austauschs ist, weil die KI als „Gesprächspartnerin“ über kein Bewusstsein verfügt, aber auch und gerade, weil von der menschlichen Seite her diverse Fehlprojektionen unvermeidlich sind. So ist z. B. die Grundsituation des Sozialen und damit auch jeder Kommunikation die einer doppelten Kontingenz, die zugleich das Wissen um die doppelte Kontingenz auf beiden Seiten impliziert (was hier nicht der Fall ist, die KI

verfügt über kein Wissen über und interessiert sich auch nicht für ihr Gegenüber). Auf affektiver Ebene setzt dies voraus, dass die Kommunikationspartner:innen die affektive Grunddisposition des Sozialen – Empathie und Scham – grundsätzlich teilen. Ähnliches gilt für gewisse soziale Befürchtungen, etwa der, nicht für voll genommen zu werden oder überhaupt den für Menschen so fundamentalen Wunsch nach Anerkennung.

Anders die KI: Ohne Empathie keine Scham und damit keine soziale Affektivität. Eine KI (oder ein Mensch mit einem entsprechenden hirnorganischen Defekt, der *keinerlei* Empathiefähigkeit aufwiese), ist nicht in der Lage, so etwas wie Scham zu empfinden, und wer diese Möglichkeit nicht kennt, frei von Schamangst ist, steht vollständig jenseits dessen, wie Menschen soziale Interaktion erleben.

Anders gesagt: Wir können gar nicht anders als in das, was mit uns spricht, bereits auf vorbewusster Ebene dieselben sozialen Affektierungen hineinzudenken, die uns leiten. Das aber ist im Chat mit der KI eine vollkommen unangemessene Rahmung mit ggf. erheblichen Konsequenzen.

Selbstverständlich ist eine generative KI wie ChatGPT im Zuge des *affective computing* mit einem Selbstbewertungsmechanismus und gewissen „Schamgrenzen“ insbesondere hinsichtlich gängiger identitätspolitischer Empfindlichkeiten ausgestattet, auf welche die US-Programmierer:innen besonders stark abgehoben haben, dennoch wäre es ein Trugschluss, dies mit dem Prozessieren des menschlichen Schamapparats zu identifizieren; sowenig die Maschine denkt, sowenig schämt sie sich oder besitzt auch nur die Voraussetzung dafür, die Fähigkeit der Empathie – auch hier handelt es sich bestenfalls um den Fall eines Simulakrums dieser Grundvoraussetzungen von menschlicher Interaktion und Kommunikation.⁹⁹

Eine weitere vorbewusste Fehlperspektivierung auf der menschlichen Seite, von der oben bereits die Rede war, spielt im Bereich der Bilder mutmaßlich eine andere Rolle als in dem des Texts: Völlig unberechtigt, aber als kognitive Rahmung schwer zu eliminieren ist der bereits angesprochene *automation bias*, d. h. das Vorurteil, dem folgend Menschen Maschinen grundsätzlich eine Objektivität zusprechen, die sie anderen Menschen nicht ohne Weiteres zubilligen. Dies greift in Bezug auf Texte (bzw. Antworten auf Sachfragen) voll, wohingegen ein ähnlicher Effekt bei fotorealistischen Bildern entfällt: Wir wissen alle, dass digitale Fotografien nicht „die Realität“ zeigen brauchen, sondern vielfach manipuliert sein können und dies in der Regel auch sind. Das gilt selbstverständlich für gänzlich von Computerprogrammen wie Midjourney generierte Bilder noch viel mehr. Das Problem hier wie da ist, dass ein solches Bild, etwa von der vermeintlichen Inhaftierung Trumps Anfang 2023, in bestimmten Filterblasen für ein Wirklichkeitsbild verkauft, hier durchaus als Beweis gelesen werden wird, ähnlich wie zuvor auch all die vermeintlichen Beweise der „gestohlenen Wahl“

geglaubt wurden und werden. Die Tatsache, dass es von einer KI stammt, wird dabei gerade unterschlagen.

Die Frage, ob es sich bei diesen Schöpfungen subjektiv um Fake News oder aber Tatsachenbilder handelt, ist zu einer technischen Frage der Simulationsqualität der generativen KI geworden, und zu einer des Wissens der Rezipient:innen solcher Bilder um deren aktuelle Simulationskapazität, pragmatisch also eine des Vertrauens in die vermittelnden Instanzen (*CNN*, *Fox News*, Freunde und Bekannte, die eigene Filterblase in den sozialen Medien). Das ist eine beunruhigende Perspektive.

Beispiele – KI und Kunst

Mit Blick auf die künftige Rolle von KI im kreativen Bereich ist wichtig zu wiederholen, was eingangs über die unterschiedlichen sozialen Felder einer Implementation, vom Militär über die Verwaltung bis hin zum Gesundheitswesen gesagt wurde: Weder sind die Möglichkeiten, die Grenzen oder die Gefahren einer Verwendung generativer KI für das Feld des Roman- oder Drehbuchschreibens, beim Komponieren, in der Architektur, im Grafikdesign (z. B. *Firefly* in Photoshop oder Programme wie *Lensa* für Porträts), in der Videoproduktion (KI-erzeugte Pre- und Post-Production) usw. durchweg dieselben noch sind sie überhaupt vergleichbar; für jeden dieser Bereiche muss diesbezüglich vielmehr eine eigene Einschätzung getroffen werden.

Im Feld der zeitgenössischen Kunst und besonders im medienexpansiven Segment zwischen (später) Medienkunst, net.art und Post-Internet-Art werden zu den vielen klassischen Arbeiten, die sich thematisch und gegebenenfalls auch kritisch mit dem Phänomen KI und seinen Auswirkungen auf die Gesellschaft beschäftigen, recht bekannt sind etwa Kate Crawfords und Vladan Jolers Infografiken „Anatomy of an AI System“ (2018) oder Trevor Paglens Blick auf die Trainingsdaten und die Biases generativer KI in „Behold these glorious Times“ (2017), erwartbar noch zahlreiche Werke hinzukommen.¹⁰⁰

Die in Bezug auf die ersten Gehversuche generativer KI im Kunstfeld am häufigsten diskutierte Frage lautet, etwas flapsig formuliert, aber oft durchaus besorgt vorgetragen: „Kann KI Kunst?“ Obwohl diese Frage offensichtlich naiv ist, möchte ich sie hier wenigstens kurz streifen.¹⁰¹ Es wird bei der genannten Frage selbstverständlich nicht um lediglich technisch avanciertere Bilder in der Tradition Harold Cohens gehen, der bereits 1973 mit seinem Programm AARON maschinell gefertigte Zeichnungen herstellte, die indes niemand ernsthaft als Kunstwerke aufgefasst hat. Denn der Effekt

des Staunens, dass eine Maschine „so etwas“ hervorbringen kann, ist spätestens mit der Popularität von Midjourney, Stable Diffusion, Dall-E und Co. weitgehend passé. Dass sich generative KI recht offensichtlich mit Kunst schwertut, emphatisch und nicht im vagen Alltagssinn, der Tätowierer:innen oder die Hersteller:innen von Konzeptskizzen für Animationen u. v. a. m. „artist“ nennt, liegt zunächst einmal an der Spezifik dieser Erfindung (im modernen Sinne) durch das ästhetische Regime der Klassiker des Felds aus dem ausgehenden 18. und 19. Jahrhundert, von Baumgarten bis Nietzsche, ist Kunst *per definitionem* an den menschlichen Geist gekoppelt. Hegel etwa deklariert einigermaßen kategorisch: „Kunstwerk ist es nur, insofern es, aus dem Geiste entsprungen, nun auch dem Boden des Geistes angehört, die Taufe des Geistigen erhalten hat und nur dasjenige darstellt, was nach dem Anklang des Geistigen gebildet ist.“¹⁰² Und an dieser Zuschreibung wurde, vielfach säkularisiert und von diversen antiidealistischen Revolten der Avantgarden ungebrochen, im Kern festgehalten. Bei Adorno heißt es: „Wodurch die Kunstwerke, indem sie Erscheinung werden, mehr sind als sie sind, das ist ihr Geist. [...] Was in den Kunstwerken erscheint, nicht abzuheben von der Erscheinung, aber auch nicht mit ihr identisch [...], ist ihr Geist.“¹⁰³ Und noch Christoph Menke schreibt, für die heutige ästhetische Theorie: „Die Kunst existiert [...] in Werken, weil ihre Tätigkeit in sich negativ oder allgemein ist; negativ, weil sie die Bestimmungsmacht des Natürlichen, bloß Individuellen bricht; allgemein, weil ihre Tätigkeit für andere da ist und zu gelten beansprucht. [...]. Die Kunst ist werkhaf – das hat nicht den äußerlichen Sinn, dass ihr eine Tätigkeit der Herstellung vorhergegangen ist (die auch lediglich in einem Akt des Hierher- oder Ausstellens bestehen kann). Die Kunst ist vielmehr werkhaf, weil sie die Wirklichkeit einer normativen Tätigkeit ist, weil sie Geist ist.“¹⁰⁴

Es ist nach dem bisher Skizzierten leicht zu sehen, dass die Schöpfungen bewusstseinsloser (schwacher) KI, von Algorithmen hervorgebracht, sozusagen *grundsätzlich geistfrei* sind. Diese Kompositionen, Bilder und Texte, die insofern keine Werke sind, aber dennoch wirken können, genügen lediglich dem zweiten notwendigen von Menke angeführten Kriterium einigermaßen, nämlich dass sich kunstbezogene Kommunikationsakte durchaus auf sie richten können; mit einer treffenden Formulierung Albrecht Wellmers gesagt: dass auch sie „in einem öffentlichen Raum zwischen ästhetisch erfahrenden Subjekten lokalisiert [...] [sein können, J. R.], ein ‚Objekt‘, worauf diese Subjekte im ästhetischen Diskurs sich beziehen und gleichsam zurückkommen können“.¹⁰⁵ Auf das konkrete Machen dieser besonderen Objekte bezogen, genügen sie mit Paul Valéry damit zwar dem Kriterium der *Unwahrscheinlichkeit* eines Kunstwerks – niemand kann vorhersagen, was die Black Box der generativen KI an Formen, Wortgebilden, Tönen etc. zusammenkombinieren

wird –, aber nicht dem zweiten Kriterium seiner *Unverhältnismäßigkeit*; genauer kehrt sich dieses Kriterium hier um: Während zum Beispiel ein großes Gedicht insofern schon unverhältnismäßig ist, als in einer Minute rezipiert werden kann, woran der Autor über Jahrzehnte hinweg gearbeitet hat, besteht bspw. bei einem Roman, den eine generative KI verfasst hat, umgekehrt das Problem, dass sie um zu fertigen, wofür menschliche Leser:innen viele Stunden brauchen werden, nur einige Sekunden gebraucht hat.¹⁰⁶ Nicht nur fällt es insofern schwer, eine KI für ihre „Meisterschaft“ zu preisen (die sie, ganz jenseits der Kriterien des fälschlicherweise Oscar Wilde zugeschriebenen Bonmots „Talent borrows. Genius steals“, wenn überhaupt aus der Meisterschaft zahlloser Menschen zusammenkombiniert), und sie verfügt ja an sich auch weder über „Talent“ noch über „Genialität“, ja noch nicht einmal über eine „künstlerische Intention“. All dies sind für ihr Prozessieren gänzlich sachfremde Kriterien – wie damit eben auch das der Kunst? Jedenfalls versteht die KI noch nicht einmal ihre diesbezügliche Produktion, könnte sie aber jederzeit interpretieren, ohne freilich wiederum diese Interpretation zu verstehen oder als die ihre zu behaupten ... Was auch immer eine solche Produktion vorstellt, Kunst im überkommenen Sinne einer besonderen geistigen Schöpfung in sinnlich wahrnehmbarer Gestalt ist sie nicht.

Auch ist hier ganz unmöglich, das Machen des „Werkes“ zu kennen oder nachzuvollziehen (und so kann sich auch die für Kunst, wiederum mit Menke, konstitutive Paradoxie nicht entfalten, wonach man das Werk als Interpret natürlich nie ganz kennen kann).¹⁰⁷ Eine Interpretation der Hervorbringungen der KI ist überhaupt müßig, etwa sich darüber Gedanken zu machen, warum hier jenes Wort steht und da dieses, oder hier jener Farbton und dort ein anderer, weil sich *niemand* etwas dabei gedacht hat.

Ganz pragmatisch und ohne derartige Erwägungen kann sich jedermann im Selbstversuch leicht der Tatsache vergewissern, dass beim derzeitigen Stand der Entwicklung die Schöpfungen der generativen KI kaum 1:1 als Kunstwerke im weitesten Sinne deklariert werden können, und das liegt zunächst einmal an ganz offensichtlichen qualitativen Mängeln. (Vom ggf. offen als medienkünstlerisches Experiment deklarierten einmaligen Akt der Transgression hinein ins Feld der Nicht-Kunst einmal abgesehen: Bekanntlich kann seit Duchamps Readymades prinzipiell jedes Ding der Welt exemplarisch ein Fall von Kunst sein, um deren Begriff zu hinterfragen, ohne dass deshalb die jeweilige Gattung dieser Dinge dadurch zur potentiellen Kunst würde – weshalb also nicht auch etwas, das eine KI hervorgebracht hat?)¹⁰⁸

Weder der „surrealistische Alptraum mit tanzenden Mäusen und einer toten Jungfrau“ von Dall-E noch das Sonett im Stile Shakespeares, das Drehbuch im Stile Ingmar Bergmans oder der Romananfang in dem Thomas Manns von ChatGPT können rein qualitativ, also inhaltlich wie ästhetisch, ernsthaft mit Kunstwerken, literarischen Werken oder professionellen Drehbüchern verwechselt werden (was nicht heißt, dass die KI Tools hier nicht ausgesprochen nützliche Produktionswerkzeuge darstellen).¹⁰⁹ Allein im Bereich der Tonkunst scheint dies insofern anders zu sein, als zumindest das Niveau eher mediokrer Komponist:innen, etwa irgendwelcher Bach-Epigonen des 17. Jahrhunderts, durchaus erreichbar zu sein scheint.

Völlig anders sieht es aber aus, wenn menschliche Künstler:innen generative KI als Werkzeug nutzen. Alle eben gemachten Beobachtungen und Einwände fallen in sich zusammen und es wäre pedantisch bis lächerlich, die Nutzung generativer KI für diesen Zweck unterbinden zu wollen, so als ob man einem modernen Maler wie Frank Stella Zirkel und Lineal untersagt hätte oder dem frühen Gerhard Richter die Nutzung von Fotografien als Vorlage seiner Gemälde oder der Comiczeichnerin Alison Bechdel ihre Verwendung derselben für ihr autobiografisches Werk „Fun Home: A Family Tragicomic“ (2006).¹¹⁰ Und man braucht kein Prophet zu sein, um zu sehen, dass dieses aktuell gerade neue Hilfsmittel für zahlreiche Werke der Zukunft genutzt werden wird, was wiederum die poetische Frage stellt, ob es Kriterien einer sinnvollen Nutzung geben kann. Einige wenige, willkürlich ausgewählte Beispiele KI-basierten Kunstschaffens mögen diesen Punkt und seine Virulenz unterstreichen; viel genannt werden etwa Anna Ridler, Helena Sarin, Mario Klingemann, Sofia Crespo, Mary Flanagan, die Malerinnen Sougwen Chung und Cornelia Es Said, OBVIOUS, Supercomposite (Steph Maj Swanson), Robbie Barrat, Scott Eaton und Refik Anadol.¹¹¹ Für den einfachsten Fall – den, die KI schlicht als Ideengeberin zu nutzen –, kann der in Polen geborene und in Berlin arbeitende Maler Roman Lipski als Beispiel stehen.¹¹² Er nutzt seit einigen Jahren eine vom Programmierer Florian Dohmann eigens für ihn hergestellte KI mit dem Namen A.I.R. als „digitale Muse“.¹¹³ Dohmann beschreibt ihre Funktionsweise folgendermaßen: „Im Prinzip ist es ein künstlich maschinell lernendes System, das in der Lage ist, Lipskis Bilder zu analysieren, zu dekonstruieren, neu zusammensetzen und am Ende ganz Neuartiges zu generieren. Das sind im Wesentlichen erstmal digitale Dateien, die Roman Lipski derzeit auf einem Interface zu sehen bekommt, die wir aber zwischenzeitlich schon in verschiedenen Größen und Formaten gedruckt haben.“¹¹⁴ Wenn Lipski der KI beispielsweise Bilder kalifornischer Landschaften vorlegt, skizziert sie für ihn Ansätze, wie man diese Region anders, d. h. im abstrakten Stil Lipskis, darstellen könnte. Der Künstler nutzt die Vorschläge der

Software als Inspiration, um auf deren Basis seine finalen Interpretationen zu entwickeln. Er sagt: „Der Vorteil dieser Systeme ist die Schnelligkeit und die Fähigkeit, eine hohe Anzahl an Kombinationen zu generieren. Dadurch bin ich schneller an mein Ziel gekommen, abstrakt zu malen. [...] Die digitale Muse hilft mir dabei, mein eigenes Potenzial zu steigern und – entgegen ihrem Hintergrund der beschleunigten Digitalisierung – zu entschleunigen, denn die andauernde Beschäftigung mit dem eigentlichen Ursprungsmotiv hilft mir, die Essenz unterschiedlichster malerischer Ansätze herauszukristallisieren. Ich habe mich befreit von erzählerischen und visuellen Symbolen und Metaphern. Das ist aus meiner Sicht ein großer Schritt ...“¹¹⁵ Im hier diskutierten Zusammenhang wichtig ist Lipskis Überzeugung: „Wahre Kreativität [...] stammt nur vom Menschen. Künstliche Intelligenz kann das Humane ‚nur‘ unterstützen und nicht ersetzen.“¹¹⁶ Die aktuelle Entwicklung ist bezüglich dieser These so etwas wie ein Großexperiment mit ungewissem Ausgang.

Ein anderer Weg, mit KI zu arbeiten, besteht, darin, die Schwächen der KI in eine Stärke umzuwandeln; den Apparat also mit Flusser gewissermaßen gegen sich selbst zu wenden. Die Merz Akademie-Absolvent:innen Edna Syed Gulam und Oliver Weiss haben so im Wintersemester 2021/22 als Bachelorprojekt unter dem Titel DAYS IN A LIFE einen experimentellen Kurzfilm aus dem Genre des Autorenfilms gedreht, wobei das Drehbuch weitestgehend aus Textversuchen einer eigens trainierten KI stammt. Als vielversprechend hatte sich der Ansatz erwiesen, die KI zunächst einen riesigen Pool von Geschichten schreiben zu lassen und nun Fragmente dieser ganz unterschiedlichen Stories nach dem Collageprinzip miteinander kollidieren zu lassen. Gerade wenn die Zuschauer:innen um diese Genese nicht wissen, ist schlussendlich der Reichtum des mäandernden Gedankenstroms der namenlosen Protagonistin (Cora Kneisz) frappierend – auch der an (vermeintlichen) Referenzen auf bedeutende filmische und literarische Werke. Es besticht zudem der, paradoxerweise, ausgesprochen authentisch wirkende Charakter der intimen Innenschau, die sich hier über die Interaktion von KI und menschlichen Drehbuchautor:innen entfaltet, ein Effekt, der umso verstörender wird, wenn man weiß, dass hier buchstäblich nichts persönlich, intim oder authentisch sein kann.¹¹⁷

Höchst produktiv für die Absicht der beiden Filmemacher:innen ist auch der für diese Technik typische Heilige Ernst der KI, personifiziert in der Stimme des Erzählers, der einigermaßen penetrant die Filmbilder kommentiert; die Sinnlosigkeit dieser Dopplung erinnert an das absurde Theater der klassischen Moderne.

Ein kleiner Auszug aus dem Drehbuch mag hier das Gesagte verdeutlichen:

4. Szene: INTERIEUR KÜCHE - TAG

[..]

ERZÄHLER (Voiceover)

.....

„Sie geht zurück auf den Teppich, ihr kleines Reich. Sie fühlt etwas, ist sich aber nicht sicher, was es ist. Verloren, leer, gelangweilt?

+Wir hören eine surrende FLIEGE, begleitet von dem Schmatzen DES MENSCHEN beim Essen von Toast. Sie hört auf zu schmatzen. Nichts. Sie macht weiter: das summende Geräusch kehrt zurück. Das geht eine Weile so weiter.+

.....

ERZÄHLER

.....

„Aus einem anderen Zimmer ist das Geräusch eines Staubsaugers zu hören, aber sie steht nicht auf, um das Fenster zu schließen. Stattdessen beobachtet sie eine Fliege, die durch das Fenster hereinkommt.

+Sie schaut sich um, seufzt und steht auf.+

.....

ERZÄHLER

....

„Sie verfolgt sie mit einer Zeitschrift, gibt aber sofort auf, als sie bemerkt, was sie tut.“

+DER MENSCH geht aus dem Bild, mit der Zeitschrift in der Hand.+

5. Szene: INT. WOHNZIMMER - TAG

+DER MENSCH liegt auf dem Teppich im Wohnzimmer. In der Hand hält sie eine Fernbedienung für eine Stereoanlage. Aus der Stereoanlage dröhnt ein elektrisches Summen.+

.....

ERZÄHLER

.....

„Sie trifft selten jemanden, den sie kennt. Sie sieht sie durch einen Film, von dem sie weiß, dass er da ist, den sie aber nicht durchdringen kann.

6. Szene: INT. BAD - TAG

+Sie schaut sich selbst im Spiegel an.+

.....

ERZÄHLER

„Sie wurde zu der Art von Frau, die, wenn sie an den Schaufenstern vorbeikommt, stehen bleibt und sich in der Spiegelung betrachtet.“

+DER MENSCH dreht den Wasserhahn auf. Sie beginnt, sich das Gesicht zu waschen, starrt dann aber nur in den Spiegel. Sie fängt an, mit ihrem Spiegelbild zu reden.+

DER MENSCH

„Ich bin mir nicht sicher, ob ich das bin.“

DER MENSCH (SPIEGELBILD)

„Was meinst du?“

DER MENSCH

.....
„Ich weiß es nicht! Ich habe nur das Gefühl, dass es da draußen noch jemanden gibt, der mir ähnlichsieht.“

.....

ERZÄHLER

„Sie wurde zu der Art Frau, die in den Friseursalon geht und wartet, bis sie an der Reihe ist.“

+Sie blättert in Zeitschriften und schneidet Bilder von Hüten, Flaschen, Taschen und Schuhen aus. Sie steht auf, putzt sich die Zähne, betrachtet sich im Spiegel und reibt ihr Gesicht mit einem Reinigungstuch.+

.....

ERZÄHLER

„Sie geht in den Flur, angezogen von dem warmen Licht, das von der Veranda kommt und in den Gang geschickt wird.“

+DER MENSCH läuft aus dem Bild.+

7. Szene: INT. WOHNZIMMER - NACHMITTAG

+DER MENSCH sitzt auf dem Tisch mit einem Stift in der rechten Hand und einem großen Blatt Papier, welches sie von einem Notizblock abgerissen hat, in der linken.+

ERZÄHLER

„In der rechten Hand hältst du einen Stift, in der linken ein großes Blatt Papier, das du von einem Notizblock abgerissen hast. Du wartest darauf, dass etwas kommt, dass etwas anfängt. Du wartest auf ein Zeichen, auf eine Idee, auf einen Ausweg, auf Mut, auf jemanden, der dir sagt, was du tun sollst. Du hast keine Erinnerungen. Du weißt nicht einmal mehr, wie du hierhergekommen bist, warum du jetzt ein Kartenspiel in die Hand genommen hast, das du nacheinander vor dir ausbreitest. Du weißt nicht, wie man mit Geduld spielt, das hast du nie gewusst. Du weißt nicht,

wer du bist und wohin du gehst, aber die Sonne scheint dir
in die Augen.“

*+DER MENSCH sitzt auf dem Bett mit einem Kartenspiel vor
sich. Er spielt mit sich selbst.+*

Für den Ansatz der künstlerischen Forschung im Feld kann das großangelegte Projekt „AI isn't Artificial but Human“ von Anna Ridler und Caroline Sindors stehen, die 2020 die Residency des von der Ars Electronica initiierten *European ARTificial Intelligence Lab (AI Lab)* in Edinburgh und Linz gewonnen haben. Sie versuchen hier, um KI zu demystifizieren, den menschlichen Einfluss auf generative KI – von der Auswahl der Trainingsdaten, den zugehörigen Biases bis zur Rolle der *click worker* – aufzuarbeiten.¹¹⁸ Unter dem programmatischen Motto „entanglements – [a] fair, moral and transparent AI“ wurden hier diverse zentrale Fragen der Entwicklung generativer KI aufgegriffen und in zwei Kunstwerke übersetzt, welche simulative Naturerfahrungen während des Lockdowns 2020 ermöglichen sollten; die beiden Arbeiten, „mechanized cacophonies and cypress trees“ mit KI-generierten Soundscapes, wurden dann auf der Ars Electronica 2021 präsentiert.¹¹⁹

Vielleicht der größte Publikumserfolg unter den kontemporären KI-generierten Kunstwerken ist Refik Anadol's „Machine Hallucinations“ (2022) im MoMa. Er nutzt die Metadaten zu den über 138.000 Werken des Museumsbestands, um eine eigens konstruierte KI aus diesen Beschreibungen neue Bilder generieren zu lassen. Diese werden dann in dem „Echtzeit-Software-Kunstwerk“ auf verschiedenen Displays in permutierende abstrakte Formwelten umgerechnet, die in der morpho-(un-)logischen Prozesshaftigkeit ihres Flows entfernt an die Arbeiten der psychedelischen Kunst der 1960er Jahre erinnern.¹²⁰ Die Rhetorik Anadols von der „Imagination der Maschine“ oder von ihren vermeintlichen „Träumen“ ist zwar reines Marketing und, wie er selbst weiß, haarsträubend unpräzise; auch der Fachbegriff der „Halluzination“ wird hier ganz gegen seinen KI-bezogenen Sinn im Alltagsverständnis verwendet, und der von ihm für seine Arbeit in Anspruch genommene Bezug auf „kollektive visuelle Erinnerungen“ ist nur insofern ein solcher, als kollektive Erinnerungen in Form von Tags und Metadaten – also englischsprachigem Text – zu einem Bestand eines Museums mit allen Facetten dezidiert Moderner Kunst überhaupt repräsentierbar sind, von welchen die Formberechnungen der KI Refiks ja ausgehen. Aber unstrittig entsteht hier ein KI-generierter Fluss nie gesehener Bilder, die von berückender Schönheit sind und trotz ihrer algorithmischen Genese ausgesprochen real und natürlich wirken.

Das sind nur vier willkürlich ob ihrer Unterschiedlichkeit gewählte Beispiele, und im Moment ist es sicher verfrüht, etwas Abschließendes über das Verhältnis von KI und Kunst sagen zu wollen, weil, mit Luhmann gesagt, bei der diesbezüglichen Evolution des Kunstsystems gerade erst die Phase einer *Selektion* eingesetzt hat, welche innersystemisch der *Variation* der zugehörigen Kommunikation (durch KI-generierte „Kunst“) folgt; Motor dieses kommunikativen Aushandlungsprozesses ist immer die nun inflationär gestellte Frage, ob derlei Kunst sein kann.

Es ist bei derart multikomplexen Kommunikationsprozessen nicht möglich, auf das Ergebnis, die *(Re-)Stabilisierung* zu schließen.

Dass es sich tatsächlich um einen offenen Prozess handelt, lässt sich am besten vielleicht am Beispiel des in Paris lebenden und arbeitenden „KI-Künstlers“ ALAgrApHY verdeutlichen: Er experimentiert bereits seit 2016 transmedial mit den unterschiedlichsten Möglichkeiten der bildgenerierenden KI, die sich häufig der Tradition des Porträts – hier als gewissermaßen „Nicht-Porträt“, d. h. als Bild von nicht-existenten Personen – bedienen oder Fragen der Diversität und des notorischen Bias der KI thematisieren.¹²¹ Seine künstlerische Forschung erstreckt sich dabei auf Zeichenmaschinen, aber auch auf die Nutzung von KI im Modedesign oder im Filmbereich.¹²² Gerade die Popularität dieser mit einigen Preisen ausgezeichneten Arbeiten im Netz führte aber dazu, dass seiner Arbeit allzu rasch das Label einer „KI-Kunst“ aufgesteckt wurde, das zunächst einmal, wie einst bei der diffamierenden Etikettierung „Frauenkunst“, immer auch ausdrückt, dass es sich dabei irgendwie nicht um „richtige Kunst“ handelt.¹²³

Einer solchen vom Reiz des Neuen bestimmten Momentaufnahme steht eine Perspektive gegenüber, der gemäß generative KI künftig ein völlig gängiges Hilfsmittel der Kunstgenerierung unter vielen anderen sein könnte, wie eben eine Rakel in der abstrakten Malerei spätestens seit den 1950er Jahren oder der eigene Körper in der Kunst spätestens seit den 1960er Jahren ganz typisch geworden ist, sodass derlei, anfangs skandalisiert, schon bald keiner besonderen Erwähnung mehr bedurfte. Ähnlich dürfte es schon bald in Bezug auf die Nutzung generativer KI zu Kunstzwecken sein. Zu vermuten ist an diesem Punkt: Gerade die erwartbare Alltäglichkeit der KI in allen Lebensbereichen dürfte sich umgekehrt proportional zu ihrer Prominenz im Kunstfeld verhalten, in deutlicher Differenz zu den angewandten Künsten der Massenkultur. Dies würde dann recht exakt dem sich für manche Beobachter:innen bereits abzeichnenden Differenzierungsprozess im Feld der ästhetischen Fiktionen folgen, z. B. fürs Feld der Literatur gesagt: Die aus Verlagssicht kommerziell ohnehin weniger ökonomisch als symbolisch bedeutende Literatur mit Ausrufezeichen stammt weiterhin von menschlicher Hand, mit echten Autor:innenlesungen und womöglich gar einer

Fetischisierung handschriftlicher Manuskripte für betuchte Sammler:innen, der Trash der sogenannten Unterhaltungsliteratur hingegen wird größtenteils kostengünstig von KIs serienproduziert.¹²⁴

Schluss: Ende der ästhetischen Fiktion?

Foucault hat darauf hingewiesen, dass ästhetische Fiktionen aus Sicht der Macht prinzipiell gefährlich sind, „eine große Gefahr [...] für unsere Welt“ (so wie sie ist) darstellen, da mit ihnen eine „krebsartige und gefährliche Vermehrung der Bedeutungen“ einhergehe, was mit potentiell zersetzender oder mit Deleuze und Guattari „deterritorialisierender“ Kraft allzeit den Status quo latent infrage stellen und jedenfalls das Tagesgeschäft der Macht erschweren würde.¹²⁵ Selbstredend plädierte ein Foucault deshalb vehement für diese „Gefahr“.

Er attestierte weiter dem gesellschaftlichen Prinzip der Autor:innenschaft, d. h. dem diskursiven Konstrukt Autor, das nur vermeintlich auf den „genialen Schöpfer des Werks“ reduzier- bzw. zurechenbar ist, eine grundsätzliche Kompliz:innenschaft mit der Macht, nämlich als ein Prinzip, das „die freie Zirkulation, die freie Handhabung, die freie Komposition, Dekomposition und Rekomposition von Fiktion behindert“.¹²⁶ Das Letzte war damals schon eine fragwürdige Idee und ist es, obgleich notorisch wiederholt, bis heute geblieben, und seine abschließende Prognose, wonach „die Autorfunktion verschwinden wird“, hat sich so auch, zumindest bisher, nicht erfüllt – entgegen der Erwartung so vieler in Hinsicht auf die Auswirkungen der digitalen Revolution.¹²⁷ Allerdings sind nun die Produktionen der KI und eine algorithmengesteuerte Memesphere ubiquitär-anonymer kultureller Produktion neben die altbekannten, von der Autor:innenfunktion bestimmten Diskurse getreten, ohne sie zu verdrängen.¹²⁸

Auch in diesem Zusammenhang befinden wir uns gegenwärtig vielleicht an einem Kipppunkt; ein solcher ist mit den Revolutionen generativer KI wohl auch hinsichtlich der (vermeintlich) ewigen Tradition der ästhetischen Fiktionen erreicht. In unserer, durch und durch von ästhetischen Fiktionen geprägten Kultur (ästhetische Fiktionen nehmen hier ungefähr die Rolle ein, die in archaischen Kulturen der Mythos eingenommen hat, auch unser Bild der KI ist, wie gesagt, von einschlagenden Fiktionen fehlgeprägt, was uns immerzu zuerst die ganz verkehrte Frage stellen lässt, ob KI bald „die Macht übernehmen“ wird ...) ist die Figur eines *Endes der Fiktion* zwar nur schwer zu denken, aber durch die rasante Entwicklung, welche gegenwärtig generative KI-Anwendungen in Bezug auf die Simulation von ästhetischen Fiktionen machen, ist dieses Szenario zum ersten Mal in greifbare Nähe gerückt.

Zum jetzigen Zeitpunkt muss als offen eingeschätzt werden, ob das, was an den ästhetischen Fiktionen reizt, eher die dialogisch-sympraktische Verschmelzung auktorialer und rezeptiver Vorstellungen ist, d. h. der kommunikative Aspekt, oder aber der reine Unterhaltungswert, die mit McLuhan immer auch „Massage“ ist, was die Algorithmen mittelfristig sicher besser bedienen werden können. Sollte Letzteres der Fall sein, würde das Eindringen der KI in dieses Feld vor allem die Konsequenz haben, dass der letzte Konnex der ästhetischen Fiktion zur Kunst gekappt wäre. Sollte hingegen die im weitesten Sinne menschliche Autor:innenschaft einen Wert an sich darstellen, der virtuelle Kontakt mit einem identifizierbaren, ggf. gemeinschaftlichen menschlichen Bewusstsein mithin den rezeptiven Reiz erhöht, wäre infolge des Eindringens der KI ins Feld der ästhetischen Fiktion eine Spaltung desselben wahrscheinlich.¹²⁹

Neu dürfte auch, ganz im Jenseits des Felds der ästhetischen Fiktionen, die *Conditio* einer Politik sein, die fortlaufend auf öffentliche Erregungswellen, die auf erfundene Bilder zurückgehen, reagieren muss, die also ständig die Nichtrealität fiktionaler Bildschöpfungen mit Geltungsanspruch auf Tatsächlichkeit beweisen können sollte bzw. mit dem von solchen Bildern in irgendwelchen Filterblasen generierten Hass umgehen muss. Der Grund, auf dem die kommunikativen Aushandlungsprozesse darüber, was real, was faktisch ist, und was Fake oder Verschwörungstheorie, wird durch diese Entwicklungen noch schwankender werden, und die derzeit zu beobachtende soziale Spaltung dürfte sich dank der generativen KI und ihrer Erfindungsmacht gegen die künftig zu erwartende noch eher behaglich ausnehmen.

LITERATUR

Stefan Achleitner: „Generative Adversarial Networks: Wenn künstliche Intelligenz künstliche Daten erschafft“ (2020), online unter <https://intelligente-welt.de/generative-adversarial-networks-wenn-kuenstliche-intelligenz-kuenstliche-daten-erschafft/>.

Theodor W. Adorno: Negative Dialektik. Jargon der Eigentlichkeit [1966], in: ders.: Gesammelte Schriften, hg. Rolf Tiedemann, Bd. 6. Frankfurt a. M.: Suhrkamp 2003.

Theodor W. Adorno: Ästhetische Theorie. Gesammelte Schriften, hg. Rolf Tiedemann, Bd. 7. Frankfurt a. M.: Suhrkamp 2003.

Steffen Albrecht: „ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotentiale und mögliche Auswirkungen“ (2023), online unter <https://www.bundestag.de/resource/blob/944148/30b0896f6e49908155fcd01d77f57922/20-18-109-Hintergrundpapier-data.pdf>, S. 1-112 [= Hintergrundpapier Nr. 26 vom 21. April 2023 des Büros für Technikfolgenabschätzung beim Deutschen Bundestag, Ausschussdrucksache 20(18)109].

Hannes Bajohr: „Die ‚Gestalt‘ der KI. Jenseits von Atomismus und Holismus“, in: Zeitschrift für Medienwissenschaft 23, Heft 2 (2020), S. 168-183.

Alison Bechdel: Fun Home: A Family Tragicomic. Boston: Mariner Books 2006.

Emily M. Bender; Tamnit Gebru; Angelina McMillan-Major; Margaret Mitchell: „On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?“ (2021), in: FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021, online unter <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>.

Carl T. Bergstrom: „Your chatbot is not ‚hallucinating‘. It’s bullshitting“ (2023), online unter <https://undark.org/2023/04/06/chatgpt-isnt-hallucinating-its-bullshitting/>.

Ali Borji: „A Categorical Archive of ChatGPT Failures“ (2023), online unter <https://arxiv.org/pdf/2302.03494.pdf>, S. 1-43.

Ben Buchanan; Andrew Lohn; Micah Musser; Katharina Sedova: Truth, Lies, and Automation. How language models could change disinformation. Hg. vom Center for Security and Emerging Technology. Washington, DC 2021, online unter <https://cset.georgetown.edu/publication/truth-lies-and-automation/>.

Jean Burgess; Aaron J. Snoswell: „The Galactica AI model was trained on scientific knowledge – but it spat out alarmingly plausible nonsense“, in: The Conversation vom 30. 11. 2022, online unter <https://theconversation.com/the-galactica-ai-model-was-trained-on-scientificknowledge-but-it-spat-out-alarmingly-plausible-nonsense-195445>.

Patrick Butlin; Robert Long; Eric Elmoznino; Yoshua Bengio; Jonathan Birch; Axel Constant; George Deane; Stephen M. Fleming; Chris Frith; Xu Ji; Ryota Kanai; Colin Klein; Grace Lindsay; Matthias Michel; Liad Mudrik; Megan A. K. Peters; Eric Schwitzgebel; Jonathan Simon; Rufin VanRullen: „Consciousness in Artificial Intelligence: Insights from the Science of Consciousness“ (2023), online unter <https://doi.org/10.48550/arXiv.2308.08708>.

Stanley Cavell: „Welt durch die Kamera gesehen“ [1974], in: Dieter Henrich; Wolfgang Iser (Hg.): Theorien der Kunst. Frankfurt a. M.: Suhrkamp 1992, S. 447-490.

Eva Cetinic; Jams She: „Understanding in Creating Art with AI: Review and Overlook“ (2021), online unter <https://arxiv.org/pdf/2102.09109.pdf>, S. 1-17.

Kent K. Chang; Mackenzie Cramer; Sandeep Soni; David Bamman: „Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4“ (2023), online unter https://www.researchgate.net/publication/370442472_Speak_Memory_An_Archaeology_of_Books_Known_to_ChatGPTGPT-4, S. 1-15.

Noam Chomsky; Ian Roberts; Jeffrey Watumull: „The False Promise of ChatGPT“, in: New York Times vom 8. 3. 2023, online unter <http://www.nytimes.com/2023/03/08/opinion/noam-chomskychatgpt-ai.html>.

Devin Coldewey: „OpenAI’s Dall-E creates plausible images of literally anything you ask it to“, online unter: <https://techcrunch.com/2021/01/05/openais-dall-e-creates-plausible-images-of-literally-anything-you-ask-it-to>.

Marc Engenhardt; Sebastian Löwe: Design und künstliche Intelligenz. Theoretische und praktische Grundlagen der Gestaltung mit maschinell lernenden Systemen. Basel: Birkhäuser 2022.

Michel Foucault: „Das Denken des Außen“, in: ders.: Schriften in vier Bänden. Dits et Ecrits, Bd. I 1954-1969, Frankfurt a. M.: Suhrkamp 2001, S. 670-697.

Jörg Phil Friedrich: Degenerierte Vernunft. Künstliche Intelligenz und die Natur des Denkens. München: Claudius 2023

Detlev Garz; Uwe Raven: „Lebenspraxis – Krise und Routine“, in: dies. (Hg.): Theorie der Lebenspraxis. Einführung in das Werk Ulrich Oevermanns. Wiesbaden: Springer VS 2015.

Dirk von Gehlen: Meme. Muster digitaler Kommunikation. Berlin: Wagenbach 2020.

Anna Grubauer: „Women in Media Arts: Does AI think like a (white) man?“ (2023), online unter <https://ars.electronica.art/aeblog/en/2020/04/10/women-in-media-arts-ai/>.

Georg Wilhelm Friedrich Hegel: Vorlesungen über die Ästhetik I [=Werke Bd. 13]. Herausgegeben Eva Moldenhauer und Karl Michel. Frankfurt a. M.: Suhrkamp 1997.

Will D. Heaven: „Why Meta’s latest large language model survived only three days online“ (2022), online unter <https://www.technologyreview.com/2022/11/18/1063487/meta-largelanguage-model-ai-only-survived-three-days-gpt-3-science/>.

Melissa Heikkilä, Will D. Heaven: „Was uns 2024 in der Künstlichen Intelligenz erwartet“ (2024), online unter https://www.heise.de/hintergrund/Was-uns-2024-in-der-Kuenstlichen-Intelligenz-erwartet-9590341.html?utm_source=pocket-newtab-de-de.

Alex Hern: „Sci-fi Publisher clarkesworld halts pitches amid deluge of AI-generated series“, in: The Guardian vom 21. 2. 2023, online unter <https://www.theguardian.com/technology/2023/feb/21/sci-fi-publisher-clarkesworld-halts-pitches-amid-deluge-of-ai-generated-stories>.

Aaron Hertzmann: „Computers do not make art, people do“, in: Communications of the ACM 63, 5 (2020), S. 45-48.

Jennifer Jacquet: Scham: Die politische Kraft eines unterschätzten Gefühls. Frankfurt a. M.: S. Fischer 2015.

Andrian Kreye: „Was die KI liest“, in: Süddeutsche Zeitung vom 24./25. 6. 2023, S. 20.

Jakob von Lindern: „Braucht die deutsche Vorzeige-KI mehr Erziehung?“, in: ZEIT Online vom 11. 9. 2023, online unter https://www.zeit.de/digital/2023-09/aleph-alpha-luminous-jonas-andrulis-generative-ki-rassismus?utm_source=pocket-newtab-de-de.

Niklas Luhmann; Karl Eberhard Schorr: „Das Technologiedefizit der Erziehung und die Pädagogik“, in: dies. (Hg.): Zwischen Technologie und Selbstreferenz. Fragen an die Pädagogik. Frankfurt a. M.: Suhrkamp 1996, S. 11-40.

Niklas Luhmann: Die Gesellschaft der Gesellschaft. Frankfurt a. M.: Suhrkamp 1999.

Sebastian Meineck: „Der Anfang von etwas Großem“ (2022), online unter https://netzpolitik.org/2022/text-zu-bild-generierung-der-anfang-von-etwas-groessem/?utm_source=pocket-newtab-global-de-DE.

Christoph Menke: Die Kraft der Kunst Berlin: Suhrkamp 2013.

Dieter Mersch: „Kreativität und künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität“, in: ZfM 21, 2 (2019), S. 65-74.

Roland Meyer: „Es schimmert, es glüht, es funkelt – zur Ästhetik der KI-Bilder“, in: 54Books vom 21. 3. 2023, online unter <https://www.54books.de/es-schimmert-es-glueht-es-funkelt-zur-aesthetik-der-ki-bilder/>.

Albrecht L. Merz: Manifest d. Erziehung. Stuttgart: o. V. 1947.

Merzmann (Vladimir Alexeev): KI-Kunst. Kollaboration von Mensch und Maschine. Berlin: Klaus Wagenbach 2023.

ders. (Vladimir Alexeev): „Kann Dall-E fotorealistische Bilder erzeugen?“ (2022), online unter: <https://the-decoder.de/kann-dall-e-2-fotorealistische-bilder-erzeugen/>.

Cade Metz: „The Godfather of A.I.’ Leaves Google and Warns of Danger Ahead“, in: New York Times vom 1. 5. 2023, online unter <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>.

Arthur I. Miller: The Artist in the Machine. The World of AI-powered Creativity Cambridge, Mass. MIT Press 2019.

Hans Moravec: Computer übernehmen die Macht. Vom Siegeszug der künstlichen Intelligenz. München: Hoffmann & Campe 1999.

Ulrich Oevermann: „Biographie, Krisenbewältigung und Bewährung“, in: Sylke Bartmann; Axel Fehlhaber; Sandra Kirsch (Hg. u. a.): „Natürlich stört das Leben ständig.“ Perspektiven auf Entwicklung und Erziehung. Wiesbaden: VS Verlag 2009, S. 35-55.

Ulrich Oevermann: „Krise und Routine‘ als analytisches Paradigma in den Sozialwissenschaften“ [=Abschiedsvorlesung am 28. 4. 2008, gehalten an der Goethe Universität Frankfurt a. M.], in: Roland Becker-Lenz; Andreas Franzmann; Matthias Jung (Hg. u. a.): Die Methodenschule der Objektiven Hermeneutik. Wiesbaden: VS Verlag 2016, S. 43-114.

Elena Panagiotidis: „Klara Indernach‘ – die Autorin ohne Schreibblockade und ohne Lohn“, in: NZZ vom 21. 9. 2023, online unter <https://www.nzz.ch/panorama/ki-im-journalismus-klara-inder-nach-die-autorin-ohne-schreibblockade-und-ohne-lohn-ld.1757246>

Aditya Ramesh; Prafulla Dhariwal; Alex Nichol; Casey Chu; Mark Chen: „Hierarchical Text-Conditional Image Generation with CLIP Latents“ (2022), online unter <https://arxiv.org/abs/2204.06125>, S. 1.27.

Richard Restak: „Mind over Matter“ (2006), online unter <http://www.washingtonpost.com/wp-dyn/content/article/2006/12/14/AR2006121401554.html>.

Markus Reuter: „Die schöne neue Welt der virtuellen Influencer“ (2023), online unter https://netzpolitik.org/2023/kuenstliche-intelligenz-die-schoene-neue-welt-der-virtuellen-influencer/?utm_source=pocket-newtab-de-de.

Jürgen Riethmüller: Kalkül der Scham. Der soziale Affekt und das Politische. Berlin: Kadmos 2020.

Daniel Rubinstein: Fotografie nach der Philosophie. Repräsentationsdämmerung. Übersetzt von Tom Wohlfarth. Berlin: Merve 2020.

Marie-Laure Ryan: „Fiktion, Kognition und nichtverbale Medien“, in: Gertrud Koch; Christiane Voss: (Hg.): ‚Es ist, als ob‘. Fiktionalität in Philosophie, Film- und Medienwissenschaft. München: Fink 2009, S. 69-86.

Marcus du Sautoy: Der Creativity-Code. Wie künstliche Intelligenz schreibt, malt und denkt. München: C. H. Beck 2021.

Stephan Scheuer: „KI macht auch ‚weniger Talentierte‘ zu Trickbetrügern“ (2023), online unter <https://app.handelsblatt.com/technik/ki/wormgpt-und-fraudgpt-ki-macht-auch-weniger-talentierte->

Konstanze Schütze: „Bildlichkeit nach dem Internet – Kunstvermittlung am Bild als Gegenwartsbewältigung“, in: Zeitschrift Kunst Medien Bildung | zkmb – Oktober 2019: Postdigital Landscapes (hg. Kristin Klein, Willy Noll), online unter: <http://zkmb.de/bildlichkeit-nach-dem-internet-kunstvermittlung-am-bild-als-gegenwartsbewaeltigung/>.

Limor Shifman: Meme. Kunst, Kultur und Politik im digitalen Zeitalter. Aus dem Englischen von Yasemin Dinçe. Berlin: Suhrkamp 2014.

Wolfgang Stieler: „Anzeichen von Bewusstsein bei ChatGPT und Co.“ (2023), online unter https://www.heise.de/hintergrund/Anzeichen-von-Bewusstsein-bei-ChatGPT-und-Co-9295425.html?utm_source=pocket-newtab-de-de.

Robert Strasser: KI-Kunst. Die neue Ära der Bildgenerierung. Wrocław: o. V. 2023.

Albrecht Wellmer: Versuch über Musik und Sprache. München: Hanser 2009.

Harald Willenbrock: „Künstliche Intelligenz und Kunst: Wann ist Kunst eigentlich Kunst?“ (2021), online unter <https://www.goethe.de/prj/zei/de/art/22094490.html>.

Dorothea Winter: KI, Kunst und Kitsch: Ein philosophischer Aufreger. Heidelberg: Carl Auer 2023.

Marty J. Wolf; Kate W. Miller; Frances S. Grodzinsky: „Why We Should Have Seen That Coming“, in: The ORBIT Journal 1, 2 (2017), S. 1-12.

¹ Der sprechende Begriff von Emily Bender etwa in Emily M. Bender; Tamnit Gebru; Angelina McMillan-Major; Margaret Mitchell: „On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?“ (2021), in: FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021, online unter <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>, S. 610-623, hier S. 617. Mitchell publizierte hier unter dem Pseudonym „Shmargaret Shmittell“, um sich vor Repressalien ihres Arbeitgebers zu schützen – ohne Erfolg, Google entließ sie und Gebru infolge des kritischen Papers.

² ChatGPT, „Beschreibe dich selbst“ am 25. 7. 2023.

³ Auch im größeren Teil der Welt, der bereits von autoritären Regimes beherrscht wird, vom Iran, Saudi-Arabien über Russland, China u. v. a. m., haben diese Entwicklungen gewiss nicht zu überschätzende Auswirkungen, die etwa der Formel folgen werden: mehr Unterdrückung, effektivere Überwachung, tödlichere Waffen; es werden nur eben mutmaßlich ganz andere Effekte sein als hierzulande, über die ich an dieser Stelle nicht spekulieren möchte.

Es ist nicht einfach, alternativ zum normativen und vielfach vorbelasteten „westlich“, „westliche Welt“ etc. ein zutreffendes Adjektiv zu finden, das einerseits die fortbestehende Wirkmacht des *digital divide* berücksichtigt und andererseits die nahezu maximalen Unterschiede der Digitalisierung und des WWW in den diversen autoritären politischen Systemen berücksichtigt; kurz gesagt: dass die global gesehen absolut minoritäre und zunehmend prekäre Position dessen, was sich selbst hegemonial zur „Weltöffentlichkeit“ aufbläst, also der lose Bund der immer weniger werdenden liberalen Demokratien, nicht von vorneherein übersieht. Eine sehr gute Darstellung zum derzeitigen Stand der KI-Diskussion findet sich bei Steffen Albrecht: „ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotentiale und mögliche Auswirkungen“ (2023), online unter <https://www.bundestag.de/resource/blob/944148/30b0896f6e49908155fcd01d77f57922/20-18-109-Hintergrundpapier-data.pdf>, S. 1-112 [= Hintergrundpapier Nr. 26 vom 21. April 2023 des Büros für Technikfolgenabschätzung beim Deutschen Bundestag, Ausschussdrucksache 20(18)109].

⁴ Jörg Phil Friedrich: Degenerierte Vernunft. Künstliche Intelligenz und die Natur des Denkens. München: Claudius 2023, S. 8 f. folgert an diesem Punkt, dass die jetzt mögliche Spiegelung durch die kreative Potenz generativer KI es endlich ermögliche, ein uraltes, aber ganz falsches Stereotyp hinsichtlich der Besonderheit der menschlichen Vernunft zu korrigieren. Es verortete in der Schiller-Tradition das Besondere des Menschen und seiner *ratio* in seiner Schaffenskraft und -freude, seiner Geschicklichkeit und Kreativität. Als Folie bei diesem Gedanken eines *animal fingens* diene offensichtlich das nichtmenschliche Tier: „Dabei kann sich herausstellen, dass die Dynamik, die in der Vergangenheit gewisse Produkte als Krönung menschlicher Leistungs- und Schöpfungskraft erscheinen lassen hat, die Menschheit womöglich in ein Missverständnis über das eigene Wesen geführt hat, dass sich Verfahren und Ziele verselbstständigt haben, weil sie für besonders kreativ oder meisterhaft gehalten wurden. Die Tatsache, dass auch eine Maschine [...] Vergleichbares hervorbringen kann, lässt uns erkennen, dass es sich vielleicht um etwas gehandelt hat, was zunächst als etwas ganz Besonderes erschien, am Ende aber eine Spielerei, eine Ablenkung ein Irrweg war, der verdeckte, was uns als Menschen tatsächlich ausmacht.“ Deutlich werde, dass dieses Besondere in der prinzipiell empathischen und moralischen Grundierung des menschlichen Denkens liegt.

Dass Kunst selbstverständlich niemals eine Ablenkung (von was?) bzw. ein Irrweg (wohin?) war oder dazu geworden ist, und dass umgekehrt das bloße Hervorbringen von ganz Erstaunlichem oder sogar Wunderbarem (man denke an die alten *ludi naturae*-Debatten) bereits seit dem ausgehenden 18. Jahrhundert in philosophisch-ästhetischer Perspektive für eine Identifizierung als Kunst nicht ausreicht – siehe etwa Hegels rigide Ausklammerung des Naturschönen aus dem Reich der Ästhetik – sind demgegenüber Aspekte, welche die vorliegende Skizze andeuten soll.

⁵ Deziert nicht geht es mir um die technische oder ökonomische Perspektive, d. h. das notorische Loblied auf „Innovation“ und „Chancen“.

⁶ Albrecht, ChatGPT (2023), S. 71. Gerade die sehr rasche Aneignung durch manche Schüler:innen und Studierende für eigene Arbeiten ist durch die technischen Möglichkeiten nicht gedeckt. Diesbezüglich kann es weniger um ein (ganz unrealistisches) Verbot der Nutzung gehen als um eine vertiefte Auseinandersetzung, welche Möglichkeiten und Grenzen z. B. ChatGPT im jeweiligen Kontext hat. Vgl. dazu die Übersicht ebd. S. 71-82. Für den Hochschulbereich, allerdings vor der ChatGPT-Revolution verfasst und einigermaßen affirmativ, vgl. etwa den Band von Tobias Schmohl; Alice Watanabe; Kathrin Schelling: Hochschulbildung: Lehre und Forschung. Künstliche Intelligenz in der Hochschulbildung. Chancen und Grenzen des KI-gestützten Lernens und Lehrens. Bielefeld: transcript 2023. Instruktiv sind auch viele der fast durchgängig ausgesprochen liberalen Handreichungen größerer Hochschulen zum ChatGPT-Einsatz in der Lehre, etwa der TU München (https://www.prolehre.tum.de/fileadmin/w00btq/www/Angebote_Broschueren_Handreichungen/prolehre-handreichung-chatgpt-v2.2.pdf) oder der TU Stuttgart (<https://www.uni-stuttgart.de/universitaet/aktuelles/dokumente/handreichung-fuer-pruefende-KI-werkzeuge-und-pruefungen.pdf>).

⁷ Ulrich Oevermann: „Biographie, Krisenbewältigung und Bewährung“, in: Sylke Bartmann; Axel Fehlhaber; Sandra Kirsch (Hg. u. a.): „Natürlich stört das Leben ständig.“ Perspektiven auf Entwicklung und Erziehung. Wiesbaden: VS Verlag 2009, S. 35-55, hier S. 36 f.: „In der Routine macht man keine neuen Erfahrungen, man wendet sie darin nur an.“ (Ebd. S. 27).

⁸ Detlev Garz; Uwe Raven: „Lebenspraxis – Krise und Routine“, in: dies. (Hg.): Theorie der Lebenspraxis. Einführung in das Werk Ulrich Oevermanns. Wiesbaden: Springer VS 2015, S. 25-59, hier S. 25.

⁹ Der Begriff „Technologiersatztechnologien“ bei Niklas Luhmann; Karl Eberhard Schorr: „Das Technologiedefizit der Erziehung und die Pädagogik“, in: dies. (Hg.): Zwischen Technologie und Selbstreferenz. Fragen an die Pädagogik. Frankfurt a. M.: Suhrkamp 1996, S. 11-40, hier S. 21; sie verweisen auf das basale Technologieproblem der Pädagogik, die von der Kreidetafel an permanent versucht ist, den Erziehungsalltag an „neue“ technische Medien auszulagern.

¹⁰ Ob der psychologisierende Begriff der Halluzination gut gewählt ist, scheint mit Carl T. Bergstrom: „Your chatbot is not ‚hallucinating‘. It’s bullshitting“ (2023), online unter <https://undark.org/2023/04/06/chatgpt-isnt-hallucinating-its-bullshitting/> fraglich: „The terms ‚hallucination‘ and ‚confabulation‘ both share one big problem: As used in medicine, they each refer to states that arise as a consequence of some apparent malfunction in an organism’s sensory or cognitive machinery. (Importantly, perspectives on what hallucinations and confabulations are — and how they manifest — are profoundly shaped by cultural and social factors.“

¹¹ Ulrich Oevermann: „‚Krise und Routine‘ als analytisches Paradigma in den Sozialwissenschaften“ [=Abschiedsvorlesung am 28. 4. 2008, gehalten an der Goethe Universität Frankfurt a. M.], in: Roland Becker-Lenz; Andreas Franzmann; Matthias Jung (Hg. u. a.): Die Methodenschule der Objektiven Hermeneutik. Wiesbaden: VS Verlag 2016, S. 43-114.

¹² Vgl. das ernüchternde Fazit bei Allbrecht, ChatGPT (2023), S. 58 ff. (Informationssuche und Journalismus) u. S. 79 ff. (Wissenschaft und Forschung)

¹³ Elena Panagiotidis: „‚Klara Indernach‘ – die Autorin ohne Schreibblockade und ohne Lohn“, in: NZZ vom 21. 9. 2023, online unter <https://www.nzz.ch/panorama/ki-im-journalismus-klara-inder-nach-die-autorin-ohne-schreibblockade-und-ohne-lohn-ld.1757246> berichtet von zwei aktuellen Beispielen aus Deutschland; im Kölner Boulevardblatt „Express“ verfasst die KI sogar mit von *Midjourney* generiertem Autor:innenbild diverse Artikel unter dem Pseudonym „Klara Indernach (KI)“, im Burda-Verlag erschien ein Sonderheft „Lisa Kochen und Backen“ mit 99 Pasta-Rezepten, bei denen bis zu den Abbildungen alles KI-generiert war – ohne Kennzeichnung.

¹⁴ Zum Beispiel des Verlags *Clarkesworld* vgl. etwa Alex Hern: „Sci-fi Publisher clarkesworld halts pitches amid deluge of AI-generated stories“, in: The Guardian vom 21. 2. 23, online unter <https://www.theguardian.com/technology/2023/feb/21/sci-fi-publisher-clarkesworld-halts-pitches-amid-deluge-of-ai-generated-stories>.

¹⁵ In einer ohnehin an der für Gesellschaftlichkeit notwendigen Aufgabe der kommunikativen Konstruktion *einer* gemeinsamen Wirklichkeit zunehmend scheiternden Epoche könnte dies hinsichtlich der immer offensichtlicheren Spaltungstendenzen („alternative Fakten“ usw.) wie ein sozialer Brandbeschleuniger wirken.

¹⁶ Albrecht, ChatGPT (2023), S. 53-57 bilanziert, dass etwa in der Tumordiagnostik die Resultate ernüchternd sind, und im therapeutischen Kontext werden sogar Fälle kolportiert, bei denen Chatbots Patient:innen zum Selbstmord geraten haben. Insbesondere die Probleme, die Chatbots bei länger andauernden Interaktionen mit einer Person haben (vgl. dazu ebd. S. 39), und die offensichtlichen Datenschutzprobleme dürften gegen solche Einsätze sprechen.

¹⁷ Diese Fragestellung soll nicht implizieren, dass Kunst prinzipiell Fiktion sei; die vormoderne Nähe von Fiktion und Kunst ist längst passé. Gewiss können auch nichtfiktionale Ansätze der Gegenwartskunst, etwa dokumentarische oder postkonzeptuale, von einer Nutzung generativer KI profitieren, Erstere etwa analog zum Datenjournalismus, nur liegt der Fokus meiner Skizze nicht auf diesem Bereich.

¹⁸ Michel Foucault: „Das Denken des Außen“, in: ders.: Schriften in vier Bänden. Dits et Ecrits, Bd. I 1954-1969, Frankfurt a. M.: Suhrkamp 2001, S. 670-697, hier S. 673 f.

¹⁹ Ebd., S. 674.

²⁰ Ebd. im Titel. Georg Wilhelm Friedrich Hegel: Vorlesungen über die Ästhetik I [=Werke Bd. 13]. Herausgegeben Eva Moldenhauer und Karl Michel. Frankfurt a. M.: Suhrkamp 1997, S. 18: „Vor der Strenge des Gesetzmäßigen und der finsternen Innerlichkeit des Gedankens suchen wir Beruhigung und Belebung in den Gestalten der Kunst, gegen das Schattenreich der Idee heitere, kräftige Wirklichkeit.“

²¹ Korrekt zitiert heißt der Satz: „Die Schlüsselposition des Subjekts in der Erkenntnis ist Erfahrung, nicht Form.“ So Theodor W. Adorno: Negative Dialektik. Jargon der Eigentlichkeit [1966], in: ders.: Gesammelte Schriften, Bd. 6. Frankfurt a. M.: Suhrkamp 2003, S. 752.

²² Zum „derzeitigen Hype“ Albrecht, ChatGPT (2023), S. 17.

²³ Wohingegen auf GANs beruhende generative KIs problemlos miteinander in Dialog treten könnten. Stefan Achleitner: „Generative Adversarial Networks: Wenn künstliche Intelligenz künstliche Daten erschafft“ (2020), online unter <https://intelligente-welt.de/generative-adversarial-networks-wenn-kuenstliche-intelligenz-kuenstliche-daten-erschafft/> erklärt das Konzept folgendermaßen: „Generative Adversarial Networks‘ bestehen aus zwei neuronalen Netzwerken, die sozusagen ein Spiel miteinander spielen: Das erste Netzwerk heißt ‚Generator Network‘ und erzeugt Daten – zum Beispiel ein Bild, das am Anfang nur aus zufällig statistischem Rauschen besteht. Das zweite neuronale Netzwerk nennt man ‚Discriminator Network‘. Es analysiert die vom ersten Netzwerk generierten Daten. Um diese Daten bewerten zu können, trainieren die Programmierer das ‚Discriminator

Network‘ zuvor mit echten Daten.“ In allergrößter Vereinfachung besteht das so entstehende System damit aus drei Elementen: den neuronalen Netzen, dem Discriminator und dem Input inklusive der Daten, mit denen die KI „trainiert“ wurde, d. h. ebendiesen Daten, den Prompts und das menschliche Feedback usw.

²⁴ Friedrich, *Degenerierte Vernunft* (2023), S. 40.

²⁵ Hier ist ein infiniter Regress latent angelegt, denn der Begriff des Denkens wird traditionell an das Subjekt geknüpft. Ansonsten gilt: Vielleicht ist der Logos nur eine antike Projektion, und was nicht existiert, kann nicht denken. Ein Medium wie die Sprache, die hier am häufigsten erhalten muss, denkt nicht, es basiert nur über die Etymologie auf einer unendlichen Spur von vergangenen, zu Wörtern kondensierten Bewusstseinsakten, welche die „sprechenden“ Begriffe gebildet haben. Besonders deutlich ist dies bei Metaphern wie etwa im Deutschen „Zweck“, „begreifen“ oder „Himmelszelt“. Insofern „denken“ die Sprache (als mit McLuhan auch nur ein Medium mit einer Botschaft) wie auch der Mythos und, weit stärker, das Unbewusste, aufgefasst mit C. G. Jung als Archetyp oder mit Lacan als Sprache, a) nur metaphorisch und b) über Begriffe, Schemata, Assoziationen, Metaphern usw. auch ein Stück weit in mir oder durch mich (*es denkt* mich statt *ich denke*), aber sie denken nie faktisch, in der Weise einer von einem bestimmten Reiz ausgelösten mentalen Proposition bzw. einer rekursiven Folge von propositionalen Bewusstseinsakten, was man in geisteswissenschaftlicher Perspektive gemeinhin als Denken bezeichnet.

Das Unbewusste ist ohnehin, entgegen Lacans Idee, gewiss nicht subjektlos, sondern vielmehr immer auch radikal subjektiv, sodass nur der Mythos und die an diesem Punkt strukturell verwandte Kultur als mögliche Fälle eines tatsächlich subjektlosen Denkens verbleiben. Beide müssen aber, um wirken zu können, immer im konkreten Denken eines Menschen bzw. eines Kollektivs aktualisiert werden, sodass die Differenz des „reinen“ Mythos bzw. der „reinen“ Kultur und ihrer jeweiligen Aktualisierung am Ende gar nicht zu ziehen ist und beide somit nur das Postulat einer beobachtenden Instanz sind, das *in toto* nicht fass- oder beschreibbar ist, schon gar nicht als selbstdenkende Instanz.

²⁶ Gerade diese Trennung von Oberfläche und Tiefe, wie sie bereits für Computer allgemein typisch ist, gepaart mit der Reaktivität legt beim menschlichen Interaktanten die Metapher des Denkens nahe (das man beim menschlichen Gegenüber auch nicht beobachten kann). Vgl. Noam Chomsky; Ian Roberts; Jeffrey Watumull: „The False Promise of ChatGPT“, in: *New York Times* vom 8. 3. 2023, online unter <http://www.nytimes.com/2023/03/08/opinion/noam-chomskychatgpt-ai.html>.

²⁷ Vgl. Albrecht, *ChatGPT* (2023), S. 20 f.: „Transformer weichen von früheren Formen neuronaler Netzwerke insofern ab, als sie Daten nicht sequentiell abarbeiten. Stattdessen sind sie in der Lage, eine Vielzahl an Daten, beispielsweise einen ganzen Satz oder gar Absatz, gleichzeitig zu verarbeiten. Im Unterschied zu früheren Modellarchitekturen können mit Transformern Verbindungen auch zwischen weit entfernt stehenden Wörtern beachtet und die Erkennung sprachlicher Muster verbessert werden [...]“. Technisch beruhen sie auf der Leistungsfähigkeit heutiger Grafikprozessoren.

²⁸ Und zwar die insgesamt noch plausibelste *Aufmerksamkeitsschema-Theorie des Bewusstseins*, die arg panpsychische *Integrated Information Theory*, die *Global Workspace Theory*, die *Predictive Processing Theory* und die *Higher-Order Theories*. Die in meinen Augen allesamt fragwürdigen Bewusstseinstheorien sind überwiegend materialistisch und argumentieren teilweise sogar quantitativ; insofern sind sie nicht allzu erklärungsstark. Noch nicht einmal die Frage, ob höhere Säugetiere ein Bewusstsein haben, ist einhellig geklärt. Das Zitat im Abstract des Papers von Patrick Butlin; Robert Long; Eric Elmoznino; Yoshua Bengio; Jonathan Birch; Axel Constant; George Deane; Stephen M. Fleming; Chris Frith; Xu Ji; Ryota Kanai; Colin Klein; Grace Lindsay; Matthias Michel; Liad Mudrik; Megan A. K. Peters; Eric Schwitzgebel; Jonathan Simon; Rufin VanRullen: „Consciousness in Artificial Intelligence: Insights from the Science of Consciousness“ (2023), online unter <https://doi.org/10.48550/arXiv.2308.08708>. Wolfgang Stieler: „Anzeichen von Bewusstsein bei ChatGPT und Co.“ (2023), online unter https://www.heise.de/hintergrund/Anzeichen-von-Bewusstsein-bei-ChatGPT-und-Co-9295425.html?utm_source=pocket-newtab-de-de merkt hierzu kritisch an: „Die Autoren sind zu Ergebnissen gekommen, die in der sorgfältig ausgewählten Fragestellung bereits angelegt waren.“

²⁹ Friedrich, *Degenerierte Vernunft* (2023), S. 107 formuliert dies ohne jede Konnotation des Luhmann’schen Sinnbegriffs, auf den ich hier anspiele – „Sinn“ oder „Bedeutung“ (meaning) ist immer Resultat des ihn prozessierenden sozialen oder psychischen Systems, kein objektiv feststehender Bezug –, sehr klar: „[...] Steckt denn kein Sinn in den Erzeugnissen der künstlichen Intelligenz? Ist das, was sie erzeugt, etwas sinnlos? Natürlich nicht. Aber es hat keinen Sinn für das KI-System [...]“.

³⁰ Beim maschinellen Lernen unterscheidet man für gewöhnlich drei Stufen, das *supervised learning* (der Mensch „füttert“ die Maschine mit Daten und gibt an, was richtig ist und was falsch), das *unsupervised learning*, bei dem die Maschine eigenständig Zusammenhänge erschließt, und ggf. auch ein *reinforcement learning* über „Belohnungen“ o. Ä.

Friedrich, *Degenerierte Vernunft* (2023), S. 109 merkt zu diesem gängigen Sprachgebrauch zu Recht kritisch an: „Zum Training eines Menschen gehört immer, dass er selbst sich motiviert, dass er ein Ziel erreichen will [...]. Insofern ist es schon eine ungenaue Metapher, davon zu sprechen, dass ein KI-System trainiert wird, wenn durch Dateneingaben seine Parameter optimiert und justiert werden. Zum Training gehören Ziele, gehören Motivation und Ausdauer, gehört der Wille, das Ziel zu erreichen. All das hat das KI-System nicht; ob es eine Leistung verbessert, ist ihm nicht einmal gleichgültig, weil ihm die Selbstbezüglichkeit der Besinnung völlig fehlt.“ Die oben zitierte ChatGPT-Aussage zur Subjektlosigkeit von ChatGPT ist nur die Simulation von reflexiver Selbstbezüglichkeit auf algorithmischer Basis in sprachlicher Ich-Form; verfasst von einer Entität, die weder wissen kann noch will, was Ich-Sein meint, z. B. Rückenschmerzen haben.

³¹ Niklas Luhmann: Die Gesellschaft der Gesellschaft. Frankfurt a. M.: Suhrkamp 1999, Bd. 1, S. 303 wies hier darauf hin, dass Computern nicht ohne Weiteres die „gesellschaftskonstituierende Leistung der Kommunikation möglich sei. Sie müssten, heißt es hier exakt, auch wissen, „was andere Computer nicht wissen“. Hintergrund ist: „Kommunikation ist ein laufendes Prozessieren der Differenz von Wissen und Nichtwissen [...]“ (Ebd.)

³² Zu den Limitationen sehr viel ausführlicher, als mir dies möglich ist, Albrecht, ChatGPT (2023), S. 38-43.

³³ Vgl. dazu die Beispiele bei Chomsky et al., False Promise (2023). Es ist anzunehmen, dass ein anderes Prozedere beim Training, welches nicht auf sprachbasierter Interaktion, sondern auf Biofeedback-Systemen beruht, also auf der direkten Verknüpfung mit z. B. Hirnscans menschlicher Probanden, was den Maschinen gewissermaßen ein rudimentäres *moral brain* antrainierte, zu besseren Ergebnissen kommen könnte. Daran wird verschiedenorts gearbeitet. Angesichts enormer affektbezogener Defizite der Neurowissenschaften sollten die Hoffnungen hier allerdings nicht allzu groß sein.

³⁴ Frühere Chatbots wie 2016 Microsofts „Tay“ auf Twitter waren diesbezüglich extrem negativ aufgefallen – Vgl. dazu Marty J. Wolf; Kate W. Miller; Frances S. Grodzinsky: „Why We Should Have Seen That Coming“, in: The ORBIT Journal 1, 2 (2017), S. 1-12, hier S. 1 ff.

³⁵ Er sagt, zitiert bei Jakob von Lindern: „Braucht die deutsche Vorzeige-KI mehr Erziehung?“, in: ZEIT Online vom 11. 9. 2023, online unter https://www.zeit.de/digital/2023-09/aleph-alpha-luminous-jonas-andrulis-generative-ki-rassismus?utm_source=pocket-newtab-de-de: „Wir haben Filter und Sicherheitstools, aber die sind da, wo unsere Technologie zum Einsatz gebracht wird. Unsere Kunden und Partner nehmen diese Technologie und richten sie nach ihren Vorstellungen aus. Wir wollen Kunden wie SAP oder Bosch nicht unsere Vorstellung davon vorgeben, was richtig und falsch ist.“

³⁶ Beleg abgedr. ebd.

³⁷ Die generative KI „ist reines Instrument eines menschlichen Willens [...]. Neuronale Netze sind nicht mehr und nicht weniger als leistungsstarke Generalisierungsmaschinen“. So Marc Engenhardt; Sebastian Löwe: Design und künstliche Intelligenz. Theoretische und praktische Grundlagen der Gestaltung mit maschinell lernenden Systemen. Basel: Birkhäuser 2022, S. 48.

³⁸ Neuerlich handelt es sich um eine perfekte technische Umsetzung poststrukturalistischer Theorie in der Wirklichkeit: Die Signifikatlosigkeit, die Derrida (fälschlicherweise) fürs menschliche Sprechen bzw. Schreiben annimmt, trifft hier tatsächlich weitgehend zu.

³⁹ Insbesondere die experimentellen Python-Skripte, die man AutoGPT nennt und die man gewissermaßen als Interface für alle KI-Anwendungen auf Open-AI-Basis nutzen kann, und in einem eingeschränkten Maß und weniger wild als Alltagshelfer auch Gerätschaften wie der R1-KI-Assistent von *Rabbit* oder der AI Pin von *Humane* (beides sogenannte Large Action Models, die auf kurze Sprachbefehle hin selbstständig Arzttermine vereinbaren, Restaurantplätze oder Kinokarten bestellen oder Taxis oder Essen ordern können) suchen selbstständig auch nach Lösungswegen der gestellten Aufgaben, was zwar nicht heißt, dass die KI hier *denkt*, aber dass sie ggf. gegen diverse auch rechtliche Regeln und Etikette verstößt.

Dahingestellt sei, ob man hier von „degenerierter Vernunft“ sprechen sollte, wie es Friedrich, Degenerierte Vernunft (2023) tut, der eher unterstellt als zeigt, „wir“ verwendeten in Bezug auf KI die Begriffe Intelligenz und Vernunft synonym (ebd., S. 33).

⁴⁰ Das instruktive Beispiel stammt von meinem Kollegen David Quigley. Bei der verbesserten Version GPT4 lautet die Antwort dann „Yes, Buddha had a son named Rahula.“ Die KI hat kein Wissen und arbeitet auch nicht über symbolische Repräsentation, sie berechnet Wahrscheinlichkeiten. Insofern läuft mit Hannes Bajohr: „Die ‚Gestalt‘ der KI. Jenseits von Atomismus und Holismus“, in: Zeitschrift für Medienwissenschaft 23, Heft 2 (2020), S. 168-183, hier S. 178 auch die alte philosophische Unterscheidung atomistischer und holistischer Konzepte ins Leere: „Künstliche neuronale Netze können also weder als vollständig atomistisch noch als vollständig holistisch konzeptualisiert werden.“

⁴¹ Insofern schlägt hier auch der komplette Bias des netzbasierten Sich-Äußerns durch, etwa hinsichtlich Maskulinität, *digital divide*, Marginalisierung bestimmter Perspektiven u v. a. m.

⁴² Mit der Frage: „Was liest die KI?“ hat sich die diesbezüglich sehr erhellende Studie von Kent K. Chang; Mackenzie Cramer; Sandeep Soni; David Bamman: „Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4“ (2023), online unter https://www.researchgate.net/publication/370442472_Speak_Memory_An_Archaeology_of_Books_Known_to_ChatGPTGPT-4, S. 1-15 beschäftigt; der entscheidende Faktor, dass die KI einen Text kennt, ist, dass dieser (ggf. in Auszügen) gratis im Netz zugänglich ist, nicht sein literarischer Rang.

⁴³ Andrian Kreye: „Was die KI liest“, in: Süddeutsche Zeitung vom 24./25. Juni 2023, S. 20.

⁴⁴ Chang et al., Archaeology of Books (2023), S. 8 umschreiben den diesbezügl. Bias folgendermaßen: „Our work here has surfaced not only that ChatGPT and GPT-4 have memorized popular works, but also what kinds of narratives count as ‚popular‘—in particular, works in the public domain published before 1928, along with contemporary science fiction and fantasy.“

⁴⁵ Ebd., bereits im Abstract des Beitrags: „We find that OpenAI models have memorized a wide collection of copyrighted materials ...“ Folgerichtig haben u. a. Sylvia Day, Roxana Robinson, John Grisham, George R. R. Martin und andere prominente US-Schriftsteller:innen gemeinsam mit dem US-Schriftstellerverband *Authors Guild* im September 2023 den ChatGPT-Entwickler OpenAI wegen Urheberrechtsverletzungen in einer bei einem Bundesgericht in New York eingereichten Sammelklage verklagt. Der Ausgang des Prozesses dürfte eine längere Phase juristischer Aushandlungen zur Urheberrechtsfrage einläuten.

⁴⁶ Die deutlichen Warnungen vor der ungebremsten Implementation der KI etwa seitens Geoffrey E. Hinton, der Bahnbrechendes im Bereich der Theorie der neuronalen Netze geleistet hatte und führend an der Entwicklung der Google-Alternative zu ChatGPT „Apprentice Bard“ beteiligt war und vieler anderer, kommen recht spät; u. a. können die ethischen Limitationen der generativen KI, die das Schlimmste bei selbstaktiven Anwendungen wie Auto GPT verhindern sollen, gehackt und somit umgangen bzw. eliminiert werden. Vgl. dazu Cade Metz: „The Godfather of A.I.’ Leaves Google and Warns of Danger Ahead“, in: New York Times vom 1. 5. 2023, online unter <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>. Einige Wochen zuvor hatte bereits das Bostoner *Future of Life Institute* in einem offenen Brief mit prominenten Unterzeichner:innen für eine sechsmonatige Entwicklungspause geworben.

⁴⁷ Vgl. etwa meine Zusammenfassung dieser These aus der *Kontrollgesellschaft außer Kontrolle* in Jürgen Riethmüller: *Kalkül der Scham. Der soziale Affekt und das Politische*. Berlin: Kadmos 2020, S. 83-126.

⁴⁸ Vgl. die redaktionelle Meldung „Influencer:innen sind zunehmend keine echten Menschen mehr“, in: Der Standard vom 7. 8. 2023, online unter https://www.derstandard.at/story/300000181901/influencerinnen-keine-echten-menschen?utm_source=pocket-newtab-de-de; ausf. und mit vielen Beispielen zu V-Tuber:innen und Influencer:innen auf Instagram, *Twitter/X* und *TikTok* wie „Kyra“ und „Shudu“ etwa bei Markus Reuter: „Die schöne neue Welt der virtuellen Influencer“ (2023), online unter https://netzpolitik.org/2023/kuenstliche-intelligenz-die-schoene-neue-welt-der-virtuellen-influencer/?utm_source=pocket-newtab-de-de.

⁴⁹ Ebd. (Reuter).

⁵⁰ So zu lesen in CROSSROADSEVOLUTION. Der Trend-Atlas für den deutschen Kommunikationsmarkt 2023, hg. Ströer Core GmbH & Co. KG, online unter

https://www.stroeer.de/media/02_downloads/studien_whitepaper/stroeer_crossroads_evolution_2023_01.pdf, S. 44.

⁵¹ Zitat Reuter, *Schöne neue Welt* (2023). Es ist fraglich, ob die virtuelle Simulation einer Kunstfigur, sozusagen das Simulakrum des Simulakrums der Authentizität, allzu viel mit einer ästhetischen Fiktion wie etwa einem Roman oder Spielfilm zu tun hat. Die zynische Haltung der Schöpfer:innen von Aww Inc., der Firma, die hinter der in Japan erfolgreichen Influencerin „imma“ steht, korrespondiert allerdings ähnlich wie die Chuzpe der sogenannten „alternativen Fakten“, auf die sich der global erstarkende Rechtspopulismus beruft, in auffälliger Weise mit poststrukturalistischen Ideen, welche schon lange die gesamte Realität als von „Fiktionen“ durchsetzt ansehen usw.: „Alles, was wir online und in den Medien sehen, ist eine Erzählung. Und es ist jedermanns eigene Entscheidung, ob er es glaubt oder nicht.“ (Zit. ebd.)

⁵² Friedrich, *Degenerierte Vernunft* (2023), S. 71 f.

⁵³ Chomsky et al., *False Promise* (2023).

⁵⁴ Albrecht, *ChatGPT* (2023), S. 11.

⁵⁵ Jennifer Jacquet: *Scham: Die politische Kraft eines unterschätzten Gefühls*. Frankfurt a. M.: Fischer 2015, S. 178.

⁵⁶ (Selbstversuch) Ali Borji: „A Categorical Archive of ChatGPT Failures“ (2023), online unter <https://arxiv.org/pdf/2302.03494.pdf>, S. 1-43, hier S. 13 [Übers. J. R.] bilanziert: „ChatGPT hat Schwierigkeiten, zwischen faktischen Informationen und Fiktion zu unterscheiden, und erzeugt frei erfundene Informationen.“

⁵⁷ Vgl. dazu den Bericht im MIT Technology Review von Will D. Heaven: „Why Meta’s latest large language model survived only three days online“ (2022), online unter <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/>. Ebenso frei ist die KI, wissenschaftliche Begründungen für diskriminierende oder menschenverachtende Thesen zu liefern. Vgl. dazu Aaron J. Snowell; Jean Burgess: „The Galactica AI model was trained on scientific knowledge – but it spat out alarmingly plausible nonsense“, in: *The Conversation* vom 30. November 2022, online unter <https://theconversation.com/the-galactica-ai-model-was-trained-on-scientificknowledge-but-it-spat-out-alarmingly-plausible-nonsense-195445>.

⁵⁸ Johanna Bryson im Interview mit Marie Kilg, in: *taz* vom 26. 7. 2017, S. 13. Sie plädiert hier für eine unabhängige Prüfstelle, die ähnlich wie im Fall von Medikamenten die ethische Zulässigkeit eines Algorithmus prüfen soll, ehe er zugelassen wird. „KI hat auf unsere Gesellschaft einen Einfluss, der mindestens so groß ist wie der von Medikamenten.“ (Ebd.)

⁵⁹ Umso perfider, dass die Unternehmenslobby beim derzeit ausverhandelten EU-Gesetz zur Regulierung darauf drängt, den Schutz der Menschenrechte aus dem Gesetz zu streichen. Im September 2023 wurde dazu von [avaaz.org](https://www.avaaz.org) eine Petition gestartet. Vgl. zum AI Act des EU-Parlaments, auf den ich hier nicht eingehen kann, die ausführlichen Meldungen beim Deutschlandfunk (<https://www.deutschlandfunk.de/ai-act-eu-kuenstliche-intelligenz-gefahr-regulierung-100.html>) und beim SRF (<https://www.srf.ch/news/wirtschaft/europaeisches-ki-gesetz-wie-die-eu-kuenstliche-intelligenz-regulieren-will>) oder auch beim ORF (<https://orf.at/stories/3316790/>). Das Gesetzgebungsverfahren ist derzeit noch offen.

⁶⁰ Hegel, *Vorlesungen über die Ästhetik* (1997), S. 18.

⁶¹ An diesem Punkt kann an die Mahnung hellstichtiger Beobachter:innen älterer Generationen erinnert werden wie etwa die des Gründervaters der Merz Akademie, der 1947 notierte: „Desgleichen müsste klar ins Bewusstsein treten, dass die Technik für die Menschheit [...] weit gefährlicher geworden ist als ihre letzte Auswirkung: die Atombombe je zu werden vermag [...], weil die

internationale Überwachung der Technik, die in Zukunft genauso gefordert werden müsste wie die der Atomenergie, bisher eine unvollständige war ...“ Albrecht L. Merz: *Manifest d. Erziehung*. Broschüre Stuttgart o. V. 1947, S. 21.

⁶² Bereits Luhmann, Gesellschaft der Gesellschaft (1999), S. 303 merkte zu dieser Idee an: „Wie verzaubert durch eine lange humanistische Tradition hatte man das Problem zunächst in der Frage gesehen, ob die Computer und ihre ‚artificial intelligence‘ dem Bewusstsein Gleichwertiges oder Überlegenes leisten [können] [...]. Das Fluchtziel der Geisteswissenschaft war und blieb dann das menschliche Subjekt.“

⁶³ Offenbar hat die russische Armee allein aufgrund der KI-Panzerabwehrrakete „Javelin“ aus US-Fertigung bereits im Oktober 2022 1.500 Panzer verloren. Hervorzuheben ist auch das System „Phoenix Ghost“, sogenannte *loitering munition*, also fliegende Drohnen, die lange über einem Zielgebiet auf der Suche nach einem Ziel kreisen können. Tuck sagt: „In der Ukraine findet die Erprobung neuer Waffensysteme unter Bedingungen des Ernstfalls statt. Die ukrainischen Soldaten entwickeln und modifizieren neue US-Waffensysteme. Es ist eine Zusammenarbeit. Die Patente halten selbstverständlich die Unternehmen aus den Vereinigten Staaten. Aber die Ukrainer profitieren ebenso davon.“ So im Interview von Marc von Lüpke mit Jay Tuck: „Eine Demütigung der russischen Streitkräfte“ (2023), online unter https://www.t-online.de/nachrichten/ausland/krisen/id_100321436/putin-und-der-ukraine-krieg-eine-demuetigung-der-russischen-streitkraefte-.html. Tuck, der aktuell ein Buch unter dem Titel „KI und der moderne Krieg. Wie künstliche Intelligenz die russische Armee besiegen kann“, vermarktet, hat über 35 Jahre für das deutsche Fernsehen gearbeitet und ist heute Keynote Speaker und CEO von *Airtime Dubai Ltd.*, einem internationalen Medienunternehmen mit engen Kontakten in der arabischen Welt, wo die neuen Waffensysteme auf großes und kaufkräftiges Interesse stoßen.

⁶⁴ Vgl. die Meldung mit den Bildern, die ein Journalist „gemacht“ hat, unter https://correctiv.org/faktencheck/2023/03/22/kuenstliche-intelligenz-generiert-bilder-von-donald-trumps-festnahme-die-gab-es-aber-gar-nicht/?utm_source=pocket-newtab-global-de-DE.

⁶⁵ Albrecht, ChatGPT (2023), S. 11. Größere Mengen dürfte noch ein Euphemismus sein im Hinblick auf die Flut der zu erwartenden Desinformation. In jedem Fall gilt: „Gegenmaßnahmen, wie die automatisierte Erkennung von KI-generierten Texten oder die Verbesserung von Medienkompetenzen, gelten als bislang nicht ausreichend wirksam.“ (Ebd.)

⁶⁶ Vgl. ausf. zur Frage der Desinformation Ben Buchanan; Andrew Lohn; Micah Musser; Katharina Sedova: Truth, Lies, and Automation. How language models could change disinformation. Hg. vom Center for Security and Emerging Technology. Washington, DC 2021, online unter <https://cset.georgetown.edu/publication/truth-lies-and-automation/>. Vgl. dazu das Fazit ebd. S. 47 bereits für das noch eher harmlose GPT 3: „Systems like GPT-3 offer reason for concern about automation in disinformation campaigns. Our tests show that these systems are adept at some key portions of the content generation phase of disinformation operations. As part of well-resourced human-machine teams, they can produce moderate-quality disinformation in a highly scalable manner. Worse, the generated text is not easily identifiable as originating with GPT-3, meaning that any mitigation efforts must focus elsewhere, such as on the infrastructure that distributes the messages. [...] We think GPT-3’s most significant impact is likely to come in scaling up operations, permitting adversaries to try more possible messages and variations [...]“

⁶⁷ Es kann nicht ernsthaft davon die Rede sein, dass die Qualität KI-generierter belletristischer oder journalistischer Texte, Musikstücke oder Comics denen aus Menschenhand in irgendeiner Weise überlegen wären, und da dies an strukturellen Voraussetzungen liegt, wie fehlender sinnlicher Erfahrung und Empathie, ist hier auch kaum ein Quantensprung durch weiterentwickelte Sprachmodelle zu erwarten; der lockende Faktor ist eher der (idealerweise) nahezu minimale Zeit- und Kostenaufwand, was in vielen konkreten Fällen, nötige Überarbeitungen etc. miteingerechnet, die rein betriebswirtschaftliche Kalkulation gleichwohl ins Negative umschlagen lässt.

⁶⁸ In der EU wird derzeit an einem Gesetz zur umfassenden Regulierung der Entwicklung und des Einsatzes von KI gearbeitet; es befindet sich gerade nach einer ersten Einigung von Rat und Parlament am 9. 12. 2023, mitten im Gesetzgebungsverfahren, weshalb sich an dieser Stelle hierzu noch nichts sagen lässt. Vgl. dazu die Verlautbarungen hier:

<https://www.europarl.europa.eu/news/de/headlines/society/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz>.

⁶⁹ Es wird vom privaten Haushalt über Großunternehmen bis zum Pentagon notwendig werden, für alle relevanten Bereiche funktionierende Offlinestrukturen in der Hinterhand zu haben.

⁷⁰ Vgl. dazu etwa den Artikel von Ende August 2023 im *Handelsblatt* von Stephan Scheuer: „KI macht auch ‚weniger Talentierte‘ zu Trickbetrügnern“ (2023), online unter https://app.handelsblatt.com/technik/ki/wormgpt-und-fraudgpt-ki-macht-auch-weniger-talentierte-zu-trickbetruegnern/29333912.html?utm_source=pocket-newtab-de-de.

⁷¹ Online unter

https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2023.pdf?__blob=publicationFile&v=4, vgl. hier bes. S. 40-49 u. S. 70-74.

⁷² Vgl. dazu beispielsweise die Auslassungen von Stuart Russell, Gründer des größten KI-Labors der Welt, dem *Center for Human-Compatible Artificial Intelligence* (CHAI), der viele Regierungen zu Fragen der KI berät: „Wir sprechen nicht von Robotern, die die Welt übernehmen. Wir sprechen von neuen Massenvernichtungswaffen“, so zitiert im Interview mit Jan Christoph Wechmann vom 16. 3. 2023, online unter <https://www.stern.de/digital/chatgpt-sieht-er-skeptisch-ki-experte-stuart-russell-im-gespraech-33259834.html>. Eine vom

Bostoner *Future of Life Institute* angestoßene freiwillige Selbstverpflichtung, nicht an autonomen Waffensystemen forschen zu wollen, haben zwar 2018 mehr als 2 500 Forscher:innen und 176 Organisationen unterzeichnet, darunter fehlen aber viele wichtige Player: „[...] So unterzeichneten weder *Google* noch die Google-Holding *Alphabet* die Selbstverpflichtung; auch *Boston Dynamics* – Hersteller berühmter Robotermodelle – finden sich bislang nicht auf der Liste. Ebenso fehlen einschlägige Rüstungskonzerne.“ So Jan Dönges: „Tausende KI-Forscher geloben Verzicht auf Waffenentwicklung“, in *Spektrum* vom 19. 7. 2018. Überhaupt dürften andere Akteur:innen, z. B. in Russland oder Nordkorea, hier geringere Skrupel haben. Bereits im November 2017 hatte das Institut in einem Video das fiktive Szenario einer Terroristengruppe, die mit massenhaft produzierten Minidrohnen nach ihren Gegnern fahndet und diese vollautomatisch exekutiert, thematisiert (online unter https://www.youtube.com/watch?v=HipTO_7mUOW).

⁷³ Vgl. dazu Albrecht, ChatGPT (2023), S. 11 u. S. 31 (Konzentration auf wenige Big Player).

⁷⁴ Von der ästhetischen Fiktion abzugrenzen ist in jedem Fall der ebenso traditionsreiche und gleichfalls auf die Antike zurückgehende Begriff einer juristischen Fiktion. Alltagssprachlich bezeichnet „Fiktion“ zudem alles, das als nicht tatsächlich markiert werden soll im Sinne von z. B. „das ist nicht wirklich“. Im pejorativen Sinn („das ist bloß eine Fiktion“), im Poststrukturalismus bei Lacan, Derrida, Baudrillard u. v. a. m. auch durchaus positiv, taucht der Begriff häufig als Synonym für Trugbild, Ideologem, Simulakrum oder sonst etwas Täuschendes auch in der Philosophie auf, was bis zur *panfiktionalen* Verwendung geht, für die dann praktisch alles – als letztlich eben doch nicht wahr – als „Fiktion“ dekonstruiert werden kann. Dieser Sprachgebrauch ist epistemologisch redundant und über die implizite Analogisierung zur juristischen bzw. ästhetischen Fiktion auch irreführend.

⁷⁵ Die etwas avanciertere medientheoretische Frage, ob Traumbilder und/oder reine Vorstellungsbilder *überhaupt* a-medial sind, lasse ich hier offen; sofern sprachlich verfasste Proposition hinzutreten, ist ein Gedanke bzw. ein Traumfragment sicher medial. Ähnlich argumentierte bereits Stanley Cavell: „Welt durch die Kamera gesehen“, in: Dieter Henrich; Wolfgang Iser (Hg.): *Theorien der Kunst*. Frankfurt a. M.: Suhrkamp 1992 [1974], S. 447-490, hier S. 478. Er wehrt sich hier dezidiert gegen die v. a. psychoanalytische Analogisierung von Film und Traum, die auch bei Christian Metz anklingt. Insofern lägen zwei Fiktionen vor, eine der Rahmenhandlung und eine der Traumsequenz, die auch *kein* Traum ist.

⁷⁶ Das Gesicht (der Schauspieler:innen) ermöglicht als Medium, *fiktive* Gefühlsregungen auszudrücken, wobei diskutiert werden könnte, ob dies als eigenständige Fiktion (und nicht Teil einer solchen, eines Films oder Theaterstücks oder inszenierten Fotografie) aufgefasst werden kann.

⁷⁷ Viele der eher als Spielerei generierten Bilder gängiger Programme wie *Midjourney* erinnern aber stark an diese phantastischen Vorbilder der Kunst und sind insofern rezeptiv als fiktionale Bilder anzusehen, d. h. als Bilder, die als „erfundene Bilder“ ihren Betrachter:innen einen fiktionalen Kosmos offerieren, in den mit Marie-Laure Ryan „rezentriert“ werden kann. Vgl. dazu etwa Marie-Laure Ryan: „Fiktion, Kognition und nichtverbale Medien“, in: Gertrud Koch; Christiane Voss: (Hg.): „Es ist, als ob“. Fiktionalität in Philosophie, Film- und Medienwissenschaft. München: Fink 2009, S. 69-86, hier S. 86. Entscheidend ist für sie dabei das *recentering* als das Sich-in-eine-imaginäre-Welt-Hineinversetzen, eine Art Weltensprung.

⁷⁸ So werden etwa in Recherchen wissenschaftliche Papers aufgelistet, die eine bestimmte Koryphäe durchaus geschrieben haben könnte, aber eben nicht geschrieben hat (Kevin B. Lee machte einen solchen Vorfall am 30. 4. 2023 auf *Facebook* öffentlich, wo auf Anfrage eines Studierenden auf ein Paper über den Videoessay hingewiesen wurde, das er gar nie verfasst hat). In einem Selbstversuch am 2. 5. 2023 wurde bei zwei von drei Fragen – „Wer ist Jürgen Riethmüller?“, „Wer ist Johann Adolph Dori?“, „Was ist Honey Tower?“ – halluziniert; nur im letztgenannten Fall wurde korrekt angegeben, dass ChatGPT das Gesuchte, eine eher unbekannte Elektromusikerin, nicht kennt; in den ersten beiden Fällen wurde kurzerhand eine Fantasieperson samt Biographie erfunden. Drei Monate später wurde auf alle drei Fragen die nichtssagende, aber immerhin korrekte Antwort gegeben, das System kenne die Person nicht.

⁷⁹ Ob ein Algorithmus *genauso* kreativ sein kann, wie ein Mensch, wie immer wieder zu lesen ist, sei einmal dahingestellt. Der Punkt scheint eher, dass er es auf eine ganz andere Weise ist. Über mögliche Folgen spekuliert in seinem Essay Hanno Rautenberg: *Die Kunst der Zukunft: Über den Traum von der kreativen Maschine*. Berlin: Suhrkamp 2021. Die zugehörigen Werke von Mathematikern wie etwa Marcus du Sautoy: *Der Creativity-Code. Wie künstliche Intelligenz schreibt, malt und denkt*. München: C. H. Beck 2021 krankten an einem sehr konventionellen Verständnis von Kunst und einem allzu weiten Begriff der Kreativität, der locker auch Schachcomputer inkludiert. Aus der Masse der zuletzt erschienen Titel über Kunst und KI ragen gerade von hier aus gesehen nur ganz wenige heraus, etwa Arthur I. Miller: *The Artist in the Machine. The World of AI-powered Creativity* Cambridge, Mass. MIT Press 2019; Dorothea Winter: *KI, Kunst und Kitsch: Ein philosophischer Aufreger*. Heidelberg: Carl Auer 2023; Merzmensch (Vladimir Alexeev): *KI-Kunst. Kollaboration von Mensch und Maschine*. Berlin: Klaus Wagenbach 2023. Der (auch hier) gern gezogene Vergleich mit der Musik ist diesbezüglich womöglich irreführend: Die kreative Potenz generativer KI in den verschiedenen Künsten ist ausgesprochen unterschiedlich einzuschätzen; selbst nur im Bereich der Musik ist die KI z. B. eine bessere Komponistin als sie Kompositionen interpretieren kann. Vgl. zu dieser mutmaßlich verfrühten Diskussion Harald Willenbrock: „Künstliche Intelligenz und Kunst: Wann ist Kunst eigentlich Kunst?“ (2021), online unter <https://www.goethe.de/prj/zei/de/art/22094490.html>.

⁸⁰ Chat GPT ist insofern in diesem Kontext am ehesten ein beachtliches Fabuliertool. Viel diskutiert wird der Einsatz durch Studierende und Forschende beim Erstellen von Hausarbeiten und Papers. Hierzulande hat sich etwa die Uni Tübingen sich Anfang 2022 genötigt gesehen, den Einsatz stark einzuschränken. Überall werden nun mehr oder minder restriktive KI-Policies der einzelnen Hochschulen und Fakultäten verfasst.

⁸¹ Es gibt im Netz diverse Tutorials, wie man Comics per KI generieren kann, meist als schlichte Kombination von ChatGPT und Dall-E, z. B. <https://towardsdatascience.com/using-ai-to-create-new-comic-strips-without-writing-any-code-cc669bb317a7>, wobei Ersteres derzeit noch deutlich besser funktioniert. In den USA gibt es erste Tendenzen bei den Verleger:innen in Richtung einer Kennzeichnungspflicht bzw. freiwilliger Selbstverpflichtung der Unterlassung einer Publikation (teilweise) KI-generierter Comics.

⁸² Beispielsweise <https://huggingface.co/spaces/dalle-mini/dalle-mini>.

⁸³ Dazu Sebastian Meineck: „Der Anfang von etwas großem“ (2022), online unter https://netzpolitik.org/2022/text-zu-bildgenerierung-der-anfang-von-etwas-groessem/?utm_source=pocket-newtab-global-de-DE. Zumindest Dall-E basierte auf der (stärkeren) Textseite bereits 2022 auf dem „Generative Pre-trained Transformer 3“ (GPT-3).

⁸⁴ So nachzulesen bei Melissa Heikkilä; Will Douglas Heaven: „Was uns 2024 in der Künstlichen Intelligenz erwartet“ (2024), online unter https://www.heise.de/hintergrund/Was-uns-2024-in-der-Kuenstlichen-Intelligenz-erwartet-9590341.html?utm_source=pocket-newtab-de-de.

⁸⁵ Roland Meyer: „Es schimmert, es glüht, es funkelt – Zur Ästhetik der KI-Bilder“, in: 54Books vom 21. 3. 2023, online unter <https://www.54books.de/es-schimmert-es-gluet-es-funkelt-zur-aesthetik-der-ki-bilder/>.

⁸⁶ Daniel Rubinstein: Fotografie nach der Philosophie. Repräsentationsdämmerung. Übersetzt von Tom Wohlfarth. Berlin: Merve 2020, S. 72.

⁸⁷ Dazu Robert Strasser: KI-Kunst. Die neue Ära der Bildgenerierung. Wroclaw: o. V. 2023, S. 1-4; Engenhardt; Löwe, Design und künstliche Intelligenz. Theoretische und praktische Grundlagen (2022), S. 38-44.

⁸⁸ Vgl. Aditya Ramesh; Prafulla Dhariwal; Alex Nichol; Casey Chu; Mark Chen: „Hierarchical Text-Conditional Image Generation with CLIP Latents“ (2022), online unter <https://arxiv.org/abs/2204.06125>, S. 1-27, hier S. 4: „Autoregressive (AR) prior: the CLIP image embedding z_i is converted into a sequence of discrete codes and predicted autoregressively conditioned on the caption y . Diffusion prior: The continuous vector z_i is directly modelled using a Gaussian diffusion model conditioned on the caption y .“

⁸⁹ „Our approach allows us to encode any given image x into a bipartite latent representation ($z_i; x_T$) that is sufficient for the decoder to produce an accurate reconstruction. The latent z_i describes the aspects of the image that are recognized by CLIP, while the latent x_T encodes all of the residual information necessary for the decoder to reconstruct x .“ (Ebd., S. 6.)

⁹⁰ Ebd., S. 4. Vgl. auch bereits im Abstract des Beitrags: „Given an image x , we can produce related images that share the same essential content but vary in other aspects, such as shape and orientation.“ Die Autor:innen stellen die Verbesserungen ihres Verfahrens gegenüber vorangegangenen Entwicklungen folgendermaßen heraus: „Our decoders conditioned on image representations can also produce variations of an image that preserve both its semantics and style, while varying the non-essential details absent from the image representation. [...]“

⁹¹ Einige Beispiele mit fiktiven Skifahrern ebd., S. 14.

⁹² Auch dieses Beispiel ebd., S. 2.

⁹³ Dieses Beispiel verweist am ehesten in Richtung eines tatsächlichen Nutzens, der v. a. im Design oder im Animationsbereich liegen dürfte.

⁹⁴ Vladimir Alexeev: „Kann Dall-E fotorealistische Bilder erzeugen?“, online unter: <https://the-decoder.de/kann-dall-e-2-fotorealistische-bilder-erzeugen/>. Vgl. auch Devin Coldewey: „OpenAI’s Dall-E creates plausible images of literally anything you ask it to“, online unter: <https://techcrunch.com/2021/01/05/openais-dall-e-creates-plausible-images-of-literally-anything-you-ask-it-to>.

⁹⁵ Vgl. dazu die Erläuterungen unter <https://openai.com/blog/dall-e/>.

⁹⁶ Meyer, Es schimmert (2023). [Herv. J. R.] Mit Blick auf den zugehörigen Remediation-Prozess seit 1839 ist dies wahrlich ein interessanter Epilog.

⁹⁷ Interessant an genanntem Tool ist, dass die Entwickler:innen sehr detailliert ihre Vorgehensweise bezüglich des Konzepts der verwendeten neuronalen Netzwerke oder auch dahingehend, wie die KI trainiert wurde usw. vorstellen – <https://wandb.ai/dalle-mini/dalle-mini/reports/Dall-E-Mini-Explained-with-Demo--Vmlldzo4NjlxODA#the-dall-e-experiment>. Vgl. auch die Vorstellung des eigentlichen Programms <https://openai.com/dall-e-2/>.

⁹⁸ Die eine KI ist dann trainiert darauf, perfekte Prompts für die andere, etwa Midjourney oder Dall-E, zu generieren.

⁹⁹ Gleich also, wie viel von den Träumen eines Marvin Minsky oder Hans Moravec gerade Realität wird; zu hoffen scheint, dass die künftigen Forscher:innen im Bereich des *affective computing* sich an Moravec erinnern, wenn er schrieb: „Ich sehe diese Maschinen als unsere Nachkommen. Im Augenblick glaubt man das kaum, weil sie eben nur so intelligent sind wie Insekten. Aber mit der Zeit werden wir das große Potential erkennen, das in ihnen steckt. Und wir werden unsere neuen Roboterkinder gemhaben, denn sie werden angenehmer sein als Menschen. Man muss ja nicht all die negativen menschlichen Eigenschaften, die es seit der Steinzeit gibt, in diese Maschinen einbauen. Damals waren diese Eigenschaften für den Menschen wichtig. Aggressionen etwa brauchte er,

um zu überleben. Heute, in unseren großen zivilisierten Gesellschaften, machen diese Instinkte keinen Sinn mehr. Diese Dinge kann man einfach weglassen ...“ und zudem bedenken, dass die so oft vergessene Scham seitens der KI-Designer:innen der Zukunft nicht vergessen wird – denn ohne den genuin sozialen Affekt wären diese „Kinder“ vermutlich ziemlich unsoziale Gesellen, vor denen sich auch ohne Aggressionsfähigkeit das Fürchten lohnte. So Hans Moravec: Computer übernehmen die Macht. Vom Siegeszug der künstlichen Intelligenz. München: Hoffmann & Campe 1999, S. 136. Über den aktuellen Stand des *affective computing* siehe die sehr umfangreiche Literatur des Forschungsbereichs am MIT unter <http://affect.media.mit.edu/publications.php>. Gegenwärtig sind die Hoffnungen diesbezüglich eher vage, da bereits bei der Messung menschlicher Emotionen über Tracking von Mimik und Stimme und beim dahinterstehenden Emotionsbegriff nicht selten eher fragwürdige Konzepte verfolgt werden; Minsk selbst etwa hat behauptet, Emotionen seien „not especially different from the processes that we call ‚thinking‘“, was einerseits richtig ist, aber eben auf der psychologischen Erkenntnisstufe Descartes' festhängt. So zit. in: Richard Restak: „Mind over Matter“ (2006), online unter <http://www.washingtonpost.com/wp-dyn/content/article/2006/12/14/AR2006121401554.html>.

¹⁰⁰ Diese sollen hier als Beispiele nicht Gegenstand sein, wo es nur um die Fälle gehen soll, da die KI als „Autor:in“ oder wenigstens als so etwas wie ein Co-Autor:in in Erscheinung tritt, ein auch vor 2022 schon ab und an beleuchtetes Feld, wobei nach der öffentlichkeitswirksamen Versteigerung des „Portrait of Edmond Belamy“ 2018 viele vor allem einmal an ganz Unterschiedlichem arbeitende Künstler:innen als Beispiele herangezogen wurden. Vgl. dazu die Richtigstellung von A. Hertzmann: „Computers do not make art, people do“, in: *Communications of the ACM* 63, 5 (2020), S. 45-48.

¹⁰¹ Die verwandten Fragen „Kann KI Gestaltung?“ oder „Kann KI Architektur?“ usw. klammere ich aus, weil man diesbezüglich jeweils zu einer definitiv anderen Antwort wird kommen müssen.

Bereits Dieter Mersch: „Kreativität und künstliche Intelligenz. Einige Bemerkungen zu einer Kritik algorithmischer Rationalität“, in: *ZfM* 21, 2 (2019), S. 65-74, hier S. 73 f., wies im Fazit seines Beitrags darauf hin, dass die oftmals euphorischen Einschätzungen von Informatiker:innen bezüglich der „künstlerischen“ Fähigkeiten der KI von einem naiven Kunstbegriff ausgehen. Das gilt auch für viele der KI-basierten quantitativen Untersuchungen von Kunstwerken mit kunsthistorischem Bezug, über die etwa Eva Cetinic; Jams She: „Understanding in Creating Art with AI: Review and Overlook“ (2021), online unter <https://arxiv.org/pdf/2102.09109.pdf>, S. 1-17 berichten. So kann die KI selbstverständlich bspw. so etwas wie den Grad der Repräsentativität eines einzelnen Werks für den einem Künstler:in zuberechneten Stil messen (als Fall automatischer Klassifikation) oder auch diverse etwas originellere *latent features* aller surrealistischer Kunstwerke aufdecken und ggf. benennen, wobei sich hier durchaus die Sinnfrage stellt, zumal im Hintergrund auch immer die alten Irrtümer der mathematischen Theorie der Information oder gar der sogenannten „exakten Ästhetik“ etwa bei George D. Birkhoff oder Max Bense (oder, für die Literatur und besonders schräg, Wilhelm Fucks) virulent sind. Ein tatsächlicher Nutzen besteht am ehesten in der Verwaltung großer Bestände, z. B. in der Programmierung von „smart retrieval systems that can help organize and analyze large collections of artworks in an efficient way“, so ebd. S. 4, die dann als multimodale Systeme sowohl die Bild- als auch die Beschreibungsebene inkludieren und so z. B. alle Bilder mit Schafen darauf im Bestand identifizieren könnten. Ob derlei tatsächlich etwas zum Verständnis von Kunst beitragen kann, wird sich weisen müssen.

¹⁰² Hegel, Vorlesungen über die Ästhetik, Band 1, S. 48.

¹⁰³ Theodor W. Adorno: Ästhetische Theorie. Gesammelte Schriften, hg. Rolf Tiedemann, Bd. 7. Frankfurt a. M.: Suhrkamp 2003, S. 134.

¹⁰⁴ Christoph Menke: Die Kraft der Kunst Berlin: Suhrkamp 2013, S. 17 f.

¹⁰⁵ Albrecht Wellmer: Versuch über Musik und Sprache. München: Hanser 2009, S. 134.

¹⁰⁶ Mit dieser Betrachtung steigt Friedrich, Degenerierte Vernunft (2023) in seine Überlegungen ein. Er übersieht hier, dass die KI gerade das nicht kann: ein Werk produzieren. Sie kann nur etwas produzieren, was unter Umständen, die gegenwärtig in der Regel noch nicht gegeben sind, ein Werk simuliert.

¹⁰⁷ Menke, Kraft der Kunst (2013), S. 24 ff., hier auch eine Zusammenfassung der bekannten poetischen Reflexionen Valéry's.

¹⁰⁸ Interessant in diesem Zusammenhang ist sicher auch die anhaltende Diskussion, ob Duchamp überhaupt der Autor hinter ‚Fountain‘, dem meistgenannten Beispiel, ist – oder aber Elsa von Freytag-Loringhoven die Autorin. Vgl. dazu etwa die Anmerkungen Philipp Meiers in der NZZ, online unter <https://www.nzz.ch/feuilleton/marcel-duchamps-fountain-hinter-pissoir-steckt-eine-frau-ld.1777480>: „Ob ‚Fountain‘ wirklich von Marcel Duchamp stammt, wird wohl nie geklärt werden können.“

¹⁰⁹ Etwa für die zuletzt genannten Felder: Sicher sind die literarischen Versuche der KI nicht 1:1 verwendbar; professionelle Autor:innen können aber ChatGPT bei der Arbeit als Dialogpartner nutzen, um auf diese Weise vielerlei Probleme vor dem finalen Text zu lösen; so können etwa Thrillerautor:innen fragen, was im Kommandozentrum eines Atomkraftwerkes oder im Cockpit eines neuen Airbus an täglichen Routinen zu tun ist, was unter gar keinen Umständen passieren oder vergessen werden darf etc., oder man spielt diverse Varianten ein und desselben Dialogs durch, oder lässt das problematische Passagen des eigenen Werks kommentieren und auf Konsistenz, sachliche Richtigkeit etc. prüfen.

¹¹⁰ Das Comicbeispiel stammt von meiner Kollegin Barbara M. Eggert.

¹¹¹ Insgesamt wird für die sogenannte „AI art“ derzeit noch ein beträchtlicher *gender gap* konstatiert, vgl. dazu etwa den Blogeintrag vom 12. 9. 2023 „Female artists in AI: Pioneering a bright new Frontier?“, online unter <https://www.krautart.de/female-artists-in->

subjektlos gedachten Informationsübertragung, auf dem das Konzept, kommunikationstheoretisch gewendet, immer noch besteht, argumentativ prekär. Anhänger:innen dieser Thesen, die inzwischen oft mit posthumanen Konzepten von Barad, Haraway und Latour usw. bezüglich nichthumaner Akteure/Aktanten kombiniert werden, definieren dann, wie z. B. Konstanze Schütze: „Bildlichkeit nach dem Internet – Kunstvermittlung am Bild als Gegenwartsbewältigung“, in: Zeitschrift Kunst Medien Bildung | zkmb – Oktober 2019: Postdigital Landscapes (hg. Kristin Klein, Willy Noll), online unter: <http://zkmb.de/bildlichkeit-nach-dem-internet-kunstvermittlung-am-bild-als-gegenwartsbewaeltigung/>, S. 2, das Bild als „geschäftsführende Entität in Verbreitung“. Schütze erläutert: „Das geschäftsführend beziehe ich dabei auf die von Torsten Meyer in die Kunstpädagogik eingeführte und auf den Soziologen und Kulturtheoretiker Dirk Baecker zurückgehende These: Die *nächste Kunst* sei die Kunst der nächsten Gesellschaft und diese nächste Gesellschaft basiere auf dem Computer als geschäftsführender Medientechnologie [...]. Diese Überlegungen schließen nahezu nahtlos an Fragen einer agentiellen Bildlichkeit an und lassen sich produktiv wenden, indem *Bilder generell als in Verbreitung* begriffen werden. Die soziologisch und anthropologisch informierte Memtheorie nach Richard Dawkins, Susan Blackmore und Limor Shifman kennt Vorgänge ‚kultureller Übertragung‘ [...], die sich nicht zuletzt auch an den jeweiligen Artefakten (z.B. Bildern) festmachen lassen.“

¹²⁹ Die Erfahrungen während der Pandemie mit virtuellen Partys, die kaum angenommen wurden, könnten auf eine gewisse Persistenz überkommener Schemata hindeuten. Voraussetzung einer Spaltung etwa in ein KI-generiertes Low und ein weitgehend KI-freies High wäre eine Kennzeichnungspflicht.

Kurzbiografie

Dr. Jürgen Riethmüller, lehrt nach einer mit dem Förderpreis für junge Wissenschaftler und Künstler (Rotary-Club) ausgezeichneten Dissertation über die Anfänge des demokratischen Denkens in Deutschland seit 2002 als Wissenschaftlicher Mitarbeiter für Kulturwissenschaften, Kunstgeschichte und Text an der Merz Akademie in Stuttgart. Diverse Publikationen zu Ästhetik, Medien und Politik, u. a. »Die Aufgabe der Ästhetik« (2008) und zu Fragen der Wissensgesellschaft »Der Graue Schwan« (2012) erschienen Wilhelm Fink Verlag, sowie zuletzt »Kalkül der Scham« (2019) im Kadmos Verlag. Aktuell arbeitet er an einer größeren Abhandlung zu den Perspektiven transmedialer Fiktionstheorie.

Impressum

Dr. Jürgen Riethmüller: „Was tun mit einem „stochastischen Papagei“, der malen, dichten und komponieren kann? Generative KI und die Künste: Eine kulturtheoretische Perspektive“, Mai 2024, alle Rechte vorbehalten.

Hrg. Dr. Barbara M. Eggert, Merz Akademie, Hochschule für Gestaltung, Kunst und Medien, Stuttgart, Teckstraße 58, 70190 Stuttgart; www.merz-akademie.de