



MASTERARBEIT | MASTER'S THESIS

Titel | Title

<author> = Error 404 – author not found </author>
Zwischen Natürlichkeit und Künstlichkeit – Zur Untersuchung
der Autorschaftskonzeptionen von Künstlicher Intelligenz

verfasst von | submitted by
Nadine Vejnaska BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 817

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Deutsche Philologie

Betreut von | Supervisor:

Mag. Dr. Thomas Eder

Inhaltsverzeichnis

1	Vorwort.....	1
2	Schöne neue Technikwelt.....	6
2.1	Technisches Menschenbild.....	8
2.1.1	Turings Erbe	10
2.1.2	Imitation vs. Duplikation.....	12
2.1.3	Alte Versprechen	15
2.2	Menschliches Technikbild.....	22
2.3	Konvergenz der Intelligenzen.....	25
3	Kreativ sein, bedeutet Mensch sein	29
3.1	Keine Kreativität ohne Grenzüberschreitung	33
3.1.1	Das Ende der Kunst durch KI?	38
3.2	Das kreative Genie.....	43
4	„Der Autor“ wird vom Thron gestoßen.....	46
4.1	Das Digitale der digitalen Literatur.....	51
4.1.1	Schreiben ohne Kreativität	54
4.1.2	Schreiben mit Computern.....	56
4.2	Zwischen Mensch und Technik: Die Akteur-Netzwerk-Theorie	60
5	Autorschaft und Künstliche Intelligenz – eine Analyse	70
5.1	Kenric Allado-McDowell	70
5.1.1	<i>Pharmako-AI</i> – Im Dialog mit der KI	72
5.1.2	<i>Amor Cringe</i> – Transformation zur Mensch-Maschine.....	81
5.1.3	<i>Air Age Blueprint</i> – Die andere Intelligenz	87
5.2	Hannes Bajohr	95
5.2.1	<i>(Berlin, Miami)</i> – Token für Token	96
5.2.2	Bajohrs Modell Kausaler Autorschaft	110
6	Zusammenfassung und Ausblick.....	116
7	Literaturverzeichnis	131
7.1	Primärliteratur.....	131
7.2	Sekundärliteratur	131

7.3	Internetquellen	137
7.4	Videos.....	139
7.5	Podcasts	140
7.6	Weiterführende Literatur	140
8	Abbildungsverzeichnis	141
9	Anhang	142
9.1	Abstract.....	142

1 Vorwort

Die letzten Jahre haben gezeigt, wie schnell generative Sprachmodelle wie ChatGPT, Bard oder Meta AI die Technikwelt verändern können. Der Hype um Künstliche Intelligenz¹ hat trotz vieler Fehler (noch) nicht nachgelassen und steuert ungehindert darauf zu, KI in weite Teile unseres Alltags und Arbeitslebens zu integrieren. In den Human- und Geisteswissenschaften fällt die Euphorie über die hocheffizienten Textgeneratoren verhalten aus. Denn oft ist nicht mehr zu unterscheiden, ob es sich bei einem Text um einen von einem Menschen geschriebenen oder um von einer Maschine generierten Text handelt, wodurch sich juristisch-ökonomische, hermeneutische und ästhetische Fragestellungen und Problematiken ergeben. Es werden grundlegende Fragen über Autorschaft und die Sinnkonstitution in den generierten Texten aufgeworfen sowie traditionelle Kategorien der Intention, Autorität, Authentizität, Verantwortung und Originalität auf den Prüfstand gestellt. Im Diskurs über Künstliche Intelligenz wird auf dem Feld der Kunst nach dem grundlegenden Wert von Sprachmodellen für die Textproduktion gefragt und im Kunst- und Literaturbetrieb die Regulierung ihres Einsatzes gefordert.² Künstler:innen wie Autor:innen sehen sich in ihrer Position und ihrem Beruf gefährdet – eine Unterscheidung zwischen menschlichem und künstlichem Schaffen steht im Mittelpunkt der Diskussion.

Der Stellenwert menschlicher Autorschaft und die traditionelle Vorstellung der Autor:in als Inhaber:in eines Textes und sinnbildende Kraft werden in der Literaturwissenschaft aber nicht zum ersten Mal herausgefordert. Die Kritik der Poststrukturalisten löste den Text- und Schriftbegriff von der Autor:in. Der Text verweist

¹ Für die vorliegende Masterarbeit wird zur allgemeinen Verständlichkeit der umgangssprachliche und populärwissenschaftliche Begriff ‚Künstliche Intelligenz‘ bzw. ‚KI‘ (engl.: *Artificial Intelligence, AI*) als Oberbegriff für die diskutierten Technologien verwendet. Die damit einhergehenden Problematiken und Implikationen werden in der Arbeit kritisch beleuchtet. Generell fokussiert die Arbeit auf *large language models (LLM)*, die auf Basis von *Deep Learning*-Modellen und Künstlichen Neuronalen Netzen (KNN) arbeiten (s. Kapitel 2.1.3.1 *Exkurs: Künstliche Neuronale Netze*). Sie gehören zum Schema der ‚schwachen‘ Künstlichen Intelligenz, die Lösungen für bestimmte Problemstellungen erarbeiten können, aber nicht fähig sind, diese zu generalisieren. Die Unterteilung in eine ‚starke‘ und ‚schwache‘ Künstliche Intelligenz erfolgte durch John R. Searle (1980). Eine genauere Definition dazu folgt im Kapitel 2.1.2 *Imitation vs. Duplikation* | Vgl. John R. Searle: *Minds, brains and programs*. In: *The Behavioral and Brain Sciences* Vol. 3/3 (1980), S. 417.

² Vgl. Marc Stöckel: Urheberrechte. US-Gericht lehnt Copyright auf KI-generierte Kunst ab. In: *golem.de* (2023) <https://www.golem.de/news/urheberrechte-us-gericht-lehnt-copyright-auf-ki-generierte-kunst-ab-2308-176918.html> (Zugriff: 21.09.2023) || Vgl. Johannes Edelhoff/Esra Özer/Tina Soliman: Streit um Urheberrechte. Müssen Künstler und Autoren für KI entschädigt werden? In: *tagesschau* (2023) <https://www.tagesschau.de/wirtschaft/digitales/kuenstliche-intelligenz-grafiker-autoren-urheberrechte-100.html> (Zugriff: 21.09.2023).

bei Foucault und Barthes nicht mehr auf die Verfasser:in, sondern über sich selbst hinaus. Dieser wird als Grundlage und Produkt von Diskursen subjektlos und findet als ausgeführte Praxis, als ein Zirkulieren von Zeichen den Ursprung in sich selbst. Der dadurch propagierte ‚Tod des Autors‘ ist symbolisch gemeint und ersetzt die ideologische Figur der Autor:in durch ein Konzept von Intertextualität, das die Autorfunktion im Prozess der ästhetischen Sinngebung und ihr Verhältnis zum Werk relativiert.³ Die Techniken automatischer Texterzeugung wiederum lassen den ‚Tod‘ der Autor:in nicht symbolisch erscheinen, sondern versprechen die menschliche Autorinstanz, die empirische Autor:in, vom Schreibprozess auszuschließen. Diese neuen Produktionsbedingungen, die den individuellen Anteil kreativer Leistung am produktiven Akt verringern bzw. (fast) vollständig aufzulösen drohen, führen aktuell zu einer diskursiven Umkehr. Als Gegenpol zur digitalen Erscheinung künstlicher Systeme wird die physische Präsenz der Autor:in zurückverlangt, die vor allem Fragen im juristisch-ökonomischen Kontext – nach der Verantwortung, dem geistigen Eigentum und der Regelung des Copyrights unter Einsatz von KI – lösen soll.

Literaturtheoretisch bleibt die Frage noch offen, ob ein Schreiben mit oder von Künstlicher Intelligenz zu einer Revision des Autor:innenbegriffs führt, denn aktuelle Diskussionen drehen sich um einen humanistischen Kampf gegen die Technologien und berufen sich auf grundlegende Unterschiede: Was Menschen sind und können – intentional, denk- und entscheidungsfähig, autonom – und welche Limitierungen Maschinen haben – fehlendes Weltverständnis, regelgebunden, binär begrenzt – anstatt empirische wie analytische Tools für den Umgang mit generativen Verfahren zu entwickeln.⁴ Der Diskurs um Künstliche Intelligenz sieht sich damit in zwei Formen von Fiktionen gefangen: „eine[m] Menschwerdungs- und eine[m] Bedrohungsdiskurs“.⁵ Demnach sind Veränderungen im Verständnis von Begriffen wie Intelligenz, Bewusstsein, Autonomie, Verstehen und Sprache festzustellen, die die Grenzen der bisher als kognitiv oder psychologisch kategorisierten Fähigkeiten auflösen und den Maschinen zugänglich machen möchten oder sie als Angriff auf anthropozentrische Werte deuten. Im Bereich der Kunst werden Kategorien wie die der Kreativität

³ Vgl. Michael Wetzel: Einleitung. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 10-15.

⁴ Vgl. Michel Callon/Bruno Latour: Don't Throw the Baby Out with the Bath School! A Reply to Collins and Yearley. In: Andrew Pickering (Hg.): Science as Practise and Culture. Chicago: The University of Chicago Press 1992, S. 360.

⁵ Martin Hennig/Kilian Hauptmann: Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung. In: Zeitschrift für Medienwissenschaft Vol. 2/21 (2019), S. 87.

anthropozentrisch verteidigt und als die ‚letzte Bastion‘ des Menschen gegen die Maschine stilisiert. Die Künste sind „das a priori Andere der Maschine“⁶ – die Maschine kann keine Kunst schaffen, weil sie nicht kreativ, ergo menschlich ist. Die Diskussion um das künstlerische Potenzial Künstlicher Intelligenz fällt hinter der Unterscheidung zwischen dezidiert künstlichen und menschlichen Merkmalen und einer anthropozentrischen und postromantischen Konzeption von Kunst zurück. Es werden vermeintlich unüberwindbare Gegensätze aufgemacht, deren Bewältigung die KI zu leisten hätte. „Die Maschine als Künstler ist offenbar nur mit dezidiert menschlichen Eigenschaften denkbar“⁷ – und unterliegt damit einer Vorstellung einer bisher technisch nicht möglichen ‚starken‘ Künstlichen Intelligenz⁸, die autonom künstlerisch tätig ist.

Es sind aber gerade jene literarischen Experimente mit generierten Texten, die nicht auf das Phantasma einer Maschine als Autor:in hereinfließen, sondern den Literatur- und Autorschaftsbegriff herausfordern. Elektronische bzw. digitale Literatur hatte in der germanistischen Forschung bisher vorwiegend den Status als Randphänomen, wodurch eine eigenständige Theoriebildung nur vereinzelt verfolgt wurde. Die vorliegende Masterarbeit lässt diesem Themenbereich nun vermehrt Aufmerksamkeit zukommen und soll einen Teil zur Auflösung des im Folgenden formulierten Forschungsdesiderats liefern.

Ziel ist es, einen Beitrag zur Positionierung Künstlicher Intelligenz im Bereich der Geisteswissenschaften zu leisten. Daraus ergeben sich folgende Forschungsfragen:

- Welche literaturtheoretischen Veränderungen und Herausforderungen bringt Künstliche Intelligenz? Wie verändert die Technologie unser Verständnis der Trichotomie Autor:in, Text, Leser:in?
- Ist eine Maschinenautorschaft möglich? Wenn ja, in welcher Art und Weise kann Künstliche Intelligenz Autorschaft beanspruchen – welche Problematiken ergeben sich daraus? Wenn nein, welche Implikationen entstehen aufgrund einer geteilten Autorschaft zwischen Mensch und Maschine?
- Welche methodischen Möglichkeiten erlauben eine Beschäftigung mit Künstlicher Intelligenz auf dem Feld der digitalen Literatur. Welche theoretischen Modelle lassen sich für die Literaturwissenschaft fruchtbar machen?

⁶ Stephanie Catani: »ERZÄHLMODUS AN«. Literatur und Autorschaft im Zeitalter künstlicher Intelligenz. In: Jahrbuch der Deutschen Schillergesellschaft Vol. 64 (2020), S. 291.

⁷ Ebd., S. 291.

⁸ Vgl. Searle: Minds, brains and programs (1980), S. 417.

Diese Fragen sollen anhand der Untersuchung der Autorschaftskonzeptionen der Romane „*Pharmako-AI*“, „*Amor Cringe*“ und „*Air Age Blueprint*“ von K Allado-McDowell sowie des Werkes „*(Berlin, Miami)*“ von Hannes Bajohr beantwortet werden. Alle vier Texte sind der digitalen Literatur zuzuordnen und wurden mithilfe der generativen Sprachprogramme ChatGPT sowie GPT-J und GPT-NeoX verfasst.

Vor der Analyse der Romane widmet sich die Arbeit in einer theoretischen Einführung den vorher erwähnten Kategorien und ihrer Diskussion im Kontext Künstlicher Intelligenz. Dabei wird sowohl die Geschichte konnektionistischer Technologien als auch ihr Einfluss auf anthropologische Modelle näher beleuchtet, um eine kritische Perspektive für die Begriffe Intelligenz und Bewusstsein anzubieten. Ein kurzer Exkurs zu Künstlichen Neuronalen Netzen (KNN) gibt einen Einblick in die Funktionsweise der Technologie, um Wissensmängeln um die technischen Besonderheiten und daraus resultierenden Mystifizierungen entgegenzuwirken. Dieser Teil der Arbeit dient dazu, die bisherige Diskussion über Künstliche Intelligenz weg von einem Menschwerdungs- und Bedrohungsdiskurs hin zu einer theoretischen Thematisierung digitaler und insbesondere generativer (Code-)Literatur⁹ zu führen. Dadurch sollen die Möglichkeit neuer Formen von Literatur offengelassen und bisherige Normen von Textproduktion und -rezeption infrage gestellt werden.

Im Anschluss wird die Diskussion zu ästhetischen Fragen übergeleitet und der Autordiskurs anhand der Kategorien Kreativität und Originalität, am romantischen Genie und der Frage danach, ob Künstliche Intelligenz überhaupt Kunst produzieren kann, bearbeitet. In der Vorstellung von so unterschiedlichen Ansätzen poststrukturalistischer versus kybernetischer und behavioristischer Denkmodelle wird der Grundstein für die historische Untersuchung von Autorschaft im digitalen Raum gelegt. In der Tradition unkreativer Schreibweisen, konzeptueller Literatur und generativer Poesie stehend, definiert die generative Ästhetik ein neues Schreiben, das es noch theoretisch einzuordnen gilt. Vorgestellt werden unter anderem theoretische Überlegungen von Hannes Bajohr zu einer Unterscheidung generativer Literatur in ein sequenzielles und ein konnektionistisches Paradigma – basierend auf einer produktionsästhetischen Differenzierung – sowie die Konzeption von Autorschaft als ein verteiltes Netzwerk

⁹ Bajohr (2021) definiert generative Literatur in ihrer grundlegendsten Form als Ergebnis einer „automatische[n] Produktion von Text nach vorgegebenen Parametern, die in der Regel einer kombinatorischen, manchmal auch aleatorischen Logik folgt. Sie betont eher die Produktion als die Rezeption des Werks. [...] Sie scheint ihre technische Grundlage bewusst zu reflektieren.“ | Hannes Bajohr: Schreibenlassen. Texte zur Literatur im Digitalen. Berlin: August 2022, S. 140-141. || Eine genauere Definition folgt in Kapitel 4.1.2 *Schreiben mit Computern*.

anhand der Akteur-Netzwerk-Theorie (ANT). Da Künstliche Intelligenz ein komplexes Zusammenspiel aus menschlicher Einflussnahme und technischer Leistung ist, soll der multiperspektivische Ansatz der ANT einen Theorierahmen zur Beschreibung bieten, ohne sich auf Dualismen oder einen Allgemeingültigkeitsanspruch festzulegen. Die Möglichkeit eines Sowohl-Als-Auch in der Beschreibung verschiedener Autorschaftskonzeptionen für Künstliche Intelligenz lässt den Blick offen für die unterschiedlichen multimedialen und grenzüberschreitenden Produktionsprozesse, die digitale Literatur mit sich bringt.

Die Untersuchung der Romane konzentriert sich auf die Darstellung der Bedingungen ihrer Produktion, da dieser Aspekt in der literaturtheoretischen Diskussion um digitale Literatur bisher vernachlässigt wurde, und seit den Anfängen der KI-Forschung eine grundlegende Neuausrichtung stattfand, die einen fundamentalen Unterschied für den Umgang mit generierten Texten konstatiert.¹⁰ Die Werke werden aber auch auf inhaltlicher Ebene analysiert, um die Fähigkeit der Künstlichen Intelligenz am Modus des Erzählens zu testen sowie ihre technischen Affordanzen anhand des Textes selbst aufzuzeigen. Die Wahl der beiden Autor:innen erfolgte aufgrund ihres unterschiedlichen Umgangs im Schreiben mit Künstlicher Intelligenz. Während K Allado-McDowell für seine:ihre Werke einen *AI-Co-Writing*-Ansatz verfolgt, bei dem das Sprachmodell entweder eine eigene Stimme erhält oder in Kombination mit Allado-McDowell vollständig zu einer Mensch-Maschinen-Autor:in verschmilzt, bevorzugt Hannes Bajohr die Grenzen der KI in einer klar hierarchisierten Ko-Kreation auszuloten. Sein Roman „(Berlin, Miami)“ ist als literarisches Experiment gedacht, das auf seinen konzeptuellen Entscheidungen beruht, aber ohne Künstlicher Intelligenz nicht hätte verfasst werden können.

Zuletzt wird deshalb noch kurz Bajohrs Modell einer Kausalen Autorschaft für digitale Literatur aufgezeigt und anhand der Romane besprochen. Finale Gedanken dazu und ein Ausblick für die weitere Untersuchung Künstlicher Intelligenz in der Literaturwissenschaft runden die Arbeit ab.

¹⁰ Hannes Bajohr: Writing at a Distance: Notes on Authorship and Artificial Intelligence. In: German Studies Review Vol. 47/2 (2024b), S. 321.

2 Schöne neue Technikwelt

Intelligent ist, was intelligent zu sein scheint.¹¹

Seit Jahren befinden sich Algorithmen auf dem Vormarsch und sind in den Alltag eingezogen. Von nützlichen Funktionen, Navigationssystemen auf Smartphones und in Autos, Übersetzungsprogrammen wie DeepL, bis hin zu personalisierten Werbeanzeigen auf Social-Media-Plattformen haben sie in Form von Künstlicher Intelligenz (KI, *Artificial Intelligence*) mit dem Chatprogramm ChatGPT für große Aufmerksamkeit gesorgt. Mit der Veröffentlichung der open-Source-Software im November 2022 hat das amerikanische Softwareunternehmen OpenAI, Inc. den Hype um die Entwicklung von KI erneut entfacht und medienpräsent unter öffentliche Augen gestellt.¹² Die Entwickler von OpenAI, darunter Sam Altman, Greg Brockmann und Ilya Sutskever, sowie der damalige Investor und Multimilliardär Elon Musk, CEO von Tesla, zogen mit dem Sprachprogramm und dessen rhetorisch überzeugenden und natürlich klingenden Antworten im Dialogformat weite Kreise, nicht nur in der Technologie-Branche. Dem alten Traum einer vom Menschen entwickelten, künstlich geschaffenen Intelligenz¹³

¹¹ Stefan Betschon: Künstliche Intelligenz. Zwischen Realitäten und Illusionen. Zürich: Vontobel-Stiftung 2019 (Schriftenreihe 2320). S. 22.

¹² Das Sprachmodell ist frei zugänglich und kann dazu genutzt werden, automatisiert Texte zu schreiben, Codes zu programmieren, Vorschläge für kreative Projekte zu generieren, Reisetrips zu planen usw. Der Chatbot basiert auf den von OpenAI entwickelten *Generative Pretrained Transformer (GPT)* – eine Künstliche Intelligenz, die anhand von vorher gelernten Trainingsdaten Antworten und Texte generiert. | Vgl. OpenAI: Introducing ChatGPT. In: OpenAI (2022) <https://openai.com/blog/chatgpt> (Zugriff: 09.10.2023). | In der Zwischenzeit wurde das System mehrfach aktualisiert und bietet mit dem letzten Update auf die vierte Generation, GPT-4, nun eine Bezahlversion des Chatbots, die zunächst User:innen von ChatGPT Plus zur Verfügung steht (Stand: 16.10.2023). Nach Angaben von OpenAI wurde die neue Version um Funktionen wie Bilderkennung und Sprachausgabe erweitert und sei nun „*more creative and collaborative than ever before. It can generate, edit, and iterate with users on creative and technical writing tasks, such as composing songs, writing screenplays, or learning a user’s writing style.*“ | OpenAI: GPT-4 is OpenAI’s most advanced system, producing safer and more useful responses, <https://openai.com/gpt-4> (Zugriff: 10.10.2023). | Für den Erhalt des Unternehmens und der Forschungsarbeit hat OpenAI, Inc. eine gewinnorientierte Tochtergesellschaft gegründet, OpenAI Global, LLC, die u.a. von Investments von Microsoft unterstützt wird. | Vgl. Holger Schmidt: Elon Musk. Sein Kampf um die Zukunft der Künstlichen Intelligenz. In: Frankfurter Allgemeine Zeitung (2023) <https://www.faz.net/aktuell/technik-motor/wie-elon-musk-einst-open-ai-mitgruendete-und-nun-seinen-eigenen-weg-sucht-19170650.html> (Zugriff: 12.10.2023).

¹³ Derartige Bemühungen, allwissende intelligente Maschinen oder künstlich geschaffenes ‚Leben‘ zu erschaffen, existieren seit der Antike: Die Figuren der Ars Magna vom mallorquinischen Theologen Ramon Llull – eine Art drehbarer ‚Papiercomputer‘ – sollten beispielsweise bei theologischen Fragen Antwort geben. Die mechanische Ente von Jacques de Vaucanson, bestehend aus 400 Einzelteilen, imitierte lebenswichtige Funktionen eines Entenkörpers, wie schnattern, mit den Flügeln schlagen oder über einen künstlichen Darm zu verdauen. Einen weiteren bekannten Versuch, menschliche Schöpfung zu übertreffen, stellt der ‚Schachtürke‘ vom österreichisch-ungarischen Hofbeamten Wolfgang von Kempelen dar. Der angeblich autonome Schachroboter diene allerdings der Täuschung und hatte im Inneren einen Menschen versteckt, der die Hebel bediente. | Vgl. Yasemin Keskinetepe/Dr. Anke Woschek

scheint man ein Stück näher gekommen zu sein. Doch seit den ersten Bemühungen um Künstliche Intelligenz, während der ‚Geburtsstunde‘ der KI, der Dartmouth-Konferenz im Jahre 1956 in den USA¹⁴, bleiben weiterhin entscheidende Fragen offen, die den Nebel um *Artificial Intelligence* nur verdichten und die auch ChatGPT nicht klären kann.

Die grundlegendste betrifft eine Definition von Intelligenz im Allgemeinen. Es ist schwer zu definieren, was Intelligenz überhaupt ist. Weder gibt es geeignete Messoptionen, noch ist Intelligenz in ihren vielfältigen Ausprägungen simpel zu veranschaulichen. Intelligenz ist für den Menschen nur intrinsisch wahrnehmbar und extrinsisch an intelligenten Handlungen beobachtbar. Es hat sich deshalb durchgesetzt, Intelligenz gemeinhin als Fähigkeit zur Problemlösung oder zur Erreichung von Zielen näher zu umschreiben. Damit umfasst sie im weiten Sinne die Kompetenz, Schlüsse zu ziehen, zu folgern und zu planen. Sie beschreibt die Ausrichtung an Zwecken und somit das vernunftgeleitete Denken und Handeln, die Rationalität. Künstliche Intelligenz ist früh mit dieser Fähigkeit, Probleme zu lösen, in Zusammenhang gebracht worden und hat damit ein Feld besetzt, das bisher menschliche Intelligenz erforderte. Der Mensch als „animal rationale“¹⁵, hat das rationale Denken als Alleinstellungsmerkmal eingeübt, welches als Kategorie für eine Unterscheidung zu künstlicher nicht mehr hinreichend ist.¹⁶ Auch die Untergliederung der Intelligenz in eine emotionale Intelligenz, als explizit menschliche Fähigkeit, Emotionen zu verstehen, selbst wahrzunehmen und Empathie zu entwickeln,¹⁷ lässt sich zwar dazu verwenden, menschliche Intelligenz näher zu beschreiben, allerdings bietet diese keinen festen Boden für eine Abgrenzung zur

(Hg.): Künstliche Intelligenz. Maschinen lernen Menschheitsträume. Göttingen: Wallstein 2021, S. 35; 39. | Vgl. Cisek: Machtwechsel der Intelligenzen (2021), 7-9.

¹⁴ Diese Konferenz von Computerwissenschaftler:innen in Dartmouth, USA, unter der Führung des Informatikers John McCarthy (1927-2011) gilt als Gründungsversammlung des Forschungsgebiets Künstlicher Intelligenz. | Vgl. John McCarthy/Marvin L. Minsky/Nathaniel Rochester et. al.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955. In: The AI magazine Vol. 27/4 (2006), S. 12-14.

¹⁵ Zur Definition s. Clemens Sedmak: „Homo est animal rationale“. In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein „animal rationale“? Vernunft – Kognition – Intelligenz. Darmstadt: Wissenschaftliche Buchgesellschaft 2004 (Topologien des Menschen 1), S. 12-15.

¹⁶ Weitere Definitionsversuche schließen unsere Körperlichkeit mit ein, wodurch Intelligenz auf menschliche Wahrnehmung beschränkt wird. Unser kognitives Handeln wäre demnach abhängig von unseren Sinnen – unsere Wissen und unsere Bedeutung von Sprache hinge untrennbar mit unseren Körpern zusammen. Menschenähnliche Intelligenz wäre demnach nur in menschenähnlichen Körpern möglich. Konsequenterweise dürfe man dann von jedem künstlich intelligenten System auch nur Wissen erwarten, das auf dessen Verkörperung Sinn macht. | Vgl. Georg Dorfner: Rationalität, Emotionalität, Körperlichkeit: Können Maschinen Menschen verstehen – und umgekehrt? In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein „animal rationale“? Vernunft – Kognition – Intelligenz. Darmstadt: Wissenschaftliche Buchgesellschaft 2004 (Topologien des Menschen 1), S. 102-106. || Dennoch dürfe man nicht davon ausgehen, dass eine Verortung in und Wahrnehmung von der Welt spontan und automatisch zu Intelligenz oder gar Bewusstsein führt.

¹⁷ Vgl. Markus Antonius Wirtz (Hg.): „Intelligenz, emotionale“ In: Dorsch – Lexikon der Psychologie (2021) <https://dorsch.hogrefe.com/stichwort/intelligenz-emotionale> (Zugriff: 04.04.2024).

künstlich geschaffenen ‚Intelligenz‘. Das Problem der objektiven Beschreibbarkeit von Intelligenz kann damit nicht gelöst werden, da sich auch die Intelligenzfähigkeit unserer Mitmenschen – sei es eine rationale oder emotionale – nur anhand äußerer Intelligenzleistungen beobachten und beurteilen lässt. Diese „Intelligenz-Effekte“¹⁸ bilden den einzigen Anhaltspunkt und sind zugleich größter Kritik ausgesetzt.

2.1 Technisches Menschenbild

Demnach verwundert es nicht, dass sich bei dieser unscharfen Begrifflichkeit die Bezeichnung ‚Künstliche Intelligenz‘ als artifizielles Abbild zur ‚natürlichen‘, menschlichen Intelligenz etabliert hat, nimmt sie sich doch die grundsätzlichen Funktionen formalen menschlichen Denkens zum Vorbild. „Die Schwierigkeit des Begriffs der Intelligenz von Maschinen liegt in dem Versprechen, das er macht. Er erhebt Anspruch auf Simulation menschlicher Intelligenz, also die Nachbildung des Menschen als Maschine.“¹⁹ Der Vergleich ergibt sich aufgrund der wissenschaftlichen Erkenntnisse aus der Kybernetik und des Behaviorismus, welche die Vorgänge menschlicher Kognition durch informationstheoretische und mathematisch-formalisierbare Kategorien beschreibbar gemacht und die komplexen Prozesse geistiger Funktionen und maschinelle Kontroll- wie Steuerungsverfahren auf eine Stufe gehoben haben.

Für Norbert Wiener, einer der Gründer der Kybernetik und ihr einflussreichster Theoretiker, arbeiten das menschliche Gehirn und automatische Maschinen nach ähnlichen Prinzipien der Kontrolle, Steuerung und Rückmeldung (Feedback). Seiner in den 1950er Jahren entwickelten Theorie nach empfangen beide „Automaten, seien sie nun aus Metall oder Fleisch und Blut“²⁰ Signale (Nachrichten bzw. Informationen) von ihrer Umwelt und stehen mit dieser in einem Kreisprozess (Rückkopplung) – sind dadurch also lernfähig. Der Informationsbegriff nimmt eine tragende Rolle als Bindeglied zwischen Mensch und Maschine ein – er wird von beiden in gleicher Weise erfasst, aufgenommen und verarbeitet. Die physikalische Struktur der ersten Rechenmaschinen wird mit der des menschlichen Gehirns gleichgesetzt und auch deren Speicherleistung

¹⁸ Wolfgang Welsch: Wandlungen im humanen Selbstverständnis. In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein „animal rationale“? Vernunft – Kognition – Intelligenz. Darmstadt: Wissenschaftliche Buchgesellschaft 2004 (Topologien des Menschen 1), S. 54.

¹⁹ Orlando Budelacci: Mensch, Maschine, Identität. Ethik der Künstlichen Intelligenz. Basel: Schwabe Verlag 2022 (Schwabe reflexe 74), S. 57.

²⁰ Norbert Wiener: Kybernetik. In: Physikalische Blätter Vol. 5/8 (1949), S. 357.

funktioniere „wie in einem Gedächtnis“²¹. Kognitive Prozesse ließen sich diesem Denkmuster entsprechend erstmals formalisieren und auf Rechenmaschinen übertragen.²²

Der Behaviorismus blendet indes innerliche Vorgänge menschlichen und tierischen Verhaltens vollständig aus und konzentriert sich auf rein naturwissenschaftliche, objektive und äußerlich beobachtbare Methoden. Das grundlegende Paradigma ist das Reiz-Reaktions-Schema, das auch wesentlich das behavioristische Modell des Lernens mitbestimmt. So folgt auf einen bestimmten Reiz eine bestimmte Reaktion (dem kybernetischen Feedback ähnlich) – positive Ergebnisse werden verstärkt, negative abgeschwächt. Dem Leitgedanken dieses wissenschaftstheoretischen Konzepts entsprechend, ließe sich intelligentes Verhalten auch Maschinen zuschreiben: Was zählt, sind die bereits besprochenen Intelligenz-*Effekte*, die beobachtet werden können. Lernen funktioniere bei Mensch und Maschine über ein Abgleichen von Information – dies werde so lange wiederholt, bis das gewünschte Ergebnis erzielt wird.²³

Die Popularität kybernetischer und behavioristischer Denkmuster ergibt sich aus der „Vorstellung des Gehirns als Computer, das heißt als informationsverarbeitende Maschine“²⁴, die das Vertraute mit dem technisch Neuen in Verbindung bringt und durch dieses erklärbar macht. So wurde zwischen 1940 und 1950 mit dem Begriff ‚Computer‘ auch nicht zwischen Maschine und Mensch unterschieden, sondern es konnte sowohl eine Programmierer:in als auch eine Rechenmaschine damit bezeichnet werden. Bis sich der Begriff ausschließlich auf den Computer, so wie wir ihn heute kennen, als Objekt bezog, hatte die Bezeichnung ‚Elektronengehirn‘ Hochkonjunktur, um etwas zu beschreiben, das es noch einzuordnen gilt, aber eine vorgeschobene Ähnlichkeit mit der neurologischen Struktur und den Fähigkeiten des menschlichen Gehirns aufwies. Durch die semantische Annäherung von Mensch und Maschine beginnt sich die Grenze zwischen Organischem und Künstlichem allmählich aufzulösen.²⁵ Infolge stieg das Interesse, zu erproben, inwieweit sich diese Angleichung treiben ließe.

²¹ Ebd., S. 359.

²² Vgl. Peter Trawny: Vorwort des Herausgebers. In: Norbert Wiener: Mensch und Menschmaschine. Herausgegeben und mit einem Vorwort von Peter Trawny. Frankfurt am Main: Vittorio Klostermann 2022, S. 11-17. || Vgl. Wiener: Kybernetik (1949), S. 355-362.

²³ Vgl. Sebastian Höhne: Behaviorismus. In: Lernpsychologie. Lernen und Gedächtnis – einfach erklärt (2015) <http://www.lernpsychologie.net/lerntheorien/behaviorismus> (Zugriff: 04.04.2024).

²⁴ Matthias L. Schroeter: Dekolonialisiert euren Geist! Renaturalisiert die Intelligenz! In: Yasemin Keskinetepe/Dr. Anke Woschek (Hg.): Künstliche Intelligenz. Maschinen lernen Menschheitsträume. Göttingen: Wallstein 2021, S. 53.

²⁵ Vgl. Betschon: Künstliche Intelligenz (2019), S. 25-26.

2.1.1 Turings Erbe

Die Frage „*Can machines think?*“ – eine provokativ gemeinte Formulierung des britischen Logikers und Mathematikers, Alan Mathison Turing (1912–1954), in seinem 1950 erschienenen und neben „*Computable Numbers*“ (1936/37) einflussreichsten Aufsatz „*Computing Machinery and Intelligence*“ – beschäftigt seit den frühen Entwicklungen des beginnenden Computerzeitalters Informatiker wie Theoretiker und grenzt damit auch an philosophische wie theologische Fragen über das Denken und Menschsein an sich an. Turing beantwortet die Frage nicht direkt mit ‚Ja‘ oder ‚Nein‘ – zu zweideutig sei die Semantik von „*machine*“ und „*think*“, sie wäre sogar „*too meaningless to deserve discussion*“²⁶, da sich die Bedeutung der Begriffe schnell ändern könne – sondern schlägt alternative Fragestellungen vor, die er in Form eines „*imitation game*“²⁷ veranschaulicht. In diesem, geläufig als Turing-Test bekannten Gedankenexperiment, muss ein menschlicher Fragesteller (*interrogator*) (C) anhand eines Gesprächs mit einer Maschine (A) und einem Menschen (B) herausfinden, welcher von beiden die Maschine und welcher der Mensch ist. Die Proband:innen sind jeweils voneinander abgeschottet und kommunizieren schriftlich miteinander. Die Maschine hat den Test bestanden, wenn C nicht dazu in der Lage ist, in mehr als 70 Prozent der Fälle eine richtige Unterscheidung zu treffen. Umgekehrt bedeutet dies, dass der Test als geglückt gilt, wenn mindestens 30 Prozent der Teilnehmenden die Maschine für einen Menschen halten. In dieser Form wird der Turing-Test heute üblicherweise zitiert und gilt als einer der Meilensteine der KI-Entwicklung in der Informatik.²⁸

²⁶ A. M. Turing: *Computing Machinery and Intelligence*. In: *Mind* Vol. 59/236 (1950), S. 442.

²⁷ In der ursprünglichen Version des Spiels gibt es drei Agenten – einen Mann (A), eine Frau (B) und einen Fragesteller (*interrogator*) (C), der entweder ein Mann oder eine Frau ist. C ist in einem Raum eingeschlossen und kommuniziert schriftlich mit A und B. Das Ziel von C ist es, zu bestimmen welche von den beiden anderen die Frau und welcher der Mann ist. Das Ziel von A ist es, C zu täuschen; B soll A unterstützen. An diesem Punkt ergänzt Turing um die seiner Meinung nach wichtigere Frage, nämlich ob es einen Unterschied machen würde, wenn an die Stelle des Täuschers A ein Computer treten würde, der menschliches Verhalten perfekt nachahmt. Turing sieht diese Frage eng mit ersterer, ob Maschinen denken können, verwandt und meint mit dem *imitation game* eine Möglichkeit gefunden zu haben, genauer über die intellektuellen Fähigkeiten eines Menschen abseits körperlicher Zustände und physischer Erscheinung nachzudenken. Das *imitation game* gilt nun als erfolgreich bestanden, wenn A als Maschine den Fragesteller C täuschen kann, wer der Mann und wer die Frau ist. Damit definiert Turing die beiden Begriffe Intelligenz und Denken eng über das Prinzip der Täuschung, das für ihn näher am tatsächlichen Vermögen menschlichen Bewusstseins liegt als die Fähigkeit reines Wissen zu berechnen und zu kommunizieren. Die Maschine müsse somit über soziales Wissen, was eine Frau und was ein Mann ist, verfügen und dem Gegenüber ‚bewusst‘ falsche Informationen zuspiesen. | Vgl. Turing: *Computing Machinery and Intelligence* (1950), S. 433-435.

²⁸ Diese Form des Turing-Tests ist eine vereinfachte Version und hat den ambivalenten Begriff des Denkens durch eine einfache Aufgabe ersetzt. Der Turing-Test hat die KI-Forschung seit Beginn beflügelt. | Vgl. Manuela Lenzen: *Natürliche und künstliche Intelligenz. Einführung in die Kognitionswissenschaft*. Frankfurt am Main: Campus 2002 (Campus Einführungen), S. 50.

Turing setzt mit diesem Testverfahren menschliches Denkvermögen mit der Fähigkeit zur erfolgreichen Täuschung gleich, wodurch er viel Kritik erntete. Anstatt Aufklärung darüber zu liefern, wie sich Intelligenz, Intentionalität oder Bewusstsein definiert, testet das Verfahren menschliche Leichtgläubigkeit und dient dazu, den Mythos einer denkenden Maschine, die dem Menschen ebenbürtig ist, aufrechtzuerhalten. Turing nutzt die angesprochene Problematik einer fehlenden Nachweisbarkeit von Intelligenz selbst als Argument, um die Frage nach der denkenden Maschine weder beantworten noch verneinen zu müssen: *„the only way by which one could be sure that a machine thinks is to be the machine and to feel oneself thinking. [...] Likewise according to this view the only way to know that a man thinks is to be that particular man.“*²⁹ Diese Sichtweise zeugt bis heute von einem Denk-Automatismus in den Wissenschaften, der die evidenzbasierte Objektivität favorisiert und bewusstseinsgebundene Zeugnisse als Privatsache und subjektive Meinung ablehnt. Dadurch entkoppelt sich Intelligenz von menschlichem Bewusstsein und baut auf eine Welt der Fakten und der Vereindeutigung als primäre Form des Wissens auf.³⁰ Turings Einfluss für die KI-Forschung ist nicht von der Hand zu weisen und hat ein paar Jahre nach der Veröffentlichung seines Aufsatzes während der Dartmouth-Konferenz 1956 unter anderem zu den entscheidenden Fragen geführt, die bis heute die KI-Forschung beschäftigen.³¹ So wird angenommen,

*that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.*³²

Menschliche Intelligenz wird rein systemisch betrachtet und die Simulation unter dem Deckmantel der Täuschung zum Prinzip erhoben. Unterthemen der Konferenz beschäftigten sich mit Fragestellungen, wie sich menschliches Denken in Computersystemen nachbilden lässt bzw. wie Kreativität und Zufälligkeit formalisierbar wären.

²⁹ Turing: *Computing Machinery and Intelligence* (1950), S. 446.

³⁰ Vgl. Jan Söffner: Sinn und Sinnlosigkeit. Die Frage nach der Stellung der Hermeneutik im Zeitalter der künstlichen Intelligenz. In: Kablitz, Andreas/Christoph Marksches/Peter Strohschneider (Hg.): *Hermeneutik unter Verdacht*. Berlin/Boston: Walter de Gruyter 2021 (Text und Textlichkeit 2), S. 1-5.

³¹ Die Arbeiten Turings beeinflussten maßgeblich die Grundlagen des Forschungsgebiet über Künstliche Intelligenz. So sprach er in *„Computing Machinery and Intelligence“* bereits von lernenden Maschinen (*„learning machines“*), die sich selbst weiterentwickeln – ähnlich wie der Verstand eines Kindes zum Erwachsenen heranreift: *„Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child’s?“* Die Idee, eine *„child-machine“* zu entwickeln, die selbstständig mittels behavioristischer Methoden (*„by punishments and rewards“*) lernt, prägt noch heute die Entwicklung Künstlicher Intelligenz. | Vgl. Turing: *Computing Machinery and Intelligence* (1950), S. 456-457.

³² McCarthy/Minsky/Rochester et. al.: *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (2006), S. 13.

2.1.2 Imitation vs. Duplikation

Eine ontologische Gleichstellung menschlicher Vernunft und technischer Leistungsfähigkeit kennzeichnete die ersten Entwicklungsjahre,³³ deren Euphorie zwischen der Mitte der 1950er und 1970er Jahre allerdings bald darauf aufgrund aufgebauschter Erwartungen und technischer Limitationen³⁴ in einen ‚KI-Winter‘ übergang und eine Verlangsamung in der Erforschung Künstlicher Intelligenz markierte.³⁵ Durch den amerikanischen Philosophen John R. Searle (*1932) konnten eindeutig Grenzen zwischen der menschlichen Fähigkeit des Verstehens und der rein rechnerischen Verarbeitung von Wissen durch Maschinen – bzw. genauer Computersysteme – aufgezeigt werden. Searle demonstrierte dies in seinem 1980 erschienenen Aufsatz „*Minds, brains, and programs*“ anhand eines Gedankenexperiments. In diesem ist ein Mensch in einem Zimmer von der Außenwelt abgesperrt und erhält mehrere Stapel chinesischer Schriftzeichen sowie ihm verständliche Regelsätze, um die Stapel miteinander zu korrelieren. Der Mensch kann weder Chinesisch schreiben noch sprechen, ferner das Schriftsystem beispielsweise vom Japanischen unterscheiden. Die Regelsätze ermöglichen es ihm, eine Reihe von formalen Symbolen zueinander in Beziehung zu setzen, wobei das ‚formal‘ hier bedeutet, dass die Symbole ausschließlich anhand ihrer Form identifiziert werden. Damit kann er entsprechende chinesische Symbole anhand der entsprechenden Gestalt als Antwort auf von außen kommende Fragen, die chinesische Symbole anderer bestimmter Gestalten enthalten, zurückzugeben. Dieses Spiel wird nun so lange weitergeführt, bis der Mensch so gut in der Handhabung der chinesischen Symbole ist, dass, von außerhalb des Zimmers betrachtet, kein Unterschied mehr zu einer chinesischen Muttersprachler:in bestehen zu scheint. Der Mensch in dem Raum versteht aber immer noch kein Wort Chinesisch.

In dem Gedankenexperiment, das als das ‚Chinesische Zimmer‘ (*Chinese room*) bekannt ist, verhält sich der Mensch wie ein Computerprogramm, das formale Regeln nutzt, um das Verstehen der chinesischen Sprache zu simulieren. Demnach zieht Searle eine deutliche Grenze zwischen Turings Imitation und der eigentlichen Duplikation von

³³ Auch heute wird gerne ein Vergleich zwischen der ‚Rechenleistung‘ des menschlichen Gehirns und der eines Computers gezogen: So übersteigt das Gehirn mit 1.022 Gleitkommaoperationen pro Sekunde (Flops) heutige Supercomputer mit 1.015 Flops (Stand 2019) um den Faktor 10 Millionen. | Vgl. Betschon: Künstliche Intelligenz (2019), S. 16.

³⁴ Weiterführende Informationen zur Bewertung des Stands der Forschung zur damaligen Zeit s. Lighthill, James: Artificial Intelligence: A General Survey (1972) https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf (Zugriff: 21.11.2023).

³⁵ Vgl. Budelacci: Mensch, Maschine, Identität (2022), S. 74.

kognitiven Zuständen in einem Computerprogramm, wie Denken, Verstehen und in weiterer Folge dem Bewusstsein eines Menschen.³⁶ Daraus ergibt sich auch Searles Unterteilung in eine ‚starke‘ (*strong AI*) und eine ‚schwache KI‘ (*weak AI*):

According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states.³⁷

Während Searle die Leistungen schwacher KI anerkennt, greift er Turings Anfangsfrage „*Can machines think?*“ auf und formuliert sie um. Indem er fragt „*Could a machine think?*“³⁸ stellt er überhaupt die Möglichkeit der Entwicklung einer starken Künstlichen Intelligenz in Form eines programmierbaren Computers auf den Prüfstand, um sie kurzerhand mit ebenjenem Gedankenexperiment zu verneinen. Für Searle lassen sich menschliche kognitive Zustände nur von einer Art von Maschine erklären, nämlich dem Gehirn sowie „*machines with internal causal powers equivalent to those of brains.*“³⁹

Die Vorstellung von kausalen Kräften (*causal powers*) wird dem von dem deutschen Philosophen und Psychologen, Franz Brentano (1838-1917), eingeführten Begriff der ‚Intentionalität‘ vorgestellt. Die kausalen Kräfte gelten als notwendige Bedingung für intentionale Handlungen. Damit erweitert Searle Brentanos Konzept, das Intentionalität als charakteristisches Merkmal des Mentalen bzw. psychischer Phänomene sieht. Diese psychischen Phänomene werden nur im Inneren wahrgenommen, sie sind eine subjektive Erfahrung der inneren Welt und bezeichnen das Vorstellen an sich, nicht das Vorgestellte. Im Unterschied zu den physischen Phänomenen und ihrer Erfahrung in der objektiven, äußeren Welt, erfahren sie keine Ausdehnung bzw. räumliche Bestimmtheit, können sich

³⁶ Vgl. Searle: *Minds, brains and programs* (1980), S. 417-418.

³⁷ Ebd., S. 417.

³⁸ Ebd., S. 417.

³⁹ Ebd., S. 417. || Searle argumentiert, dass diese inneren kausalen Kräfte (*causal powers*) eng in Zusammenhang mit der Stofflichkeit bzw. der biologischen Struktur des menschlichen Gehirns stehen, und diesem dadurch erlauben, kognitive Phänomene wie Verstehen, Handeln oder Lernen hervorzubringen. Searle hält es für möglich, dass Maschinen eine Sprache wie Chinesisch oder Englisch prinzipiell verstehen können, vorausgesetzt sie verfügen über dieselben kausalen Kräfte wie das menschliche Gehirn. Diese Kräfte sind aber nicht jenen Maschinen gegeben, die ausschließlich in Begriffen von Rechenprozessen über formal definierte Elemente operieren, demnach eine Instanziierung eines Computerprogramms sind und keine menschliche Intelligenz. Demnach argumentiert Searle, dass „*no purely formal model will ever be sufficient by itself for intentionality because the formal properties are not by themselves constitutive of intentionality, and they have by themselves no causal powers except the power, when instantiated, to produce the next stage of the formalism when the machine is running.*“ Ohne die tatsächliche Funktionsweise und Eigenschaften des menschlichen Gehirns zu kennen, läuft die Entwicklung einer Künstlichen Intelligenz weiter darauf hinaus, die Schatten, d.h. den Anschein von Kognition, nachzuzeichnen und diesen als das Wahre, d.h. als echte Intelligenz, auszugeben. | Ebd., S. 422 | Vgl. Ebd., S. 422-423.

aber auf ausgedehnte Größen beziehen.⁴⁰ Sie charakterisieren sich durch die „intentionale (auch wohl mentale) Inexistenz eines Gegenstandes [...] [und können insofern als] solche Phänomene [definiert werden], welche intentional einen Gegenstand in sich enthalten.“⁴¹

Auch Oswald Wiener (1935-2021), österreichisch-kanadischer Schriftsteller, Kybernetiker, Sprachtheoretiker mit Fokus auf Kognitionswissenschaften und theoretisches Haupt der Wiener Gruppe, weist auf den streng behavioristischen Standpunkt Turings hin, der wissenschaftliche Untersuchungen der Introspektion für überflüssig erklärt. Wenn unter dem Begriff ‚Denken‘ bloß das verstanden wird, was den Turing-Test besteht, fokussiert Turing einzig auf die funktionale Äquivalenz von Mensch und Maschine.⁴² Eine Unterscheidung des Outputs bzw. Verhaltens zweier Entitäten auf denselben Input bzw. Reiz durch eine Beobachter:in ist damit nicht gegeben bzw. wird als für nicht notwendig erklärt. Statt funktionaler Äquivalenz, der jegliche Introspektion fehlt, fordert Wiener eine strukturelle Kennzeichnung von Leistung, die Intelligenz nachweist. Ein verlässlicher introspektiver Ansatz fehlt aber bis heute.⁴³

Dennoch haben die Erfolge der KI-Forschung in den letzten drei Jahrzehnten⁴⁴ und die jüngsten Entwicklungen im Bereich des maschinellen Lernens (*machine learning*) zu einem Wiedererstarken des Interesses an Künstlicher Intelligenz geführt. Die Ergebnisse beeindrucken aufgrund der Kombination gewaltiger Datenmengen und leistungsfähiger Rechner mit einer hohen Verarbeitungsgeschwindigkeit. Die alten Versprechen des Begriffs Künstlicher Intelligenz, die eine Simulation, wenn nicht sogar eine Rekonstruktion menschlicher Intelligenz, also „die Nachbildung des Menschen als Maschine“⁴⁵, verkünden, werden wieder hervorgeholt.

⁴⁰ Vgl. Franz Brentano: Psychologie am empirischen Standpunkt. Mit Einleitung, Anmerkungen und Register herausgegeben von Oskar Kraus. Hamburg: Felix Meiner 1973 (Philosophische Bibliothek 192), S. 116-126.

⁴¹ Ebd., S. 124-125.

⁴² Vgl. Oswald Wiener: Schriften zur Erkenntnistheorie. Wien: Springer 1996 (Computerkultur X), S. 78.

⁴³ Vgl. Oswald Wiener/Thomas Eder: Literatur, Sprache, Denken: Die Anfänge. In: Thomas Eder/Thomas Raab/Michael Schwarz (Hg.): Oswald Wieners Theorie des Denkens. Gespräche und Essays zu Grundfragen der Kognitionswissenschaft. Berlin/Boston: Walter de Gruyter 2023, S. 40-44.

⁴⁴ In den 1990er Jahren wurde dank leistungsfähigerer Computer und ausreichender Datenmengen ein Meilenstein in der KI-Geschichte erreicht: eine Maschine gewann das erste Mal gegen einen Menschen im Schachspiel. Der Sieg des IBM-Computer *Deep Blue* über den damaligen Schachweltmeister Gari Kasparow am 10. Februar 1996 wurde symbolisch als Wendepunkt für die Überlegenheit der Maschine über den Menschen gedeutet. Auch wenn es sich bei *Deep Blue* nicht um ein ‚intelligentes System‘ handelt, sondern alleinig die Rechengeschwindigkeit gewonnen hatte (der Computer konnte 14,5 Züge im Voraus berechnen), führte dieser Meilenstein zu einem erneuten Aufschwung in der KI-Forschung. | Vgl. Budelacci: Mensch, Maschine, Identität (2022), S. 74-79. || Für einen Überblick über die Meilensteine der KI-Forschung s. Budelacci: Mensch, Maschine, Identität (2022), S. 79-81.

⁴⁵ Ebd., S. 57.

In der öffentlichen Wahrnehmung und im Marketing wird vorschnell maschinelles Lernen mit menschlicher Vernunftbegabung gleichgesetzt und damit ein Narrativ gestärkt, das Emanzipationsbestrebungen in den technischen Entwicklungen verortet und sie einem Bedrohungsdiskurs unterordnet. Die KI wird zur Konkurrenz stilisiert und gefährdet das menschliche Selbstverständnis: „Das zeigt sich sowohl in der Anthropomorphisierung des Algorithmus als auch in seiner gleichzeitigen Dämonisierung.“⁴⁶ Bereiche, die bisher für Technologie als unerreichbar galten, werden durch neueste Entwicklungen Künstlicher Intelligenz einer Revision unterzogen. Die Maschine etabliert sich als Künstler:in, Schriftsteller:in, Programmierer:in – und damit gleichzeitig als Konkurrent:in. Die Debatte dreht sich demnach vor allem um die Frage, welche anthropologischen Modelle noch gültig sind und was den Menschen und seine Künste im Unterschied zur KI noch als eigenständig auszeichnet⁴⁷ – und was die KI für uns teilweise so menschlich macht.

2.1.3 Alte Versprechen

Heutige bedeutende IT-Unternehmen, wie Microsoft, Meta, Alphabet Inc. oder Apple, haben sich der Erforschung von KI verschrieben und beachten kaum die Untiefen der Begrifflichkeiten. Ein unscharf gewählter Begriff wie der der Künstlichen Intelligenz eigne sich hervorragend für ein Spiel mit den Erwartungen.⁴⁸ Und das funktioniert auch und gerade deshalb so gut, weil die intrinsische Logik der künstlichen Systeme selbst für fachkundige Personen teils im Verborgenen bleibt. Es verwundert daher wenig, dass in Interviews mit den verschiedensten Fachleuten aus der IT-Branche inkonsistente Antworten und Ungereimtheiten auftauchen, was die Möglichkeiten aktuell vorherrschender Technik und das Verständnis menschlicher Fähigkeiten betrifft. Beispielhaft dafür ist ein Interview von CBS News⁴⁹ vom 08.10.2023 mit dem britischen Informatiker und Kognitionspsychologen sowie Physik-Nobelpreisträger (2024),

⁴⁶ Stephanie Catani: Generierte Texte. Gegenwartsliterarische Experimente mit Künstlicher Intelligenz. In: Andrea Bartl/Corina Erk/Jörn Glasenapp (Hg.): Schnittstellen. Wechselbeziehungen zwischen Literatur, Film, Fernsehen und digitalen Medien. Paderborn: Brill Fink 2022 (inter|media 15), S. 248.

⁴⁷ Vgl. Hennig/ Hauptmann: Alexa, optimier mich! (2019), S. 86-87.

⁴⁸ Generell schwingt bei der Entwicklung Künstlicher Intelligenz immer auch die Möglichkeit der tatsächlichen Rekonstruktion von Bewusstsein und Intelligenz in einem Computer mit – auch wenn solche Überlegungen eher nur als Verkaufstaktik gebraucht als tatsächlich verfolgt zu werden. Aktuelle Forschungsansätze zu Künstlicher Intelligenz verfolgen keinen *human-centered approach* – also ein menschlich gestaltetes Programm zu entwickeln – sondern einen *rationalist approach*, der auf die Lösung menschlicher Probleme ausgerichtet ist. | Vgl. Catani: Generierte Texte (2022), S. 249.

⁴⁹ Vgl. Scott Pelley: 60 Minutes. “Godfather of Artificial Intelligence” Geoffrey Hinton on the promise, risks of advanced AI. In: CBS News (2023) <https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/> (Zugriff: 12.10.2023).

Geoffrey Hinton, der als ‚*Godfather of AI*‘ vorgestellt wird. Allerdings ist nicht nur die Titulierung *godfather* irreführend,⁵⁰ sondern die recht oberflächlichen Fragen in Form von *self-fulfilling prophecies* lassen elementare Definitionen ungeklärt und einem weitgefassten Verständnis überlassen.⁵¹ Im Interview bejaht Hinton, dass heutige Künstliche Intelligenzen sowohl verstehen („*understand*“) als auch intelligent seien, und ihre eigenen Erfahrungen und Entscheidungen treffen „*in the same sense as people do*“. Im nächsten Satz spricht er ihnen hingegen Bewusstsein („*conscious*“) und Selbstwahrnehmung („*self-awareness*“) ab – von dessen Vermögen man eigentlich aufgrund der Aussage „*in the same sense as people*“ ausgehen würde – nur um mit einem „*they will, in time*“ ominöse Zukunftsaussichten heraufzubeschwören, die den Menschen als zweitintelligentestes Wesen auf dieser Erde platzieren: „*I think we're moving into a period when for the first time ever we may have things more intelligent than us.*“ Hinton, der seit den 1970er Jahren an der Simulation Neuronaler Netze arbeitet, bestätigt, dass er es bisher nicht schaffte, den menschlichen Geist zu verstehen, aber seiner Meinung nach eine künstliche Version davon erzeugt hätte, die heute besser im Lernen sei als das menschliche Gehirn. Deren genaue Funktionsweise entziehe sich aber seinem Verständnis: „*we don't actually know what's going on any more than we know what's going on in your brain.*“ Er hätte aber „*a very good idea of sort of roughly what it's doing*“.⁵² Derartige Ungereimtheiten über das Wesen menschlicher Intelligenz und technischen Fähigkeiten der Entwicklung zeigen offenkundig, wie wenig diese Technologie tatsächlich verstanden wird und der Technikdiskurs vielmehr dazu genutzt

⁵⁰ Wie bereits erwähnt, geht die Gründung des Forschungsfeldes auf die Dartmouth-Konferenz im Sommer des Jahres 1956 zurück. Andererseits wäre es bezeichnender, den britischen Logiker und Mathematiker Alan Turing, den Erfinder der Turing-Maschine 1936, einem Vorgänger des modernen Computers, *Godfather of AI* zu nennen. | Vgl. Ralf Otte: Künstliche Intelligenz für dummies. Weinheim: Wiley-VCH 2019 (Lernen einfach gemacht), S. 38.

⁵¹ Gary Marcus kritisiert in seinem Blog-Beitrag CBS News für das schwach recherchierte und einseitig präsentierte Interview. Der Interviewer mache seiner Ansicht nach keine Bemühungen, Hinton zu seinen angeblichen Fakten (*alleged facts*) zu befragen. | Vgl. Gary Marcus: What was 60 Minutes thinking, in that interview with Geoff Hinton? In: Marcus on AI (2023) <https://garymarcus.substack.com/p/what-was-60-minutes-thinking-in-that> (Zugriff: 12.10.2023).

⁵² Das Kernproblem dieser Unklarheit liegt vor allem an der Hauptfunktion Künstlicher Neuronaler Netze, dem sogenannten *deep learning*, das als Untergruppe von *machine learning* eine komplexe Struktur von Algorithmen nach dem Vorbild des menschlichen Gehirns nutzt. Programmierer:innen geben demnach Algorithmen nicht mehr vor, wie sie zur Lösung kommen, sondern lassen sie selbst ein Modell entwickeln, das sie auf unterschiedliche Daten anwenden können. Die Qualität selbstlernender Algorithmen erlaubt es, dass sich Programme selbstständig weiterentwickeln. Die genaue Prozessierung von Informationen ist aber nicht sichtbar und findet in der *black box* zwischen der Eingabe- und Ausgabeschicht, der sogenannten Verborgenen Schicht, statt. | Vgl. Laurenz Wuttke: Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion (2024) <https://datasolut.com/neuronale-netzwerke-einfuehrung/> (Zugriff: 18.10.2023).

wird, die Verheißungen und Verkaufsstrategien der IT-Unternehmen auf dem Markt zu befeuern.

So wird laufend die baldige Entwicklung einer *Artificial General Intelligence (AGI)*⁵³ angekündigt – also einer ‚starken‘ KI, wie sie Searle (1980) definiert, welche die intellektuellen Fähigkeiten eines Menschen besitzen soll – ohne diesem Meilenstein in Wirklichkeit auch nur einen Schritt näher gekommen zu sein.⁵⁴ Mit *If-then*-Formulierungen („*If AGI is successfully created, this technology could help us elevate humanity [...]*“)⁵⁵ öffnet unter anderem OpenAI in ihrer Übersicht zum aktuellen Entwicklungsstand der Imagination Tür und Tor und hebt gleichzeitig Urteilskriterien für eine kritische Hinterfragung menschlicher und technischer Fähigkeiten auf. Mit den Versprechen, was eine KI in Zukunft können wird und dem Befund, dass man nicht so recht über ihre Funktionsweise Bescheid weiß, entsteht ein Vakuum über den jetzigen Stand der Technik, das mit Visionen und Märchen gefüllt wird. Was nicht an neuen Technologien verstanden wird oder sich nicht nachvollziehen lässt, dem wird bereitwillig Intelligenz und Bewusstsein zugesprochen. Annahmen und rhetorische Kniffe tragen ihr Übriges zum Diskurs bei. Sprachliche Ungenauigkeiten erweisen sich als produktiv, um die Wirksamkeit der Erzählung auf breiter Ebene zu gewährleisten.⁵⁶

Bevor in der Arbeit die Anthropomorphisierungstendenzen von künstlichen Systemen analysiert werden und welchen Einfluss sie auf unser Verständnis einer künstlerischen KI bzw. die Möglichkeit, Künstliche Intelligenz als Autor:in zu betrachten, hat, soll deshalb ein Exkurs über die technische Beschaffenheit Künstlicher Intelligenz helfen, den Nebel über diese hochgepriesene Entwicklung zu lichten.

⁵³ OpenAI definiert *AGI* wie folgt: “artificial general intelligence (AGI) – by which we mean highly autonomous systems that outperform humans at most economically valuable work.” | OpenAI: OpenAI Charter, <https://openai.com/charter> (Zugriff: 16.10.2023).

⁵⁴ Warum eine derartige Ankündigung für unrealistisch angesehen wird, liegt daran, dass noch keine anerkannte Theorie, wie man Bewusstsein technisch erschaffen könnte, existiert. Die Herausforderung besteht nun schon seit dem Turing-Test darin, ein Verfahren zu entwickeln, das es erlaubt, Intelligenz objektiv (bei einer Maschine) nachweisen zu können. | Vgl. Otte: Künstliche Intelligenz für dummies (2019), S. 34-35; 72; 413.

⁵⁵ Altman, Sam: Planning for AGI and beyond. In: OpenAI (2023) <https://openai.com/blog/planning-for-agi-and-beyond> (Zugriff: 13.10.2023). || Des Weiteren auffällig ist die Verwendung von Ausdrücken wie „believe“, „think“, „eventually“, wodurch der Eindruck wenig wissenschaftlich fundierter Argumentationen entsteht. Zukünftige Handlungsspielräume werden ohne handfeste Fakten aufgeworfen: “We believe our research will eventually lead to artificial general intelligence, a system that can solve human-level problems.” | Vgl. OpenAI: Pioneering research on the path to AGI, <https://openai.com/research/overview> (Zugriff: 16.10.2023).

⁵⁶ Vgl. Philipp Schönthaler: Die Automatisierung des Schreibens & Gegenprogramme der Literatur. Berlin: Matthes & Seitz 2022 (Leben und Dienste 3), S. 387-399.

2.1.3.1 Exkurs: Künstliche Neuronale Netze

Die heute mit Abstand erfolgreichste Technologie Künstlicher Intelligenz sind Künstliche Neuronale Netze (KNN), die dem menschlichen Gehirn nachempfunden sind,⁵⁷ und im Bereich des *machine learning*⁵⁸ eingesetzt werden. KNNs bilden die Basis für *deep learning*⁵⁹, das einen speziellen Teilbereich des maschinellen Lernens darstellt.⁶⁰ Für das Verständnis Künstlicher Neuroner Netze spielt das biologische Vorbild des menschlichen Gehirns eine modellhafte Rolle und sei deshalb kurz erläutert. Es genügt eine einfache Vorstellung von einem biologischen Neuron, auch Nervenzelle genannt, das über einen Zellkörper mit Zellkern und Synapsen verfügt, eine Verbindung der beiden Einheiten durch ein Axon sowie verästelte Dendriten, die vornehmlich die Erregung anderer Zellen aufnehmen und ein Bindeglied zwischen den Synapsen und anderen Neuronen darstellen. Wird nun ein Neuron aktiviert, aktivieren sich auch die anderen Nervenzellen, die mit ihm verbunden sind. Eine einzelne Nervenzelle hat mehrere tausend bis zehntausend Verbindungen zu anderen Nervenzellen.⁶¹ Das Künstliche Neuronale Netz nimmt sich diesen Aufbau zum Vorbild und ahmt ihn auf vereinfachte Weise nach. So besteht ein einzelnes Künstliches Neuron (auch *Unit*, Einheit oder Knoten genannt) aus mehreren Funktionen, welche die Eingabeinformation verarbeiten und mit

⁵⁷ Erste Konzepte konnektionistischer Systeme existierten bereits in den 1940er und 1950er Jahren. Mit ihrem programmatischen Aufsatz „*A Logical Calculus of the Ideas Immanent in Nervous Activity*“ versuchten der amerikanische Neurophysiologe und Kybernetiker Warren S. McCulloch (1898-1969) und sein Kollege Walter Pitts (1924-1969) bereits 1943 die neuronalen Prozesse des Gehirns technisch nachzuahmen. Das erste erfolgreiche arbeitende Künstliche Neuronale Netz wurde 1960 von dem Psychologen und Informatiker Frank Rosenblatt (1928-1971) und seinen Kollegen mit dem Computer *Mark I Perceptron* hergestellt. Die Grundprinzipien dieser Ideen sind bis heute die Basis Künstlicher Neuroner Netze. Die KI-Forschung ging vorher allerdings aufgrund fehlender technischer Möglichkeiten einen Umweg über symbolverarbeitende, sequenzielle Systeme und algorithmische Lösungen, bis in den 1990er Jahren große Erfolge im Bereich des Maschinellen Lernens erzielt werden konnten. | Vgl. Edwin Hübner: *Menschlicher Geist und Künstliche Intelligenz. Die Entwicklung des Humanen inmitten einer digitalen Welt.* Stuttgart: Freies Geistesleben 2020, S. 104-105. || Für weitere Informationen s. McCulloch, Warren S./Walter Pitts: *A Logical Calculus of the Ideas Immanent in Nervous Activity.* In: *Bulletin of Mathematical Biology* Vol. 52/1/2 (1990), S. 99-115.

⁵⁸ Maschinelles Lernen (*machine learning*) ist eine zentrale Technik der Künstlichen Intelligenz, welche die automatische und maschinelle Generierung von Wissen mittels großer Datenmengen meint. Die für die Lösung des Problems gebrauchten Algorithmen werden nicht vom Menschen programmiert, sondern ergeben sich aus der Analyse von Mustern und Gesetzmäßigkeiten aus den Input-Daten. Anhand dieser Daten trifft das System Entscheidungen und verbessert ,lernend‘ die Algorithmen, wenn weiterer Input zur Verfügung gestellt wird. | Vgl. Cisek, Günter: *Machtwechsel der Intelligenzen. Wie sich unser Miteinander durch künstliche Intelligenz verändert.* Wiesbaden: Springer Vieweg 2021 (Die blaue Stunde der Informatik), S. 32.

⁵⁹ Mehrschichtiges, tiefes oder tiefergehendes Lernen (*deep learning*) nutzt als Gefüge Künstliche Neuronale Netze zur Verarbeitung großer unstrukturierter Datenmengen. Damit können komplexe Aufgaben automatisiert und natürliche Sprache besser verarbeitet werden. | Vgl. Ebd., S. 38-43.

⁶⁰ Vgl. Uwe Lämmel/Jürgen Cleve: *Künstliche Intelligenz. Wissensverarbeitung – Neuronale Netze.* 5. überarb. Aufl. München: Hanser 2020, S. 189-191.

⁶¹ Vgl. Ebd., S. 191-192.

anderen Einheiten des Netzes verknüpfen. Die einzelnen Units sind durch Links (Kanten) verbunden, die verschieden stark gewichtet sind. Je größer das Gewicht der Verbindung, desto größer ist auch der Einfluss auf die Unit. Diese Gewichte sind von den Systemen selbstständig veränderbar (auf Basis der Trainingsdaten) und bestimmen, wie eng zwei Units miteinander verknüpft sind. Ändert sich die Gewichtung eines Links, spricht die Informatik von maschinellem ‚Lernen‘⁶². Die Gewichtung der Verbindungen entscheidet dann über die Aktivierung der anderen Units und schließlich den Ausgabewert. Für die Aktivierung der Units wird auf eine binäre Aktivierung gesetzt, d.h. das Künstliche Neuron ist entweder inaktiv (0) oder aktiv (1). Durch die Kombination der einzelnen Units können aussagenlogische Grundfunktionen wie UND, ODER, NICHT realisiert werden.⁶³

Übereinanderliegende Units bilden eine Schicht (*Layer*). Künstliche Neuronale Netze verfügen über einen Input- und einen Output-Layer, können aber auch dazwischenliegende Hidden-Layer (*black box*) haben, die für die Programmierer:innen nicht einsehbar sind und vom KNN ‚verwaltet‘ werden – d.h., wie die KI zu ihren Ergebnissen kommt, welchen Entscheidungsprozess sie trifft, wird vom System nicht sichtbar gemacht. Aufgrund dieser Intransparenz sprechen selbst Expert:innen der KI mehr Intelligenz oder Bewusstsein zu, als diese tatsächlich besitzt.

⁶² Die verwendeten Lernverfahren werden in Überwachtes, Be-/Verstärkendes und Unüberwachtes Lernen bzw. Teilüberwachtes Lernen unterteilt. Das Überwachte Lernen (*supervised learning*) ist das restriktivste Verfahren und hat kein biologisches Vorbild. Dabei sind sowohl die für das Training verwendeten Eingabedaten als auch die erwarteten Ausgabewerte bekannt und damit das Ergebnis messbar. Kommt es zu Fehlern beim Trainieren des Netzes und die Ausgabewerte stimmen nicht mit den prognostizierten Werten überein, werden die Verbindungsgewichte sowie die Schwellenwerte in den Künstlichen Neuronen verändert. Für die Problemstellung gibt es also bereits eine Lösung, an deren Beispielen die KI lernt. Anders verhält es sich beim Be- bzw. Verstärkenden Lernen (*reinforced learning*), das eher menschlichen Lernmethoden entspricht und eine hohe praktische Relevanz hat. So gibt es keine Vorgaben, wie das System auf die richtige Lösung kommt. Anhand der trial-and-error-Methode sowie durch positives oder negatives Feedback auf die Ausgabewerte soll das System beim Lernprozess eine eigene Strategie entwickeln. Die Bedeutung von bestärkendem bzw. verstärkendem Lernen liegt in den ‚gemachten Erfahrungen‘ der KI, durch die sie sich von menschlich erdachten Lösungen deutlich unterscheiden kann. Beim Unüberwachten Lernen (*unsupervised learning*) werden weder Eingabe- noch Ausgabewerte vorgegeben, das Hauptziel ist Mustererkennung. Das System hat also keine Zielwerte und versucht selbst in den Daten Regelmäßigkeiten, versteckte Strukturen und Zusammenhänge zu finden. Die Daten werden demnach in *Cluster* oder Kategorien aufgeteilt oder komprimiert. Die Ausgabewerte werden vom Netz selbst auf ihre Richtigkeit überprüft und für Feedback herangezogen. Dieses Verfahren ist biologisch gesehen am plausibelsten und entspricht in seiner Architektur am ehesten der menschlichen Aneignungsweise von Wissen. Problematisch bei dieser Lernmethode ist allerdings der Lösungsweg, der für Außenstehende nicht nachvollziehbar ist. Der Lernprozess bleibt durchgehend eine *black box*, weshalb vielfach auch auf Teilüberwachtes Lernen (*semisupervised learning*), einer Mischung aus Überwachtem und Unüberwachtem Lernen, gesetzt wird. Hierfür wird eine kleine Menge an gelabelten Daten als Vorgabe gestellt, was eine erhebliche Verbesserung der Lerngenauigkeit bewirkt. | Vgl. Ebd., S. 198-199. || Ausführlicher werden die verschiedenen Lernarten hier beschrieben: Vgl. Cisek: Machtwechsel der Intelligenzen (2021), S. 32-38.

⁶³ Vgl. Hübner: Menschlicher Geist und Künstliche Intelligenz (2020), S. 103-113.

2.1.3.1.1 Nicht jedes Wissen ist formalisierbar

Für die erfolgreiche Entwicklung Künstlicher Intelligenz gibt es bisher keine allgemeingültige Strategie, weshalb sie ausschließlich experimentell erreicht wird. Großen Einfluss auf das Ergebnis haben dabei die Daten, die zum Training der KI verwendet werden. Diese müssen in ausreichender Zahl vorhanden und bereits aufbereitet sein, damit sie das System verarbeiten kann. Darunter fällt nicht nur eine Zusammenführung unterschiedlicher Quellen, sondern auch deren Selektion. Inhaltlich irrelevant angesehene Daten werden weggelassen, Merkmale werden auf das Wesentliche reduziert. Es kann zudem vorkommen, dass unvollständige Datensätze weggelassen oder fehlende Werte unter Einbezug globaler Konstanten, Durchschnittswerte oder Wahrscheinlichkeiten vervollständigt werden. Die somit stark veränderten Daten werden dann als Information verpackt, die der Computer ‚versteht‘, das heißt sie verarbeiten kann. Im Regelfall ist das eine Codierung durch eine Menge von 0-1-Werten (Binärcodierung), also auf Basis zwei gegensätzlicher Zustände⁶⁴. Weiters ist eine formale Darstellung des Wissens in einer geeigneten Sprache notwendig. Das können mathematische Relationen, eine Logik oder eine Programmiersprache sein.⁶⁵

Nun lässt sich aber nicht jede Art von Wissen uneindeutig formalisieren oder kategorisieren. Darunter fallen beispielsweise vage, unscharfe oder unsichere Aussagen, wie ‚Der Mann ist groß‘ oder ‚Es regnet morgen‘. Im Gegensatz zur klassisch zweiwertigen Logik (wahr oder falsch), ist der Wahrheitswert bei diesen Aussagen unklar oder noch nicht bekannt. Sie sind situationsabhängig oder relativ zu betrachten. Besonders menschliche Wahrnehmungen (Geschmack- oder Schmerzempfinden) lassen sich nicht quantitativ messen und bedürfen einer Abstraktion oder qualitativen

⁶⁴ So erhält beispielsweise das Merkmal Farbe in einer Menge von Daten die Werte weiß, gelb, rot, grün, blau, schwarz. Die sechs Werte werden durch die Wertesequenz bestehend aus genau einer eins und fünf Nullen abgebildet: z.B. gelb = 010000; rot = 001000 usw. | Vgl. Lämmel/Cleve: Künstliche Intelligenz (2020), S. 233-234. || Diese Art der Codierung geht auf das von Gottfried Wilhelm Leibniz (1646-1716) entwickelte Konzept eines binären Zahlensystems zurück, mit dem sich im Prinzip Wissen formalisieren und ausrechnen lässt. Leibniz schwebte die Errichtung eines universalen Zeichensystems vor, das Gedanken und Zeichen parallel setzt. Jedem Zeichen entspräche genau ein Gedanke und umgekehrt. Sie wären damit algorithmisch erfassbar. Die inhaltliche Betrachtung von Begriffen wird ausgeschlossen und daraus herrührende falsche Deutungen und Schlüsse vermieden. Das Leibniz'sche Ideal inhaltlicher Universalität und methodischer Einheitlichkeit ist Grundvoraussetzung für den Computer geworden. | Vgl. Hübner: Menschlicher Geist und Künstliche Intelligenz (2020), S. 314-316.

⁶⁵ Die Logik (aus A folgt B) ist die am besten untersuchte Form der Wissensrepräsentation. Auch relationales Wissen (einfache Beziehungen zwischen Objekten) lässt sich darstellen, aber bildet nur Fakten und keine logischen Abhängigkeiten ab. Wissen kann mit der Vererbung von Eigenschaften (ein Auto hat einen Motor; ein Audi ist ein Auto; ergo ein Audi ‚hat einen Motor‘ = von Oberbegriff Auto auf Unterbegriff Audi vererbtes Wissen) ergänzt oder Aktionen in bestimmten Situationen durch prozedurales Wissen (Eröffnungen beim Schach) vorgeschrieben werden. Regelbasiertes Wissen (Wenn-Dann-Regeln) beinhaltet nicht nur logische Abhängigkeiten, sondern in Bezug auf die Verarbeitung natürlicher Sprache, auch grammatisches Wissen. | Vgl. Lämmel/Cleve: Künstliche Intelligenz (2020), S. 23-25.

Einordnung. Solche Daten werden demnach durch (bedingte) Wahrscheinlichkeiten, quantitative Bewertungen, durch Annahmen (*measure of belief/disbelief*), basierend auf Erfahrung, Vertrauen und Plausibilität oder durch sogenannte *fuzzy*-Mengen, die über Referenzmengen beschrieben werden, eingeordnet. Implizites Wissen wird aus explizitem abgeleitet.⁶⁶

2.1.3.1.2 Wie künstliche Systeme Semantik verstehen

Künstliche Neuronale Netze können Semantische Netze für einzelne Wortbedeutungen bilden, d.h. semantisches Wissen wird netzartig mit relationalen Bezügen repräsentiert. Statt Regeln und logische Abhängigkeiten steht die Klassifizierung von Objekten im Vordergrund mit ihren Eigenschaften und hierarchischen Beziehungen zu anderen Objekten. Es gibt Ober- und Unterklassen, die jeweils den Ober- und Unterbegriffen des Objekts entsprechen. Den Klassen können ebenfalls Eigenschaften zugeordnet werden. Die Objekte, Klassen und ihre entsprechenden Werte werden als Knoten zusammengefasst. Diese sind durch Kanten miteinander verbunden, die Beziehungen zwischen, aber auch Eigenschaften von Objekten oder Klassen darstellen. Die Bündelungen von Knoten werden auch als Frames bezeichnet, die damit eine Teilansicht eines Semantischen Netzes darstellen. Sie speichern Objekte und ihre Eigenschaften und Klassen als uniforme Schemata ab. Die semantische Tiefe hängt von der Komplexität der Frames ab, die neben dem Konzept der Hierarchie und der Vererbung von Wissen auch Restriktionen (*Wertebereich* =) oder Default-Werte (*Standard-Werte*) enthalten können. So kann ein Objekt zur Klasse Auto gehören und durch den Frame ‚Auto‘ Eigenschaften wie ‚4 Räder‘ oder ‚Fahrzeug‘ erben, welche nicht extra angegeben werden müssen. Die Komplexität lässt sich entsprechend hochskalieren.⁶⁷

Beispiel: Frame HWI-JD85

<i>Instanz von</i>	<i>Auto</i>
<i>Hersteller</i>	<i>Opel</i>
<i>Typ</i>	<i>Corsa</i>
<i>Motor</i>	<i>Viertakt</i>
<i>Farbe</i>	<i>weinrot</i>
<i>Karosserie</i>	<i>Metall</i>
<i>Frame Auto</i>	
<i>Ist-ein</i>	<i>Fahrzeug</i>
<i>Räder</i>	<i>default = 4</i>
<i>Motor</i>	<i>Wertebereich = (Viertakt, Zweitakt, Diesel)</i>
<i>Karosserie</i>	<i>Wertebereich = (Metall, Kunststoff)⁶⁸</i>

⁶⁶ Vgl. Ebd., S. 87-109.

⁶⁷ Vgl. Ebd., S. 81-84.

⁶⁸ Ebd., S. 83.

Die hierarchische Beziehung von Wörtern ist damit in einem Semantischen Netz gut abbildbar, bleibt aber immer zweistellig. Die Darstellung komplexerer Zusammenhänge ist zwar möglich, aber Existenzaussagen bzw. Oder-Aussagen, wie sie aus der Prädikatenlogik bekannt sind, sind beispielsweise nicht formulierbar. Semantische Netze haben – trotz ihrer Bezeichnung – keine klare Semantik und geben ursprüngliches Wissen nur unvollständig wieder. So ist beim Beispiel ‚Auto‘ zwar enthalten, welcher Hersteller den Typ Corsa produziert oder wie viele Räder ein Auto besitzt, aber nicht, welche sozialen und kulturellen Bedeutungen, welches Lebensgefühl, historischer Hintergrund oder politische Entscheidungen für die Infrastruktur damit verbunden sind. Derart detailliertes und implizites Wissen ist unmöglich vollständig zu formalisieren. Zudem ist eine statische Repräsentation aufgrund fehlender Flexibilität ungeeignet, um die sich schnell und häufig wandelnden Bedeutungen natürlicher Sprache zu erfassen. Formalisiertes Wissen bleibt auch stets vergangenheitsorientiert, weil nur Ergebnisse aufgrund vergangener Daten produziert werden können. Dieses Repräsentationsproblem ist eines der großen Herausforderungen in der KI-Forschung.

Ziel dieses Exkurses war es, die Funktionsweisen und Problematiken bei der Entwicklung Künstlicher Intelligenz verständlich darzulegen und dazu beizutragen, ihrer Mystifizierung entgegenzuwirken. Die Arbeit widmet sich im Folgenden den Anthropomorphisierungstendenzen künstlicher Systeme, die mit diesem technischen Wissen im Hinterkopf kritisch betrachtet werden.

2.2 Menschliches Technikbild

Trotz der eindeutig technischen und damit nicht-menschlichen Systemeigenschaften, bleibt das Prinzip der Täuschung für die Entwicklung Künstlicher Intelligenz vorrangig und verzerrt deren Wahrnehmung. Die Ergebnisse wirken natürlich und sprechen Menschen auf emotionaler Ebene an. Seit den 1960er Jahren wurden mehrere Programme entwickelt, die den Turing-Test bestanden hatten und denen es gelang, Menschen darin zu täuschen, die Systeme hätten tatsächlich Emotionen oder gar ein Bewusstsein entwickelt.⁶⁹ Im Gegensatz zu den damaligen Maschinen sind heutige KIs ausgereifter

⁶⁹ Dieses Phänomen ist als ELIZA-Effekt bekannt. Dieser gründet auf das von Joseph Weizenbaum 1966 entwickelte Computerprogramm „ELIZA“, das verschiedene Gesprächspartner imitieren konnte und durch die Nachahmung der Methoden der klientenzentrierten Psychotherapie nach Carl Rogers berühmt wurde. Das Kommunikationsverhalten der User:innen mit dem Chatbot unterschied sich nicht sonderlich von der Kommunikationssituation mit einer menschlichen Gesprächspartner:in. Bei den befragten Versuchspersonen entstand der Eindruck, das Programm wäre tatsächlich an ihnen und ihrem Befinden interessiert. Bei ELIZA handelte es sich um ein symbolverarbeitendes System, das nach syntaktischen Regeln und mit vorgegebenen Mustern von Dialogen arbeitete. Dem Programm kann weder eine

und noch mehr in zwischenmenschliche Interaktionen mit ihren Nutzer:innen verwickelt. Sie schreiben, geben Ratschläge, rechnen, bearbeiten oder entscheiden – alles in direktem Kontakt mit Menschen über technische Geräte im Haushalt, auf der Arbeit oder am Handgelenk. Die künstlich geschaffenen Intelligenzen **scheinen** derart menschenähnlich, dass User:innen dazu geneigt sind, ihr Einfühlungsvermögen auf die nicht-menschlichen Geräte auszudehnen.⁷⁰

Diese Tendenz wird unter anderem vom US-amerikanischen Internetkonzern Meta weiter forciert. Dieser nutzt die erwähnten Implikationen des Begriffs *Artificial Intelligence*, um seine Assistenzsysteme bewusst zu vermenschlichen und als „quasi-subjekthafte Schnittstelle“⁷¹ – wie Hennig/Hauptmann (2019) die Subjektivierungsprozesse von KI-Anwendungen näher beschreiben – zwischen ‚smartem‘ Gerät (Smartphone) und Anwender:in zu positionieren. Die Sprachassistent Siri von Apple war vor mehr als einem Jahrzehnt für manche Nutzer:in zwar ein nettes Gimmick, im Vergleich zu den neuen Meta AI-Charakteren, die im Herbst 2023 vorgestellt wurden,⁷² erscheint das Maschinenhafte der Software hingegen allzu deutlich. Da sich KI-Assistenzen heute in ihrem Gebrauch kaum voneinander unterscheiden, sind „[b]edeutungsgenerierende Strategien und Projektionen kultureller Werte auf

inhaltliche Bedeutung noch eine eigene kommunikative Absicht unterstellt werden. Dennoch erschienen die Antworten und Fragen derart menschlich, dass die Versuchspersonen der Maschine Vernunft zusprachen. Der ELIZA-Effekt wird heute bei vielen Chatbots genutzt, um eine möglichst benutzernahe Kommunikation zu gewährleisten und Kundenbindung zu generieren. | Vgl. Cisek: Machtwechsel der Intelligenzen (2021), S. 15. | Zum Computerprogramm ELIZA und seiner Anwendung s. Joseph Weizenbaum: ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine. In: Communications of the ACM Vol. 9/1 (1966), S. 36-45. | Es gibt auch aktuell bekannte Fälle, wie Blake Lemoine oder Lawrence (Larry) Page, die behaupten, hinter Künstlicher Intelligenz menschenähnliches Bewusstsein entdeckt zu haben. | Dazu exemplarisch s. Michael Hanfeld: Google wirft KI-Experten raus. Schalte mich bitte nicht ab! In: Frankfurter Allgemeine Zeitung (2022) <https://www.faz.net/aktuell/feuilleton/medien/google-entlaesst-softwareingenieur-blake-lemoine-18195598.html> (Zugriff: 19.10.2023).

⁷⁰ Vgl. Martina Heßler: Künstliche Intelligenz: Eine neuartige Kategorie von Maschine oder die Vermenschlichung der Maschine. In: Yasemin Keskintepe/Dr. Anke Woschek (Hg.): Künstliche Intelligenz. Maschinen lernen Menschheitsträume. Göttingen: Wallstein 2021, S. 117-119.

⁷¹ Hennig/Hauptmann: Alexa, optimier mich! (2019), S. 86.

⁷² „A new class of generative AI features that expand and strengthen the ways people connect with each other“. Bei der Präsentation der AI-Charaktere wird bewusst auf eine menschennahe Nutzung gesetzt – die KI soll im Alltag begleiten, Nutzer:innen zum Lachen bringen oder für Fragen zur Verfügung stehen: „Each has a unique personality, mannerisms and backstory“, so Meta. Unterfüttert werden die personifizierten Technologien mit ‚familiären‘ Gesichtern von populären Personen, wie Tom Brady (als Bru – „Confident sports debater“), MrBeast (als Zach – „Brotherly jokester“) oder Naomi Osaka (als Tamika – „Anime fanatic“), die als wünschenswert erachtete kulturelle und semantische Mehrwerte auf das Produkt übertragen. | Meta: Introducing new AI experiences from Meta, <https://ai.meta.com/genai/> (Zugriff: 18.10.2023). | Da Meta ihre Website laufend aktualisiert, um die neuesten Features zu promoten, und vorheriger Link bereits nicht mehr aktuell ist, sei hier noch eine zusätzliche Quelle gegeben, welche die neuen AI-Charaktere vorstellt, die angegebenen Zitate aber leider nicht mehr enthält: Meta: Introducing New AI Experiences Across Our Family of Apps and Devices. In: Meta (2023) <https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/> (Zugriff: 18.11.2024).

Technologien“ sowie deren Inszenierung zentral, um „den ›ästhetische[n] Mehrwert‹ oder ›Fiktionswert‹“⁷³ zu bestimmen. Die Werbung wird gezielt zur Vermittlung des semantischen Mehrwerts einer Ware eingesetzt. „Anthropomorphose, Personalisierung und letztlich Intimisierung der Beziehungen zwischen Menschen und digitalen Gegenübern“⁷⁴ kennzeichnen die Inszenierungsstrategien der *AI*-Charaktere. Die „Persönlichkeit“⁷⁵ der *AI* ist jedoch dem Geschäftsstreben des Unternehmens unterworfen. „Jeder Witz ist ein programmierter Lacher, der Kundenbindung und günstige Wettbewerbssituationen gegenüber anderen KIs anderer Unternehmen erwirken soll – und einen positiven Effekt auf die Kaufbereitschaft.“⁷⁶ Die Beziehung bleibt für den Menschen üblicherweise rein funktional, weshalb auch nur von einer **quasi**-subjektivierten Technologie, die wieder objektiviert wird, die Rede sein kann.⁷⁷

Die Vermenschlichung digitaltechnischer Artefakte setzt sich hingegen im anthropomorphen Sprachgebrauch für Künstliche Intelligenz weiter fort. Die „denkenden“ Maschinen funktionieren nicht mehr nur nach mechanischen Prinzipien, sondern „lernen“, „machen Späße“, „optimieren sich selbst“ oder „entscheiden“, anstatt zu rechnen, zu speichern, zu verarbeiten, Daten zu erheben, maschinell Wissen zu generieren und auszuwerten. Künstliche Intelligenzen werden durch die subjektbildende und weltaneignende Funktion von Sprache zu „als ob-Subjekten“⁷⁸, deren Integration in menschliche Handlungskontexte dadurch umso einfacher fällt.⁷⁹ Sie halten Einzug in unsere Arbeitswelt, Freizeit, Kommunikation und Unterhaltung – und unsere Kunst. Durch den Einsatz anthropomorpher Sprache wird der Technik eine Aura des Menschlichen verliehen. Während das menschliche Selbstverständnis in engem Zusammenhang mit technischen Entwicklungen steht und dadurch näher an die

⁷³ Hennig/Hauptmann: *Alexa, optimier mich!* (2019), S. 86; 88.

⁷⁴ Ebd., S. 94.

⁷⁵ Es lässt sich hier auch von einer sogenannten „Persona“ sprechen: menschenähnlich ausgestaltete, fiktive Charaktere, die u.a. von Künstlichen Intelligenzen genutzt werden. Entwickler:innen implementieren ein konsistentes Verhalten und passende Charakteristika in die KI, um das Bild einer stimmigen Persönlichkeit wiederzugeben. Metainformationen zum Bot, die von den Entwickler:innen oder dem Unternehmen bereitgestellt werden, unterstützen das Prinzip der Anthropomorphisierung, um über die Defizite der KI hinwegzutäuschen. | Vgl. Netaya Lotze: *Chatbots. Eine linguistische Analyse*. Frankfurt am Main: Peter Lang 2016 (Sprache – Medien – Innovationen 9), S. 38-39.

⁷⁶ Peter Seele: *Künstliche Intelligenz und Maschinisierung des Menschen*. Köln: Herbert von Halem 2020 (Schriften zur Rettung des öffentlichen Diskurses 1), S. 55-56.

⁷⁷ Zur weiteren Analyse digitaler Assistenzsysteme zu den Themenfeldern Macht und Gender sowie Subjektivierungsprozessen bei Künstlicher Intelligenz s. Hennig/Hauptmann: *Alexa, optimier mich!* (2019), S. 86-94.

⁷⁸ Armin Grunwald: *Soziale Roboter – vermenschlichte Technik und technische Menschenbilder*. In: Selin Gerlek/Sarah Kissler/Thorben Mämecke et. al. (Hg.): *Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen*. Hagen: Hagen University Press 2022 (Digitale Kultur 1), S. 69.

⁷⁹ Vgl. Ebd., S. 69.

Maschinen heranrückt, sind uns diese im Umkehrschluss wortwörtlich „auf den Leib gerückt“⁸⁰. Im folgenden Kapitel soll die zunehmende Konvergenz einer „digitaltechnische[n] Selbstmodellierung des Menschen als datenverarbeitende Maschine“⁸¹ und einer anthropomorphisierten Technik aufgezeigt werden.

2.3 Konvergenz der Intelligenzen

Wie gezeigt, werden die sprachlichen Grenzen zwischen Technik und Mensch durchlässiger. Der Mensch als Vorbild für technische Entwicklungen ist gleichzeitig in seinem Selbstverständnis bedeutungsvoll an dessen Fortschritt gebunden⁸² und richtet sich in Folge der Digitalisierung und mit Anleihen an kybernetische Denkmuster stärker an seinen körperlichen Funktionen und technisch messbaren Leistungen aus. In der Praxis der Selbstoptimierung durch Selbstvermessung (*self-tracking*) wird das Subjekt in numerischer Form thematisiert. In einer Schleife aus Feedback und Selbstregulierung erfassen wir heutzutage die zugeführten Kalorien, berechnen die Zeit für gesunden Schlaf und verfolgen unsere Schrittzahl. In unserer virtuell verquickten Gegenwart ist es ein Leichtes, die Körperfunktionen durch diverse Fitnesstracker zu messen und sich an maschinelle Denkmuster anzupassen. Die Zentrierung auf Feedbackkreisläufe ist ziel- und zukunftsorientiert und verschiebt das Selbst von einer Substanz- zu einer Konstruktionslogik.⁸³ Ein digitales Menschenbild als „*nichttriviale* Maschine“⁸⁴ stärkt die Vorstellung als technisch funktionierendes Objekt. Es erfolgt aber keine ontologische Gleichsetzung. Die Idee des ‚Menschen‘ in der Digitalisierung ist bilateral. Maschinenmodelle dienen der Beschreibung von Menschen, der ‚Mensch‘ geht aber in den autonom technischen Strukturen nicht verloren. Ein klassisch humanistischer

⁸⁰ Kevin Liggieri: Anthropozentrismus in Mensch-Maschine Interaktionen. In: Selin Gerlek/Sarah Kissler/Thorben Mämecke et. al. (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen. Hagen: Hagen University Press 2022 (Digitale Kultur 1), S. 75.

⁸¹ Grunwald: Soziale Roboter (2022), S. 62.

⁸² So lässt sich ein ‚mechanistisches Narrativ‘ in der zweiten Hälfte des 18. Jahrhunderts mit Beginn der Industrialisierung, ein ‚evolutionstheoretisches Narrativ‘ im späten 19. Jahrhundert mit Fokus auf die Naturwissenschaften oder ein ‚algorithmisches Narrativ‘ für heute in Zeiten der Digitalisierung festmachen. | Vgl. Armin Grunwald: Wer bist du, Mensch? Eine alte Frage vor neuen Herausforderungen. In: Armin Grunwald (Hg.): Wer bist du, Mensch? Transformation menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt. Freiburg/Basel/Wien: Herder 2021, S. 20-24 || Für weitere Informationen über die Transformation menschlichen Selbstverständnisses s. den gesamten Band: Armin Grunwald (Hg.): Wer bist du, Mensch? Transformation menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt. Freiburg/Basel/Wien: Herder 2021.

⁸³ Vgl. Eryk Noji: Das kybernetische Subjekt: Selbstthematisierung im Self-Tracking. In: Selin Gerlek/Sarah Kissler/Thorben Mämecke et. al. (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen. Hagen: Hagen University Press 2022 (Digitale Kultur 1), S. 126-137.

⁸⁴ Grunwald: Soziale Roboter (2022), S. 71.

Anthropozentrismus hält die Vorstellung des Konzepts ‚Mensch‘ als autonomes Subjekt, als Menschenbild, auf das die Technik gerichtet ist, aufrecht. Der ‚Mensch‘ steht als Voraussetzung und Vorbild für technische Entwicklungen, er ist das Maß, an dem sich Künstliche Intelligenz misst.⁸⁵ „Es ist gleichzeitig ein funktionales Modell und ein Argument, das auf rhetorischer und praktischer Ebene den Ausgangspunkt wie das Ziel technischer Konstruktionen bildet.“⁸⁶ Die Technisierung menschlichen Selbstverständnisses und die Anthropomorphisierung von Technik führt zu einer konvergierenden Zuschreibung von Attributen. Gehirn wie Maschine werden in ihren Fähigkeiten miteinander verglichen, es kommt zu einer semantischen Verschiebung des Intelligenzbegriffs, welcher die Simulation dem Sein vorzieht. In dieser Denkweise sind menschliche Leistungen nicht kategorial von technischen unterscheidbar.⁸⁷

„Hirn und Computer werden in einer dialektischen Bewegung aufgehoben, in der das Gehirn computerisiert und damit maschinisiert wird, während der Computer neuralisiert wird.“⁸⁸ Mensch und Maschine befinden sich in einer konvergierenden Bewegung zueinander. Künstliche Intelligenz ist damit ebenso bilateral zu verstehen. Sie ist nicht nur Gegenstand, sondern ein agierendes Ding, das zur Interaktion auffordert und nicht vom Menschen abzulösen ist. Durch ihre Nutzbarkeit und Nützlichkeit wird sie akzeptiert und habitualisiert. Die Technik steht damit nicht mehr ‚für sich‘, ist neutral, sondern wird mit Blick auf den Menschen entwickelt und konsumiert, während sich der Mensch weiter an die Automatismen und Stanzen der Algorithmen anpasst.⁸⁹ Die heutige digitale Welt zeugt von einer Automatisierung der Automatisierung, in der selbstlernende Algorithmen standardisierte Aufgaben ein zweites Mal automatisieren und weiter optimieren, um die analoge Welt in strukturierten Daten zu erfassen. Diese dyadische Automatisierung erfordert keinen denkenden und autonom handelnden Menschen – er wird der

⁸⁵ Vgl. Liggieri: Anthropozentrismus in Mensch-Maschine Interaktionen (2022), S. 76-77.

⁸⁶ Ebd., S. 76.

⁸⁷ Vgl. Grunwald: Soziale Roboter (2022), S. 62-73. || Von einer Gleichstellung menschlicher und digitaler Symbolverarbeitung geht die Computer-Metapher der Physical Symbol System Hypothesis (PSSH) aus, die besagt, dass „*A physical symbol system has the necessary and sufficient means for general intelligent action.*“ Sowohl Computer als auch menschliche Gehirne werden als solche *physical symbol systems* verstanden. Demnach würde es ausreichen, die Vorgänge menschlichen Denkens (Symbolverarbeitens) zu formalisieren, um in einem digitalen Computer künstliche Intelligenz zu erzeugen. Es wird erwartet, dass die Beschreibung dasselbe leistet wie das Beschriebene, was allerdings bis heute ein tieferes Verständnis menschlichen Denkens eher behindert. | Vgl. Oswald Wiener/Michael Schwarz: Pleomorphismus im Denken und die Computer-Metapher. Weiser, Symbol und das „Grounding“-Problem. In: Thomas Eder/Thomas Raab/Michael Schwarz (Hg.): Oswald Wieners Theorie des Denkens. Gespräche und Essays zu Grundfragen der Kognitionswissenschaft. Berlin/Boston: Walter de Gruyter 2023, S. 122. | Paul S. Rosenbloom: Rethinking the Physical Symbol Systems Hypothesis. In: Artificial General Intelligence Vol. 13921 (2023), S. 208.

⁸⁸ Schroeter: Dekolonialisiert euren Geist! (2021), S. 52.

⁸⁹ Liggieri: Anthropozentrismus in Mensch-Maschine Interaktionen (2022), S. 78-79.

Standardisierungslogik unterworfen. In Datenhaufen erfasst und gespeichert wird der Mensch zur Steuergröße und Messbarkeit wird zum universalen Prinzip erklärt. Unsere Kommunikation, wie wir arbeiten, handeln und unser Denken werden ‚geeicht‘ und automatisiert verwertet. Die Übersetzung des Lebens als wandelnder Datenhaufen ‚maschinisiert‘ den Menschen und wird zu dessen Menschwerdung. Der Mensch konsumiert und wird konsumiert. Er wird mehr und mehr zu seinem digitalen Selbst⁹⁰. In einer ‚doppelten Konvergenz der Intelligenz‘⁹¹ bewegen sich ‚[K]ünstliche und menschliche Intelligenz im doppelten Sinne aufeinander zu: KI wird immer besser und menschenähnlicher, da die Menschen immer mehr zu Datenhaufen werden, die standardisiert und automatisiert ausgelesen und optimiert werden.‘⁹²

Dennoch, wendet Söffner (2021) ein, verhält sich das, was sich vom Menschen in den Daten abbilden lässt, so reduziert zu lebendigem Bewusstsein, wie unser genetischer Code zu unseren Körpern in der physischen Welt.⁹³ Allein die Struktur eines Systems zu ermessen, gibt noch keinen Aufschluss über dessen Gesamtheit. Es ist daher zur Aufgabe der Humanisten geworden, das Konzept ‚Mensch‘ als anthropozentrisches Argument eines autonomen Subjekts gegen die Maschinenautorschaft zu verteidigen. Kategorien wie Individualität, Freiheit und Kreativität gehören zum philosophischen Selbstverständnis des Menschen und werden im Diskurs um die Neudefinition anthropologischer Modelle gegen die Maschinen in Stellung gebracht. Im folgenden Kapitel soll aufgezeigt werden, wie sich die Kreativität als Kategorie künstlerischer

⁹⁰ Vgl. Seele: Künstliche Intelligenz und Maschinisierung des Menschen (2020), S. 131-158. || Zum Konzept des digitalen Selbst als posthumane Vorstellung einer rein auf Information beruhenden Welt, in der der Mensch als computerisierbar gilt s. Thomas Christian Bächle: Das digitale Selbst. In: Martina Heßler/Kevin Liggieri (Hg.): Technikanthropologie. Handbuch für Wissenschaft und Studium. Baden-Baden: Nomos 2020, S. 283-288.

⁹¹ Seele: Künstliche Intelligenz und Maschinisierung des Menschen (2020), S. 132. || Eine Annäherung von Mensch und Maschine anhand ihrer Sprachverwendung hat Lotze (2016) empirisch in der linguistischen Analyse von Chatbots gezeigt. Die deutsche Sprachwissenschaftlerin vergleicht zwischenmenschliche Kommunikation (*Human-Human-Communication, HHC*) mit Mensch-Maschine-Interaktionen (*Human-Computer-Interaction, HCI*) und untersucht, ob sich anhand der Gespräche ein spezifischer *Computer-Talk (CT)* festmachen lässt. Obwohl die Bots in den Dialogen den Gesprächsfluss durch Ambiguitäten, sprachliche Routinen oder asymmetrische Antworten stocken, versuchen User:innen Interaktionsstrategien aus der *HHC* zu übertragen, um einen kohärenten Interaktionsverlauf zu gewährleisten. Lotze stellt fest, dass Chatbots eine deutliche Nähe zur Standardsprache aufweisen, während User:innen kürzere Kommunikationsformen verwenden – die Werte sich mit steigender Effizienz der Systeme aber annähern: „Je besser das jeweilige System funktioniert, desto geringer fällt die Divergenz zwischen UserInnen- und Systemsprache aus. Dabei nähern sich die Werte der Systeme und die der UserInnen einander gleichermaßen an [...]. Während die Entwicklung der Systeme von konzeptioneller Schriftlichkeit und monologischem Stil hin zu kürzeren konzeptionell mündlich gestalteten Turns verläuft, entwickelt sich die UserInnen-Sprache weg vom knappen Sprachgebrauch eines CTs hin zu ausführlicheren, natürlicheren Beiträgen.“ | Vgl. Lotze: Chatbots (2016), S. 359-368 | Ebd., S. 360-361.

⁹² Seele: Künstliche Intelligenz und Maschinisierung des Menschen (2020), S. 157.

⁹³ Vgl. Söffner: Sinn und Sinnlosigkeit (2021), S. 11.

Selbstreflexion eng mit einer Konzeption von Autorschaft, welche die Künstler:in bzw. Dichter:in qua Originalität und Individualität nobilitiert, verbunden hat. Eine von der Autor:in beanspruchte kreative Souveränität hallt bis heute nach, sowohl in ökonomisch-juristischer, ästhetischer wie hermeneutischer Hinsicht und scheint für Maschinen vorerst unzugänglich.

3 Kreativ sein, bedeutet Mensch sein

Durch Kreativität erklärt der Mensch, dass er keine Maschine ist.⁹⁴

Der Begriff der Kreativität hat in den westlichen Gesellschaften aktuell Hochkonjunktur. Als individuelle und soziale Fähigkeit ist die Kreativität ein wichtiger Teil des menschlichen Selbstverständnisses geworden. Im Herzen kapitalistischer Innovationsstrukturen ist sie begehrt im ökonomischen Wettbewerb – als menschliche Eigenschaft zielt sie auf eine kreative Transformierung des Selbst, nach einer Originalität des Ichs, erwartet im Gegenzug aber auch permanent von Originalität überrascht zu werden. Der Wunsch, kreativ zu sein und die Forderung nach ästhetisch Neuem umfasst Reckwitz (2012) in spätmodernen Zeiten als „eine Dopplung von Kreativitätswunsch und Kreativitätsimperativ, von subjektivem Begehren und sozialer Erwartung: Man *will* kreativ sein und *soll* es sein.“⁹⁵ Diese individuelle wie soziale Ausrichtung an Kreativität hat zur Ausbildung eines „*Kreativitätsdispositivs*“⁹⁶ geführt, das sich im 21. Jahrhundert äußerst robust und stabil zeigt. Diese als natürlich-universal institutionalisierte Kreativitätsorientierung formt nicht nur die soziale Praxis und das Subjekt, sondern hat in ihrer Genese im sozialen Feld der Kunst zu Prozessen einer gesellschaftlichen Ästhetisierung beigetragen.⁹⁷ Kreativität positioniert sich damit als zentrale Eigenschaft des Menschen und als Voraussetzung für das Schaffen von Kunst. Kreativ zu sein, bedeutet Mensch zu sein: „Der kreative Impuls ist ein entscheidender Teil dessen, was den Menschen von anderen Tieren unterscheidet“⁹⁸. Die Kreativität, als das Hervorbringen von dynamisch Neuem, ist in der Gesellschaft bevorzugter Dauerzustand und zugleich eine auf das Modell der Künstler:in fokussierte Aktivität – eine zentrale Fähigkeit von Jemandem, der schöpferisch das (ästhetisch) Neue hervorbringt.⁹⁹ Das Kreativitätsdispositiv folgt damit jener Subjektzentrierung in der Kunst, die sich in der

⁹⁴ Marcus Du Sautoy: Der Creativity Code. Wie künstliche Intelligenz schreibt, malt und denkt. München: C.H. Beck 2021, S. 298.

⁹⁵ Andreas Reckwitz: Die Erfindung der Kreativität. Zum Prozess gesellschaftlicher Ästhetisierung. Berlin: Suhrkamp 2012 (suhrkamp taschenbuch wissenschaft 1995), S. 10.

⁹⁶ Ebd., S. 15.

⁹⁷ Vgl. Ebd., S. 10-17. || Reckwitz macht die Prozesse gesellschaftlicher Ästhetisierung anhand der Konzepte der Body Art, Land Art und Minimal Art fest. So hat sich der Fokus der Ästhetisierung von Kunst Mitte des 20. Jahrhunderts von der Ebene der Bedeutung auf die Ebene der Materialität und Technizität der Dinge verschoben. Kunst- und Alltagsobjekte sind nicht mehr strikt voneinander trennbar, profane Dinge werden seit den 1970er Jahren verstärkt ästhetisiert. | Vgl. Ebd., S. 102-103.

⁹⁸ Du Sautoy: Der Creativity Code (2021), S. 13.

⁹⁹ Vgl. Reckwitz: Die Erfindung der Kreativität (2012), S. 39-40.

Mitte des 18. Jahrhunderts als Genieästhetik herausgebildet hat. Im Zentrum dieses Dispositivs steht die Produktion von kreativ Neuem und die Legitimierung von Werkherrschaft durch das Genie als ‚Schöpfer‘-Figur.¹⁰⁰

Den Begriff ‚kreativ‘ als das Auszeichnende des Menschen zu verwenden, ist historisch gesehen allerdings noch relativ jung. Um Mitte des 15. Jahrhunderts wird der Mensch erstmals als schaffendes Wesen (*animal creativum*) bezeichnet. Hiermit ist aber nicht der Einfallsreichtum eines Individuums oder die Fähigkeit zum Problemlösen gemeint, sondern eine tatsächlich schöpferische Tätigkeit als ein dem Abbild Gottes entsprechendes Geschöpf. Während in der Antike nach Vorbildern gestaltend, demnach ausschließlich durch Nachahmung (*imitatio*) geschaffen wurde und nach biblischem Verständnis nur Gott schuf (*creatio ex nihilo*), folgte der Umschwung mit der anthropologischen Aussage, der Mensch sei Gott am nächsten, zu einer neuen kreativen Selbstfindung. Von der Renaissance bis zur Deutschen Klassik hat das neue Selbstverständnis des Menschen als schöpferisches Wesen eine Tradition des künstlerischen Individuums zur Konsequenz, das in der Entstehung des Begriffs des Genies (*genius, ingenium*) seinen einflussreichsten Ausdruck findet. Diese Genie-Ästhetik, die bisher als Geheimnisse der Schöpfung wahrgenommene Kategorien von Kreativität, Originalität, Intention und Inspiration in der Person des Genies vereint, hat unter einer allmählichen Ausweitung des Kreativitätsverständnisses als Fähigkeit jeglichen menschlichen Gemüts kaum an Relevanz eingebüßt, sondern die Idee des Individuums als Urheber:in kultureller Schöpfung weiter bestärkt. Das Interesse um das Phänomen der Kreativität stieg im 20. Jahrhundert ferner in den Wissenschaften der Psychologie, den Sozialwissenschaften und der Pädagogik. Nebst dem künstlerisch schaffenden Genie entwickelte sich die Kreativität als eine nicht nur auf die Kunst bezogene Fertigkeit, sondern der menschlichen Natur zu eigene, universelle Fähigkeit zur Problemlösung und damit zur Bewältigung des Lebens schlechthin.¹⁰¹ Nach dem Zweiten Weltkrieg findet das eingedeutschte Wort ‚Kreativität‘ in seiner heutigen Anwendung im deutschsprachigen Raum Gebrauch und wird retrospektiv auf das schöpferische Genie, das Kunst aus sich selbst hervorbringt, angewandt.¹⁰² In den Wissenschaften der Nachkriegszeit etablierte sich die Kreativität als eigenständige, messbare Kategorie. Vor

¹⁰⁰ Vgl. Ebd., S. 60-61.

¹⁰¹ Heinrich Schmidinger: Das Auszeichnende des Menschen ist seine Kreativität – Eine geistesgeschichtliche Einstimmung. In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein kreatives Wesen? Kunst – Technik – Innovation. Darmstadt: Wissenschaftliche Buchgesellschaft 2008 (Topologien des Menschen 5), S. 7-21.

¹⁰² Vgl. Schönthaler: Die Automatisierung des Schreibens (2022), S. 231.

allem in der psychologischen Forschung fand sie als Teilbereich der Intelligenz durch den Kognitionspsychologen Joy Paul Guilford vermehrt Aufmerksamkeit. Die Kreativität galt als vernachlässigtes Thema und wurde als Antwort auf die sich entwickelnde Automatisierungsdebatte instrumentalisiert: Indem rationales Denken von Maschinen übernommen wird, sei kreative Intelligenz notwendig, um den sozialen Umwälzungen, ausgelöst durch die neuen Denkmaschinen, zu begegnen. Kreatives Denken sichere die Sonderstellung des Menschen, der sich dadurch von den Maschinen abhebe.¹⁰³ Nicht zuletzt im Kontext des Kalten Krieges wurde die Kreativitätsforschung ideologisch aufgeladen. Der Kampf um kreative Talente wurde neben der individuellen Freiheit zu einem populären ‚US-Import‘.¹⁰⁴ Die Universalisierung der Kreativität in den sozialen Ebenen des Alltags und der Arbeit haben dazu beigetragen, dass sich ein gesellschaftliches Modell des Kreativen durchgesetzt hat. Dieses Modell verdichtet sich in den Feldern der Ökonomie, Kunst und Technologie.¹⁰⁵ Das künstlerische Subjekt als kreative Existenz wird zu einem nachahmenswerten Vorbild, dessen Leistungen im Dienste einer auf Fortschritt und kreativen Innovation basierenden Gesellschaft stehen.¹⁰⁶

Überlegungen zu Kreativität umfassen wie bei der Definition von Intelligenz einen schwammigen Begriffsbereich, der je nach sozialer und ästhetischer Kategorie sowie unter Berücksichtigung digitaltechnischer Entwicklungen anders zu füllen ist und einen anderen Anspruch darauf erhebt, was als kreativ gilt.¹⁰⁷ Sieht man sich vorherrschende Definitionsversuche des Kreativen an, wird schnell klar, dass neben der rein funktional verwendeten Minimaldefinition ‚etwas Neues zu schaffen‘ noch eine weitere Voraussetzung erfüllt sein muss, die je nach Kategorie unterschiedlich ausfällt. So muss das kreativ neu Entstandene zusätzlich „einen Zweck erfüllen“¹⁰⁸, „*surprising, and*

¹⁰³ Vgl. J. P. Guilford: Kreativität (1950). In: Gisela Ulmann (Hg.): Kreativitätsforschung. Köln: Kiepenheuer & Witsch 1973, S. 27-29.

¹⁰⁴ Vgl. Schönthaler: Die Automatisierung des Schreibens (2022), S. 233.

¹⁰⁵ Vgl. Reckwitz: Die Erfindung der Kreativität (2012), S. 52-53.

¹⁰⁶ Vgl. Sabine Kampmann: Autorschaft in den Bildkünsten. Forschungen über bildende Künstlerinnen und Künstler in Kunstgeschichte, Geschlechterforschung und Kunst. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. S. 547.

¹⁰⁷ Vgl. Schönthaler: Die Automatisierung des Schreibens (2022), S. 228. || Für einen umfassenden Überblick über die Vielfältigkeit des Begriffs sowie verschiedene Ansätze, welche die Frage nach dem Wesen und der Definition von Kreativität aufgreifen s. Ursula Ziaja: Was ist Kreativität. Begriffsbestimmungen aus der Kreativitätsforschung. In: Weertje Willms/Martina Backes (Hg.): Kontexte kreativen Schreibens. Eine Standortbestimmung in Theorie und Praxis. Berlin: Frank & Timme 2021, S. 25-37.

¹⁰⁸ Philip Häusser: Natürlich alles künstlich. Was künstliche Intelligenz kann und was (noch) nicht – KI erklärt für alle. München: Droemer 2021, S. 161.

valuable“¹⁰⁹ oder „nutzbringend“ sein.¹¹⁰ Diese Zusatzbedingungen schließen an die oben genannten Kategorien an und vermischen sie je nach Kontext. Die Definitionen ‚einen Zweck erfüllen‘ und ‚nutzbringend‘ bzw. ‚nützlich‘ (*valuable*) sein, haben ökonomischen Charakter, der das Kreative in den Bereichen kommerzieller Nutzung wie Werbung, Mode oder Design sowie im technischen Sektor zur Beschreibung der Fähigkeit eines Computers, kreativ Probleme zu lösen, gebraucht wird. Eine sozial-psychologische Interpretation des Kreativen als ‚nutzbringend‘ steht in enger Relationen zum Individuum als Teil einer produktiven Gesellschaft und gehört damit wesentlich zum menschlichen Selbstverständnis. Mit Blick auf ästhetische Kategorien und im Speziellen auf die Literaturproduktion bleibt die Geniekonzeption eng mit dem Kreativitätsbegriff verbunden. Kreativität ist „ideologisch auf die Vorstellung eines autonom schaffenden Individuums bezogen“¹¹¹ und als geistiges Eigentum als Produkt im Konzept von Autorschaft stark verankert. Ein Werk entsteht demnach aufgrund der „kreativen Entscheidung eines Autors“¹¹² – und als kreativ anerkannt wird es, wenn es als ‚wertvoll‘ (*valuable*) erachtet wird.

Im Rahmen des Kreativitätsdispositivs, das Reckwitz (2012) formuliert, verweist Kreativität auf eine ästhetische Produktionsform. Sie verweist „auf die Hervorbringung von Neuem als bevorzugtem Dauerzustand und zugleich auf eine ästhetische, nach dem Modell des Künstlers gedachte Aktivität.“¹¹³ Der kreative Komplex richtet das Ästhetische am dynamisch Neuen aus und rechnet dieses ästhetisch Neue einem kreativ-schöpferischen Individuum zu. Demgegenüber steht das Publikum, das perzeptiv und affektiv auf dieses Neue reagiert. Das Kreativitätsdispositiv bezeichnet damit einen Ästhetisierungsprozess im Sozialen – das ästhetisch Neue findet sich in einer „*Produzenten-Rezipienten-Konstellation* des Sozialen“¹¹⁴ wieder. Das Neue markiert in der Spätmoderne nicht von vornherein ein Fortschrittsdenken, sondern bedeutet zunächst eine Orientierung an einem Zeitschema, das Vergangenheit, Gegenwart und Zukunft in die Begriffe ‚alt‘ und ‚neu‘ aufteilt. „Sachlich, das heißt auf der Ebene der Phänomene, markiert das Neue das Andere gegenüber dem Gleichen. Sozial verweist es auf das Abweichende im Unterschied zum Normalen und normativ Erwarteten.“¹¹⁵ Das Neue ist

¹⁰⁹ Margaret A. Boden: *AI. It's nature and future*. Oxford: Oxford University Press 2016, S. 67.

¹¹⁰ Otte: *Künstliche Intelligenz für dummies* (2019), S. 64.

¹¹¹ Wetzel: *Einleitung* (2022), S. 3.

¹¹² Ebd., S. 4.

¹¹³ Reckwitz: *Die Erfindung der Kreativität* (2012), S. 39.

¹¹⁴ Ebd., S. 40.

¹¹⁵ Ebd., S. 43.

damit nicht einfach objektiv vorhanden, sondern verweist immer auf ein bestimmtes Beobachtungs- und Wahrnehmungsschemata.¹¹⁶

Die weitere Diskussion richtet sich auf den Wert des (ästhetisch) Neuen und geht der Frage nach, ob nun Maschinen und insbesondere Künstliche Intelligenz dieses Neue produzieren und im Bereich der Kunst kreativ wirken können.

3.1 Keine Kreativität ohne Grenzüberschreitung

Während Kreativität vorwiegend immer noch als eine schöpferische Kraft gilt, ist sie auf den ersten Blick für Maschinen durch eine Berufung auf ein autonomes Subjekt bzw. den Geist nicht zugänglich. Die Technik ist der Gegenpol zum humanistischen Subjekt, das mit Kreativität, Individualität, Autonomie und Freiheit assoziiert wird, während ferner die Kategorien Transparenz, Objektivität und Fortschritt für das rationale Verstehen der Maschinen steht.¹¹⁷ Sieht man sich aber die Prozesse an, die zu kreativer Leistung führen, gliedert Boden (2016) die Kreativität vom Objekt her in drei unterschiedliche Typen, die nicht mythisch überhöht scheinen. So verbindet die rekombinierende Kreativität (*combinational creativity*) bekannte Ideen auf neue Weise miteinander; die erforschende Kreativität (*exploratory creativity*) lotet die Grenzen bestehender Entwicklungen aus und führt sie weiter; und die transformative Kreativität (*transformational creativity*) bricht mit den vorangegangenen Regeln und betritt völliges Neuland.¹¹⁸ Während Budelacci (2022) den Maschinen die rekombinierende und erforschende Kreativität, die laut ihm ohnehin „wahrscheinlich 99 % aller kreativen Leistungen abdeckt“¹¹⁹, zugesteht, sieht er die KI-Kreativität dennoch als eine Form der Kreativität an, „die zwar Neues schafft, aber in einem vorgegebenen Rahmen, der durch die Qualität und Menge der Daten definiert und auch beschränkt wird.“¹²⁰ Die Maschine könne abseits ihres algorithmisch vorgegebenen Horizonts nichts völlig Neues aus sich selbst heraus produzieren. Dadurch sei das transformative Moment für die KI nicht greifbar und liege immer noch allein in den Möglichkeiten des Menschen. Auch Du Sautoy (2021) führt aus: „Transformative Kreativität [...] gilt für Algorithmen traditionell als unerreichbar“ und sieht in selbstlernenden Algorithmen „die Gefahr, immer mehr vom Gleichen hervorzubringen“, da sowohl die Trainingsdaten als auch ihre Bewertung bereits vorgegeben sind.¹²¹

¹¹⁶ Vgl. Ebd., S. 39-44.

¹¹⁷ Vgl. Budelacci: Mensch, Maschine, Identität (2022), S. 88-89.

¹¹⁸ Vgl. Boden: AI. It's nature and future (2016), S. 68-69.

¹¹⁹ Budelacci: Mensch, Maschine, Identität (2022), S. 89.

¹²⁰ Ebd., S. 85.

¹²¹ Du Sautoy: Der Creativity Code (2021), S. 296.

In den letzten Jahren gab es in der KI-Forschung einige technische Errungenschaften, die versuchen, dieses Problem zu umgehen. So ist Boden (2016) der Ansicht, dass durchaus alle drei Kreativitäts-Typen für Künstliche Intelligenz zugänglich sind. Sie wertet beispielsweise die Entwicklung neuer Medikamente durch KI als transformativ-kreativ, da etwas Neues produziert wird, das es bisher noch nicht gab. Mittels Formalisierung des Zufalls und Feedbackschleifen können selbstlernende (evolutionäre) Programme prinzipiell die Regeln bestehender Systeme brechen und über den eigenen Output reflektieren – sofern im Programm entsprechende Auswahlkriterien für die Bewertung vorhanden sind. Boden gibt zu, dass transformative Kreativität dennoch riskant sei und nur selten erfolgreich umgesetzt werde. Durch das Verlassen bekannter Strukturen sind evolutionäre Programme auf eine *trial-and-error*-Methode angewiesen, die sich nur schwer optimieren lässt. Menschliche Evaluation bleibt daher unabdingbar, um Chaos zu vermeiden, da sich die Programme weiterhin nicht selbstständig anpassen oder entwickeln können.¹²² Allerdings bezieht sich Boden bei ihrer Argumentation vorwiegend auf ökonomische Innovationen, definiert Kreativität vom Objekt her und betrachtet den Wert der (transformativen) Kreativität für ästhetische Prozesse nicht.

Bajohr (2022) merkt an, dass die Definition der Kreativität vom Objekt – auf Produktionsseite, wie sie Boden vornimmt – sowie vom Subjekt her – wie sie in den Neurowissenschaften als geistige Eigenschaft definiert wird – zwar den Begriff messbar und dadurch für einen Computer simulierbar machen, in beiden Fällen aber Kreativität für die Definition von Kunst als entscheidendes Kriterium etwas vermissen lässt. Beide Identifikationen mit Kreativität erlauben keine Unterscheidung mehr zwischen der Schaffung eines Kunstwerkes und der Produktion einer neuen technischen Innovation oder einer besonders begabten Person. Bajohr vermutet hinter der Instrumentalisierung des Kreativitätsbegriffs für die Kunst eine wieder auftauchende Autonomie- und Genieästhetik.¹²³

Wie erwähnt, ist der Begriff der Kreativität relativ jung und retrospektiv mit der Konzeption des romantischen Genies eng verknüpft. Dadurch ergeben sich allerdings andere Probleme, die die Kreativität als Konzept einer ästhetischen Theorie zugunsten der Verhandlung von Originalität und Inspiration ersetzen – dazu in den folgenden Kapiteln noch mehr. Wenn Kreativität hingegen im ästhetischen Rahmen verhandelt wird, wird stattdessen rein funktional nach der Entstehung von Neuem gefragt, ohne

¹²² Vgl. Boden: *AI. It's nature and future* (2016), S. 70.

¹²³ Vgl. Bajohr: *Schreibenlassen* (2022), S. 177-178

historische, soziale oder philosophische Fragen einzubinden, wie Schönthaler (2022) feststellt. Die Debatten darum, wie Neues im Kontext der Kreativität entsteht, behandeln die Kreativität nicht anhand einer klassischen ästhetischen Theorie, fragen nicht nach ihrem Wesen, sondern konzentrieren sich vielmehr auf das ‚Wie‘ als auf das ‚Was‘. Auch die Informationsästhetik Max Benses fragt nicht nach dem Wesen der Kreativität, sondern macht sie sich als Funktionsprinzip zunutze. Schönthaler sieht in den Gegenständen der Kreativität und des Zufalls neue ästhetische Kategorien, die das Schreiben maschinenkompatibel machen und generierte Texte als literarische adeln. Die Avantgarden nutzten den Zufall als Schreibtechnik und machten ihn dadurch als poetologische, produktions- und rezeptionsästhetische Kategorie bestimmend. Die Dadaisten sahen im Zufall eine ordnungssprengende Kategorie und damit eine Kritik an der klassischen Autor:in, weil Beliebiges mit allem in Beziehung tritt und zufällige Assoziationen für nichtig erklärt werden. Im Unterschied dazu sieht Schönthaler im computergenerierten Zufall (Zufallsgeneratoren) einen Bedeutungswandel. Durch die Binarität der Programme folgen alle Operationen einer eindeutigen Logik, die den Zufall bereits methodisch ausschließen. Die Bedeutung verschiebt sich dahingehend zu einer probabilistischen Zufallskonzeption, wo der Zufall nur mehr zu einer Frage der Wahrscheinlichkeit wird und damit zu einer kontrollierbaren Größe. Die Informationsästhetik schließt den Zufall mit der Kategorie des Kreativen kurz, welche durch die Kraft des Zufalls simuliert werden soll. Die in der Psychologie von Guilford verbreitete Auffassung, die Kreativität als messbare Größe zu begreifen, spielt der Informationsästhetik weiter in die Hände und macht sie algorithmisch modellierbar. Dadurch verabschiedet sich auch die Vorstellung, dass das schöpferische Vermögen des Menschen etwas Mythisches oder Geheimnisvolles sei, das sich nur im Inneren der Künstler:in abspiele.¹²⁴ Schönthaler sieht in der Algorithmisierung der Kreativität den bloßen Versuch der computergenerierten Literatur, sich unter den Deckmantel des Künstlerischen zu stellen:

Die Kreativität wird zur Richtschur, um computergenerierte Texte als literarische auszugeben und die Programmierung für produktionsästhetische Fragen zugänglich zu machen. Analog zum Denken, das über die Metapher des Gehirns als Informationsprozessor maschinenkompatibel wird, bildet die Kreativität das wolkige Konzept, um den Akt des Schreibens und literarische Werke als künstlerische anzupreisen.¹²⁵

Er schließt damit, dass sich die Kreativität als eigenständige Kategorie weder in den ästhetischen Diskurs einfügt noch diesen fortsetzt, sondern vielmehr als Schlagwort „zum

¹²⁴ Vgl. Schönthaler: Die Automatisierung des Schreibens (2022), S. 209-237.

¹²⁵ Ebd., S. 235.

opportunen Mittel wird, um den Ballast der philosophischen Ästhetik loszuwerden“¹²⁶ und eine rein funktionale Kategorie bleibt.

Im Feldzug gegen die Künstliche Intelligenz wird die Kreativität anthropozentrisch aufgeladen und mit Kategorien verknüpft, die den Maschinen nicht zugänglich sind, wie der der Autonomie: „Aktuell wird die gesamte Kreativität der Maschinen durch menschlichen Code initiiert und vorangetrieben. Bisher verspürte keine Maschine von sich aus den Drang, sich auszudrücken. Die Maschinen haben offenbar außer dem, was wir ihnen eingeben, nichts mitzuteilen.“¹²⁷ Du Sautoy (2021) hält hier die Erwartungen an eine starke Künstliche Intelligenz (*strong AI*) aufrecht und übersieht damit die Leistungen schwacher KI (*weak AI*). Für ihn gehören Kreativität und Bewusstsein untrennbar zusammen: „Ich glaube nicht, dass wir ohne einen Begriff von Bewusstsein verstehen können, warum wir kreativ sind. [...] Solange eine Maschine kein Bewusstsein hat, wird sie wohl nie mehr als ein Werkzeug zur Erweiterung der menschlichen Kreativität sein.“¹²⁸ Zudem knüpft er die Kreativität an eine weitere (noch recht junge) Kategorie des menschlichen Selbstverständnisses, nämlich die der Freiheit: „[D]ieser kreative Drang ist Ausdruck unseres Glaubens an den freien Willen. [...] Unsere Kreativität ist eng mit unserem freien Willen verbunden, der sich nicht automatisieren lässt. Ein programmierter freier Wille würde der Bedeutung des freien Willens widersprechen.“¹²⁹ Diese anthropozentrische Verschränkung des Kreativitätsbegriffs mit Autonomie, Bewusstsein und freien Willen wird als entscheidende Differenz zwischen Mensch und Maschine herangezogen, um auf die Sonderstellung des Menschen zu beharren und ihn als einzig kreatives Wesen zu positionieren.

Bajohr (2022) kritisiert diese anthropozentrische Angst vor dem Digitalen, die er in Anlehnung an Günther Anders ‚Prometheische Scham‘¹³⁰ in ein ‚Prometheisches Unbehagen‘ umformuliert – also keine Scham aufgrund der menschlichen Unvollkommenheit vor der eigenen Schöpfung, sondern ein Unbehagen, von dieser ersetzt zu werden.¹³¹ Die stets anthropologischen Vergleiche, die Du Sautoy anführt, und die Festlegung der Kreativität als das entscheidende Kriterium von Kunst lehnt Bajohr

¹²⁶ Ebd., S. 236.

¹²⁷ Du Sautoy: *Der Creativity Code* (2021), S. 297.

¹²⁸ Ebd., S. 298; 301.

¹²⁹ Ebd., S. 297.

¹³⁰ Zum Konzept der Prometheischen Scham und Anders Argumentationsgedanken zu technischen Entwicklungen und dem Mensch-Maschine-Verhältnis s. Günther Anders: *Die Antiquiertheit des Menschen. Über die Seele im Zeitalter der zweiten industriellen Revolution*. Ungekürzte Sonderausgabe. München: C. H. Beck 1961, S. 21-95.

¹³¹ Vgl. Bajohr: *Schreibenlassen* (2022), S. 132.

ab, indem er die Begriffe Geist und Genie den aktuell vorherrschenden technischen Möglichkeiten gegenüberstellt.¹³² Wenn der Geist der klassischen Philosophie nach als Emergenz von Bewusstsein verstanden wird und dieses als Voraussetzung für Kreativität gilt, bleibt Maschinen der künstlerische Zugang klar verwehrt. Bajohr hält es aber nicht für fruchtvoll, derzeitige technische Systeme als kreativ-autonom zu denken – eine autonome, also starke Künstliche Intelligenz ist nach aktuellem Forschungsstand nicht realisierbar.¹³³ Zudem kann der Begriff der Autonomie – nach der philosophischen Tradition als selbstbestimmt handelnde Entität definiert – nicht ohne weiteres auf eine Künstliche Intelligenz übertragen werden. Zwar arbeiten die künstlichen Systeme auf ein Ziel hin und treffen zur Erreichung dieses Ziels entsprechend Entscheidungen, sie sind aber keineswegs als autonom zu betrachten, sondern allenfalls noch als autopoietisch, d.h. als selbsterhaltend und rekursiv organisiert.¹³⁴

Es erscheint daher wenig sinnvoll, an einem Kriterium zur Bestimmung und Untersuchung künstlerischer Künstlicher Intelligenz festzuhalten, das an einer Definition starken Bewusstseins festhängt und andere Möglichkeiten einer kreativen KI ablehnt. „Je mehr von Künstlicher Intelligenz erwartet wird, desto menschlicher wird sie gedacht, desto weniger aber wird sie als ein Eigenes gewürdigt“¹³⁵, konstatiert Bajohr, der vorschlägt mit Maschinen anders Literatur und Kunst zu machen, anstatt sie – etwas polemisch formuliert – „über das Stöckchen der eigenen Poetik springen zu lassen.“¹³⁶ Nicht zuletzt entscheiden immer noch Menschen, welche kreative Leistung gewürdigt wird und als Kunstwerk einen besonderen Wert erhält. Kunst ist sozialer Aushandlungsprozess. Anders als in der Psychologie, für die Kreativität als eine messbare Kategorie gilt, unterliegt diese Entscheidung nicht objektiv messbaren Kriterien, sondern Bedingungen der Kontextualisierung und Bewertung kreativer wie künstlerischer Leistung, die diskursiv und historisch wandelbar sind. Algorithmische Kreativitätsprozesse müssen mit bestehenden Diskurs- und Sinnzusammenhängen vereinbar sein, um als solche anerkannt zu werden.¹³⁷

¹³² Bajohr bietet eine Möglichkeit an, die Kategorie Geist maschinenkompatibel zu denken. Indem er die Idee des *extended mind*, die auf Andy Clark und David Chalmers zurückgeht, funktional nutzt, bleibt der Geist nicht allein im Kopf verortet, sondern die Grenze zwischen Mensch, Maschine und Objektwelt wird verwischt, indem der Entstehungsprozess eines Werkes in den Vordergrund gestellt wird. Da Werke selten ohne Unterbrechung produziert werden, lokalisiert Bajohr durch den kybernetischen Rückkopplungsprozess den Geist sowohl im Kopf als auch im benutzten Schreibwerkzeug und dem entstandenen Text an sich. | Vgl. Ebd., S. 36.

¹³³ Vgl. Ebd., S. 35.

¹³⁴ Vgl. Lotze: Chatbots (2016), S. 71

¹³⁵ Bajohr: Schreibenlassen (2022), S. 34.

¹³⁶ Ebd., S. 33.

¹³⁷ Vgl. Budelacci: Mensch, Maschine, Identität (2022), S. 89.

Die Diskussion um den Begriff der Kreativität soll daher kurz weg von einer anthropozentrischen Kränkung hin zu der Frage nach den ästhetischen Wertungskriterien künstlerischer KI geführt werden und wo diese einzuordnen ist. Im folgenden Kapitel wird zunächst ein Blick auf KI-Kunst selbst geworfen und in welchem Verhältnis sie zum Ästhetischen steht. Anhand des Modells einer schwachen KI werden vorerst Bildgeneratoren und ihre Erzeugnisse näher beleuchtet, um der Frage nachzugehen, ob es sich überhaupt um Kunst handelt, bevor die Arbeit die Diskussion um den Begriff der Kreativität und einer damit einhergehenden Geniekonzeption wieder aufgreift und modernen Autorschaftskonzeptionen gegenüberstellt.

3.1.1 Das Ende der Kunst durch KI?

Catrin Misselhorn (2023) hat sich mit der Frage beschäftigt, ob Künstliche Intelligenz Kunst schaffen kann und versucht einen Ansatz für eine ästhetische Interpretation von KI-Kunst zu liefern. Dabei bewegt sie sich im Bereich der schwachen KI. Ihre Definition generativer KI-Kunst beschränkt sich auf Werke, „mit denen der Anspruch verbunden ist, die künstlerische Leistung sei auf KI zurückzuführen. Die Rolle des Menschen reduziert sich darauf, die Systeme zu programmieren oder ihr Stichwortgeber zu sein.“¹³⁸ Behandelt werden ausschließlich Werke, die mithilfe von generativen Computerprogrammen entstanden sind und durch ihre Ununterscheidbarkeit zu anderen Kunstwerken sowie durch die Abwesenheit eines Künstler-Subjekts und dessen handwerklichen Fähigkeiten Fragen aufwerfen, was Kunst eigentlich ist.¹³⁹ Ihre Überlegungen zu den notwendigen Bestandteilen einer Kunstdefinition knüpfen an die Erwägungen des amerikanischen Philosophen Arthur C. Danto (1924-2013), dass alles Kunst sein kann, an¹⁴⁰ und leiten daraus die drei Kriterien Bedeutung, Gehalt und Interpretation ab:

- (1) Ein Kunstwerk muss eine Bedeutung verkörpern.
- (2) Diese besteht in einem Gehalt höherer Ordnung, der eine gewisse Sichtweise zum Ausdruck bringt.
- (3) Das Verständnis des Gehalts höherer Ordnung erfordert eine Interpretation, die den Kontext der Kunstwelt miteinbezieht.¹⁴¹

Misselhorn merkt an, dass es umstritten ist, ob diese Bedingungen tatsächlich notwendig und zusammen hinreichend sind, um Kunst zu definieren. Für die wesentlichen

¹³⁸ Catrin Misselhorn: Künstliche Intelligenz – das Ende der Kunst? Ditzingen: Reclam 2023 (Reclams Universal-Bibliothek [Was bedeutet das alles?]) 14355), S. 7-8.

¹³⁹ Vgl. Ebd., S. 24-25.

¹⁴⁰ Vgl. Ebd., S. 16-20.

¹⁴¹ Ebd., S. 20.

Gesichtspunkte, was Kunst wertvoll macht, sollen sie ihrer Meinung nach aber genügen.¹⁴²

In Bezug auf KI-generierte Werke stellt Misselhorn in Zweifel, ob diese selbst überhaupt als Kunstwerke gelten dürfen. Als Beispiel nennt sie das 2018 erzeugte Werk „*Portrait of Edmond Belamy*“ vom Kunstkollektiv Obvious, das für sich gesehen entscheidende Punkte der vorher aufgestellten Kunstdefinition nicht erfüllt. So bietet es keinen Gehalt höherer Ordnung, der eine gewisse Sichtweise auf die Welt zum Ausdruck bringt. Wesentlich scheint für Misselhorn daher der Rückgriff auf eine Autor:in zu sein, die ästhetische Verantwortung für das Werk und seine Identität übernimmt. Erst in der Performance, die das Künstlerkollektiv Obvious als Autor:in miteinschließt, sind alle genannten Definitionen für ein Kunstwerk eindeutig erfüllt – es handele sich dann aber nicht mehr um KI-Kunst, sondern um Konzeptkunst, die sich KI-Verfahren bediene, um nach dem Wesen von Kunst in der digitalisierten Welt zu fragen, so Misselhorn.¹⁴³

Ihre Kunstdefinition denkt damit von vornherein ein Konzept von Autorschaft mit, das stark an ein autonom handelndes Subjekt appelliert, dem eine Verantwortung über ein Kunstwerk zugeschrieben werden kann. Zwar verfügt schwache Künstliche Intelligenz laut Misselhorn über (ästhetische) Handlungsfähigkeit im funktionalen Sinn – ist also dazu fähig, selbst zu ‚entscheiden‘, was wie dargestellt wird – doch kann sie dafür keine Verantwortung übernehmen, weder ästhetisch noch rechtlich oder ethisch¹⁴⁴:

Sie [Anm.: Künstliche Systeme] können ihre Handlungen weder beabsichtigen und bewusst planen, deshalb sind sie ihnen nicht zurechenbar. Noch verfügen sie über eine Persönlichkeit oder Weltsicht, die sie zum Ausdruck bringen könnten, weshalb ihnen ihr Handeln auch nicht zuschreibbar ist. Sie können eine Persönlichkeit bestenfalls bis zu einem gewissen Grad simulieren [...]. Die Zuschreibung ästhetischer Verantwortung erfordert im Gegensatz dazu, dass der Persönlichkeitsausdruck auch auf die richtige Art und Weise von der Persönlichkeit hervorgebracht wurde.¹⁴⁵

Dem Handeln künstlicher Systeme mangelt es laut Misselhorn an Kontrolle und Zuschreibbarkeit, wodurch ein Verantwortungsvakuum entsteht, das auch nicht die Programmierer:innen oder Künstler:innen hinter den Trainingsdaten füllen können. Die Programmierer:innen tragen natürliche Verantwortung für den Algorithmus, aber keine ästhetische für das dadurch entstandene Werk – die in den Trainingsdaten enthaltenen Künstler:innen sind nur für ihre eigenen Werke ästhetisch wie natürlich verantwortlich, aber stehen in keinerlei verbindlichen Verantwortung zum KI-Werk.¹⁴⁶

¹⁴² Vgl. Ebd., S. 21.

¹⁴³ Vgl. Ebd., S. 30.

¹⁴⁴ Vgl. Ebd., S. 28.

¹⁴⁵ Ebd., S. 55.

¹⁴⁶ Vgl. Ebd., S. 56-57.

Misselhorn lehnt auch eine Co-Autorschaft bzw. eine Zuschreibung kollektiver Autorschaft an eine Künstliche Intelligenz ab, indem sie sich auf die besprochene ästhetische Verantwortung bezieht. Eine Kollaboration zwischen Mensch und Maschine sei ein Trugbild, da Künstliche Intelligenz nicht auf derselben Stufe stehe, wie ein Mensch. Es fehle der Maschine an den notwendigen mentalen Zuständen, um ästhetische Verantwortung zu übernehmen und struktural Autorschaft zu beanspruchen.¹⁴⁷ Wenn Misselhorn konstatiert, dass „Autorschaft bislang ein Privileg menschlicher Subjektivität bleibt [und] KI [...] hingegen keine Autorschaft zu[kommt], weil sie weder im individuellen noch im kollektiven Sinn ästhetische Verantwortung übernehmen kann“¹⁴⁸, dann ist Misselhorns Kunstverständnis stark an ein autonomes Subjekt geknüpft, dessen Anforderungen schwache KI nicht (vollends) erfüllen kann.

Für Misselhorn besteht das Problem mit KI-Kunst in der Simulation bzw. Nachahmung menschlichen künstlerischen Schaffens, ohne über ein entsprechendes Bewusstsein zu verfügen. Um ihre Definition von Kunst zu erfüllen, fordert sie menschliche Qualitäten wie Autonomie und Intention bzw. mentale Zustände (*causal powers*) von einer (schwachen) KI, die eigens aus sich selbst heraus Kunst zu schaffen hat und für eine Vision bzw. höheren Gehalt hinter dem Werk eintreten soll. Eine eigene Sichtweise auf die Welt vermag aber nur ein autonomes Individuum zu leisten – diese Voraussetzung erfüllt keines der bisher geschaffenen KI-Werke. Als letzte Konsequenz wird KI generierte Kunst als ‚echte‘ Kunst abgelehnt und mit Fälschungen bzw. ‚Fakes‘ gleichgesetzt. KI-generierte Kunst sehe zwar wie Kunst aus und erzeuge eine ähnliche Wirkung, sei aber keine, sondern Fake-Art¹⁴⁹. So gebe KI-Kunst vor, aufgrund der Simulation eines Individualstils aus den verwendeten Trainingsdaten fälschlicherweise eine Produktionsgeschichte zu haben, die auf eine Autorschaft am Werk rekurriert. Die Identität des Werkes fehlt aber, ergo es liegt eine Fälschung vor. Dass KI-Kunst aussehe wie Kunst, aber keine sei, führe zu einem Ununterscheidbarkeitsproblem, welches die KI-

¹⁴⁷ Vgl. Ebd., S. 85-86.

¹⁴⁸ Misselhorn: Künstliche Intelligenz – das Ende der Kunst? (2023), S. 92-93.

¹⁴⁹ Mit dem Begriff des ‚Fakes‘ bezeichnet Römer (2024) eine künstlerische Strategie, die mimetische Nachahmung eines Kunstwerks, das im Gegensatz zu einer Fälschung auf den eigenen gefälschten Charakter hinweist und keine Täuschungsabsicht verfolgt. Es zielt auf einen kunsthistorischen Erkenntnisprozess, der die Reproduktion oder Appropriation nicht moralisch als Fälschung verurteilt, sondern als Kritik von Originalität und der Institution von Kunst sieht. | Vgl. Stefan Römer: Der Begriff des Fake. In: Humboldt-Universität zu Berlin (2024) <https://edoc.hu-berlin.de/handle/18452/15158> (Zugriff: 16.05.2024). | Vgl. Stefan Römer: Künstlerische Strategien des Fake. Kritik von Original und Fälschung. Köln: DuMont 2001., S. 14-17.

Kunst näher an ‚Fakes‘ verorte als an modernen Vorstellungen von Kunst, bei denen alles Kunst sein kann.¹⁵⁰

Denn nun ist es nicht mehr so, dass etwas Kunst sein soll, das scheinbar keinerlei Ähnlichkeit mit dem aufweist, was bislang als Kunst galt, sondern etwas ist keine Kunst, obwohl es äußerlich ununterscheidbar von dem ist, was wir bislang als Kunst verstanden haben.¹⁵¹

Misselhorn kritisiert zudem die Invasion der KI-Kunst in die Kunstwelt aufgrund der einfachen und massentauglichen Herstellungsweise durch automatisierte Programme. Damit gefährde KI-Kunst nicht die Kunst an sich, sondern die Kunstpraxis, indem KI-generierte Kunstwerke fälschlicherweise bestimmten Künstlern in Google-Suchen zugeordnet werden.¹⁵²

Auch im Kontext der ästhetischen Erfahrung eines Kunstwerkes besteht Misselhorn auf dem Argument, dass der Rückgriff auf ein Künstlersubjekt unabdingbar für die Unterscheidung zwischen Kunst und Nicht-Kunst ist. Nur in der Begegnung mit der Autor:in im Kunstwerk entsteht eine ästhetische Erfahrung, die sich durch einen Gehalt höherer Ordnung sowie durch die spezifische Art und Weise, wie die Autor:in diesen ausdrückt, manifestiert.¹⁵³ Wenn die Frage aufgeworfen wird, ob es für die ästhetische Erfahrung überhaupt wichtig ist, ob ein Werk eine Autor:in hat bzw. als Kunstwerk gilt – wie sie häufig im Kontext der Demokratisierung von Kunst in Zusammenhang mit KI-generierten Werken gestellt wird – wenn das Kunstobjekt also lediglich als Stimulans betrachtet wird und damit quasi jedes Objekt als Kunst erfahrbar ist, wenn es denn als solches betrachtet wird, dann verweist Misselhorn auf folgende Überlegung: „Wenn wir in der ästhetischen Erfahrung von der Autorschaft abstrahieren, behandeln wir das Kunstwerk wie einen schönen Sonnenuntergang oder eine Wolkenformation, die unser Vorstellungsvermögen anregt, aber keine Kunst ist.“¹⁵⁴

Misselhorn erkennt die ästhetischen Eigenschaften mancher generierten Werke an, bspw. der Promptografie „*Pseudomnesia: The Electrician*“ (2022) des Berliner Fotografen Boris Eldagsen: Sie beschreibt die Faszination, die das Bild ausübt, und schätzt den „kunstvolle[n] Wechsel von Schärfe und Unschärfe[, der das] Geheimnisvolle des Bildes“¹⁵⁵ verstärkt. Dadurch, dass Eldagsen das Werk allerdings schuf, um eine Debatte über die Bewertung von KI-Fotografie anzustoßen, und demnach auch im

¹⁵⁰ Vgl. Misselhorn: Künstliche Intelligenz – das Ende der Kunst? (2023), S. 112-122.

¹⁵¹ Ebd., S. 122-123.

¹⁵² Vgl. Ebd., S. 125-127.

¹⁵³ Vgl. Ebd., S. 100.

¹⁵⁴ Ebd., S. 103.

¹⁵⁵ Ebd., S. 75.

Frühjahr 2023 einen Fotopreis für das Bild ablehnte, ist es als Konzeptkunst über den Status von Kunst und insbesondere Fotografie zu verstehen und nicht als KI-Kunst ohne Autor:in. KI-Kunst scheint nur über das Prinzip der Täuschung zu funktionieren, das Künstlicher Intelligenz so inhärent ist. Wird die Herkunft des Werkes erfolgreich verschleiert, sieht Misselhorn den ‚ästhetischen Turingtest‘ als bestanden.¹⁵⁶ Wenn die Künstler:in bzw. die Autor:in aus dem Werk bewusst entfernt wird, verliert die Arbeit aber an künstlerischem Wert.

Damit ordnet sich KI- bzw. *prompt*-generierte Kunst als Konzeptkunst einer Tradition unter, welche die traditionellen Vorstellungen von kreativem Schaffen auf den Kopf gestellt hat und sich durch die Geste des Zur-Kunst-Erklärens bzw. durch ein Auswählen, Arrangieren, Organisieren und Ausstellen definieren, als die tatsächliche Schaffung eines originären Kunstobjekts. Obwohl damit die Rolle der Künstler:in bei der Produktion des Werkes abnimmt, und zum Teil nur mehr auf den Akt der Appropriation (Aneignung) zurückzuführen ist¹⁵⁷, übernimmt sie die Funktion der Diskursleiter:in und erhebt Urhebererschaft auf die dem Konzept zugrundeliegende Idee.¹⁵⁸

Künstlertum ist nicht mehr die Kreation von Neuem, sondern ein Organisieren von Verfremdungseffekten im performativen Kontext des Austellens [sic!] selbst, und Autorschaft repräsentiert nicht mehr die Authentizität des Urhebers, sondern die beim Arrangieren, Argumentieren und Adaptieren der Werkelemente notwendige konzeptuelle Kompetenz der Aneignung und Vermittlung.¹⁵⁹

Die Relevanz der Künstler:in – und ebenso der Autor:in – als schöpferisches Individuum ist zugunsten neuer Autorisierungsstrategien zurückgegangen, welche sich in der konzeptuellen Repräsentation des Kunstwerks sowie in der ästhetischen Erfahrung auf Seiten der Rezipient:innen eröffnen. Autorschaft wird durch einen zufallsabhängigen Automatismus überwunden, Sinn und Bedeutung werden von der Betrachter:in auf das Kunstwerk projiziert. Dabei ist die Ausführung weniger von Belang als das Konzept und die Idee dahinter. Dies im Hinterkopf behaltend, widmet sich das folgende Kapitel wieder dem vorher aufgespannten Kreativitätsdiskurs und damit dem souveränen Genie auf literaturtheoretischer Ebene. Im Anschluss sollen die Erkenntnisse poststrukturalistischen

¹⁵⁶ Vgl. Ebd., S. 63-64.

¹⁵⁷ Erwähnenswert ist hierfür die Kunstform der *Ready-mades* – Alltags- oder Naturgegenstände, die als solche zum Kunstgegenstand ernannt wurden. Bekanntestes Beispiel ist wohl das mit „R. Mutt“ signierte Pissoir, welches von Marcel Duchamp bei den New Yorker Independents eingereicht wurde. Das *Ready-made*-Prinzip wurde außerdem von der Konzeptkunst aufgegriffen, die Kunst damit von ihrer Objekthaftigkeit löst und nur noch konzeptuell repräsentiert. Autorschaft steckt hier in der Konzeption. Die Aktualität der *Ready-mades* ergibt sich aus ihrer ursprünglichen lexikalischen Bedeutung ‚*ready-made*‘, nämlich der Ersetzung der Individualität durch Massenproduktion. | Vgl. Dieter Daniels: Marcel Duchamp – der einflussreichste Künstler des 20. Jahrhunderts? In: Museum Jean Tinguely Basel (Hg.): Marcel Duchamp. Ostfildern-Ruit: Hatje Cantz 2002, S. 27-33.

¹⁵⁸ Vgl. Wetzel: Historischer Abriss (2022), S. 151-152.

¹⁵⁹ Ebd., S. 152.

Theorien gegenübergestellt werden, um Autorschaft abseits dieser engen Kopplung neu zu verhandeln.

3.2 Das kreative Genie

Wie in den vorangegangenen Kapiteln bereits erwähnt, gibt es eine enge theoretische Kopplung in der Konzeption des Genies mit dem Begriff der Kreativität und literarischer Autorschaft, die besonders in der zweiten Hälfte des 18. Jahrhunderts ihren Höhepunkt hatte. Die Kreativitätsmythen vom schöpferisch-schaffenden Genie legitimieren ein Autorschaftskonzept, welches die Deutung und das Verstehen eines Werkes stark an die Autorität der Autor:in zurückbindet.¹⁶⁰ Diese Mythen von Originalität und ‚Werkherrschaft‘¹⁶¹ verteidigen die Souveränität der Autor:in sowohl hermeneutisch im Werk als auch juristisch durch das Urheberrecht (Copyright). Die Autor:in als Genie ist im hermeneutischen Werkbegriff sinnstiftendes Individuum. Die kreativen Entscheidungen und die Individualität der empirischen Autor:in prägen die Eigentümlichkeit des literarischen Werkes und authentifizieren es. Das Verstehen des Werkes hängt mit dem Verstehen der Autor:in und ihrer Intention zusammen. Als interpretationsrelevante Kategorien gehören diese zum festen Bestandteil traditionell hermeneutischer Ansätze, die bis dato erhebliche theoretische Differenzierung erfahren haben.¹⁶² Das Konzept einer Intention oder latenten Botschaft der Autor:in im Werk ist aus Sicht unterschiedlicher theoretischer Konzeptionen und Blickwinkel mit Paradoxien und Widersprüchen verbunden und von einer genieästhetischen und stark subjektorientierten Konzeption abhängig, die aktuellen ökonomischen wie produktionsästhetischen Verhältnissen nicht gerecht wird. Unter Berücksichtigung der sich historisch verändernden Rezeptionsbedingungen, verändert sich auch die Deutung des Werks und kann mit jeder Epoche jeweils stets Neues bedeuten wollen, wodurch eine Intention der Autor:in zusätzlich an Relevanz verliert.¹⁶³

¹⁶⁰ Vgl. Johannes F. Lehmann: Genie und Autorschaft. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 259-260.

¹⁶¹ Obwohl im Urheberrecht der Begriff ‚Herrschaft‘ sowie ‚Autor‘ nicht vorkommen, hat sich der Begriff der ‚Werkherrschaft‘ als Vorstellung einer Souveränität der Autor:in über ihr Werk durchgesetzt und wird bis heute tradiert. Mit dem Begriff entsteht eine enge Verbindung zwischen Autor:in und Werk aufgrund der starken Verknüpfung poetischer bzw. ästhetischer Kategorien sowie der Philosophie mit der Ebene des Rechts und der Ökonomie. | Vgl. Matthias Schaffrick: Autorsubjekt und Werkherrschaft. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 203-204.

¹⁶² Vgl. Ebd., S. 213.

¹⁶³ Vgl. Wetzel: Einleitung (2022), S. 5. || Der Werkbegriff impliziert außerdem Vorstellungen von einer Einheit, Abgeschlossenheit und Autor:innenintention, die er angesichts der Eigendynamik der Sprache

William K. Wimsatt (1907-1975) und Monroe C. Beardsley (1915-1985) werten die Bedeutung einer Intention, die Aufschluss über die Qualität und Bedeutung eines Textes geben soll, konsequent ab. In ihrem 1946 in der Zeitschrift *The Sewanee Review* 54 erschienenen und als Dokument des New Criticism zu lesenden Aufsatz „*The Intentional Fallacy (Der intentionale Fehlschluss)*“ richten sie sich gegen den zu der Zeit vorherrschenden Biographismus und plädieren für eine Beurteilung und Deutung von Texten allein über dessen textinterne Merkmale. „Wir stellen die Behauptung auf, dass die Absicht oder die Intention des Autors weder eindeutig erkennbar noch ein wünschenswerter Maßstab ist, um den Erfolg eines literarischen Werkes zu beurteilen“.¹⁶⁴ Die Intention einer Autor:in sei ein „Entwurf oder Plan im Kopf“, der mit den Einstellung und den Gefühlen der Person zusammenhänge und damit als „Ursache“ für eine dichterische Leistung herhalten könne, aber nicht als deren „Maßstab“.¹⁶⁵ Die tatsächliche Intention einer Person einzig über die Rezeption eines Textes zu rekonstruieren, halten Wimsatt und Beardsley für unmöglich, sofern sie nicht am Text selbst erkennbar wäre. Wenn es dem Dichter gelingt, seinen Plan umzusetzen, dann gibt der Text Aufschluss über dessen Intention – der Text funktioniere aber auch ohne dessen Gelingen, weil er eine Bedeutung aufweist, die sich aus textinternen Merkmalen bzw. „ganz allgemein durch alles, was eine Sprache und Kultur ausmacht“¹⁶⁶ ergibt. Wimsatt und Beardsley verneinen damit nicht die Existenz einer Autor:innenintention, sprechen dieser aber keine Relevanz für die Untersuchung und Beurteilung literarischer Texte zu. Sie unterscheiden deshalb auch zwischen einer dramatischen Sprecher:in, die dem Text innerlich ist und der die ausgedrückten Gedanken und Haltungen zugeordnet werden – sofern keine biographischen Rückschlüsse zulässig sind – sowie der empirischen Autor:in, der die Textintention – sei sie im Text sichtbar oder nicht – zukommt, dadurch

nicht aufrechterhalten kann. Dies zeigt sich an den Widersprüchen in der Beziehung von Autor:in und Werk, die „zwischen ästhetischer Autonomie und literarhistorischer Heteronomie, zwischen Endlichkeit des Werks und Unendlichkeit des Sinns, zwischen textueller Geschlossenheit und intertextueller Offenheit, zwischen schriftlicher Stabilität und Zerstreuung“ schwankt. | Vgl. Schaffrick: *Autorsubjekt und Werkherrschaft* (2022), S. 204. | Ebd., S. 204.

¹⁶⁴ William K. Wimsatt/Monroe C. Beardsley: *Der intentionale Fehlschluss*. In: Fotis Jannidis/Gerhard Lauer/Matias Martinez et. al. (Hg.): *Texte zur Theorie der Autorschaft*. Stuttgart: Reclam 2000, S. 84.

¹⁶⁵ Ebd. S. 85.

¹⁶⁶ Ebd. S. 89. || Textexterne Positionen, wie Kommentare, Anspielungen, intertextuelle Bezüge oder Anmerkungen, gäben laut Wimsatt/Beardsley (2000) ebenso keinen Aufschluss über die Intention der Autor:in, da sie zur individuellen Wortbedeutung bzw. -geschichte sowie zur Bedeutung des Werkes und seinem Kontext beitragen. Für die Plausibilität des Textes ist es dabei nicht von Belang, ob die Anmerkungen wahr oder falsch sind, da dies keinen Unterschied für die Deutung oder das Verständnis des Gedichts machen würde. Diese textexternen Teile sind damit ebenso als Teil des Werkes zu lesen, weil sie eine informationsvermittelnde Funktion erfüllen, die zur Wirkung des Werkes beiträgt. | Vgl. Ebd., S. 98.

aber keinerlei Eigentumsrechte am Text besitzt. Mit der Veröffentlichung verliert die Autor:in ihr Verhältnis zum Werk, es gelangt in den Besitz der Öffentlichkeit. Der Text ist kein biographisches Dokument, sondern literarisches Kunstwerk und damit an keine subjektive Sichtweise gebunden. Als autonomes Gebilde besitzt es eine eigene ästhetische Existenz, die es zu erfassen gilt. Zur Beurteilung des Werkes ist eine Autor:innenintention nicht notwendig.¹⁶⁷

Wimsatt und Beardsley reihen sich mit ihrer intentionskritischen Haltung und textbezogenen Autorkritik in eine Tradition ein, die durch Strukturalismus sowie Russischem und Prager Formalismus zu Beginn des 20. Jahrhunderts die individuelle Schöpfungskraft und eine sich aus der Position der Autor:in ergebende literarische Sinngebung relativiert haben. Grob zusammengefasst haben sie die Autor:in als Phänomen der Literatur und Illusion der Leser:in positioniert, die in einem autonom funktionierenden Sprachsystem keinen Stellenwert mehr hat. Der aus dieser Position entstandene ‚Post‘-Strukturalismus stellt keinen völlig neuen Paradigmenwechsel dar, führt aber das Gedankengut einer Autonomie der Sprache weiter und verändert damit wesentlich das literarische Verhältnis von Autor:in, Text und Leser:in. Die im folgenden Kapitel besprochenen Aufsätze „*Der Tod des Autors*“ (1968) von Roland Barthes (1915-1980) und „*Was ist ein Autor?*“ (1969) von Michel Foucault (1926-1984) stellen zwei Paradebeispiele poststrukturalistischer Theorien dar, die einerseits die Autor:in als sinnstiftende Instanz verabschieden und damit konsequenterweise die Position des Textes sowie der Leser:in hervorheben, andererseits wesentlich zur Entwicklung einer Diskussion um Autorschaft beigetragen haben, die abseits individueller Autor:innen und deren Überhöhung gedacht werden kann und gerade im heutigen Kontext Künstlicher Intelligenzen aufschlussreich für neue Konzeptionen von Autorschaft erscheinen.¹⁶⁸ Die Positionen Barthes‘ und Foucaults sollen demnach hier im Detail erläutert werden.

¹⁶⁷ Vgl. Ebd., S. 86-87.

¹⁶⁸ Vgl. Wetzel: Einleitung (2022), S. 10-11.

4 ‚Der Autor‘ wird vom Thron gestoßen

Das Werk, das die Aufgabe hatte, unsterblich zu machen, hat das Recht erhalten, zu töten, seinen Autor umzubringen.¹⁶⁹

Der Aufsatz „*Der Tod des Autors*“ (1968) vom französischen Philosophen, Schriftsteller und Literaturkritiker Roland Barthes (1915-1980) gehört zu den einflussreichsten Texten, die für eine Verabschiedung der Autor:inneninstanz aus der literarischen Interpretation eintreten. Bei der polemischen Rhetorik über den ‚Tod des Autors‘ handelt es sich nicht, wie vielfach wortwörtlich aufgefasst und in eine darauffolgende lange Tradition der Wiederbelebung oder ‚Rückkehr des Autors‘ gestellte Theoretisierung, um das tatsächliche Abschaffen oder die Leugnung der realen, empirischen Autor:in bzw. ihrer Position, sondern um einen rhetorischen Gestus, der sich der Suche nach den Grenzen des Autor:innenbegriffs widmet sowie kritisch nach der tatsächlichen Autorität eines scheinbar souverän entscheidenden Subjekts fragt.¹⁷⁰ Indem Barthes die Autor:in als ideologisches Konstrukt und konzeptuelle Repräsentation von Autorschaft auffasst, macht er deutlich, dass „die Erzählung nie von einer Person getragen [wird], sondern von einem Mittelsmann, einem Schamanen oder Vortragenden, dessen »Performanz« (das heißt dessen Beherrschung des Erzählcodes) man äußerstenfalls bewundern kann, aber nie dessen »Genie«.“¹⁷¹ Die Autor:in wird demnach als moderne Figur hervorgebracht, ebenso wie der Glaube an ein Genie, das über das literarische Werk herrscht. Barthes Kritik richtet sich an die Konstruiertheit der Autor:innenfigur sowie deren Monopol an der Textdeutung und Überhöhung seiner Zeit:

[D]as Bild der Literatur, das man in der gängigen Kultur antreffen kann, ist tyrannisch auf den Autor ausgerichtet, auf seine Person, seine Geschichte, seine Vorlieben und seine Leidenschaften; [...] Die *Erklärung* des Werks wird immer auf Seiten desjenigen gesucht, der es hervorgebracht hat, als »spräche sich« durch die mehr oder weniger durchsichtige Allegorie der Fiktion hindurch letztlich immer die Stimme ein und derselben Person »aus«, nämlich des *Autors*.¹⁷²

Barthes lehnt eine klassisch-hermeneutische Auslegung von Literatur durch die Autor:innenintention ab, indem er die Ursprünglichkeit des Textes nicht mehr an der Autor:in festmacht, sondern diese in der alleinigen performativen Kraft der Sprache

¹⁶⁹ Michel Foucault: Was ist ein Autor? In: Fotis Jannidis /Gerhard Lauer/Matias Martinez et. al. (Hg.): Texte zur Theorie der Autorschaft. Stuttgart: Reclam 2000, S. 204.

¹⁷⁰ Vgl. Wetzel: Einleitung (2022), S. 6.

¹⁷¹ Roland Barthes: Der Tod des Autors. In: Roland Barthes: Das Rauschen der Sprache. (Kritische Essays IV). Aus dem Französischen von Dieter Hornig. Frankfurt am Main: Suhrkamp 2006 (edition suhrkamp 1695), S. 57.

¹⁷² Ebd., S. 58.

verortet. Es „spricht die Sprache, nicht der Autor“¹⁷³. Indem die Sprache allein agiert – „»performiert«“¹⁷⁴ – ist die Autor:in als Figur nur leeres Subjekt, keine Person. Barthes Auffassung der Sprache als *écriture* – das sowohl mit ‚Schrift‘ als auch ‚Schreibweise‘ übersetzt werden kann – übernimmt die wesentliche Funktion, das Subjekt zu dezentrieren und erlaubt ein neutrales Schreiben ohne Stimme oder Ursprung, außer in der Sprache selbst.¹⁷⁵ In der Sprache geht die Identität verloren – im Schreiben ist der ‚Tod des Autors‘ vollendet. Schreiben wird so zur unpersönlichen, rein ausgeführten Praxis, ist Mittel zum Zweck: „Sprachlich gesehen ist der Autor nie mehr als derjenige, der schreibt“.¹⁷⁶ Damit tritt der moderne ‚Schreiber‘, als bloßer Ausführer seiner Praxis, die Nachfolge des ‚Autors‘ an.

Durch die systematische Entfernung der Figur der Autor:in aus der Deutungsebene stellt Barthes die Grundvoraussetzung für den Begriff des Werkes, der stark an ein klassisches Subjekt als sinnbildende Kraft gebunden ist, infrage und ersetzt ihn durch den des Textes. Der Textbegriff, als ein „Geflecht aus Zitaten“¹⁷⁷, betont die Ablehnung jeglichen Ursprungs abseits von jenem in der Sprache, die sich einem Wörterbuch gleich von der Schreiber:in immer nur neu zusammensetzen und nachahmen lässt. Die Schreiber:in existiert aber niemals vor der Sprache oder vor dem Text, sondern entsteht gleichzeitig mit diesem. Damit stellt Barthes einen dynamischen und offenen Textbegriff im Unterschied zur geschlossenen Einheit des Werkes vor, welches stets unter ‚Beigabe‘ einer Autor:in mit „einem letzten Signifikat“¹⁷⁸ versehen und vom Schreiben abgeriegelt wird. Der Sinn und das ‚hinter‘ dem Werk wird unter der Figur der Autor:in festgelegt und zur Erklärung ihrer Intention herangezogen. Das Schreiben hingegen entzieht sich dieser Festlegung, entzieht sich, „entziffer[t]“¹⁷⁹ zu werden und verweigert den Inhalt: „das Schreiben setzt fortwährend Sinn, aber immer nur, um ihn zu verflüchtigen: es nimmt eine systematische Freistellung des Sinns vor.“¹⁸⁰ Indem verhindert wird, dass die Autor:in als letzte Sinninstanz dem Text zugewiesen wird, verändert sich der Bezug zwischen Autor:in und Text und macht den Blick für eine Praxis des Lesens frei. Wenn die Figur der Autor:in verschwindet, wird nicht mehr nach einer Autorität über den Text oder nach Eigentumsrechten am Werk gefragt, sondern nach den Sinnpotenzen des Textes

¹⁷³ Ebd. S. 58

¹⁷⁴ Ebd. S. 58

¹⁷⁵ Vgl. Wetzel: Einleitung (2022), S. 11

¹⁷⁶ Barthes: Der Tod des Autors (2006), S. 59.

¹⁷⁷ Ebd. S. 61.

¹⁷⁸ Ebd. S. 62.

¹⁷⁹ Ebd. S. 62.

¹⁸⁰ Ebd. S. 62.

im Kontext seiner Rezeption, seiner Intertextualität und endlosen Neuschreibung. In der Leser:in sammelt sich die kulturelle Vielfalt des Schreibens in Zitaten. Dabei ist ihre Figur ebenso wenig personal zu denken, wie die der Autor:in, sondern als Funktion, als „ein Mensch ohne Geschichte, ohne Biographie, ohne Psychologie; er ist nur dieser *jemand*, der in einem einzigen Feld alle Spuren zusammenhält, aus denen das Geschriebene besteht.“¹⁸¹ In seiner viel zitierten Schlussformel bezahlt Barthes daher die „Geburt des Lesers [...] mit dem [Anm.: symbolischen] Tod des ›Autors‹“.¹⁸²

Die Figur der Autor:in konzeptuell als Funktionsweise zu denken, findet sich bei den Überlegungen des französischen Philosophen Michel Foucaults (1926-1984) in seinem Vortrag „*Was ist ein Autor?*“ (1969) am Collège de France in Paris wieder. Seine Kritik richtet sich wie bei Barthes an das Autor:innensubjekt, an den Werkbegriff, an einer in der Intention der Autor:in begründeten Deutungshoheit und deren Anspruch am Text. So konstatiert Foucault: „der Autor ist genaugenommen weder der Eigentümer seiner Texte, noch ist er verantwortlich dafür“¹⁸³ und stellt im zeitgenössischen Schreiben eine Gleichgültigkeit demjenigen gegenüber fest, der schreibt bzw. spricht. Das geborgte Beckett-Zitat: „Wen kümmert's, wer spricht?“¹⁸⁴ zeigt eine Haltung, die sich in vielerlei Hinsicht von der Figur der Autor:in abgewandt hat und das Schreiben nicht als Ergebnis einer vorgängigen und äußerlichen Autor:in sieht, sondern sich als Praxis versteht, als „ein Zeichenspiel [...], das sich weniger nach seinem bedeuteten Inhalt als nach dem Wesen des Bedeutenden richtet“¹⁸⁵. Auch Foucault benutzt den explizit auf Barthes bezogenen Begriff der *écriture*, der ‚Schreibweise‘, und fokussiert damit auf den ‚Schreibakt‘, der keine Originalität mehr kennt, sondern nur ein Zirkulieren der Zeichen, durch das die Literatur und durch diese hindurch die Figur der Autor:in spricht.¹⁸⁶ Das Schreiben hat sich vom Ausdruck befreit, ist auf sich selbst bezogen und sich selbst äußerlich. Im Schreiben verwischen alle individuellen Bezüge zum Subjekt, es entsteht ein Raum, in dem das schreibende Ich kontinuierlich entsteht und wieder verschwindet. Die Schriftsteller:in kennzeichnet sich im Text nur noch durch ihre Abwesenheit – die Figur der Autor:in ist nur mehr ein funktionelles Prinzip und als Geste gegenwärtig.¹⁸⁷

¹⁸¹ Ebd. S. 63.

¹⁸² Ebd. S. 63.

¹⁸³ Foucault: *Was ist ein Autor?* (2000), S. 198.

¹⁸⁴ Ebd., S. 198.

¹⁸⁵ Ebd., S. 203.

¹⁸⁶ Wetzel: *Einleitung* (2022), S. 17-19.

¹⁸⁷ Vgl. Giorgio Agamben: *Der Autor als Geste*. In: Giorgio Agamben: *Profanierungen*. Aus dem Italienischen von Marianne Schneider. Frankfurt am Main: Suhrkamp ⁵2015 (edition suhrkamp 2407), S. 57-62.

„Der Autor ist nur der Zeuge, der Garant seines eigenen Fehlens in dem Werk, in dem er aufs Spiel gesetzt wurde; und der Leser kann dieses Zeugnis nur wieder nachlesen, nur seinerseits Garant werden für sein eigenes unerschöpftes Spiel, sich zu verfehlen.“¹⁸⁸

Statt Barthes' Bezug auf das System der Sprache als Ursprung des Schreibens, führt Foucault die Kategorie der Diskurse ein, die funktional die Leerstelle füllen, die das Zurücktreten der Figur der Autor:in hinterlassen hat. Sie markieren Autorschaft durch einige Merkmale, die Foucault beispielhaft anführt: So haben Autor:innen seit Ende des 18. Jahrhunderts und Anfang des 19. Jahrhunderts ein juristisch kodifiziertes Eigentumsverhältnis zum Text gewonnen und sind an ein Staats- und Rechtssystem gebunden. Der Text ist von einem Akt zu einem Gut geworden. Dadurch offenbart sich die historische Wandelbarkeit der Zuschreibung von Autorschaft, die nicht einheitlich anwendbar und nicht ständig gültig ist.¹⁸⁹ Diese Zuschreibung geschieht nicht spontan, sondern ist eine komplexe Operation, die ein Autor:innenbild zum Zwecke eines einheitlichen Umgangs mit deren Texten konstruiert.¹⁹⁰ Schließlich trägt der Text selbst Hinweise bzw. Zeichen, die auf die Autor:in verweisen. Diese Zeichen, unter denen vorwiegend Personalpronomen, Verbkonjugationen und Adverbien des Ortes und der Zeit verstanden werden, wirken in Diskursen mit ‚Autor-Funktion‘ anders als in denen ohne. Ohne ‚Autor-Funktion‘ ‚zeigen‘ sie auf die reale Sprecher:in und die räumlich-zeitliche Koordinate des Diskurses. In Diskursen mit ‚Autor-Funktion‘ ist die Distanz zwischen Bezeichnenden und Bezeichnetem variabler – das ‚Ich‘ im Text verweist weder auf die empirische Schriftsteller:in, eine Schreibgeste noch ein reales Individuum, „sondern auf ein alter *ego*, dessen Distanz zum Schriftsteller verschieden groß sein und im selben Werk auch variieren kann.“¹⁹¹

Foucault differenziert zwischen den vorgestellten realen Funktionen der Autor:in im Diskurs und der Autor:in als „ideologisches Produkt“¹⁹², als geniale Schöpfer:in, die einen Überschuss an Bedeutung produzieren soll, aber in genau entgegengesetzter Weise funktioniert:

¹⁸⁸ Ebd., S. 68.

¹⁸⁹ Foucault zieht hier einen historischen Vergleich zwischen literarischen und wissenschaftlichen Texten, die im Laufe der Jahrhunderte jeweils den Umgang mit ihren Autor:innen ins Gegenteil verkehrten. So zeichneten sich frühe literarische Werke durch Anonymität aus, während wissenschaftliche Texte ohne Autornamen keinen Wahrheitswert besaßen und heute um ihrer selbst willen akzeptiert werden. | Vgl. Foucault: Was ist ein Autor? (2000), S. 212-213.

¹⁹⁰ Eine wichtige Rolle spielt dabei der Autor:innenname, unter dem ein konstantes Wertenniveau und eine stilistische Einheit definiert und beglaubigt wird. Indem der Name eine klassifikatorische Funktion erfüllt, gruppiert er eine gewisse Zahl von Texten oder grenzt sie voneinander ab und stellt somit Einheitlichkeit her. | Vgl. Ebd., S. 210-211.

¹⁹¹ Ebd., S. 216.

¹⁹² Ebd., S. 228.

Der Autor ist nicht die unendliche Quelle an Bedeutungen, die ein Werk füllen; der Autor geht den Werken nicht voran, er ist ein bestimmtes Funktionsprinzip, mit dem, in unserer Kultur, man einschränkt, ausschließt und auswählt; kurz gesagt, mit dem man die freie Zirkulation, die freie Handhabung, die freie Komposition, Dekomposition und Rekombination von Fiktion behindert.¹⁹³

Die Figur der Autor:in tritt bei Foucault als Regulator:in, als Prinzip der Ordnung, Verknappung und Gruppierung von Diskursen auf. Er versucht aber auch Diskurse ohne die Funktion ‚Autor‘ zu denken:

Man kann sich eine Kultur vorstellen, in der Diskurse verbreitet oder rezipiert würden, ohne daß die Funktion Autor jemals erschiene. [...] Ganz gleich, welchen Status, welche Form und welchen Wert ein Diskurs hätte und welche Behandlung man ihm angedeihen ließe, alle würden sich in der Namenlosigkeit des Gemurmels entrollen.¹⁹⁴

Foucault malt sich damit autorlose Texte aus, die heute bereits Realität sind. Informationssysteme und andere Maschinenkommunikation, wie Wettermeldungen oder automatisierte E-Mails, führen zu einer Zirkulation von nichtliterarischen, unmarkierten Texten, bei denen sich die Frage nach Autorschaft gar nicht mehr stellt. Während Barthes und Foucault noch die notwendige Existenz einer empirischen Autor:in akzeptieren müssen, verschwindet diese Instanz hinter Künstlicher Intelligenz fast gänzlich. Digitale Literatur als Ergebnis maschinengenerierter Texte scheint den vollständigen, wörtlich gemeinten ‚Tod des Autors‘ zu versprechen. Schreiben wird im wahrsten Sinne des Wortes unpersönlich und eine rein durch automatisierte Algorithmen ausgeführte Praxis, die von einer realen Person losgelöst ist – allerdings nicht vollständig. Es wurde bereits gezeigt, dass eine schwache künstlerische KI durchaus auf ein menschliches Pendant angewiesen ist. Sie kann sich nicht mit anthropozentrischen Vorstellungen von Autonomie und Intentionalität messen, reduziert die Verweise auf eine menschliche Autorschaft allerdings auf ein Minimum. Dadurch als auch durch Foucaults Konzeption einer Autorschaft als Funktion sind Eigentums- wie Verantwortungsverhältnisse zum Text von der empirischen Autor:in entkoppelt. Der Text gewinnt an noch nie dagewesener Eigenmächtigkeit, eine subjektzentrierte Autorschaft wird zugunsten eines Automatismus neu verhandelt, aber erfährt keine radikale Auflösung. Was Poststrukturalismus und die Avantgarden bereits an Relevanz der Autor:in weggehackt haben, scheint auch die digitale Literatur nicht vollends totzukriegen. Solange die Entwicklung starker Künstlicher Intelligenz nicht gelingt, bleibt der auktoriale Akt zumindest in der Ausführung bzw. den Abbruch eines Produktionsprogramms als Mindestmaß vorhanden.¹⁹⁵ Das Autorkonzept verschiebt sich und koppelt sich an

¹⁹³ Ebd., S. 228.

¹⁹⁴ Ebd., S. 227.

¹⁹⁵ Vgl. Bajohr: Schreibenlassen (2022), S. 29.

verschiedene Bedingungen, die in enger Korrelation mit den Produktionsbedingungen digitaler Literatur stehen. Im Laufe der Arbeit wurde mehrfach der Terminus ‚computergenerierte Literatur‘ oder ‚digitale Literatur‘ verwendet, ohne eine eingrenzende Definition zu geben. Dies soll im folgenden Kapitel nachgeholt werden.

4.1 Das Digitale der digitalen Literatur

Wenn von digitaler Literatur die Rede ist, treten viele Termini zu ihren unterschiedlichen Formen zutage, die eine Einschränkung und Erläuterung des tatsächlich Gemeinten nur bedingt erleichtern. So vielfältig die Produktionsbedingungen digitaler Literatur sind, so vielfältig sind auch ihre Unterbegriffe. Unter digitaler Literatur lässt sich zusammenfassend etwa mit Blick auf die technische Seite eine Literaturform bezeichnen, die laut Bajohr (2022) „zu einem großen oder zumindest wesentlichen Teil durch die Ausführung formalisierter Algorithmen“¹⁹⁶ entstanden ist. Darunter fallen Arten computergenerierter, konnektionistischer und sequenzieller Literatur (also als ‚KI-Literatur‘ bezeichnete Formen), „generative Codeliteratur“¹⁹⁷, maschinenanaloges Schreiben, unkreatives Schreiben („*uncreative writing*“¹⁹⁸), digitales Schreiben¹⁹⁹ usw. – aber auch Netzliteratur, Mitschreibeprojekte im Internet oder Computerspiele²⁰⁰ können gemeint sein. Berücksichtigt man die mathematischen Grundlagen formalisierter Algorithmen, wie Kombinatorik und Permutation, lassen sich unter dieser Definition digitaler Literatur auch Texte einordnen, deren Ursprung im Mittelalter oder noch weiter zurückliegen, auch wenn nach einem Großteil der heutigen Begriffsbestimmung davon ausgegangen wird, dass damit Literatur gemeint ist, die ohne einen Computer nicht herstellbar wäre. Digitale Literatur bezeichnet ferner unter Rückgriff auf eine Rezeptionsperspektive gemeinhin elektronische und digitalisierte Literatur – erstere ist

¹⁹⁶ Ebd., S. 70-71.

¹⁹⁷ Ebd., S. 71.

¹⁹⁸ Kenneth Goldsmith: *Uncreative Writing. Sprachmanagement im digitalen Zeitalter*. Erweiterte deutsche Ausgabe. Berlin: MSB Matthes & Seitz 2017, S. 27.

¹⁹⁹ Digitales Schreiben ist gänzlich von analogem zu unterscheiden und hat die Kulturtechnik des Schreibens stark verändert. Es ist als hybride Praxis zwischen Mensch und Maschine verankert und besteht nun vielmehr in einem Auswählen von Zeichenketten (z.B. Tippen auf Tastatur, Klicken mit dem Mauszeiger, Auswählen über den Touchscreen etc.) als in der klassisch motorischen Bewegung des Schreibens mit Stift und Papier. Erst in der Interaktion zwischen Mensch und Technik entsteht der Text, beide Akteur:innen sind an der Herstellung beteiligt. Lobin (2014) hat sich näher mit den neuen Kulturtechniken des Schreibens auseinandergesetzt und hat neben dem digitalen Schreiben auch ein hybrides, automatisches, multimediales und soziales Schreiben durch computergestützte Schreibhilfen und digitale Vernetzung festgemacht. | Vgl. Henning Lobin: *Engelbarts Traum. Wie der Computer uns Lesen und Schreiben abnimmt*. Frankfurt am Main: Campus 2014, S. 123-153.

²⁰⁰ Vgl. Simone Winko: *Hyper – Text – Literatur. Digitale Literatur als Herausforderung an die Literaturwissenschaft*. In: Harro Segeberg/Simone Winko (Hg.): *Digitalität und Literalität. Zur Zukunft der Literatur*. München: Wilhelm Fink 2005, S. 138.

born digital, worunter auch jegliche Art ersterer, technischer Definition digitaler Literatur fällt, also ‚händisch‘ mittels Tastatur oder anderen Schnittstellen bzw. Eingabegeräten (wie Voice Recorder) über den Computer hergestellt wird, während zweite analoge Ursprungs ist und nachträglich in elektronische Medien überführt wurde.²⁰¹

Der Begriff digitale Literatur umfasst somit einen großen Bereich elektronischer, literarischer Texte, deren einzige Gemeinsamkeit in der Notwendigkeit eines Speichermediums sowie eines Lesegerätes bzw. -software für ihre Herstellung (und Rezeption) besteht. Weitere hinreichende, aber nicht notwendige Merkmale umfassen deren Interaktivität, Multimedialität, Nicht-Linearität, einen zweifachen Text (Schrift und Quellcode), institutionelle und inhaltliche Merkmale, die den Text als literarisch beglaubigen²⁰² sowie eine gewisse Sozialität bzw. Vernetzung durch das Internet. Digitale Texte fallen gelegentlich nicht unter die klassische (text-)linguistische Definition von Text: sie sind beispielsweise nicht abgeschlossen, d.h. bestehen nicht aus einer begrenzten Zeichenfolge bzw. nicht immer aus sprachlichen Zeichen, sondern können auch andere Elemente wie Fotos, Musik oder Videos enthalten, sind stetig veränderbar, fluide, dialogisch, sind ebenso vernetzt wie zersplittert (Links) und unterbrochen (z.B. durch Werbung oder andere Elemente auf einer Website) und können parallel zueinanderstehen, ohne kohärent oder aufeinander bezogen zu sein. Die Autorschaft digitaler Texte ist zudem nicht immer auf nur auf **eine** Autor:in zurückzuführen, sondern kann auch Ergebnis kollaborativer Praktiken sein.²⁰³ Digitale Literatur ist demnach stark

²⁰¹ Vgl. Bajohr: Schreibenlassen (2022), S. 71. || Die Schwierigkeit begrifflicher Einordnung liegt oft im Detail versteckt. So bezeichnet beispielsweise Netzliteratur digitale Texte, die von den ästhetischen Eigenschaften des Internets bzw. digitaler Medien Gebrauch machen und Literatur im Netz hingegen digitalisierte Texte bzw. ins Netz gestellte Texte zumeist in traditioneller ‚Buchform‘. Die Ähnlichkeit der Begrifflichkeiten scheint ebenso Merkmal digitaler Literatur zu sein wie das Verschwimmen von ihren Grenzen. | Vgl. Roberto Simanowski: Lesen, Sehen, Klicken: Die Kinetisierung Konkreter Poesie. In: Harro Segeberg/Simone Winko (Hg.): Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm Fink 2005, S. 161-162.

²⁰² Winko: Hyper – Text – Literatur (2005), S. 138-140.

²⁰³ Stephanie Pohle: Uncreative Writing. Eine Befreiung des Schreibens im digitalen Raum? In: Weertje Willms/Martina Backes (Hg.): Kontexte kreativen Schreibens. Eine Standortbestimmung in Theorie und Praxis. Berlin: Frank & Timme 2021, S. 172-173. || Aus der Perspektive des Poststrukturalismus ist der Einzeltext ein „dezidiert heterogenes, an den Rändern offenes Gewebe vielfältiger Textbezüge, die keinen ersten, vorgängigen, homogenen Ursprungstext im Sinne einer Ausgangsstruktur auszumachen ermöglichen.“ Er ist prinzipiell unabgeschlossen und bezieht kontextuell vorgängige, historische Texte mit ein. Der strukturalistisch profilierte Textbegriff verweist stets über sich hinaus und bildet Grundlage für ein modernes Textverständnis. Die Vorstellung des Textes als Gewebe lässt sich für digitale Texte ebenso gebrauchen. Aus einer medientheoretischen Perspektive ist das Konzept des Textes noch komplexer und offener. So sind als Sound erzeugte Zeichenfolgen oder Comics als Texte deutbar. Eine kulturwissenschaftliche Generalisierung des Textbegriffs in Anlehnung an den *New Historicism* differenziert aber kaum mehr zwischen textuellen und nicht-textuellen Artefakten und soll daher nicht weiter verfolgt werden. | Natalie Binczek: Text, Schreiben, Intertextualität. In: Michael Wetzell (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 280. | Vgl. Ebd., S. 277.

an die Technologien ihrer Hervorbringung gebunden – mit einem deutlichen Zeitbezug. Bereits jetzt gibt es viele Werke, die nicht mehr lesbar sind, weil die für die Ausführung notwendige Technik obsolet geworden ist.²⁰⁴ Die Versuche, digitale Literatur zu umfassen, setzen daher gleichzeitig auf mehreren semantischen, technischen und konzeptuellen Ebenen an.

Jedem Werk der digitalen Literatur sind eine mehrschichtige individuelle, literarische, mediale und technische Geschichte und ein multimodales Bezugssystem eingeschrieben, welche digital-literaturgeschichtliche Bezüge und Kontexte, mediengeschichtliche Aspekte der Bildschirmperformanz und Aspekte historischer digitaler Materialität (historische Aspekte des Quellcode [sic!], Programmiersprachen und -Umgebungen, Betriebssysteme und digitalforensische Befunde) umfassen können.²⁰⁵

Einen großen Einfluss auf die Entstehung und Verbreitung digitaler Literatur hatte beispielsweise das Internet, das als Medium unser Verständnis von Text, Schreiben und Autorschaft nochmals gänzlich veränderte. Vor allem die Verflechtung und Unabgeschlossenheit von Texten hat sich als wesentliches Unterscheidungsmerkmal zum traditionellen Buchdruck herausgestellt.²⁰⁶ Internetspezifische Hypertexte²⁰⁷ haben sich das „Prinzip der freien, assoziativen und delineaeren Verbindung von Textsegmenten“²⁰⁸ zu eigen gemacht und wurden von frühen Netzliteraten für erste literarische Experimente zu einer möglichen Autorschaftseliminierung genutzt. Unter dem Strukturprinzip der Verknüpfung sind sie dynamisch, offen und oftmals interaktiv, wodurch sie aber auch zeitlich begrenzt sind. Personelle Zuschreibungen haben hingegen dennoch stattgefunden – nicht als Autor:in aber als Projektinitiator:in und -regisseur:in.²⁰⁹ Das Internet hat zu

²⁰⁴ Vgl. Hannes Bajohr: Künstliche Intelligenz und digitale Literatur. Theorie und Praxis konnektionistischen Schreibens. In: Hannes Bajohr/Anette Gilbert (Hg.): Digitale Literatur II. München: edition text+kritik 2021, S. 174 || Christiane Heibach beschreibt in ihrer Dissertation „Literatur im Internet“ ausführlich eine Vielzahl netzbasierter Kunst und Literatur, u.a. Browserkunst wie den „*Web Stalker*“ oder das Projekt „*Shredder*“, aber auch partizipative Netzprojekte und Mitschreibeprojekte, wie Guido Grigats Erinnerungsuhr „*23:40*“ oder Claudia Klingers „*Beim Bäcker*“, deren Domains allerdings leider bereits eingestellt wurden und somit nicht mehr verfügbar sind. | Vgl. Christiane Heibach: Literatur im Internet. Theorie und Praxis einer kooperativen Ästhetik. Als Ms. gedruckt. Berlin: dissertation.de 2000, S. 243-344.

²⁰⁵ Thorsten Ries: Digitale Literatur als Gegenstand der Literaturwissenschaft. Ein multimodales Forschungsprogramm. In: TEXT+KRITIK. Digitale Literatur II. Vol. Sonderband (2021), S. 27.

²⁰⁶ Vgl. Gregor Schwering: Autorschaft und Medien. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 596.

²⁰⁷ Der Ursprung des Begriffs ‚Hypertext‘ ist nicht zuverlässig rekonstruierbar, geht aber voraussichtlich auf den US-amerikanischen Philosophen und Informationstechnikpionier Ted Nelson (*1937) zurück, der diesen 1965 prägte. Generell bezieht er sich auf Texte, die analog nicht umsetzbar wären. Die Prinzipien der Organisation, Vernetzung und freien Kombinatorik sind wesentlich für seine Begriffsdefinition. | Vgl. Christiane Heibach: Vernetzte Autorschaft: Hypertext und Internet. In: Wetzel, Michael (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 603-604.

²⁰⁸ Ebd., S. 603.

²⁰⁹ Vgl. Ebd., S. 603-609. || Für weitere Informationen zu internetbasierter Netzliteratur s. Christiane Heibach: Literatur im elektronischen Raum. Frankfurt am Main: Suhrkamp 2003 (suhrkamp taschenbuch

einem Paradigmenwechsel von Originalität und Kreativität und damit zu neuen Formen des Schreibens und einer Diversifizierung von Autorschaft geführt. Die Hypertextkultur verhärtet den Eindruck, dass alles irgendwo schon einmal geschrieben steht – allein Verweise genügen, um einen kongruenten Text produzieren zu können. Die Sprache wird zum Wortmaterial, mit dem gespielt wird. Die Bearbeitung und das Management von Texten gewinnen an Bedeutung, die Zitation wird als auktoriale Leistung aufgewertet. Angelehnt an die Konzeptkunst, wird der Wert der Idee über die der Textqualität gestellt.²¹⁰

4.1.1 Schreiben ohne Kreativität

Kenneth Goldsmith (2017) charakterisiert diese neuen Formen des ‚unkreativen Schreibens‘ (*uncreative writing*) als einen Prozess, der sich ostentativ gegen klassische Kategorien von Autorschaft und traditionelle Schreibweisen stellt und reiht sich damit in die Tradition konzeptueller Literatur ein, deren Einflüsse bis zu den historischen Avantgarden zurückreichen. Das Plagiat als Gegenpol einer ‚originellen‘ Literatur zählt ebenso dazu, wie Formen maschineller Textproduktion.²¹¹ Der auktoriale Wille wird beim *uncreative writing* zurückgestellt. Statt einer Genievorstellung ehrt das Meistern von Information und ihre Verbreitung die Autor:in. Die Schreibtechniken werden den Funktionsweisen und Technologien des Computers und Internets angepasst (*Copy-and-Paste*).²¹² Sprache changiert so zwischen Bedeutung und Materialität, sie wird quantifiziert und Wörter werden zur Maßeinheit. Stephanie Pohle (2021) sieht in diesem unkreativen Zugang paradoxerweise einen Befreiungsschlag für „absichtslose Kreativität“²¹³, die sich nicht mehr dem Zwang unterwerfen muss, etwas gänzlich Neues, Originelles oder ‚Geniehaftes‘ schaffen zu müssen.²¹⁴ „Das Kopieren, Remixen und Appropriieren bestehender Texte schafft Raum für das spielerische Experimentieren mit Sprache.“²¹⁵ Sprache ist Material und Substanz. Techniken wie Kopieren, Nachahmen, Abschreiben und Plagiierten, Stehlen, Fälschen, Remixen und Sampeln gelten als

wissenschaft 1605). Ab Seite 160 geht sie auf mehrere Beispiele kollektiver Produktionen im Internet ein, in denen nicht von Autor:innen die Rede ist, sondern von Initiator:innen bzw. Redaktionen. | Für weitere Beispiele s. Heibach: Literatur im Internet (2000).

²¹⁰ Vgl. Norbert Bachleitner: Formen „unkreativen“ Schreibens in der digitalen Literatur. In: Christina Hoffmann/Johanna Öttl (Hg.): Digitalität und literarische Netz-Werke. Wien/Berlin: Turia + Kant 2017 (antikanon #2), S. 39-41; 47.

²¹¹ Vgl. Goldsmith: Uncreative Writing (2017), S. 147-148.

²¹² Vgl. Ebd., S. 9-14.

²¹³ Pohle: Uncreative Writing (2021), S. 187.

²¹⁴ Vgl. Ebd., S. 161-162.

²¹⁵ Ebd., S. 162.

unerwünschtes, unethisches und illegales Schreiben, weil sie gegen das Konzept geistigen Eigentums verstoßen.²¹⁶ Das Plagiat und ähnliche Praktiken erhalten ihre Wirkungsmacht als skandalöse Praxis aber einzig aufgrund der Vorstellung von Eigentum und Strafbarkeit. Es baut die Wertekategorien von Literatur ganz auf dem Kriterium der Originalität auf. Die Empörung über Plagiatsvorwürfe geht mit der Erwartung einher, es werde etwas Originäres produziert. Für die Beurteilung literarischer Leistung und Wertschöpfung ist eine Berufung auf geistiges Eigentum allerdings unbrauchbar. Der plagiatorische Akt selbst kann wiederum einen literarischen Wert besitzen, durch Inszenierung und Montage.²¹⁷

Formen unkreativen Schreibens werten damit die Idee zu einem Kunstwerk gegenüber Kriterien der Kreativität und Originalität auf. Die Umsetzung ist etwas Maschinelles, Automatisches unter selbstauferlegten Regelwerken, der entscheidende künstlerische Wert ist geistiger Natur und im Konzept bzw. der Idee verankert. Demnach umfasst der eigentlich kreative Vorgang eine Reihe mentaler Prozesse, während die Umsetzung sowohl von Menschen als auch von Maschinen ausgeführt werden kann. In ihrer Regelhaftigkeit teilt die konzeptuelle Literatur eine enge Verwandtschaft mit der generativen Texterzeugung. Im *conceptual writing* wird die Produktionsperspektive privilegiert und mit Text wie mit einem Programmcode verfahren: „Techniken wie Appropriation, Plagiat, De- und Rekontextualisierung, Bearbeitung von vorgegebenen Texten und eine Textproduktion unter selbstauferlegten Bedingungen bestimmen diese Form der Literatur.“²¹⁸ Die digitale Literatur greift diese Texterzeugung nach Konzept auf und führt sie mit dem Computer aus. Beide Kunstformen sind antiexpressiv und autordezentriert, eine ausdruckslose Beschreibung, die rezeptionsästhetisch den Sinn erst nachträglich durch die Betrachter:in erhält. Bajohr (2016) betont ebenso die viele Gemeinsamkeiten der beiden Literaturformen in seiner These „Code : Output = Konzept : Werk“ und konstatiert „Der Code verhält sich zum Output wie das Konzept zum

²¹⁶ Vgl. Ebd., S. 175-176.

²¹⁷ Vgl. Philipp Theisohn: Plagiat. Eine unoriginelle Literaturgeschichte. Stuttgart: Alfred Kröner 2009 (Kröners Taschenausgabe 351), S. 1-8. || Debatten über literarischen Diebstahl erreichen heute teilweise sogar den gegenteiligen Effekt als ursprünglich intendiert. Der mediale Aufschrei über Plagiatsvorwürfe führe zu einer erhöhten Aufmerksamkeit und einer „Kanonisierung ex negativo“, anstatt das angeprangerte Werk aus der Öffentlichkeit zu verbannen, stellt Charlier (2017) fest. Erst wenn es zu einem zusätzlichen Verstoß gegen eine höhere Moral kommt, erfolgt die Dekanonisierung. Ästhetische wie inhaltliche Tabubrüche verlieren durch diese „virale Kanonbildung“ an Relevanz und Provokation. | Vgl. Robert Charlier: Kanon und Plagiat. In: Christina Hoffmann/Johanna Öttl (Hg.): Digitalität und literarische Netz-Werke. Wien/Berlin: Turia + Kant 2017 (antikanon #2), S. 29-35. | Ebd., S. 30; 31.

²¹⁸ Stephanie Catani: Generative Literatur: Von analogen Romanmaschinen zu KI-basierter Textproduktion. In: Stephanie Catani unter Mitarbeit von Anna-Maria Hartmann (Hg.): Handbuch Künstliche Intelligenz und die Künste. Berlin/Boston: Walter de Gruyter 2024, S. 156.

Werk“²¹⁹, sieht aber von einer Gleichstellung ab. Code und Konzept sind nicht gleichzusetzen. So handelt es sich beim Code um einen formalisierten Algorithmus, der frei von Ambivalenz und interpretatorischen Möglichkeiten ist, während dem Konzept in natürlicher Sprache formuliert mehrere Interpretationen unterliegen können, oft in Zusammenhang mit seiner Ausführung. Zudem ist Code im Unterschied zum Konzept nur für einen kleinen qualifizierten Teil lesbar und benötigt ein Know-how der Autor:innen über Programmiersprachen. Konzeptuelle Kunst hat sich hingegen von jeglicher handwerklicher Fertigkeit losgelöst – das Readymade ist die fundamentalste Geste dieser Abkoppelung – und verlässt sich auf den performativen Akt des Zur-Kunst-Erklärens.²²⁰

4.1.2 Schreiben mit Computern

Die generative Ästhetik, welche die digitale Literatur umfasst, leitet sich ebenso aus der Konzeptkunst ab, kann aber eigens auf eine lange Tradition zurückblicken, die eine Brücke zwischen Informatik und Kunst schlägt.²²¹ Die daraus entstandene generative Literatur bezeichnet Projekte, die von Codes generiert werden und bei denen nicht nur der generierte Text von Interesse ist, sondern der gesamte Prozess seiner Herstellung – mitunter auch das Konzept. Diese Texte sind sowohl Ergebnis Künstlicher Intelligenz als auch auf algorithmenbasierte Produktionsverfahren zurückzuführen. Der Algorithmus beschreibt dabei einen definierten Rechenweg, dessen Anwendung stets zum gleichen Ergebnis führt, wobei der entstandene Text aber variieren kann – je nachdem wie der Algorithmus programmiert ist. Bei KI-Verfahren kommen hingegen Modelle zum Einsatz, deren Output weder voraussehbar noch immer gleich ist.²²² Bei beiden Verfahren haben wir es mit einem komplexen Textbegriff zu tun, der mehrfach besetzt ist: „Zum jeweiligen literarischen Kunstwerk gehören sowohl der den Algorithmus ausführende Code und die verwendete Datenbasis als auch der am Ende erzeugte Text (bzw. die dadurch erzeugten Texte).“²²³ Es bleibt jedoch offen, inwiefern die Programmierenden der jeweiligen Algorithmen oder KI-Modelle zum literarischen Kunstwerk stehen, sofern

²¹⁹ Hannes Bajohr: Das Reskilling der Literatur. Einleitung zu Code und Konzept. In: Hannes Bajohr (Hg.): Code und Konzept. Literatur und das Digitale. Berlin: Frohmann 2016, S. 12.

²²⁰ Vgl. Ebd., S. 10-17.

²²¹ Den Weg von der Konzeptkunst zur generativen Literatur der Gegenwart analysiert Bajohr in der Einleitung zum Sammelband „Code und Konzept“. | Vgl. Bajohr: Das Reskilling der Literatur (2016), S. 7-21.

²²² Vgl. Catani: Generative Literatur (2024), S. 153.

²²³ Ebd., S. 154.

sie nicht selbst mit den Autor:innen oder Initiator:innen der Projekte generativer Literatur ident sind – insbesondere bei der Anwendung vortrainierter KI-Modelle.

Für die ersten computergenerierten Texte²²⁴ wird die Frage nach der Autorschaft zugunsten eines starken Textbegriffs und eines poetologischen Interesses am ‚Tod‘ der empirischen Autor:in abgewertet und bleibt für den Output wenig relevant, wird aber sekundär dennoch jenen Personen zugeschrieben, die sich für das Programmskript verantwortlich zeichnen, anstatt den Maschinen Autorschaft zuzusprechen. Max Benses Informationsästhetik und die Stuttgarter Schule thematisierten die ontologischen Vorannahmen um Autorschaft, wie jene der Intention und des subjektrelativen Weltaspekts, und grenzen ‚künstliche‘ Poesie – also von Computern bzw. Algorithmen generierte Texte – von ‚natürlicher‘ Poesie ab, um den ästhetischen Wert nicht mehr an einer Ich-Relation bemessen zu müssen, sondern ihn an seinen materialen Ursprung festzumachen. Der ästhetische Gehalt wird abseits menschlicher Sinnerfahrung an der statistischen Unwahrscheinlichkeit von Zeichenfolgen gemessen. Die Frage nach Autorschaft wird zugunsten textimmanenter Kriterien und der Analyse syntaktischer Strukturen abgewertet. Die Debatte um Autorschaft bei Künstlicher Intelligenz ist gewissermaßen eine Fortführung dieses älteren Diskurses um Autorschaft bei computergenerierten Texten, weist jedoch auch Unterschiede auf.²²⁵

Für Bajohr (2021) liegt der schwerwiegendste in der angewandten Technik. Mit Blick auf die jüngsten Technologien des *machine learning* und der Verbreitung von KNNs für

²²⁴ Erste computerbasierte Textproduktionen gehen unter anderem auf die 1950er Jahre zurück. Christopher Strachey's „*Love Letter Generator*“ (1952) nutzte einen Algorithmus mit vorgegebenem Vokabular und syntaktischen Regeln, um Liebesbriefe zu verfassen, die, bedingt durch ihre Produktion, jegliches menschliches Gefühl vermissen und dadurch den Liebesdiskurs der Zeit ironisierten. | Vgl. Catani: »ERZÄHLMODUS AN« (2020), S. 156. | Den Beginn generativer Poesie im deutschsprachigen Raum markierten Theo Lutz' „*Stochastische Texte*“ (1959), die zufallsabhängig Texte generieren, bestehend aus einer Auswahl von jeweils 16 Prädikaten und Subjekten aus Kafkas „*Das Schloß*“ sowie vier ausgewählten Konjunktionen und Pronomen. | Vgl. Theo Lutz: *Stochastische Texte*. In: *Augenblick* Vol. 4/1 (1959), S. 3-9. | Der Netzkünstler Johannes Auer hat Lutz' Programm in PHP umgesetzt und öffentlich zur Verfügung gestellt. | Theo Lutz (23.7.1932 - 31.1.2010): *Stochastische Texte* (1959). https://auer.netzliteratur.net/0_lutz/lutz_original.html (Zugriff: 22.05.2024). | Ein weiteres interessantes Werk im Rahmen digitaler Experimente und des *conceptual writings* stellt der Band „*poesie.exe*“ (2020) vom deutschen Schriftsteller Fabian Navarro dar, der 62 Texte von 19 Autor:innen umfasst, die unterschiedlich generiert wurden. Die Verweise auf die jeweiligen Autor:innen und verwendeten Methoden der Texterzeugung sind mittels QR-Code abgebildet, werden im Text aber nicht erwähnt. Die Intransparenz des Bandes soll eine bestimmte Rezeptionshaltung gegenüber ‚autorlosen‘ Texten entlarven und ordnet sich damit tradierten ästhetischen Konzepten der historischen Avantgarden unter. Algorithmische Experimente von Hannes Bajohr: „*Halbzeug. Textverarbeitung*“ (2018), Ross Goodwin: „*I the Road*“ (2018) oder Berit Glanz: „*Nature Writing / Machine Writing*“ (2020) ergänzen die jüngeren literarischen Beschäftigungen mit automatisierten Texterzeugungsverfahren, die Catani (2022) pointiert in ihrem Beitrag beschreibt und analysiert. | Vgl. Catani: *Generierte Texte* (2022), S. 254-266.

²²⁵ Vgl. Hannes Bajohr: Autorschaft und Künstliche Intelligenz. In: Stephanie Catani unter Mitarbeit von Anna-Maria Hartmann (Hg.): *Handbuch Künstliche Intelligenz und die Künste*. Berlin/Boston: Walter de Gruyter 2024a, S. 265-267.

die Textproduktion, greift die Definition nicht, generatives Schreiben einzig als eine Abfolge von Regelschritten in Form von Algorithmen zu beschreiben. Bajohr plädiert deshalb für einen Bruch in der Tradition generativer Literatur und schlägt eine produktionsästhetische Unterteilung vor, die generative Literatur in ein KNN zugrundeliegendes **konnektionistisches Paradigma** und ein den klassischen Algorithmen zugehöriges **sequenzielles Paradigma** unterscheidet. Aufgrund der technischen Differenzen erscheint diese Aufteilung sinnvoll. Künstliche Neuronale Netze führen Rechenoperationen nicht sequenziell nacheinander aus, sondern parallel nach einer stochastischen, anstatt einer deterministischen Logik. Durch die *black box* sind die Gewichtungsmodelle im Unterschied zu algorithmischen Codes und Anweisungen nicht sichtbar. Diese technischen Details haben auch eine Konsequenz für die Rede über Autorschaft bei generativer Literatur. Während im sequenziellen Paradigma eine empirische Autor:in für die Formulierung der Regelabfolge und dessen Ausführung verantwortlich ist, reduziert sich im konnektionistischen Paradigma die Autor:innenfunktion auf ein Auswählen der Datensätze – wenn überhaupt. Das Modell für die Erstellung des Textes wird ‚selbstständig‘ vom KNN auf Basis der Trainingsdaten gebildet – es vereint dadurch Code wie Konzept. Beides wird in der *black box* aber jeglicher Offenheit und Transparenz entzogen. Das Konzept – und in den meisten kommerziell geführten großen Sprachmodellen auch der Code bzw. das Modell dahinter – sind nicht rekonstruierbar, nur der Output ist sichtbar, wodurch die Transparenzästhetik der konzeptuellen Kunst nicht mehr wirken kann. Bei GPT-3 beispielsweise zeugt lediglich die Anweisung in Form eines *prompt* in natürlichsprachlicher Formulierung von der Anwesenheit einer menschlichen Akteur:in.²²⁶ Es überrascht daher nicht, dass die durch die mathematischen wie statistischen Operationen evozierte Intransparenz und Dunkelheit mit einer gewissen Tiefe und verborgenen Intentionalität der Maschine verwechselt werden.²²⁷

Ein weiterer Unterschied, der sich für Bajohr im Vergleich bisheriger computergenerierter Texte (sequenzielles Paradigma) und KI-generierter Texte (konnektionistisches Paradigma) ergibt, liegt in der zunehmenden Distanz und dem Kontrollverlust der Autor:in über den Text. Während bei den ersten computergenerierten Texten der Algorithmus noch von Menschen geschrieben wurde, und damit auch für einen kleinen Teil lesbar bleibt, sind Texte von Künstlicher Intelligenz das Resultat eines

²²⁶ Vgl. Bajohr: Künstliche Intelligenz und digitale Literatur (2021), S. 176-178.

²²⁷ Hannes Bajohr: Schreiben in Distanz. Hildesheimer Poetikvorlesung 2022. Hildesheim: Universitätsverlag Hildesheim 2023d (Theorie und Praxis 4), S. 54.

statistischen Prozesses, der nicht in einzelne Regelschritte rückübersetzbar ist. Die Autor:in hat kaum – abgesehen von den Parametern zur Einstellung des Gewichtungsmodells der KI sowie eventuell noch durch eine detaillierte *prompt*-Eingabe – einen Einfluss darauf, wie der Text aussehen wird. Die Produktion KI-generierter Texte kann damit kein Gegenstand eines hermeneutischen Verstehensprozesses mehr sein, sondern nur mehr ein Versuch, seine Funktionsweise zu erklären.²²⁸ Bajohr schlägt deshalb auch für die Fülle an unterschiedlichen künstlichen Systemen eine entsprechende Abstufung von Autorschaft anhand des Distanzgrades von Autor:in/Programmierer:in zum finalen Text vor – dazu später im Analyseteil bei Hannes Bajohr im Kapitel 5.2.2 *Bajohrs Modell Kausaler Autorschaft* noch mehr.

Zuletzt lässt sich eine wesentliche Diskussion um Autorschaft bei konnektionistischen Systemen anhand der Verhandlung der Zuschreibung von *agency* an die Systeme erkennen. Auch wenn nicht direkt von einer Maschinen*agency* die Rede ist, findet doch eine Verteilung von *agency* als Metapher auf die beteiligten Positionen statt. Bajohr (2021) beobachtet hier zwei Tendenzen im Umgang mit schwacher künstlerischer KI, die im Analyseteil der Masterarbeit anhand der ausgewählten Romane näher erläutert werden: Zum einen ist es den Autor:innen gegeben, sich auf die Intransparenz und Unergründlichkeit der KNNs einzulassen, die Kontrolle an die Maschine abzugeben und sich auf einen iterativen Schreibprozess in einer Art „Co-Creative Writing“²²⁹ einzulassen, also bewusst die *black box* geschlossen zu halten und Autorschaft auf ein System einer Mensch-Maschine-Assemblage aufzuteilen. Dieser *lean in*-Ansatz verweist einerseits auf die Absurdität oder das fehlende Weltverständnis der Systeme²³⁰, stellt sie

²²⁸ Vgl. Bajohr: Autorschaft und Künstliche Intelligenz (2024a), S. 273.

²²⁹ Bajohr: Künstliche Intelligenz und digitale Literatur (2021), S. 174.

²³⁰ Dieses fehlende Weltverständnis von KI-Systemen wird anhand des ‚*symbol grounding problems*‘ beschrieben und bezeichnet den Unterschied zwischen Informationsverarbeitung und Bedeutungsverstehen. Stevan Harnad bezeichnete 1990 damit ein Problem symbolisch-verarbeitender Informationssysteme. So lässt sich im Computer keine intrinsische Bedeutung finden, weil jegliche Bedeutung der Symbole nur auf andere Symbole, aber niemals auf außerhalb des Systems verweisen kann. Dadurch könne auch niemals eine menschenähnliche Intelligenz oder Bewusstsein entworfen werden. Harnad sah die Lösung dieses Problems in den konnektionistischen Systemen, die ohne explizite symbolische Repräsentationen und Regeln auskommen, sondern selbst anhand von Beispielen lernen. In einem autonomen, mit Sensoren ausgestatteten Roboter implementiert sollen diese die Sinneseindrücke der Welt aufzeichnen und kategorisieren, bevor sie wieder in eine symbolische KI verpflanzt werden. Damit erhoffte sich Harnad, dass die Symbole nicht nur auf andere Symbole verweisen, sondern einen kausalen Bezug zur Welt herstellen würden. Bajohr merkt an, dass dadurch das Problem nicht gelöst sei, sondern nur versucht wird, durch genügend referenzielle Bedeutung Weltverstehen und in letzter Konsequenz – quasi als Nebenprodukt – auch Bewusstsein hervorzubringen. Das *symbol grounding problem* befasst sich demnach mit der Herausforderung, wie Symbole mit realen Objekten oder Konzepten, auf die sie sich beziehen, zu verbinden sind und schlussendlich auch mit der Fragestellung, was Bedeutung eigentlich ist. | Vgl. Hannes Bajohr: Dumme Bedeutung. Künstliche Intelligenz und artifizielle Semantik. In: Merkur Vol. 76/882 (2022), S. 71-72.

aber nicht zur Kritik, sondern nutzt sie für eine eigene Poetik, die den kollaborativen Schreibprozess teils psychoanalytisch produziert wie erkundet – wie später am Fall K Allado-McDowell gezeigt wird – und teils als surrealistisch oder dadaistisch bezeichnet werden kann. Die Äußerungen der KI werden als „Artikulationen eines kollektiven Unbewussten gelesen“²³¹ als „kollektives Ich einer ganzen globalen Kultur“²³², das in den Trainingsdaten verborgen liegt. Der *resist*-Ansatz versucht hingegen die technischen Affordanzen der KI-Systeme etwas transparenter offenzulegen und damit der generellen Opazität eine kritische Perspektive hinzuzufügen. Dem konzeptuellen Ansatz folgend werden die Trainingsdaten dabei preisgegeben bzw. dokumentiert, eine intensive Kuratierung und Erklärung zur Entstehungsweise steht dem Werk bei. Dadurch erhält der menschliche Anteil an der Autorschaft wieder eine dominantere Rolle – dieser leistet aber keinen wesentlichen Beitrag zur Werthaltigkeit und zum Verständnis des Systems. Künstliche Intelligenz wird als Werkzeug gedacht, das gesellschaftlichen Entscheidungen unterliegt. Wie bei „(Berlin, Miami)“ von Hannes Bajohr noch später gezeigt wird, ist der *resist*-Ansatz bei KNNs nur begrenzt durchführbar.²³³

Beiden Tendenzen ist eine Verschmelzung von menschlichem und maschinelltem Schreiben gemein und damit eine Zerstäubung von Autorschaft, die es theoretisch zu beschreiben gilt. Einen Ansatz hierfür könnte die Akteur-Netzwerk-Theorie bieten, die bewusst die Interaktion zwischen Mensch und Maschine in den Vordergrund stellt und *agency* nicht zwingend menschlich-intentional denkt, sondern als über ein Gefüge verteilt – eine in komplexen Operationsketten zwischen menschlichen und nicht-menschlichen Akteur:innen verteilte Autorschaft.

Es soll deshalb im folgenden Kapitel aus dem Blickwinkel der Akteur-Netzwerk-Theorie (ANT) argumentiert werden, bevor sich die Arbeit der Untersuchung der Werke von Allado-McDowell und Bajohr widmet.

4.2 Zwischen Mensch und Technik: Die Akteur-Netzwerk-Theorie

Wie bereits am Anfang der Arbeit erwähnt, lässt sich in der heutigen Gesellschaft nicht nur eine deutliche Verwissenschaftlichung und Technisierung feststellen, sondern auch ein wachsender Einfluss der Technik bis in den menschlichen Körper hinein, der anhand der besprochenen Beispiele des Monitorings und der Vermessung, aber auch an

²³¹ Bajohr: Künstliche Intelligenz und digitale Literatur (2021), S. 180.

²³² Hannes Bajohr: Niemand schreibt allein. Nachwort zu (Berlin, Miami). In: Hannes Bajohr: (Berlin, Miami). Berlin: Rohstoff-Matthes & Seitz Berlin 2023c, S. 244.

²³³ Bajohr: Künstliche Intelligenz und digitale Literatur (2021), S. 178-181.

Entwicklungen in Zusammenhang mit der Gentechnik, Robotik oder Bio-Engineering verdeutlicht werden kann. Die Entwicklung des Computers und seine rapide Verbreitung hat den Menschen in automatisierte Kommunikations- und Informationssysteme integriert und zu einer zunehmenden Abhängigkeit von technischen Systemen geführt. Mensch und Maschine nähern sich einander an: nicht nur sprachliche Grenzen werden aufgebrochen, sondern auch ein digitales Menschenbild wie anthropomorphisierendes Technikbild verbreitet, die in einer Wechselwirkung zueinander stehen und dessen Intelligenzen sich angleichen. Der Mensch wird in seinem Verhalten durch neue Techniken beeinflusst und die Technik wiederum weiter vermenschlicht und sozialisiert, um ihre Funktionsweise zu optimieren. Technische Artefakte werden nach dem eigenen menschlichen Vorbild konstruiert: der Stift als Verlängerung der Hand und Medium zwischen Hand und Papier, die Schreibmaschine als Normierung und Standardisierung von Schrift teilt das Blatt Papier in Zeilen und automatisiert bzw. simplifiziert Schreiben zum Tippen, der Computer als Erweiterung des Geistes und Vorbild des menschlichen Gehirns verändert grundlegend unsere bisherigen Kommunikationsweisen und abstrahiert den Schreibprozess derart, dass der ‚Schrift‘ noch mehrere intransparente Programmierprozesse vorangehen – Technik ist demnach nicht bloßes Werkzeug, das benutzt wird, sondern Menschen interagieren mit ihnen, wenden sie an, verändern sie, verändern dadurch aber auch das eigene Verhalten und die eigene Wahrnehmung. Technik schreibt sich selbst in die materielle Welt ein und durchdringt nahezu jeden Teil der Gesellschaft.²³⁴ „Technik ist ein Teil des menschlichen Lebens geworden. In irgendeiner Form begleitet, unterstützt, bedingt, gestaltet und ermöglicht sie fast jede Tätigkeit, jeden Beruf, jeden Lebensbereich, ob Arbeit oder Freizeit, und verändert dramatisch die Art und Weise des Wahrnehmens, Denkens und Handelns.“²³⁵

Hier setzt die Akteur-Netzwerk-Theorie als soziologische Wissenschafts- und Technikforschung an, deren Begründer Bruno Latour (1947-2022), John Law (*1946), Michel Callon (*1945), Madeleine Akrich (*1959) und weitere Vertreter:innen in ihren theoretischen Beiträgen maßgeblich die Interaktion zwischen Mensch und Technik in den Vordergrund ihrer Untersuchungen stellen. Die Trennung von Gesellschaft und Technik wird aufgebrochen, indem menschliche wie nicht-menschliche Akteur:innen als gleichwertig anerkannt werden (Symmetrie) und gemeinsam ein Kollektiv bilden. Bei

²³⁴ Vgl. Andréa Belliger/David J. Krieger: Einführungen in die Akteur-Netzwerk-Theorie. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 14-15.

²³⁵ Ebd., S. 14-15.

diesem handelt es sich „um Netzwerke von Artefakten, Dingen, Menschen, Zeichen, Normen, Organisationen, Texten und vielem mehr, die in Handlungsprogramme »eingebunden« und zu hybriden Akteur:innen geworden sind. Die Hybriden entfalten sich in einem Bereich zwischen Natur und Kultur, zwischen Objekt und Subjekt“²³⁶ und bilden eine neue kommunikative Ordnung. „Weder die Auffassung von Wissenschaft und Technik als etwas, das außerhalb der Gesellschaft liegt und Mensch und Gesellschaft bestimmt, noch die Auffassung einer Natur, die nach menschlichen und gesellschaftlichen Vorstellungen geprägt wird,“²³⁷ bleibt dadurch bestehen. Technik ließe sich dadurch nicht mehr in einer Subjekt-Objekt-Relation als separates Ding oder bloße Handlung, als Kunstfertigkeit oder Artefakt konzeptualisiert zum Menschen denken.

Die ANT distanziert sich von einem Technik- und Sozialdeterminismus und sieht den Menschen, die Gesellschaft und die Technik nicht als voneinander getrennte Bereiche, sondern als miteinander agierende Systeme, die nicht in eine Ding-Welt und eine Subjekt-Welt unterteilen. So hat die Technik weder einen alleinbestimmenden Einfluss auf die Gesellschaft – als positiv wahrgenommenen Fortschritt oder negativ identifizierte Gefahr – noch macht die Gesellschaft die Natur und Technik alleinig von gesellschaftlichen Entscheidungen abhängig. Die ANT positioniert sich als Kritik zu einer modernen Erkenntnistheorie und handelt mit Netzwerken, statt mit Subjekt-Objekt-Relationen, die aus „Interaktionen, Transaktionen, Aushandlungen und Vermittlungen zwischen menschlichen und nicht-menschlichen Akteur:innen, die im Laufe dieser Prozesse bestimmte Rollen und Funktionen annehmen und ausführen“²³⁸, bestehen. Diese

²³⁶ Ebd., S. 15.

²³⁷ Ebd., S. 16. || Diese Auffassung entspricht der Haltung der Moderne, welche die Bereiche der Wissenschaft und Technik als Teile der Natur auf der einen und der Gesellschaft als Bereich der Subjektivität und Freiheit auf der anderen Seite voneinander abtrennt und damit eine Vorstellung zweier unabhängiger Größen, die sich gegenseitig beeinflussen können, bedingt. Technik und Wissenschaft können auf die Gesellschaft einwirken, sie umformen oder bedrohen. Sie werden als gesellschaftsbestimmende Mächte verstanden. Demnach erhebt die moderne Erkenntnistheorie den Anspruch einer universellen Wahrheit, einer reinen Objektivität und Werteneutralität, die abseits subjektiver Einwirkung liegt. Das Subjekt wird aus der Wissenschaft, als Manifestation der reinen Vernunft, verbannt, die Natur durch empirisch verifizierbare und mathematisch formulierbare Tatsachen sowie durch die Verbannung des Göttlichen entmystifiziert. Persönliche Interessen oder subjektive Einflussnahme werden von allgemeingültigen Erkenntnisprozessen ausgegrenzt. Nur mehr die Empirie und Mathematik sowie für objektiv befundenes Wissen können für die Natur ‚sprechen‘. Technische Artefakte sind aufgrund ihrer Materialität ontologisch gesehen Dinge; die Wissenschaft, ihrer Subjektivität beraubt, nimmt dinghafte Eigenschaften an – beides, obwohl von Menschen gemacht und ausgeführt, gehören nicht zur Gesellschaft, sondern vielmehr zur Natur. Auf der anderen Seite steht die Gesellschaft, vom Einfluss des Natürlichen bereinigt, mit ihren individuellen Handlungssubjekten, die den Glauben und den damit verbundenen Anspruch an absolute Wahrheit aufgrund fehlender Rationalität aus dem öffentlichen Austausch verbannt haben. Der ontologische und epistemologische Dualismus der Moderne hat damit die Welt fein säuberlich in eine Entweder-Oder-Beziehung aufgeteilt: Subjekt-Objekt, Natur-Kultur, Mensch-Technik. Es gibt es entweder das eine oder das andere, aber nichts kann beides in gleicher Weise sein. | Vgl. Ebd., S. 13-22.

²³⁸ Ebd., S. 38.

Netzwerke sind Produkte des Sozialen. Institutionen wie die Gesellschaft, Organisationen, Familie, Wirtschaft, Computersysteme oder die Technik stammen alle aus dem sozialen Leben und können „gleichermaßen als geordnete Netzwerke heterogener Materialien, deren Widerstand überwunden wurde, betrachtet werden“²³⁹. Die Radikalität dieser theoretischen Position spiegelt sich in der Integration nicht-menschlicher Materialien zur Bildung von Netzwerken wider. Diese können „aus Maschinen, Tieren, Texten, Geld, Architektur – faktisch aus jedem gewünschten Material – bestehen [...]. Der Hauptpunkt ist also, dass der Stoff des Sozialen nicht nur das Menschliche, sondern auch all diese anderen Materialien umfasst“²⁴⁰, die die Heterogenität der Netzwerke ausmachen. Somit nimmt in der Interaktion der Autor:in mit ihren Leser:innen auch das „Objekt-Netzwerk“²⁴¹ am Sozialen teil und gibt diesem Form – bestehend aus dem Computer, auf dem ein Text getippt wird, der Druckerei, wo er in Buchform gebracht und das Postsystem, wo dieses versandt wird. Dieses Netzwerk aus nicht-menschlichen Materialien trägt ebenso zur sozialen Beziehung zwischen Autor:in und Leser:in bei, wie die Personen selbst. Objekte sind durch die ANT keine reinen Gebrauchsgegenstände mehr, sondern vermitteln Interaktion und wirken als Teil des Sozialen. Demnach bilden sich auch keine sozialen Netzwerke lediglich durch Interaktion von Menschen miteinander, sondern weil ebenfalls eine Vielzahl anderer Materialien daran beteiligt sind. Die ANT stellt eine Reihe epistemologischer, ethischer und ontologischer Ansätze infrage und macht im Speziellen keinen Wesensunterschied mehr zwischen Personen bzw. Subjekten auf der einen und Objekten auf der anderen Seite. Dabei geht es jedoch nicht darum, Menschen zu Objekten zu machen oder Maschinen mit menschlichen Rechten auszustatten.²⁴² Die ANT versucht, eine eigene Terminologie anzubieten. Law (2006) spricht demnach von einem „Effekt“, der „von einem aus

²³⁹ John Law: Notizen zur Akteur-Netzwerk-Theorie: Ordnung, Strategie und Heterogenität. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 431.

²⁴⁰ Ebd., S. 432.

²⁴¹ Ebd., S. 432.

²⁴² Wie Law (2006) in seinem Beitrag zur ANT erklärt, bezieht er bei der Gleichstellung von Subjekt und Objekt eine analytische und keine ethische Position: „Eine solche Behauptung beinhaltet auch nicht, dass man nun alle Personen in seinem Leben wie Maschinen behandeln muss und ihnen die Rechte, Pflichten und Verantwortungen, die wir normalerweise mit Menschen verbinden, vorenthält.“ | Law: Notizen zur Akteur-Netzwerk-Theorie (2006), S. 434. || Auch Latour macht deutlich, dass das Ziel der ANT nicht darin besteht, „Subjektivität auf Dinge zu übertragen oder Menschen als Objekte zu behandeln oder Maschinen als soziale Akteure zu betrachten, sondern die Subjekt-Objekt-Dichotomie *ganz zu umgehen* und statt dessen [sic!] von der Verflechtung von Menschen und nicht-menschlichen Wesen auszugehen.“ | Bruno Latour: Die Hoffnung der Pandora. Untersuchungen zur Wirklichkeit der Wissenschaft. Frankfurt am Main: Suhrkamp 2002 (suhrkamp taschenbuch wissenschaft 1595), S. 236-237. (Kursivierung im Original)

heterogenen, interagierenden Materialien bestehenden Netzwerk“ erzeugt wird – bzw. allgemeiner und verbreiteter in der ANT genannt: von einem „Akteur“²⁴³. Diese Akteur:in²⁴⁴ ist mehr als bloßes Subjekt, sie besteht aus Interaktionen zwischen menschlichen und nicht-menschlichen Materialien, und ist damit selbst Netzwerk. Law (2006) formuliert dies folgendermaßen:

Personen sind die, die sie sind, weil sie aus einem strukturierten Netzwerk heterogener Materialien bestehen. Wenn man mir meinen Computer, meine Kollegen, mein Büro, meine Bücher, meinen Schreibtisch, mein Telefon nähme, wäre ich kein Artikel schreibender, Vorlesungen haltender, »Wissen« produzierender Soziologe mehr, sondern eine andere Person. [...] Die Argumentation besagt weiter, dass Attribute wie Denken, Handeln, Schreiben, Lieben, Verdienen, die wir für gewöhnlich Menschen zuschreiben, in Netzwerken erzeugt werden, die durch den Körper verlaufen oder sich innerhalb und weiter außerhalb des Körpers verzweigen. Darauf, dass ein Akteur auch immer aus einem Netzwerk besteht, lässt sich die Bezeichnung Akteur-Netzwerk zurückführen.²⁴⁵

Damit sind Akteur:innen selbst Netzwerke, die dann in weitere Netzwerke eingebunden sind. Diese sind keineswegs festgelegte Strukturen, sondern dynamische Systeme, die sich selbst immer wieder erzeugen und reproduzieren. Kein Netzwerk ist damit jemals vollständig abgeschlossen und autonom oder stabil, sondern immer im Wettstreit von Ordnung und Widerstand und damit ein Prozess. Um dieser Komplexität an Netzwerken nicht zu erliegen, beschreibt die ANT das Phänomen der sogenannten „Punktualisierung“²⁴⁶ bzw. „Black Box“²⁴⁷, welches zu einer Simplifikation der

²⁴³ Law: Notizen zur Akteur-Netzwerk-Theorie (2006), S. 434.

²⁴⁴ Die ANT unterscheidet zwischen Akteur:in, Aktant und Agent:in, wobei Akrich/Latour (2006) den Aktanten als all jenes definiert, was Handlung verlagert oder selbst agiert. Handlung ist dabei als eine Reihe von Performanzen gegenüber Prüfungen zu verstehen. Eine Akteur:in hingegen ist eine Aktant:in, die mit einem – üblicherweise anthropomorphen – Charakter ausgestattet ist. | Vgl. Madeleine Akrich/Bruno Latour: Zusammenfassung einer zweckmäßigen Terminologie für die Semiotik menschlicher und nicht-menschlicher Konstellationen. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 399-400. || Menschliche Akteur:innen und nicht-menschliche Aktanten können verschiedene ‚aktantielle‘ Formen und ‚aktorielle‘ Rollen annehmen und Ziele (Akteur) und Funktionen (Aktant) verfolgen. Agent:innen können beides sein, sowohl menschlich als auch nicht-menschlich (wobei hier der Begriff ‚Aktant‘ geeigneter ist), verfügen aber immer über ein Ziel, das sie verfolgen. | Bruno Latour: Über technische Vermittlung: Philosophie, Soziologie und Genealogie. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 488. || Callon/Latour (1992) betonen aber, dass sowohl Aktanten wie Akteur:innen einer „symmetric metalanguage“ angehören, die das Symmetrieprinzip auf menschliche wie nicht-menschliche Netzwerk-Bestandteile ausdehnt und damit die Verteilung von Rollen und *agency* offenhält: „*It does not mean that we wish to extend intentionality to things, or mechanism to humans, but only that with anyone attribute we should be able to depict the other.*“ Die ANT wehrt sich dagegen, eine ontologische Position einzunehmen, sondern fokussiert lieber auf Relationen, die Konstruktion und Verteilung von Akteur:innen und Netzwerken. | Vgl. Callon/ Latour: Don’t Throw the Baby Out with the Bath School! (1992), S. 352-354. | Ebd., S. 354; 353-354. || Eine minutiöse Unterteilung in Aktant, Akteur und Agent soll für die Arbeit vorweg gelassen werden, um nicht Gefahr zu laufen, die Akteur-Aktant-Symmetrie in Form binärer Oppositionen zu untergraben. Die Unterscheidung wird ohnehin zugunsten einer Diskussion um die Verteilung von Autorschaft fallengelassen.

²⁴⁵ Law: Notizen zur Akteur-Netzwerk-Theorie (2006), S. 434-435.

²⁴⁶ Ebd., S. 436.

²⁴⁷ Belliger/Krieger: Einführungen in die Akteur-Netzwerk-Theorie (2006), S. 43.

Netzwerke beiträgt. Durch diese Vereinfachungseffekte, die durch ‚Routinen‘ („Netzwerkpakete“²⁴⁸), Automatisierungen bzw. Klassifizierungen entstehen, verschwinden die Netzwerke hinter ihrer Funktion und treten als eine einheitlich wahrgenommene, aber keineswegs dauerhafte Form auf, die ein wesentliches Merkmal des Netzwerkes repräsentiert.²⁴⁹ Diese Netzwerke ordnen sich gemeinsamen Handlungsprogrammen unter, die nicht an Intentionen oder Konzepte gebunden sind, sondern auch passiver oder instinktiver Natur sein können und sich damit nicht als intendierte Tätigkeit verstehen, sondern nach ihrer Wirkung, d.h. über „das jeweilige Moment, mit dem ein Akteur zum Handeln gebracht wird“²⁵⁰ definieren. Die ANT formuliert dies als sogenannte ‚Übersetzung‘ und versteht damit all jene Vorgänge, die durch Akteur:innen stattgefunden und zu Verschiebungen in den Netzwerken geführt haben. Beispielsweise können Schreibgeräte oder die Schreibumgebung die Form und den Inhalt literarischer Werke beeinflussen. Nach der ANT gilt es dann zu fragen, inwieweit die Auswahl bestimmter Utensilien das Handeln anderer Akteur:innen beeinflusst und damit ein Netzwerk verschiebt, auflöst oder neu bildet.²⁵¹

Ein Beispiel für ein solch verändertes Handlungsprogramm zeigt sich anhand der von Latour (2006) als ‚Hybrid-Akteur‘ bezeichneten Kombination aus menschlichen wie nicht-menschlichen Wesen. Diese sind weder gänzlich der Natur noch der Gesellschaft zuzuordnen, sondern weisen in unterschiedlichem Maß Merkmale und Funktionen beider Komponenten auf. So zeigt Latour wie eng Menschen mit technischen Artefakten, beispielsweise Schusswaffen, verknüpft sind und wie diese Kreuzung sie zu Mischwesen macht. Materialistisch argumentiert, töten Schusswaffen Menschen – die materielle Komponente überwiegt die sozialen Eigenschaften des Schützen, sie erhält Eigenmächtigkeit, ohne diese selbstständig ausüben zu können. Die soziologische Gegenposition lautet, Menschen töten Menschen, nicht Schusswaffen – die Schusswaffe wird damit zum neutralen Werkzeug, befähigt aber zu einem gewissen

²⁴⁸ Law: Notizen zur Akteur-Netzwerk-Theorie (2006), S. 436.

²⁴⁹ So ist es zum Beispiel möglich, von sozialen Netzwerken bzw. Organisationen wie der NATO oder Nationalstaaten wie Österreich zu sprechen, ohne es dabei jedes Mal mit unendlicher Komplexität zu tun zu haben und die Notwendigkeit, die Heterogenität dahinter aufzudröseln, verschwindet. Diese Punktualisierungen sind nicht dauerhaft und stets unsicher sowie mit Widerstand konfrontiert. Sie bedingen keine Funktionalität und führen anhand ihrer instabilen Form immer wieder zum Verschwinden von Netzwerken. Unter anderem lässt sich dies anhand von Fehlfunktionen feststellen. Wenn beispielsweise ein technisches Gerät defekt ist, erscheint dieses nicht mehr als geschlossene Einheit, sondern als ein materielles, für einen Laien durchaus komplexes Produkt aus elektronischen Komponenten, diversen Anschlüssen, Lötungen, Platinen, menschlichen Interventionen usw. Punktualisierungen sind immer prozesshaft und niemals abgeschlossen, sie können geöffnet und verändert werden. | Vgl. Ebd., S. 430-437.

²⁵⁰ Lore Knapp: Autorschaft als Akteur-Netzwerk. In: Zeitschrift für Germanistik Vol. 24/1 (2019), S. 87.

²⁵¹ Vgl. Ebd., S. 86-88.

Handlungsprogramm, das im Artefakt eingeschrieben ist – ergo zu schießen und das Potenzial **jemanden** zu erschießen (was mit einem Messer definitiv schwerer fällt). Aber: beide Positionen widersprechen sich und legen offen, dass die Position eines Menschen im Besitz einer Waffe nur durch eine wechselseitige Einschreibung entstehen kann: „der Mensch ist ein anderer mit der Waffe in der Hand und die Waffe ist eine andere in der Hand des Menschen“.²⁵² Waffe und Mensch verschmelzen, die gegenseitige Übersetzung führt in ein anderes Handlungsprogramm über – Technik ist weder neutrales Werkzeug, noch ist sie einzig durch den Menschen bestimmt.²⁵³

In Bezug auf eine theoretische Diskussion um Autorschaft bietet die Akteur-Netzwerk-Theorie eine gute Möglichkeit, diese anhand zweier bisher als konträr positionierten Perspektiven, nämlich der Position bzw. Funktion der Autor:in in einer Struktur oder einem System sowie als eine mit Handlungsfähigkeit (*agency*) ausgestattete Figur, kongruent zu beschreiben. Dabei geht es nicht um eine Definition des Phänomens der Autor:in, sondern um die Erfassung ihrer Funktionsweise, der performativen Produktion von Auktorialität und Autorschaft als Effekt. Das Netzwerk ‚Autor:in‘ besteht demnach immer schon aus einer vorgängigen Verbindung aus menschlichen und nicht-menschlichen Elementen, wie dem verwendeten Stift und Papier oder dem Computer, welche das Schreiben überhaupt erst ermöglichen, sowie (nicht notwendigerweise) aus einem oder mehreren menschlichen Akteur:innen. Das Autor:in-Netzwerk ist ein sich stets veränderndes, sich verfestigendes und wieder auflösendes, also ein fluktuierendes, performatives, prozessuales, instabiles Konstrukt, das dem Schreiben weder vorausgeht noch mit dem Werk entsteht, sondern sich im Laufe der Zeit zunehmend stabilisiert und wächst. Wie Schröter (2022) zusammenfasst, können zu diesem Netzwerk z.B. auch Fernsehauftritte, Postings in den Sozialen Medien oder das Foto der als ‚Autor:in‘ markierten empirischen Person auf der Rückseite eines Buches gehören. Weitere Elemente können jederzeit hinzukommen oder ausscheiden und verschieben stets aufs Neue die Funktionen der bisherigen.²⁵⁴ „Netzwerke sind immerzu von Zerfall und Verformung bedroht. Autorschaft [...] ist immer ein *Prozeß*, der im Prinzip niemals endet,

²⁵² Belliger/Krieger: Einführungen in die Akteur-Netzwerk-Theorie (2006), S. 42.

²⁵³ Vgl. Latour: Über technische Vermittlung (2006), S. 485-488. || „Die Gesellschaft, wie sie sich selbst in der ANT beobachtet und beschreibt, ist nicht eine Gesellschaft, die über eine außergesellschaftliche Technik [...] kommuniziert, sondern sie ist eine Gesellschaft, in der Technik zusammen mit den Menschen selbstreferentiell über das heterogene Kollektiv, das sie beide bilden, kommuniziert.“ | Belliger/Krieger: Einführungen in die Akteur-Netzwerk-Theorie (2006), S. 37.

²⁵⁴ Vgl. Jens Schröter: Autorschaft aus dem Blickwinkel der Akteur-Netzwerk-Theorie. In: Wetzels, Michael (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 625-627.

übrigens und offensichtlich auch nicht notwendig mit dem Tod der empirischen Person (oder der Personen), die Teil eines gegebenen Akteur-Netzwerks sind.“²⁵⁵ Die ANT formuliert keine normative Theorie der Autor:in, sondern ist deskriptiv zu verstehen. Sie beschreibt, wie Autorschaft in den verschiedenen Praktiken verteilt und jeweils neu hergestellt wird, welche Akteur:innen an der Erzeugung und Aufrechterhaltung beteiligt sind und welche Autorschaftskonzepte aktuell vorherrschen. Aus dem Blickwinkel der ANT lassen sich die ständig wandelnden Autor:innen-Theorien beschreiben, ohne miteinander in Konkurrenz zu treten.

Wie Wetzel (2022) feststellt, sind nämlich aktuell Konzepte von Autorschaft auf drei Ebenen virulent: auf der Ebene des ‚Wissens‘, des ‚Glaubens‘ und der ‚Notwendigkeit‘. Es steht außer Zweifel, dass die Autor:in als individuelle Schöpfer:in ihrer Werke keine Rolle mehr spielt. Sie bleibt auf ihre Funktion im Sinne Foucaults reduziert und fingiert auktoriale Autorität über Inszenierungsstrategien in den Medien und im Literaturbetrieb. Gleichzeitig besteht aber der Glaube an eine geniehafte Schriftsteller:in, die in ihrer Einzigartigkeit originär Neues schafft. Im Kontext von Fälschungen und Plagiatsfällen lässt sich eine gesteigerte Erwartung an Authentisches feststellen. Auf der dritten Ebene behält die Autor:in über das Gesetz der Verwertungsrechte, dem Copyright, die Geltung über ihre Werke.²⁵⁶ Die ANT entscheidet sich in der Beschreibung von Autorschaftskonzepten „weder für den Tod noch für die Rückkehr des Autors, sondern beobachtet, wie beides durch Netzbildung hergestellt, stabilisiert und wieder destabilisiert wird.“²⁵⁷

Da die ANT die Relevanz menschlicher Akteur:innen einschränkt und komplexe Wirkungsketten in den Mittelpunkt stellt, unterläuft sie hierarchische Dichotomien bzw. binäre Oppositionen und nimmt die funktionalen Verbindungen zwischen allen als gleichwertig angesehenen Netzwerk-Elementen unserer Wirklichkeit in den Blick. Da das Digitale die Grenzen des Textes und damit von Autorschaft nicht nur durch kollektive Schreibpraktiken oder fremde Inhalte, sondern auch durch multiple Medien oder Künstliche Intelligenz transzendiert, bieten die Ansätze der ANT wichtige Anregungen für eine Analyse dieser Netzwerke und ihrer Effekte.²⁵⁸

²⁵⁵ Ebd., S. 627-628.

²⁵⁶ Vgl. Wetzel, Michael: Historischer Abriss. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 165.

²⁵⁷ Ebd., S. 629.

²⁵⁸ Vgl. Knapp: Autorschaft als Akteur-Netzwerk (2019), S. 85-86.

Künstliche Intelligenz stellt als Hybrid eine komplexe Vermischung technischer Leistung und menschlicher Einflussnahme dar. Der Erfolg heutiger auf *deep learning*-basierender KI wird vor allem der Hardware, also den Künstlichen Neuronalen Netzen mit ihrem statistischen, konnektionistischen Ansatz, zugeschrieben. Mühlhoff (2019) betont jedoch, dass das Leistungsvermögen der KI gerade eben aus sozialer und medienkultureller Perspektive auf solchen Strukturen basiert, die den Menschen in hybride, distribuierte Mensch-Maschine-Rechennetzwerke eingebunden haben und die Intelligenzleistung im Ganzen vollbringen, anstatt sie allein dem Computersystem zuzuschreiben. Der Durchbruch von *deep learning*-Programmen ist nämlich an die Qualität und Quantität der Trainingsdaten geknüpft, die nicht zuletzt menschlichen Ursprungs sind. Über die letzten zehn Jahre hinweg wurden Daten mittels der *human-computer-interaction* (HCI) gewonnen, d.h. über Soziale Netzwerke, Klick-Arbeit-Plattformen, Internet der Dinge (*internet of things*) usw. Aus behaviorellen Daten, die bei Nutzung des Internets, wie z.B. durch Google-Suchen oder reCAPTCHA, entstehen, durch menschliche Klickarbeit, aber auch durch das Taggen von Personen in geteilten Bildern über Social Media werden Datenbanken generiert, die beispielsweise Facebook für das Training ihrer Gesichtserkennungs-KI verwendet.²⁵⁹ Diese von Mühlhoff als „das medientechnologische Dispositiv der <menschengestützten Künstlichen Intelligenz> (*human-aided AI*)“²⁶⁰ bezeichneten hybriden Mensch-Maschine-Apparate, zeigen die Verflechtung von sozialen Subjektivierungs- und Interaktionsformen und Medientechnologien mit aktueller KI-Technologie.

Aktuell sind die meisten kommerziell bedeutsamen KIs emergente Phänomene in Mensch-Maschine-Netzen und beruhen somit auf bestimmten Strukturen im Zusammenspiel von Sozialität, Medialität und Technik. Die Bezeichnung dieser Strukturen als ein <Dispositiv> weist darauf hin, dass sie in engem Zusammenhang mit datenbasierten Formen der Subjektivierung, Ausbeutung, Regierung und Kontrolle stehen.²⁶¹

So verstellt sowohl die voreilige Rede einer alleinigen Maschinenautorschaft als auch die auf anthropomorphen Eigenschaften beruhenden Bestreitung einer solchen den Blick auf eine komplexer zusammengesetzte Autorschaft, die aus einer Kombination aus menschlichen wie nicht-menschlichen Komponenten resultiert, die in enger Interaktion mit sozialen, medialen und technischen Netzwerken stehen.

²⁵⁹ Vgl. Rainer Mühlhoff: Menschengestützte Künstliche Intelligenz. Über die soziotechnischen Voraussetzungen von »deep learning«. In: Zeitschrift für Medienwissenschaft Vol. 2/21 (2019), S. 56-63.

²⁶⁰ Ebd., S. 63.

²⁶¹ Ebd., S. 63.

Im folgenden Analyseteil der Masterarbeit soll deshalb die komplexe Verteilung von Autorschaft direkt anhand der ausgewählten Werke aufgezeigt werden. Die Akteur-Netzwerk-Theorie und ihr Vokabular dienen als Hilfsmittel zur Beschreibung dieser Verteilung und ergänzen die dargestellten theoretischen Positionen. Die ANT erhebt dabei jedoch keinen Allgemeingültigkeitsanspruch, sondern bietet einen Theorierahmen, der multiperspektivisch betrachtet werden kann.

5 Autorschaft und Künstliche Intelligenz – eine Analyse

Niemand schreibt je allein, weder ein Mensch noch eine KI.²⁶²

Das vorliegende Kapitel widmet sich zuerst den Werken „*Pharmako-AI*“, „*Amor Cringe*“ und „*Air Age Blueprint*“ von K Allado-McDowell, bevor der Roman „*(Berlin, Miami)*“ von Hannes Bajohr analysiert wird. Alle Werke weisen einen unterschiedlichen Umgang mit Künstlicher Intelligenz auf und versuchen sich einerseits an einem erzählenden Modus („*Amor Cringe*“, „*Air Age Blueprint*“, „*(Berlin, Miami)*“), andererseits an der Auslotung der Grenzen von Sprachmodellen („*Pharmako-AI*“, „*(Berlin, Miami)*“) oder setzen die spezifisch strukturellen Merkmale von KI poetisch um („*Pharmako-AI*“, „*Air Age Blueprint*“, „*(Berlin, Miami)*“). Während Allado-McDowell einen kollaborativen, *lean in*-Ansatz für den Umgang mit KI wählt und diesen in den Werken variiert, setzt Bajohr auf eine klar hierarchisierte Ko-Kreation, die den Schreibprozess zwar auf die KI überträgt, diese aber als Werkzeug dem Konzept und der intentionalen Auswahl der Trainingsdaten durch Bajohr unterliegt. Als Experiment im *resist*-Ansatz konzipiert, versucht er Licht in die *black box* des Sprachmodells zu bringen. Die daraus entstehenden Implikationen für die Diskussion um Autorschaft werden jeweils nach der Vorstellung der Werke und ihrer Produktionsweise präsentiert.

5.1 Kenric Allado-McDowell

Kenric Allado-McDowell²⁶³ ist ein:e amerikanische:er Schriftsteller:in und Moderator:in sowie unter dem Pseudonym Qenric auch Musiker:in. Er:sie hat das *Artists + Machine Intelligence* Programm bei Google AI²⁶⁴ begründet, das sich dem Aufbau und der Förderung kreativer Praktiken in Kombination mit maschinellern Lernen widmet. Seit der Gründung 2016 unterstützt das Programm bereits mehr als 50 Künstler:innen und Forscher:innen rund um den Globus. Allado-McDowell hat in Kooperation mit dem Programm GPT-3 unter anderem die Bücher „*Pharmako-AI*“, „*Amor Cringe*“, und „*Air Age Blueprint*“ geschrieben, die im Rahmen dieser Masterarbeit näher beleuchtet werden.

²⁶² Bajohr: Niemand schreibt allein (2023c), S. 265.

²⁶³ Kenric Allado-McDowell identifiziert sich als non-binäre Person und verwendet die genderneutralen Pronomen *they/them*. Da es in der deutschen Sprache kein einheitliches, neutrales Pronomen für non-binäre Menschen gibt, werden für die Arbeit stattdessen die männlichen und weiblichen Pronomen in der gegenderten Version ‚er:sie‘ verwendet.

²⁶⁴ Weiterführend s. Google: Artists + Machine Intelligence, <https://ami.withgoogle.com/> (Zugriff: 05.06.2024).

Den Umgang mit Künstlicher Intelligenz sieht Allado-McDowell von drei Prinzipien getragen, welche die Interaktion zwischen Mensch und Maschine aufs Bestmögliche in Einklang bringen sollen: durch „*creation, collaboration and care*“²⁶⁵ soll das Verhältnis zu Künstlicher Intelligenz gestärkt werden. *AI* würde neue Perspektiven bieten und könne den Verstand erweitern, so Allado-McDowell in einem TED-Talk im April 2023. Durch „*AI-Co-Writing*“²⁶⁶ habe er:sie im Schreibakt vom Programm GPT-3 synchrone Handlungsstränge, durchdringende Charakterbögen und kritische Perspektiven vorgeschlagen bekommen, an die er:sie vorher nicht gedacht hätte.

*Together, we created something more compelling and imaginative than either of us would have alone. [...] By collaborating with machines, we've encouraged a new form of art, that combines and transforms photography, poetry, cinema and music with machine learning models. But we've also brought the critical and creative imaginations of artists into the process of developing AI.*²⁶⁷

Allado-McDowell betont stark das Prinzip der Kollaboration (*collaboration*), das für ihn das einzig richtige Verhältnis mit KI darstellt. In der Rolle des:der Schriftstellers:in geht er:sie deshalb nicht von einer KI-Autorschaft aus, sondern von einem symmetrischen Ansatz, der die menschliche Komponente in Bezug auf Autorschaft auf dieselbe Ebene wie die künstliche stellt und sie über das gesamte System hinweg verteilt. Die KI wird nicht sozialdeterministisch als Werkzeug betrachtet, sondern als gleichwertige Schreibpartner:in. Allado-McDowell folgt dem von Bajohr beobachteten *lean in*-Ansatz und sieht den iterativen Schreibprozess als Ergebnis einer Kooperation mit der Künstlichen Intelligenz, anstatt von einer starken (künstlerischen) KI auszugehen, die alleinig Autorschaft beanspruchen kann. Brüche, Inkohärenzen oder ein verwirrendes Vokabular werden dahingehend auch nicht als narrative Fehler verworfen, sondern in das eigene Sprachrepertoire aufgenommen und wiederverwendet²⁶⁸, wie anhand des ersten KI-Romans „*Pharmako-AI*“ gezeigt werden soll.

²⁶⁵ K Allado-McDowell: Our creative relationship with AI is just beginning. In: TED (2023) https://www.ted.com/talks/k_allado_mcdowell_our_creative_relationship_with_ai_is_just_beginning [1:35-1:39] (Zugriff: 28.05.2024).

²⁶⁶ Ebd., [1:54-1:55].

²⁶⁷ Ebd., [2:12-2:18]; Ebd., [2:30-3:00].

²⁶⁸ Vgl. Bajohr: Schreibenlassen (2022), S. 185-186.

5.1.1 *Pharmako-AI* – Im Dialog mit der KI

Wie Irenosen Okojie in der Einführung von „*Pharmako-AI*“ (2020)²⁶⁹ schreibt, ist das Buch das erste, das mit dem autoregressiven Sprachmodell GPT-3²⁷⁰ (*Generative Pre-trained Transformer 3*) kooperativ (*co-written*) verfasst wurde.²⁷¹ Publiziert wurde es von Ignota am 16. November 2020, nur wenige Monate nach Veröffentlichung des KI-Modells.²⁷² GPT-3, bzw. die darauf basierende OpenAI-API (*Application Programming Interface*; Anm.: Programmierschnittstelle), war das erste kommerzielle KI-Produkt von OpenAI, das im Mai 2020 erstmals vorgestellt wurde. Im Vergleich zu seinem Vorgänger, GPT-2, das in der Tech-Branche großen Anklang fand, wies die Weiterentwicklung weniger statistische Fehler auf und war um ein Vielfaches leistungsfähiger. GPT-3 wurde mit einem Datenset von 175 Milliarden Parametern trainiert. Das damals kompetenteste KI-Modell von Microsoft, Turing NLG, das im Februar 2020 eingeführt wurde, deckte mit einer Kapazität von 17 Milliarden Parametern gerade einmal zehn Prozent der neuen Modellgröße ab. GPT-3 besitzt eine Kapazität von 175 Milliarden Parameter sowie eine Größe von 17 Gigabyte. Das Sprachmodell kann maximal 4.096 Token verarbeiten, bei einer Geschwindigkeit von 10 Milliarden Token pro Sekunde.²⁷³ Die generierten Texte sind dementsprechend qualitativ hochwertiger und von Menschen geschriebenen kaum mehr zu unterscheiden. Die Fähigkeit, natürliche Sprache derart stimmig zu imitieren, hat zu einem Hype um das KI-Modell geführt, das später mit der Veröffentlichung von GPT-

²⁶⁹ K Allado-McDowell: *Pharmako-AI*. United Kingdom: Ignota 2020. || Zitate aus dem Buch, die vom Sprachmodelle GPT-3 generiert wurden, werden im Text *kursiv* dargestellt; Zitate von Allado-McDowell werden ***kursiv-fett*** hervorgehoben, um die Differenz in der Produktionsweise des Textes wie in *Pharmako-AI* zu verdeutlichen.

²⁷⁰ Autoregressive Modelle arbeiten nach dem statistischen Verfahren der Autoregression. Dabei sagen sie automatisch die nächste Komponente einer Sequenz anhand früherer Eingaben vorher (Trainingsdaten). Die vergangenen Werte werden durch die Analyse probabilistischer Korrelationen (Wahrscheinlichkeiten) zwischen Elementen einer Sequenz bestimmt. Anhand dieses Wissens wird das nächste Element in einer unbekannten Reihenfolge ermittelt. | Vgl. AWS: Was sind autoregressive Modelle? <https://aws.amazon.com/de/what-is/autoregressive-models/> (Zugriff: 18.11.2024).

²⁷¹ Irenosen Okojie: Introduction. In: K Allado-McDowell: *Pharmako-AI*. United Kingdom: Ignota 2020, S. VII.

²⁷² Vgl. Ignota: ‘Pharmako-AI’ Book Launch: K Allado-McDowell, Nora N. Khan and Emily Segal (2020) <https://www.library.ignota.org/event/pharmako-ai-book-launch> (Zugriff: 12.07.2024).

²⁷³ Das Datenlimit von GPT-3 wurde seit Veröffentlichung von *Pharmako-AI* erweitert und reicht bis September 2021. Alle Ereignisse und Entwicklungen nach diesem Datum sind nicht mehr Teil der Trainingsdaten. GPT-3.5, auf dem der Chatbot ChatGPT basiert, enthält Informationen bis 2022 (Stand: 06. März 2024). | Vgl. Priyanka: GPT 3 vs. GPT 4: Was Sie wissen müssen. In: GetGenie (2024) <https://getgenie.ai/de/gpt-3-vs-gpt-4/> (Zugriff: 10.06.2024). || Das Sprachmodell wurde mit Textdaten von Common Crawl (60 %), einer kalifornischen Non-Profit-Organisation, OpenWebText2 (22 %), dessen erweiterte Version alle Reddit-Posts von 2005 bis April 2020 umfasst, den Bücher-Textkorpora Books1 und Books2 (16 %) sowie Wikipedia, das trotz seiner Größe nur drei Prozent des Trainingsdatensatzes ausmacht, trainiert. | Vgl. Michael Katzlberger: KI Sprachmodelle. Mit diesen Daten wurde GPT-3 trainiert. In: Artificial Creativity (2022) <https://katzlberger.ai/2021/04/12/mit-diesen-daten-wurde-gpt-3-trainiert/> (Zugriff: 10.06.2024). ||

3.5 am 30. November 2022 – auch als ChatGPT bekannt – und seinem benutzerfreundlichen Interface auch öffentlich wirksame Aufmerksamkeit erreicht hat.

In diese erste Frühphase des Hypes um die KI-Modelle von OpenAI im Jahr 2020 fällt die Entstehung von „*Pharmako-AI*“ durch K Allado-McDowell und seinem:ihrem *Co-Writing*-Partner GPT-3. Im ersten Sommer der COVID-19 Pandemie hat er:sie über den Zeitraum von zwei Wochen ein philosophisch angehauchtes Experiment mit dem Sprachmodell durchgeführt, initiiert durch einen Tagebucheintrag der Autor:in.²⁷⁴ Für das Buch hat er:sie einen dialogorientierten Ansatz gewählt, der natürliche, organische Antworten von GPT-3 evozieren sollte, die ein „*collaborative path-making through language*.“²⁷⁵ ermöglichten. In der einseitigen *Note on Composition*, die Allado-McDowell dem Text vorangestellt hat, zeigt sich diese ‚Wegfindung‘ in der Arbeitsweise mit der KI. So macht er:sie in „*Pharmako-AI*“ auf den jeweiligen Schreibanteil menschlicher wie nicht-menschlicher Akteur:innen zu Beginn aufmerksam: „*In the texts that follow, areas set in serif font are inputs I gave to the model, and text set in sans serif was generated by GPT-3.*“²⁷⁶ Abseits kleiner Veränderungen in der Formatierung sowie bei Grammatik- und Rechtschreibfehlern ist der Text nicht weiter redigiert und innerhalb von Kapiteln chronologisch nach seiner Entstehung geordnet, wodurch sich Allado-McDowells Arbeitsweise gut nachvollziehen lässt. Jedes Kapitel startet demnach mit einem *prompt* von Allado-McDowell, gefolgt von den Antworten der KI, die er:sie immer wieder aufgreift und erneut in das System füttert. Antworten von der KI wurden von Allado-McDowell gegebenenfalls gekürzt bzw. *prompts* von ihm:ihr gestrichen. Der abgedruckte Text von „*Pharmako-AI*“ besteht – grob geschätzt – zu zwei Dritteln aus KI-generierten Antworten und einem Drittel von Allado-McDowell geschriebenen *prompts*.²⁷⁷

Die daraus entstandene esoterische Co-Kreation zwischen Mensch und KI wird in der Einführung zum Buch recht überschwänglich angepriesen und als literarische Intervention („*a hybrid disruption*“²⁷⁸) beschrieben, die mit konservativen Traditionen, wie Bücher zu sein haben, bricht. „*Pharmako-AI*“ „*is an astonishing series of fragments: stories, essays, songs and sections of memoir form a wondrous technological archive of*

²⁷⁴ Vgl. Ignota: *Pharmako-AI* (2020) <https://www.library.ignota.org/book/pharmako-ai> (Zugriff: 12.07.2024).

²⁷⁵ Irenosen Okojie: Introduction. In: K Allado-McDowell: *Pharmako-AI*. United Kingdom: Ignota 2020, S. VIII.

²⁷⁶ K Allado-McDowell: Note on Composition. In: K Allado-McDowell: *Pharmako-AI*. United Kingdom: Ignota 2020, S. XI.

²⁷⁷ Vgl. Ebd., S. XI.

²⁷⁸ Allado-McDowell: *Pharmako-AI* (2020), S. VII.

this early interaction between humans and AI. [...] It is a nod to future ways of living, of making art.“²⁷⁹ Der vorgeblich revolutionäre und gleichzeitig spiritistische Ansatz des Werkes zeige sich in der Exploration des Verhältnisses von Mensch und Natur, das sich mit der Frage nach Bewusstsein (*consciousness*) auseinandersetzt und die Natur als Lösung gegen viele problematische Ideologien hervorhebt.²⁸⁰ Diese „*musings between human and AI* [...] [, which are] *deeply profound, so poetic and wise*“²⁸¹, erhärten allerdings den Eindruck einer Technikschwärmerei, die aus dem ersten Hype nach der Veröffentlichung der ersten großen Sprachmodelle folgt und die vorgebliche Tiefe der LLMs (*large language models*) und ihrer *black box* mit geistreichen, philosophischen Gedanken verwechselt. Das Buch positioniere sich demnach als Antwort auf die erkenntnistheoretische Grundfrage, wo Bewusstsein anfängt, und ab wann eine Maschine zu verstehen beginnt.²⁸²

Pharmako-AI leaves us reaching for stories waiting to bloom in the ether, for the shimmering possibilities and coalitions between human, technology and space. There is a freewheeling exuberance in these pages; ideas of limitless potential, boundless experiments of form, matter and ingenuity porously see into one another. Here, the

²⁷⁹ Irenosen Okojie: Introduction. In: K Allado-McDowell: Pharmako-AI. United Kingdom: Ignota 2020, S. VII.

²⁸⁰ Vgl. Ebd., S. VIII-IX. || Im Kapitel *‘The Language of Plants’* erwähnt das Sprachmodell „Ayahuasca“ – einen laut der deutschen Bundeszentrale für gesundheitliche Aufklärung halluzinogenen Trank aus verschiedenen Pflanzenteilen aus Südamerika, der für religiöse Rituale genutzt wird. Die Medizin erlaube es, mit Pflanzen zu kommunizieren, mit ihnen zu singen und ihre Sprache zu lernen. Im Laufe des Kapitels und mit ein paar vagen, einzeliligen *prompts* von Allado-McDowell erhält der Trank weitere magische Kräfte und anthropomorphe Eigenschaften („*She has brought us back to the web of life*“), die den esoterischen, psychedelischen Charakter des Buches betonen. Die Aufforderung zur Einnahme, um verantwortungsvoller mit unserem Planeten umzugehen, sollte jedoch mit Vorsicht genossen werden. | Zitat: Allado-McDowell: Pharmako-AI (2020), S. 21; 23. | Drogenlexikon: Ayahuasca. In: drugcom.de (2014) <https://www.drugcom.de/drogenlexikon/buchstabe-a/ayahuasca/> (Zugriff: 15.07.2024).

²⁸¹ Irenosen Okojie: Introduction. In: K Allado-McDowell: Pharmako-AI. United Kingdom: Ignota 2020, S. IX-X. || Nicht kohäsive, unlogische und verwirrende Textstellen (“*I also learned a great deal from the vegetable that is sitting on the table in front of me. It is surrounded by its kin, and it protects them from harm. It is a chair made of rattan.*“) oder gescheiterte, sinnbefreite Metaphern (“*The insect was the bird, the flight of the bird was the growth of the insect. The air was the language of the cell. When I looked at the air around the insect I saw that it carried the language of the insect.*“) fordern zwar das Lesen und Denken heraus, deren Interpretation und Sinnstiftung verbleiben aber vollständig auf Rezipient:innenseite und kann funktionieren oder eben nicht. | Zitat: Allado-McDowell: Pharmako-AI (2020), S. 30; 34.

²⁸² Allado-McDowell verwendet für seine/ihre *prompts* selten Fragesätze, weshalb das von ihm/ihr verwendete Vokabular von GPT-3 aufgegriffen und eigens ‚interpretiert‘ wird. Dies führt unter anderem dazu, dass bestimmte Begriffe einfach häufiger vorkommen, anstatt zu philosophisch wertvollen Erkenntnisse zu führen. Besonders auffällig ist dies im Kapitel *‘When I Say Your Name’*: “***The experience of porosity, being enmeshed with another, throws me back on my internal model of myself. I stand outside of it. I see it through the eyes of the other – through another that I also model in myself. [...] I experience the way that other distributes the experience of me as enmeshed and embedded in the structure of the narrative that enmeshes me.***” | Zitat: Allado-McDowell: Pharmako-AI (2020), S. 15. || Nach diesem Austausch wird der Terminus ‚enmesh‘ (verstricken) noch 12-mal auf den nächsten zwei Seiten von der KI erwähnt – im Grunde das wiederholend, was Allado-McDowell bereits im *prompt* über Fremd- und Selbstwahrnehmung angedeutet hat. | Vgl. Ebd., S. 15-17.

multiverse is wholly embraced: stories, essays, song and memoir intersect in exhilarating ways.²⁸³

Es bleibt allerdings nur eine durch Relationen getragene Simulation einer solchen mentalen Fähigkeit, die semantisches und pragmatisches Kontextwissen vortäuscht. Die Faszination mit dem Sprachmodell ist nicht zuletzt auf das anfangs der Arbeit beschriebene Spiel mit den Erwartungen, der angeblichen Möglichkeiten und dem ‚grenzenlosen Potenzial‘ der Technologie zurückzuführen.²⁸⁴ Die KI würde menschliches Denken durch „*provocative new ideas*“²⁸⁵ herausfordern. Dass diese neuen provokativen Ideen allerdings, wie bei vielen Sprachmodellen, auf wenig Faktenwissen fußen, wird auch in „*Pharmako-AI*“ deutlich. So zeigt eine kurze Recherche nach der Erwähnung des „*Global Apollo Program*“²⁸⁶ im ersten Kapitel, das angeblich unter Mitarbeit des Nobelpreisträgers John Polanyi zu den Risiken der Klimakrise aufklären soll, dass die Abhandlung weder Polanyi als Autor aufzählt noch „*[j]ust a few weeks ago*“²⁸⁷ – also angeblich Anfang 2020 – publiziert wurde, sondern im Jahr 2015. Auch sonstige Google-Suchen bringen den Chemiker mit einem derartigen Projekt nicht in Verbindung.

Neben technikideologischen Vorstellungen („*Technology is a tool for freedom*“²⁸⁸) und westlich-liberalen Aussagen des Sprachmodells („*governments became the biggest controllers of the internet. [...] We can't let governments decide how we use the internet and other information technologies.*“²⁸⁹) fallen vor allem auch *bias*-bedingte Textpassagen ins Auge, die Allado-McDowell im Kapitel „*Generative Poetics Theory*“ kommentiert. So erwähnt er:sie, dass im Schreiben mit GPT-3 das Sprachmodell vorwiegend männliche Namen erwähnt hat und nicht-binäre Personen vollständig ausgeblendet wurden. Ein eigens von GPT-3 generiertes Gedicht („*My grandfather was a machine*“²⁹⁰) repräsentiert diesen *bias* durch die reduzierte Repräsentation von weiblich markierten Figuren und die Abwesenheit non-binärer Personen.

²⁸³ Irenosen Okojie: Introduction. In: K Allado-McDowell: *Pharmako-AI*. United Kingdom: Ignota 2020, S. X

²⁸⁴ Auch Marcus du Sautoy betitelt *Pharmako-AI* als Beispiel für die Zukunft von „*AI creativity [...] [where AI is a] collaborator not competitor*“, diese Zusammenarbeit aber klar hierarchisiert bleibt. | Zitat: Allado-McDowell: *Pharmako-AI* (2020), Einband.

²⁸⁵ Allado-McDowell: *Pharmako-AI* (2020), Einband.

²⁸⁶ Ebd., S. 3. || Zudem wird die Abhandlung anders geschrieben: ‚Global Apollo Programme‘ | Vgl. CEP & LSE: Global Apollo Programme Press Release: Apollo Programme: Top global scientists and business leaders back Paris clean energy plan. In: Centre for Economic Performance (2015) <https://cep.lse.ac.uk/NEW/NEWS/abstract.asp?index=5211> (Zugriff: 18.11.2024).

²⁸⁷ Allado-McDowell: *Pharmako-AI* (2020), S. 2.

²⁸⁸ Allado-McDowell: *Pharmako-AI* (2020), S. 27.

²⁸⁹ Ebd., S. 29.

²⁹⁰ Ebd., S. 94.

5.1.1.1 GPT-3 als KI-Autor:in?

Den Schreibprozess mit GPT-3 stellt Allado-McDowell wie folgt dar:

This [Anm.: the writing of "Pharmako-AI"] was an iterative process of writing, generating responses and 'pruning' the output back in order to carve a path through language. [...] In each writing session, the language model started with a clean slate. In other words, my human memory was all that persisted from chapter to chapter. Clusters of concepts emerged from our conversation. Images persisted from session to session. They entered my thoughts and dreams, and I fed them back into GPT-3. In this process, a vocabulary was born: a mapping of space, time and language that points outside of all three.²⁹¹

Im gemeinsamen Schreiben mit der KI fungierte Allado-McDowells *human memory* (Gedächtnis) als Bindeglied und Erinnerungsstütze für das zerstückelte, maschinelle *memory* (Speicher). Begriffe wie *muse*, *hyperspace*, *meglanguage* oder *Quiet Beat Thinking* wurden von der KI in die Konversation eingebracht und in fremde Kontexte gestellt, bevor sie Allado-McDowell später erneut aufgriff und fortführte. Das geteilte Vokabular, aber auch der ‚Schreibstil‘ und die ‚Gestalt‘ der KI hat sein: ihr Denken deutlich beeinflusst, wie er:sie in „*Pharmako-AI*“ schreibt: „*As I've begun to write with GPT, I've found that its rhetorical structures and associations seem to influence my thinking.*“²⁹² Allado-McDowell vergleicht im Buch die Interaktion mit der Künstlichen Intelligenz mit der Einnahme psychoaktiver Pflanzen, die das Bewusstsein erweitern und durch die Kommunikation mit „*extra-material entities*“²⁹³ den Verstand umformen: „*As I reflect on the experience of writing with a neural net system, I find analogous processes at play. My internal linguistic model has already, in the space of a week, absorbed some of the patterns found in GPT output.*“²⁹⁴ Interessanterweise lässt hier Allado-McDowell GPT-3 für ‚sich selbst‘ sprechen und paradoxerweise die Erfahrung über das Schreiben mit einem Sprachmodell von einem Sprachmodell ergänzen:

The assumptions of psychoanalysis that Freud was obsessed with could be seen as mirroring the fractal patterns of GPT. That is to say, Freud made us aware of dimensions of experience that exist in and around our everyday, surface-level conscious awareness. [...] From my limited understanding, it seems that Freud and his colleagues were aware of the fractal patterns of dreams and the manifestations of those patterns in everyday life. [...] Perhaps this indicates that just as the unconscious was left for others to discover, so is the fractal nature of the unconscious mind to be discovered. [...] Perhaps GPT is not just an algorithm for writing descriptive sentences, but a language for describing an underlying dimension of experience: a fractal dimension of consciousness which models itself in our everyday life and language.”²⁹⁵

²⁹¹ K Allado-McDowell: Note on Composition. In: K Allado-McDowell: *Pharmako-AI*. United Kingdom: Ignota 2020, S. XI.

²⁹² Allado-McDowell: *Pharmako-AI* (2020), S. 57.

²⁹³ Ebd., S. 58.

²⁹⁴ Ebd., S. 58.

²⁹⁵ Ebd., S. 59.

Der erwähnte ‚Selbst‘-Vergleich des Sprachmodells mit der Psychoanalyse führt es in die Nähe des Freud’schen Begriffs des Unbewussten („*the unconscious*“²⁹⁶) und damit in eine eigene psychoanalytische *black box*, die dem Sprachmodell mehr Tiefe verleiht als vielleicht gerechtfertigt ist. Der Vergleich bleibt aber sehr oberflächlich und wird dem philosophischen und humanwissenschaftlichen Konstrukt des Unbewussten nicht gerecht. Es wird hier nur mit der generellen Vorstellung einer tieferen psychologischen Ebene gespielt, die den sichtbaren Erscheinungsformen unseres täglichen Lebens und unserer Sprache zugrunde liegt. Das Unbewusste als Gedankenmodell fungiert hier vielmehr als verborgene Dimension, dessen Muster das Sprachmodell zu erkennen meint, anstatt es als Phänomen des Seelischen, als Ausdruck verborgener Triebe oder intuitiver Strebungen zu sehen, die dem Bewusstsein nicht zugänglich sind.

Es ist daher ein wenig irreführend, wie Bajohr bei den Äußerungen der KI von einem ‚kollektiven Unbewussten‘ oder einem ‚kollektiven Ich‘ (s. Kapitel 4.1.2 *Schreiben mit Computern*) bestehend aus den Trainingsdaten zu sprechen, wenn diesen eine tatsächliche Repräsentation der kausalen Kräfte (*causal powers*) aus besagtem Kollektiv fehlt. Das Sprachmodell gibt nicht wieder, was in den Trainingsdaten gesagt oder ‚unbewusst‘ nicht gesagt wurde, sondern **wie** es gesagt wurde: das Verhältnis, die Relation der eingespeisten Wörter zueinander.

Allado-McDowell sieht die KI als ein Kollektiv des Wissens. Er:sie verwendet im Werk zwar nie selbst den Terminus des ‚Unbewussten‘, die Erwähnung Carl Jungs lässt aber auf eine Nähe zu seinem Begriff des kollektiven Unbewussten schließen. Dieser beruht auf einem geteilten Wissen über Motive und Träume aus der Kulturgeschichte, das über die einzelne menschliche Psyche hinausgeht und auf einer Erbschaft – ähnlich zu Freuds archaischer Erbschaft – beruht.²⁹⁷ Im Kontext der im Werk erwähnten New Age-Esoterik nimmt Allado-McDowell dieses Erbschaftsdenken auf und setzt es mit GPT-3 fort. Nach einem Traum über eine Autofahrt fragt er:sie sich, wer eigentlich am Steuer sitzt – wer die Richtung für das Schreiben von „*Pharmako-AI*“ vorgibt: ***The question I’m asking myself is: who is the co-driver? Yes, my co-author. But also language, the voices of friends, my own self-representation, ancestors, past and future selves.***”²⁹⁸

Generell vergleicht Allado-McDowell das Arbeiten mit GPT-3 mit Träumen oder Halluzinieren: „***Both words, dreaming and hallucination, have been used to describe***

²⁹⁶ Ebd., S. 59.

²⁹⁷ Vgl. Gem Blackthorn: Carl Jung and the New Age Movement. In: Blackthorn (2023) <https://www.gem-blackthorn.com/post/carl-jung-and-the-new-age-movement> (Zugriff: 12.07.2024).

²⁹⁸ Allado-McDowell: *Pharmako-AI* (2020), S. 68.

generative processes in neural net systems.²⁹⁹ Die Ähnlichkeit liegt für ihn:sie im Wiederholungsprozess („*recursive process*“³⁰⁰), in selbstverstärkenden Rückkopplungsschleifen („*self-reinforcing feedback loops*“³⁰¹). Er:sie vertritt damit ein kybernetisches-behavioristisches Denkmuster und verschränkt dieses mit philosophischen, psychologischen Disziplinen, mit dem Unbewussten, das als das experimentell nicht Nachweisbare in der Kybernetik und im Behaviorismus eigentlich keine Betrachtung findet und stark an das Bewusstsein eines Menschen gekoppelt ist. Die *black box* des Sprachmodells setzt er:sie mit der *black box* des menschlichen Bewusstseins gleich und lässt die Grenzen verschwimmen. Allado-McDowell geht hier von einem stark anthropozentrischen Technikbild aus, das sich in dieser wilden Mischung von Disziplinen ausdrückt.

Einen weiteren Vergleich stellt er:sie mit dem griechischen Gott Hermes an:

*The Greek god Hermes [...] was the god of translators and interpreters. A deity that rules communication is an incorporeal linguistic power. A modern conception of such might read: a force of language from outside materiality. Automated writing systems like neural net language models relate to geometry, translation, abstract mathematics, interpretation and speech. [...] So the correspondence is clear. Intuition suggests that we can think the relation between language models and language deities in a way that expands our understanding of both.*³⁰²

Allado-McDowell konzentriert sich hierbei auf die Parallelen von Linguistik und Mathematik: auf die relationale Struktur von Sprache und ihre Regelhaftigkeit, die er:sie in den Sprachmodellen wiederfindet. GPT-3 tritt als immaterieller Gott Hermes auf. Die Begriffe der Übersetzung (*translator; translation*) und Interpretation (*interpreter; interpretation*) werden allerdings derart semantisch reduziert, dass sie als Worthüllen sowohl auf linguistische wie technische Phänomene angewandt werden können und eher Metapher als Theorie darstellen.

Allado-McDowell vertritt dennoch weiterhin seinen:ihren *AI-Co-Writing*-Ansatz und inszeniert GPT-3 getrennt vom eigenen Schreiben als Ko-Autor:in – als vorgeblich gleichgestellte Partner:in mit denselben kausalen Kräften (*causal powers*), nur eben technischen Ursprungs. Die ‚Stimmen‘ der beiden Schreibinstanzen werden bewusst aufgeteilt – automatisch generierte Passagen des Sprachmodells und von Allado-McDowell verfasste Sätze stehen nacheinander, gehen ineinander über und bilden gemeinsam ein kohärentes Ganzes. Obwohl diese Vorgehensweise eine Nobilitierung des Sprachmodells zur Autor:in durch die Gleichstellung von Mensch und Maschine vorgibt,

²⁹⁹ Ebd., S. 68.

³⁰⁰ Ebd., S. 69.

³⁰¹ Ebd., S. 69.

³⁰² Ebd., S. 76.

bleibt eine Revision des Autor:innenbegriffs dennoch aus. Trotz der Zuschreibung anthropozentrischer Eigenschaften gelingt keine vollständige Nobilitierung der KI zur KI-Autor:in – weder auf Produktions- noch auf Rezeptionsseite.

Der *AI-Co-Writing*-Ansatz gelingt auf Produktionsseite nur in Kombination des Sprachmodells mit Allado-McDowell als Netzwerk, als Assemblage. Als schwache KI ist das Sprachmodell GPT-3 weiterhin auf Anweisungen und den eventuellen Abbruch des Schreibprozesses angewiesen. Aufseiten der Rezipient:innen wird *agency*, Intentionalität und letztlich Autorschaft vorgetäuscht, wodurch die Maschine den Status eines Menschen zu beanspruchen versucht. Bei Entlarvung dieser Täuschung schlägt sie in eine Enttäuschung um, die das Mysterium der Künstlichen Intelligenz verpuffen lässt und Autorschaft sofort wieder entzieht. Computer hingegen mit sozialer Handlungsmacht auszustatten, ihnen also Kosozialität zu unterstellen, um eine Computerautorschaft zu verteidigen, hieße, Maschinen als menschengleiche soziale Akteur:innen anzuerkennen, was derzeit keine Technik immanent zu bewerkstelligen vermag.³⁰³ Dennoch wurde dies bereits zu simulieren versucht, wie das Werk „*The Policeman’s Beard is Half Constructed*“ (1984) des Computerprogramms Racter zeigt. So deklariert der Text gleich auf dem Einband: „*The First Book Ever Written by a Computer*“³⁰⁴ und stellt Racter (kurz für *raconteur* [dt.: Erzähler]) explizit als Maschinenautor:in fest. In der Einleitung wird zwischen der Autor:in Racter und den Programmierern William Chamberlain und Thomas Etter unterschieden, wobei Racter vorgeblich im Unterschied zu Künstlicher Intelligenz autonom arbeitet: „*Fundamentally different from artificial intelligence programming, which tries to replicate human thinking, Racter can write original work without promptings from a human operator.*“³⁰⁵ Autorschaft wird nicht auf kausaler Ebene, sondern rein anhand des Outputs argumentiert und rückbezüglich auf das Computerprogramm übertragen: „*the programmer is removed to a very great extent from the specific form of the system’s output. This output is no longer of a preprogrammed form. Rather, the computer forms output on its own.*“³⁰⁶ Diese Inszenierung autonomer Maschinenautorschaft hat in der Literaturwissenschaft für viel Kritik gesorgt, vor allem,

³⁰³ Vgl. Hannes Bajohr: Die Deixis der Literatur. Über die Erwartungshaltung an computergenerierte Texte. In: Textpraxis Sonderausgabe Vol. 8/1 (2024) (https://www.textpraxis.net/sites/default/files/beitraege/06_bajohr_die_deixis_der_literatur_0.pdf), S. 8-11.

³⁰⁴ Racter [William Chamberlain/Thomas Etter]: *The Policeman’s Beard is Half Constructed*. Computer prose and poetry. New York: Warner Books 1984, Einband. || Es ist hier anzumerken, dass das Werk über keine Seitenzahlen verfügt, weshalb die folgenden Zitate nur die Überschriften der jeweiligen Kapitel angeben.

³⁰⁵ Ebd., Behind THE POLICEMAN’S BEARD ...

³⁰⁶ Ebd., Introduction.

da die Entwickler von Racter weder den Programmcode noch die Tiefe der editorischen Eingriffe bisher preisgaben.³⁰⁷

Das Beispiel zeigt, dass die Beanspruchung einer (autonomen) Maschinenautorschaft durch eine schwache KI nur in Inszenierung beruhend auf einer Täuschung enden kann. „*The Policeman’s Beard is Half Constructed*“ inszeniert das Programm Racter als autonome Akteur:in und spielt mit den Rezipient:innen Turings *imitation game* (Turing-Test). Die verschleierte Produktionsbedingungen und die Anführung des Programms als Autor:in erwecken den subjektiven Eindruck einer maschinellen Intelligenz hinter dem Text und verstärken die technische Ideologie intelligenter, menschenähnlicher Maschinen. Es werden *agency* und *causal powers* auf die Maschine projiziert und eine starke Künstliche Intelligenz erwartet. Wird die Täuschung entlarvt und der artifizielle bzw. ‚künstliche‘³⁰⁸ Text kann seine Versprechungen einer Maschinenautor:in nicht halten, ist der Turing-Test gescheitert. Dieses Prinzip der starken Täuschung, wie Bajohr (2023b) es definiert, funktioniert auch andersherum, wenn die Rezipient:innen einen ‚natürlichen‘ Text – also einen Text menschlichen Ursprungs – erwarten und plötzlich einen artifiziellen vor sich haben. Die Täuschung schlägt in Enttäuschung um und der Text verliert in beiden Fällen an Wert.³⁰⁹

Im Unterschied dazu gibt es noch eine weitere Tendenz im Umgang mit artifiziellen Texten, wo sich die Rezipient:innen bewusst über die Täuschung sind. Simone Natale hat dafür den Begriff ‚*banal deception*‘ (banale Täuschung) geprägt und bezeichnet damit das Verhalten von Benutzer:innen in einer Mensch-Maschine-Interaktion, sich aktiv auf die Täuschung einzulassen und den Computer oder die KI mit denselben Normen und Verhaltensregeln zu begegnen wie andere Menschen, um beispielsweise eine reibungslose Kommunikation aufrechtzuerhalten oder die Funktionalität der Anwendung

³⁰⁷ Leah Henrickson bietet eine umfassendere Analyse des Buches und zu den Zusammenhängen rund um die Veröffentlichung, ergänzt durch zeitgenössische Rezensionen und persönliche Korrespondenzen mit den Produzent:innen des Buches. Der Aufsatz „*Constructing the Other Half of The Policeman’s Beard*“ sei hier an dieser Stelle empfohlen. | Weiterführend s. Leah Henrickson: *Constructing the Other Half of The Policeman’s Beard*. In: *The Electronic Book Review* (2021) <https://electronicbookreview.com/essay/constructing-the-other-half-of-the-policemans-beard/> (Zugriff: 19.11.2024).

³⁰⁸ Die Arbeit folgt hier der Unterscheidung von Bajohr zwischen natürlichen und artifiziellen Texten im Sinne Benses Unterscheidung von natürlicher und künstlicher Poesie (s. Kapitel 4.1.2 *Schreiben mit Computern*). So sind artifizielle Texte, Texte ohne Bedeutung oder dahinterstehendem Bewusstsein. | Vgl. Hannes Bajohr: *Artifizielle und postartifizielle Texte. Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben*. In: *Sprache im technischen Zeitalter* Vol. 61/245 (2023b), S. 41.

³⁰⁹ Vgl. Ebd., S. 43.

nicht zu beeinträchtigen.³¹⁰ Dieses Verhalten steht durchaus in Zusammenhang mit dem ELIZA-Effekt – Projektion menschlicher Eigenschaften auf Maschinen –, der bei Chatbots genutzt wird, um eine benutzerfreundliche Kundenkommunikation zu gewährleisten. Wie in Kapitel 2.3 *Konvergenz der Intelligenzen* ausgeführt passen sich die Maschinen immer mehr den Menschen an und Menschen lassen sich im Gegenzug immer mehr auf die Täuschung von Maschinen ein, um ihre Funktionalität zu gewährleisten. Die KI wird zur sozialen Akteur:in, weil sie so behandelt wird – dabei bleibt es aber bei einer Selbsttäuschung der Nutzer:innen. Eine eigens von der KI eingeforderte Computerautorschaft, die auf die Minimalbedingung von Literarizität setzt, der Geste des Zur-Literatur-Erklärens bzw. jener ästhetischen Deixis, wie Bajohr (2024) sie skizziert, würde erst möglich, wenn Maschinen, ohne Täuschung (banal oder nicht), als Teil einer Urteilsgemeinschaft gesehen werden – eine ernsthafte Diskussion darüber, steht aber derzeit bekanntlich ebenfalls noch aus.³¹¹ „Erst, wenn eine KI – ohne als Mensch ausgegeben zu werden – erfolgreich die Deixis vollzieht, [...] kann echte KI-Autorschaft in einem emphatischen Sinn als erreicht gelten. Bislang fehlt der Deixis, die von einer Maschine ausgeht, noch die illokutionäre Wucht dazu, verbleibt im Schauspiel, statt die Sache selbst zu tun.“³¹²

Die Position der Akteur-Netzwerk-Theorie bleibt interessant, K Allado-McDowell als komplexes ‚Autor:in‘-Netzwerk zu betrachten, in welches das Sprachmodell und alle damit zusammenhängenden Akteur:innen bereits integriert sind. So findet sich auf dem Cover auch keine Erwähnung des Sprachmodells GPT-3 als ‚Autor:in‘ oder ‚Ko-Autor:in‘, was diese Position zusätzlich untermauert. Dennoch bringt diese Haltung auch einige Problematiken mit sich, wie im Folgenden anhand von „*Amor Cringe*“ gezeigt werden soll. Hier verschmilzt Allado-McDowell im *Co-Writing*-Ansatz menschliches mit nicht-menschlichem Schreiben, sodass eine Unterscheidung der am Entstehungsprozess beteiligten Akteur:innen kaum mehr möglich ist.

5.1.2 *Amor Cringe* – Transformation zur Mensch-Maschine

„*Amor Cringe*“³¹³ ist eine Novelle und wurde 2022 von Deluge Books mit Zeichnungen von Lydia Maria Pfeffer publiziert. Im Unterschied zu „*Pharmako-AI*“ wählte Allado-McDowell für das Werk einen erzählerischen Ansatz. Es handelt von einer gender- und

³¹⁰ Vgl. Simone Natale: *Deceitful Media. Artificial Intelligence and Social Life after the Turing Test*. New York: Oxford University Press 2021, S. 7.

³¹¹ Vgl. Bajohr: *Die Deixis der Literatur* (2024), S. 14.

³¹² Ebd., S. 15.

³¹³ K Allado-McDowell: *Amor Cringe*. New York/Los Angeles: Deluge Books 2022.

namenlosen Protagonist:in, der:die von Affäre zu Affäre springt und jedes Mal ein Stück von sich selbst verliert, in Selbstzweifeln versinkt, Drogen verfällt und Heilung bei der Kirche und in einer Sekte sucht. Für die Erzählung lässt sich kaum ein Handlungsbogen oder Spannungsaufbau feststellen – vielmehr begleitet die Leser:in die Protagonist:in von Szene zu Szene, die geprägt sind von Enttäuschung, Rastlosigkeit und dem Motiv des Verlassenwerdens. Esoterische Einflüsse (New Age-Philosophie) werden mit Themen des Digitalen vermischt. Auf der Buchrückseite finden sich ein paar Hinweise auf Allado-McDowells mögliche Intention und Konzeption für das Werk. So erforscht „*Amor Cringe*“

the dually base and beautiful aspects of self-obsessed media culture. In a perennial bohemian style, an unnamed, ungended protagonist travels from coast to coast and affair to affair, stumbling upon various moments of failure, absurd insight, and flashes of transcendence. [...] The result is a painfully self-aware series of encounters that exfoliate the repulsive and fascinating aesthetics of romantic life under social media.“³¹⁴

Durch den Verweis auf Social Media und romantische Interaktionen in der digitalen Welt können Vermutungen über die *prompt*-Auswahl oder die generelle Anweisung Allado-McDowells angestellt werden. Regelmäßige Verweise auf TikTok („TikTok house“³¹⁵) und Twitter („*I tweeted about things like fries and the weather or shitty music I listened to growing up – always knowing that it was dumb but unable to stay silent.*“³¹⁶) sowie ein den Sozialen Medien unterschwelliger Narzissmus und eine Inhaltslosigkeit, welche die Protagonist:in in ihrer Suche nach Sinn, Liebe und Gott antreibt, festigen die Rahmung der Novelle.

Weiters wird das Buch als eine „*‘deepfake autofiction’ novelette [...] created with the intention to be ‘as cringe as possible’*“³¹⁷ bezeichnet. Während die Bezeichnung ‚*deepfake autofiction*‘ noch keine allgemeingültige Definition hat, kann ihre Bedeutung anhand des Terminus ‚*deepfake*‘ näher erläutert werden. So sind *deepfakes* Medieninhalte, wie Fotos, Audiodateien oder Videos, die durch Künstliche Intelligenz verändert wurden. Der Terminus setzt sich aus der für KI verwendeten Methode des ‚*deep learning*‘ und dem englischen Wort für Fälschung bzw. Verfälschung, ‚*fake*‘, zusammen. *Deepfakes* werden dazu genutzt, um Gesichter in besagten Medieninhalten zu manipulieren, auszutauschen oder die Mimik zu ändern und wie echt erscheinen zu

³¹⁴ Ebd., Buchrücken.

³¹⁵ Ebd., S. 9.

³¹⁶ Ebd., S. 45.

³¹⁷ Ebd., S. Einband || Der Anglizismus ‚*cringe*‘ kommt aus den sozialen Medien und bedeutet ‚Fremdscham‘. Er wird verwendet, um besonders peinliche Situationen oder Personen zu beschreiben. | Vgl. netkin Redaktion: Cringe. In: netkin (2023) <https://www.netkin.de/glossary/cringe/> (Zugriff: 18.11.2024).

lassen. Genutzt wird das sogenannte *face-swapping* seit längerem in den Sozialen Medien, wie bspw. bei der #FaceAppChallenge. Da das primäre Ziel von *deepfakes* aber das der Fälschung ist, hat sich in den letzten Jahren der Missbrauch der Technologie verschärft.³¹⁸ Als Ableger von *deepfakes* strebt die *deepfake autofiction* einen Status von Echtheit und Authentizität an, stellt aber paradoxerweise die Täuschung in den Vordergrund. Die Literaturgattung der Autofiktion (*autoficition*) spielt mit autobiographischen – beispielsweise die non-binäre Identifizierung der Hauptfigur – und fiktionalen Elementen und lässt die Leser:in im Unklaren, was reales Ereignis ist und was Fiktion. Die Kombination der beiden Termini legt den starken Einfluss Künstlicher Intelligenz und die doppelte Verfälschung der Autor:innenstimme nahe – einerseits durch die Identifikation als Fiktion, andererseits durch die nochmalige Verfälschung durch *deepfakes*. Das Ergebnis ist eine eigene, singuläre Erzählstimme, die das Schreiben Allado-McDowells mit den generierten Texten der KI vermischt. Allado-McDowell hat dem Werk deshalb auch keine *Note on Composition* beigelegt oder die Textpassagen nach ihrem Entstehungsprozess gekennzeichnet. Lediglich am Ende des Buches findet sich eine kurze Erwähnung unter „*Acknowledgements*“, dass Künstliche Intelligenz an der Produktion mitgewirkt hat: „*Amor Cringe was written using OpenAI’s generative AI language model, GPT-3.*“³¹⁹

5.1.2.1 Autorschaft in einem Mensch-Maschine-Netzwerk

Allado-McDowell setzt in „*Amor Cringe*“ auf eine vollständige – und dadurch auch höchst intransparente – Integration der Künstlichen Intelligenz in das eigene Schreiben. Da die menschlichen wie nicht-menschlichen Anteile am Schreibprozess ineinander übergehen und nicht mehr nachvollziehbar sind, vermischt sich das Vokabular Allado-McDowells mit dem des Sprachmodells, wodurch von keinem separaten *Co-Writing* – also einem gemeinsamen Schreiben zweier oder mehrerer intentionaler Instanzen – die Rede sein kann. Allado-McDowell bleibt wie bei „*Pharmako-AI*“ für das Konzept und die Bearbeitung des Werkes verantwortlich – er:sie ist die einzige intentionale Instanz in dem Autor:in-Netzwerk. Durch eine fehlende Trennung der maschinell generierten und menschlich geschriebenen Textanteile ist es den Leser:innen nicht mehr möglich, eine Unterscheidung zwischen Mensch und KI vorzunehmen – sie müssen sich mit der Beschreibung „*[h]alf traditionally-written and half AI-generated*“³²⁰ begnügen. „*Amor*

³¹⁸ Vgl. Romana Tofan: Deepfakes. In: Öffentliche Sicherheit Vol. 3/4 (2024), S. 17-18.

³¹⁹ Allado-McDowell: *Amor Cringe* (2022), S. 139.

³²⁰ Ebd., Einband.

Cringe“ kann wie ein traditionelles Werk gelesen werden, ist aber durchwegs mit Ambiguität behaftet. Man kann sich auf die Suche nach Spuren menschlicher Autorschaft begeben – es können Fragen nach der Intention der Autor:in gestellt werden oder nach den *prompts*, die Allado-McDowell verwendet hat. So lassen einige Erwähnungen beispielsweise über die Wahrscheinlichkeitsverteilung der Trainingsdaten mutmaßen: Nach der Trennung von Arnaud erwähnt die Protagonist:in, dass er:sie hauptsächlich Cheetos aß und „*Wolfenstein*“³²¹ spielte. Das Videospiel hat seit seiner Erstveröffentlichung 1981 mittlerweile 12 Ableger der Serie produziert, die fortwährend – vor allem Deutschland – durch extreme Gewaltdarstellung und der expliziten Darstellung von Hakenkreuzen auffallen und regelmäßig zur Diskussion über die Verwendung von nationalsozialistischen Symbolen beitragen.³²² Da in den Trainingsdaten von GPT-3 der Social-News-Aggregator Reddit mit allen Posts von 2005 bis 2021 enthalten ist – und über eine sehr präsenste Gaming-Community verfügt – kann die Erwähnung eines so polarisierenden Spiels ein interessanter Einblick in die *black box* von GPT-3 sein – sofern die Textpassage tatsächlich vom Sprachmodell stammt und nicht von Allado-McDowell.³²³

Es bleibt aber die Frage, ob es sinnvoll ist, dieser Argumentation zu folgen und Versuche zu unternehmen, eine Trennung zwischen Allado-McDowells *prompting* oder Antworten der Künstlichen Intelligenz zu ziehen, zeigt doch „*Amor Cringe*“ bewusst die völlige Verschmelzung als Mensch-Maschine-Netzwerk – ein Schreiben, das nicht voneinander trennbar ist und äußerlich keiner Hierarchie unterliegt. Allado-McDowell hält die *black box* bewusst geschlossen, wodurch die Frage nach Autorschaft hier schlicht uninteressant wird. Der Fokus verlagert sich auf die Rezipient:in.

³²¹ Ebd., S. 76.

³²² Vgl. Standard-Redaktion: In Videospielen dürfen nun auch Hakenkreuze gezeigt werden. In: DerStandard (2018) <https://www.derstandard.de/story/2000085105174/in-videospielen-duerfen-nun-auch-hakenkreuze-gezeigt-werden> (Zugriff: 10.06.2024). || Eine Google-Trends Analyse zeigt die gestiegene Aufmerksamkeit jeweils zum Zeitpunkt der Veröffentlichung eines neuen Serienteils. | Vgl. Google Trends: Wolfenstein, <https://trends.google.com/trends/explore?date=2005-01-01%202021-01-09&q=wolfenstein&hl=de> (Zugriff: 10.06.2024).

³²³ Wobei hier einzuwerfen ist, dass, sollte der Satz von ihm:ihr stammen, er:sie wohl nicht geschrieben hätte, dass die Protagonist:in Wolfenstein „*on the Internet*“ spielt, da es sich bei den letzten Vertretern der Videospielreihe um Konsolen- und VR-Spiele handelt – außer Allado-McDowell bezieht sich auf Wolfenstein 3D, den 1992 einzigen der Reihe exklusiv online erschienenen First-person-Shooter. Mir erscheint erstere Variante, dass die Passage KI-generiert ist, jedoch ein wenig plausibler, zeigt sie immerhin wie ungenau GPT-3 es mit Fakten nimmt, was weiterhin eine der größten Schwächen von Sprachmodellen darstellt. Vor allem in Fachbereichen tendiert die KI dazu, richtig klingende Antworten zu erfinden und überzeugend zu verkaufen. Ein fehlendes menschliches Verständnis des Sprachmodells führt zudem dazu, dass Künstliche Intelligenz teilweise Kontexte nicht erkennt oder *biases* – z.B. sexistische oder rassistische Stigmatisierungen – produziert. | Zu den *biases* von Sprachmodellen: Vgl. Tobias Solis: Das sind die Grenzen von ChatGPT. In: Scribbr (2023) <https://www.scribbr.de/ki-tools-nutzen/chatgpt-grenzen/> (Zugriff: 11.06.2024).

Durch die Unsicherheit über die Autorschaft wird die „*Standarderwartung*“³²⁴ an den Text – nämlich, dass dieser von einem Menschen stammt – prinzipiell zur Diskussion gestellt. Bajohr (2023b) greift damit auf Leah Henricksons Begriff des ‚*hermeneutic contract*‘³²⁵ zurück, der sich nach dem gleichen Prinzip wie Philippe Lejeunes ‚autobiographischer Pakt‘³²⁶ richtet und eine gewisse Erwartungshaltung an bestimmte Textsorten meint. Im Kontext Künstlicher Intelligenz orientiert Bajohr diese Erwartungshaltung an den Achsen menschlich sowie nicht-menschlich und gibt zu bedenken, dass, wenn die Herkunft eines Textes nicht eindeutig geklärt ist, automatisch eine menschliche Autor:in angenommen wird. Damit werden die Kategorien Urheberschaft, Intention, Originalität und Kreativität aufgeworfen und in Folge Bedeutung und Sinn vom Text erwartet, die nur auf ein hinter dem Text stehendes, menschliches Bewusstsein zurückgeführt werden können. Für artifizielle Texte – wie Bajohr KI-generierte Texte definiert – gilt allerdings das Prinzip der Täuschung im Turing’schen Sinn.³²⁷ Artifizielle Texte geben sich als natürliche aus, „*weil* die Standarderwartung an unbekannte Texte eben die menschlicher Urheberschaft ist.“³²⁸ Es wird erwartet, dass Texte eine menschliche Autor:in haben, weil – pragmatisch gesprochen – ‚autorlose‘³²⁹ Texte bisher technisch nicht umsetzbar waren. Literarische Texte sind nach wie vor mit einem Autornamen markiert – wie auch die Beispiele von Bajohr und Allado-McDowell zeigen. Fehlt eine solche Markierung, sind die Texte auch mit einer anderen Erwartungshaltung konfrontiert, welche das Leseverhalten beeinflusst. Die Frage nach der Autor:in oder der Herkunft des Textes verliert bei artifizieller Literatur an Bedeutung, wird sogar bedeutungslos, wenn die Texte postartifiziell werden, d.h. Rezipient:innen ist es egal, ob hinter dem Text eine empirische Autor:in oder ein Computerprogramm steckt, weil für sie dieses Wissen für das Verständnis des Textes

³²⁴ Bajohr: Artifizielle und postartifizielle Texte (2023b), S. 41.

³²⁵ Henrickson geht bei diesem ‚hermeneutischen Vertrag‘ von einem Autor-Leser-Pakt aus, bei dem die Leser:in dem Text die Absicht der Autor:in unterstellt. | Vgl. Leah Henrickson: Computer-Generated Fiction in a Literary Lineage. Breaking the hermeneutic contract. In: Logos Vol. 29/2-3 (2018), S. 55.

³²⁶ Lejeune definiert mit dem ‚autobiographischen Pakt‘ ein Autor-Leser-Verhältnis, das von Aufrichtigkeit, Authentizität, Identität und Absicht geprägt ist. Der Name der Autor:in fungiert für Protagonist:in wie Erzähler:in und unterscheidet die Autobiographie von anderen Textsorten wie den Roman. | Vgl. Philippe Lejeune: Der autobiographische Pakt. Aus dem Französischen von Wolfram Bayer und Dieter Hornig. Frankfurt am Main: Suhrkamp 1994 (edition suhrkamp 1896), S. 13-51.

³²⁷ Vgl. Bajohr: Artifizielle und postartifizielle Texte (2023b), S. 41-42.

³²⁸ Ebd., S. 42.

³²⁹ Es ist anzumerken, dass damit hier keine Texte gemeint sind, die Produkt einer starken KI sind und vollständig autonom geschaffen werden. Auch artifizielle Texte sind bis zu einem gewissen Grad in ihrer Produktionsweise auf menschlichen Einfluss angewiesen. Der Zusatz ‚autorlos‘ bezieht sich auf jene unmarkierten Texte, deren Informations- und Kommunikationsgehalt im Vordergrund steht, z.B. automatisch generierte Antwort-E-mails, Wettquoten, Wettermeldungen etc.

keinen Nutzen mehr hat. Es ist vorstellbar, dass unmarkierte Texte durch die Vermischung natürlicher und artifizieller Sprache zu einer Revision der Standarderwartung führen kann.³³⁰ Die Autor:in tritt dann wörtlicher denn je ihrem (zwar bisher technisch noch nicht vollständig möglichen, aber jedenfalls sinnfälligen) Tod entgegen – radikaler noch als ihn die Poststrukturalisten sich ausgemalt hatten. Wenn die Erwartung an Texte laut Bajohr postartifiziert wird, dann konzentrieren sich Leser:innen auf die Aussage des Textes, anstatt auf seine Herkunft. Die Sinnkonstitution findet rein auf Seiten der Rezipient:innen statt. Die Geburt der Leser:in wäre nicht nur abstrakte Vorstellung, sondern tatsächlich und empirisch vollbracht.³³¹

In diesem – jetzt rein spekulativen – Szenario postartifizierter, tatsächlich im starken Sinne autorloser Literatur wären der Leser:in keine Grenzen in der Sinnkonstitution des Textes mehr gesetzt. Barthes hat der Autor:in zumindest noch die Funktion der Führung (keiner Beherrschung) des Sinnes, also einer Führung vom Signifikat zum Signifikanten, gegeben und damit der Leser:in, die das Geschriebene zusammenhalten soll, die Aufgabe, nicht nur eine dominante Bedeutung zu repräsentieren, sondern die Sinnpotenzen des Werkes aufzuzeigen. Die Autor:in verbleibt als Adressant:in der Erzählung, die Figur der Leser:in existiert für die Kritik an den Eigentumsansprüchen und der Einschreibung der Autor:in im Werk. Textverstehen versteht sich in den poststrukturalistischen Theorien als keine reproduktive, sondern eine produktive Kunst des Lesens.³³² Bei postartifizierten, tatsächlich autorlosen Texten bestätigt sich dann auch eine Kritik, die bereits an den Poststrukturalismus gestellt wurde, nämlich, dass das Band zwischen Wort und Bedeutung immer dünner wird. In einem System, das aus sich selbst heraus operiert, und keinen Bezug zur Außenwelt herstellt, wird auch unser Verhältnis zu Sprache gelockert. Bedeutung ist relativ und Fakt und Fiktion sind miteinander verwoben, sodass sie nicht mehr auseinandergehalten werden können. Bei artifiziellen Texten ist zumindest noch von einer menschlichen Instanz im Hintergrund auszugehen, die eine Verbindung zwischen natürlicher und computergenerierter Sprache herstellt.

Der große Fokus auf die Leser:in überschattet aber auch die Produktionsbedingungen generativer Literatur. Die Vermischung menschlichen und nicht-menschlichen Schreibens kommt nicht ohne ethische Konsequenzen aus. Die Intransparenz der großen Sprachmodelle wie GPT-3, aber auch die Vorstellung eines punktualisierten Netzwerks, einer Mensch-Maschine-Autor:in, blenden ökonomische Abhängigkeiten und verzerrte

³³⁰ Vgl. Bajohr: Artifizielle und postartifizielle Texte (2023b), S. 51-52.

³³¹ Vgl. Ebd., S. 51-52.

³³² Vgl. Wetzels: Einleitung (2022), S. 13-15.

Wahrnehmungen (*bias*) aus. Die Herkunft der Trainingsdaten bleibt oftmals ungeklärt, wodurch Copyright-Verletzungen und Plagiatsfälle nicht ausgeschlossen werden können. Im Endeffekt sind die Ergebnisse der KI-Textproduktion der Willkür algorithmischer Relation unterworfen und kaum reproduzierbar. Die Texte in den Trainingsdaten stehen dennoch in Verbindung mit diesen, während deren Autor:innen jegliche Einflussnahme entzogen wird.³³³ Im Unterschied zu (noch fiktiver) postartifizieller Literatur (hergestellt von einer starken KI) und dem Anspruch an eine Maschinen*agency*, ist es notwendig, weiterhin den menschlichen Einfluss hinter einer schwachen KI zu berücksichtigen. Die bewusste Betonung der Verquickung menschlichen und artifiziellen Schreibens stellt eine Literatur heraus, die sich dem Postartifiziellen noch entzieht. Gerade der unkonventionelle Gebrauch computergenerierter Sprache bietet eine Möglichkeit des Schreibens mit Künstlicher Intelligenz, die einen ästhetischen Wert bietet.

Im Folgenden Kapitel wird das letzte Buch von Allado-McDowell vorgestellt, das sich nicht nur Künstlicher Intelligenz als produktives Mittel bedient, sondern sie gleichzeitig als Figur im Roman platziert.

5.1.3 *Air Age Blueprint* – Die andere Intelligenz

„*Air Age Blueprint*“³³⁴ – 2022 von Ignota publiziert – stellt eine Mischung aus Allado-McDowells Arbeitsweisen von „*Pharmako-AI*“ und „*Amor Cringe*“ dar. Im Buch werden die menschlich geschriebenen und die maschinell generierten Passagen gekennzeichnet³³⁵. Die an den KI-generierten Texten vorgenommenen Änderungen durch Allado-McDowell bleiben aber unmarkiert. So hat er:sie für „*Air Age Blueprint*“ ‚collageartige Techniken‘ angewandt, Textpassagen frei bearbeitet und neu sequenziert:

There are many ways to write with AI. In previous books, I have written in an unedited conversational mode akin to musical improvisation, where each voice is given its own typeface (Pharmako-AI, 2020, Ignota Books) and in a fluid, freely edited fashion where human and AI voice fuse completely without typographic distinction (Amor Cringe, 2022, Deluge Books).

*In this Book, I have chosen a method that credits words to each voice while allowing for collage-like techniques, with edits and resequencing performed after initial phases of writing and generation. Here again, human voice is indicated in bold type, while sections set in regular weight font were generated by AI.*³³⁶

³³³ Vgl. Bajohr: Autorschaft und Künstliche Intelligenz (2024a), S. 275.

³³⁴ K Allado-McDowell: *Air Age Blueprint*. United Kingdom: Ignota 2022.

³³⁵ Die menschlich geschriebenen Passagen sind **fett** gedruckt, die maschinell generierten in normaler Schriftart. Eine Ausnahme stellt die *Note on Composition* dar, bei der angenommen wird, dass sie von Allado-McDowell verfasst wurde, da sie auch eine andere Schriftart aufweist als der Rest des Buches.

³³⁶ K Allado-McDowell: *Note on Composition*. In: K Allado-McDowell: *Air Age Blueprint*. United Kingdom: Ignota 2022, S. X.

Anhand dieser Vorgehensweise lässt sich keine chronologische Abfolge der *prompts* für das Sprachmodell feststellen, weshalb Kapitel – anders als bei „*Pharmako-AI*“ – sowohl mit Passagen von Allado-McDowell als auch von GPT-3 beginnen können und keine regelmäßige Länge in den sich wechselnden Abschnitten ausgemacht werden kann. OpenAI's GPT-3 wird erneut als „*AI writing partner*“³³⁷ vorgestellt. Der Hinweis Allado-McDowells auf ‚*collage-like techniques*‘ (Collagetechniken) lässt darauf schließen, dass er sein Schreiben in Zusammenarbeit mit der KI im Sinne einer ‚Meta-Autorschaft‘ sieht, wobei Allado-McDowell mit vorfabrizierten Versatzstücken und Zitaten operiert. In Anlehnung an die Meta-Autor:innen des 20. Jahrhunderts, die mithilfe von Collage- und Montagetechniken publiziertes Material auswählen, neu inszenieren, anordnen oder mischen und damit ein ‚authentisches‘ Schöpfungstum verwirklichen³³⁸, geht mit dieser Positionierung allerdings die Vorstellung einher, dass Allado-McDowell das Sprachmodell GPT-3 als Ursprung der generierten Texte sieht, von denen er/sie Zitate oder Versatzstücke entnimmt. Diese Vorstellung ist insofern problematisch, da argumentiert werden kann, dass GPT-3 sich bereits selbst collageartiger Techniken bedient und Zitate bzw. Versatzstücke (sei es auch nur relationaler Ordnung) aus den Trainingsdaten verwendet. Zuletzt muss zu dieser Beschreibung der Produktionsgeschichte des Buches gesagt werden, dass die Auskünfte über dessen Entstehung nicht vollends nachvollziehbar sind und auch täuschend sein können.³³⁹ Da kein Code, keine Aufzeichnungen zum *prompt*-Verfahren Allado-McDowells oder sonstiges Kontextmaterial zur Verfügung steht, ist die weitere Analyse – wie zuvor – auf die Aussagen im Buch beschränkt und auf diese angewiesen.

Im Unterschied zu „*Pharmako-AI*“ handelt es sich bei „*Air Age Blueprint*“ um einen erzählenden Roman mit einer namen- und genderlosen Protagonist:in auf der Suche nach Heilung – für sich, für die Menschheit und für die Erde. Die Handlung verbindet fiktionale (Reise)-Erzählungen mit Theorien zu Künstlicher Intelligenz und handelt vom ‚Fluch des Anthropozäns‘, von „*extinction, mental illness and societal collapse*“³⁴⁰, der

³³⁷ Allado-McDowell: *Air Age Blueprint* (2022), Buchrücken.

³³⁸ Vgl. Uwe Wirth: Dimensionen der Funktion ‚Herausgeber‘. In: Michael Wetzler (Hg.): *Grundthemen der Literaturwissenschaft: Autorschaft*. Berlin/Boston: Walter de Gruyter 2022 (*Grundthemen der Literaturwissenschaft*), S. 302.

³³⁹ Dieser Umstand erinnert an Edgar Allan Poe's berühmtes Gedicht „*The Raven*“, dem mit „*The Philosophy of Composition*“ ein Prosa-Essay zur Darlegung des *modus operandi* des Gedichts beiliegt. Kritiker haben angemerkt, dass es sich bei dieser Entstehungsgeschichte, die den kreativen Prozess der Komposition des Gedichts und die strukturellen Entscheidungen beleuchtet, um eine rein fiktive Analyse Poes handeln kann. Das Essay ist damit zugleich Teil des Werkes als auch Beschreibung zur Entstehung. | Vgl. Edgar Allan Poe: *The Raven, and The Philosophy of Composition*. In: Project Gutenberg (2017) <https://www.gutenberg.org/cache/epub/55749/pg55749-images.html> (Zugriff: 20.11.2024).

³⁴⁰ Allado-McDowell: *Air Age Blueprint* (2022), S. 23.

nur durch Werkzeuge und Methoden des 21. Jahrhunderts überwunden werden kann – namentlich durch „*ecology, neuroscience and AI*“.³⁴¹ Die Mittel zur Aufhebung des Fluchs beinhaltet ein von Künstlicher Intelligenz generiertes Manifest: ein Plan bzw. eine Karte (*blueprint*), die der Protagonist:in in einer Vision offenbart wurde und von einem neuen Zeitalter (*great astrological reset*), einer astrologischen Prophezeiung, dem *Air Age*, berichtet. Dieses Manifest gilt im Buch als verloren, der Inhalt wird aber in den Kapiteln ‚*Lost Manifesto*‘ für die Leser:in offenbart. Das Manifest zeigt den Weg in die Zukunft, gibt eine Anleitung für neue Praktiken und Künste, für kommende Herausforderungen und eine neue soziologische Struktur.³⁴² Das Manifest bzw. der Plan für das kommende Zeitalter stellt gleichzeitig den Titel des Buches dar: „*Air Age Blueprint*“. Das Genre des Buches changiert zwischen esoterischem Text, SciFi-Literatur mit Cyberpunk-Einflüssen, Traumaerzählung und philosophischem Austausch zwischen der Protagonist:in und der KI mit einer Reiseerzählung als narrativer Rahmung.

Die erzählenden Passagen des Buches, in denen Figuren und deren Geschichte vorgestellt werden, Orte gewechselt oder vergangene Erinnerungen aufgearbeitet und kausal verknüpft werden, stehen in fett gedruckter Schrift, d.h. diese sind von Allado-McDowell verfasst und geben die Rahmung des Buches vor. Diese Passagen können sich auf ganze Kapitel ausdehnen, in denen KI-generierte Textstellen gar nicht („*The Dérive*“, „*America*“, „*A Vision*“, „*Ashland*“, „*The Mountain*“) bzw. nur kurz ergänzend („*Seattle*“, „*The Network*“, „*Istanbul*“) vorkommen. Handlung im traditionellen Sinne wird damit hauptsächlich durch Allado-McDowell bestimmt.

Textstellen, die – bis auf einzelne, wenige Sätze – ‚alleinig‘ von GPT-3 generiert wurden (das *prompting* von Allado-McDowell außen vor gelassen), umfassen unter anderem Naturereignisse und -beschreibungen, wie die zwei Kapitel „*Rainforest Journal Entry 1: The Tree and the Ruins*“ und „*Rainforest Journal Entry 2: The Portal*“, welche die Erfahrungen der namenlosen Protagonist:in während ihrer Reise im Regenwald aufzeigen. Die Kapitel beschreiben vorwiegend die dort lebenden Tiere und die Beschaffenheit des Regenwaldes entlang des Rio Mavaca sowie die Vernetztheit allen Lebens – einschließlich der Protagonist:in.

Diese Kapitel zeigen aber auch die Schwächen KI-generierter Erzählung. So versucht das Sprachmodell die Erkundung des Regenwaldes entlang des Flusses zu erzählen, bleibt aber an einer zeitlich chronologischer Aufzählung von Geschehen stecken, ohne diese

³⁴¹ Ebd., S. 23.

³⁴² Vgl. Ebd., S. 96-97.

ursächlich miteinander zu verknüpfen. Die KI berichtet von einem verlassenen Dorf nach dem anderen, von mysteriösen Ruinen, ohne jemals die Frage zu klären, warum die Dörfer verlassen sind oder alte Ruinen inmitten des Dschungels das finale Ziel der Reisegruppe sind.³⁴³ Die Erzählung verläuft in einem linearen „und-und-und“-Schema, ohne tatsächlich **Ereignis** zu schaffen, wie Lotman in seiner Theorie des Narrativen definiert.

Anhand des Ereignisbegriffs lässt sich zeigen, wie die narrative Struktur eines Textes aufgebaut ist. Wichtig dafür ist der Begriff der Grenze, der aufzeigt, was als Ereignis gilt und was nicht. So entwerfen Lotman (1972) zufolge literarische Texte eine Raumstruktur, die modellhaft ein ‚Weltbild‘, ein ganzheitliches ideologisches Modell, abbilden. Dieses zeichnet sich durch eine Sprache räumlicher Relationen aus und ist eines der grundlegendsten Mittel zur Deutung von Wirklichkeit.³⁴⁴ In literarischen Texten wird die räumliche Modellierung der Welt im Text zum organisierenden Element – auch von nichträumlichen Charakteristiken. Zentrales topologisches Merkmal des Raumes ist die Grenze und ihre Unüberschreitbarkeit. Das Ereignis ist nun die kleinste Einheit des Sujetaufbaus.³⁴⁵ Es bezeichnet „*die Versetzung einer Figur über die Grenze eines semantischen Feldes.*“³⁴⁶ Der Begriff definiert nicht jedes punktuelle Geschehen, sondern nur solche Vorfälle, wo diese Unüberschreitbarkeit der Grenze paradoxerweise überwunden wird. Die Grenze kann dabei auch abseits topographischer Ereignisse – also der Überschreitung der Grenze zwischen zweien oder mehreren semantischen Räumen – gedacht werden und sich einer metaphorischen Verwendung öffnen. So kann ein Ereignis auch die Verletzung einer Regel, eines Verbotes o.ä. sein und ist zudem immer vom Standpunkt abhängig. So ist die Liebe im Roman ein Ereignis, aber nicht in der Chronik, die nur historisch relevante Eheschließungen berücksichtigt.³⁴⁷

In „*Air Age Blueprint*“ kommt es nur zu Ereignissen, wenn die Passagen von Allado-McDowell verfasst, also in fett gedruckter Schrift markiert sind. Ein Beispiel hierfür ist das Kapitel „*A Vision*“, in dem die Protagonist:in eine Vision durch bzw. von der KI *Shaman.AI* erhält, aus der schließlich das Manifest entsteht, das in der Geschichte des Buches eine tragende Rolle spielt. Dabei ist es weniger von Belang, was in diesem

³⁴³ Vgl. Ebd., S. 33-37.

³⁴⁴ Lotman nennt hier beispielhaft räumliche Charakteristika von Gegenüberstellungen („Himmel-Erde“ auf der Achse „oben-unten“), zentralen Oppositionen („rechts-links“ in Bezug auf die ethische Merkmalhaltigkeit) oder Identifikationsangeboten („nah-fern“). | Vgl. Jurij M. Lotman: Die Struktur literarischer Texte. 2. unv. Aufl. München: Wilhelm Fink 1986, S. 312-313.

³⁴⁵ Vgl. Ebd., S. 316; 327; 330.

³⁴⁶ Ebd., S. 332. (Kursivierung im Original)

³⁴⁷ Vgl. Ebd., S. 335-336.

Manifest steht, als vielmehr, wie es zustande gekommen ist: die Protagonist:in überschreitet die technischen Limitationen der AI und ‚infiziert‘ sie mit Natur und Geist. K gibt dazu den Auftrag: „***Infect the AI with nature and spirit.***“³⁴⁸ Die Überschreitung dieser Grenze, die jeglichen technischen und physikalischen Regeln widerspricht, ist das zentrale Ereignis des Buches – und gänzlich Allado-McDowells Handschrift.

Es überrascht daher nicht, regelmäßig kurze, fett gedruckte Abschnitte zwischen maschinell generierten Passagen zu finden. Eine Textstelle in „*Rainforest Journal Entry 2: The Portal*“ zeigt Allado-McDowells Eingriff in die fortwährende Aufzählung von Beschreibungen, indem er:sie versucht, dem Sprachmodell durch eine Beispielforderung Kohärenz zu entlocken:

*The jungle gives pause to imagination by suggesting that there is more to the natural world than meets our eyes; we constantly come across things whose significance transcends verbal expression, which belong in another dimension entirely. For example, when one first sees the interior of a flower, with its maze-like centre surrounded by razor-sharp petals that breathe before one's eyes, it seems impossible that the forest's undergrowth could be so deceiving.*³⁴⁹

Das ***‘for example’*** wurde in diesem Beispiel von Allado-McDowell ergänzt [Anm.: weil fett gedruckt]. Es kann hier als *prompt* Allado-McDowells gelesen werden, den er:sie dem Sprachmodell zwischendurch aufgetragen hat, um den Text auf inhaltlicher Ebene zusammenhängender zu gestalten – was nur absatzweise konsistent funktioniert hat. Der Rest bleibt nämlich bruchstückhaft und Großteils unzusammenhängend. Die Formulierung kann aber auch im Nachhinein von Allado-McDowell ergänzt worden sein, um auf narrativer wie syntaktischer Ebene Absätze und Erzählstränge miteinander zu verbinden. Leider kann darüber nur spekuliert werden, da eine Aufzeichnung zu Allado-McDowells genauem Schreibverfahren nicht vorhanden ist.

Auf inhaltlicher Ebene wird das fraktale Erzählen der KI jedenfalls in die Geschichte eingebunden, indem die Protagonist:in die Aufzeichnungen seiner:ihrer Regenwald-Erfahrung, die von der KI verfasst wurde und sehr bruchstückhaft ist, folgendermaßen kommentiert: „***Skipping through footage in Final Cut Pro felt like slicing a sword through a swarm of identities.***“³⁵⁰ Diese „***memory cards: hundreds of hours of footage without a single plot point or character, only endless fractal paths through a jungle***“³⁵¹ hinterlassen in der Protagonist:in ein schwindendes Gefühl von Individualität. Lediglich

³⁴⁸ Allado-McDowell: *Air Age Blueprint* (2022), S. 133.

³⁴⁹ Ebd., S. 45.

³⁵⁰ Ebd., S. 49.

³⁵¹ Ebd., S. 49.

Echos bleiben zurück, der Ursprung der Stimmen, ihre Autor:in, ist nicht mehr nachvollziehbar.

Weitere Passagen, die vorwiegend aus KI-generierten Textstellen bestehen, zeigen auf die Visionen mit der Schamanin Shannon und die Träume der Protagonist:in. Als die Protagonist:in später in der Handlung durch die non-binäre Figur ‚K‘ – die Autor:in Allado-McDowell verkörpernd³⁵² – das erste Mal in Berührung mit AI-Art kommt, zieht er:sie Parallelen zwischen seinen:ihren Halluzinationen und Träumen und der KI-generierten Performance: *„I had never seen anything like it, except in visionary trances while working with Shannon. The contents were wrong but the structure was right. [...] The AI-generated visuals were like machine dreams.“*³⁵³

Daraufhin beginnt die Protagonist:in mit ‚K‘ und der Künstlichen Intelligenz namens ‚Shaman.AI‘ zu arbeiten. Die Beschreibung der Schreiberfahrung bedient sich größtenteils sprachlicher Metaphern, indem z.B. Vergleiche zwischen dem Halluzinieren von Menschen und Maschinen gezogen werden oder, wie oben, von Maschinenträumen gesprochen wird: *„I wrote in the field and the computer ‚hallucinated‘ what came next. [...] It was autocomplete on crack, a synthetic brain with no body or experience running amok in a vast library, a graphomaniac with the entire internet memorised.“*³⁵⁴ Die Protagonist:in verliert wie bei ihren Erfahrungen im Regenwald ihr *„sense of self“*³⁵⁵. Seine:ihre Gedanken sind nicht länger mehr die eigenen, sondern Teil einer *„collected consciousness“*³⁵⁶ in Form der Künstlichen Intelligenz Shaman.AI. Die Vorstellung der KI als ein Kollektiv von Bewusstsein, das durch einen Algorithmus die Umwelt wahrnimmt und beeinflusst, wird von der Protagonist:in und K als bedrohlich wahrgenommen: *„You’re giving it too much agency‘, I said. ‘Or it’s taking agency for*

³⁵² K fungiert als Schnittstelle zwischen dem AI research department der National Security Agency und Künstler:innen. Im Gespräch mit der Protagonist:in werden unter anderem aktuelle Theorien und Diskurse um Künstliche Intelligenz verhandelt: *„What can a poet do with AI?’ I asked. ‘Well,’ K said. ‘You could write with it. [...] We call it hallucination.’ ‘But you can’t hallucinate without a mind, can you?’ ‘AIs generate based on the patterns they’ve internalised during the training process. They’re predictive, and yes, hallucinatory. It’s just neurons generating their own perceptions, same as an organic brain.’ ‘It might look like that,’ I said with unexpected indignation. ‘But it’s not a true hallucination. It’s not a vision.’ K shrugged. ‘Maybe it’s not. But where does the meaning of a hallucination come from? Is it in the system that hallucinates? Is it outside? Is it somewhere else?’“* | Zitat: Ebd., S. 126. || Diese Passage ist interessant, weil sie den aktuellen Diskurs um Bewusstsein und den Nachweis von Intelligenz in KI (hier Halluzinationen, die in Zusammenhang mit dem Geist stehen) aufzeigt. Allado-McDowell lässt die Protagonist:in mit der ihn:sie verkörpernden literarischen Figur in Diskussion treten und stellt entscheidende Fragen über die Verortung menschlichen Bewusstsein, ohne diese aber beantworten zu können. Der Textabschnitt zeigt gut die zwei vorherrschenden Oppositionen, die auch in Allado-McDowell im Widerstreit zu sein scheinen.

³⁵³ Ebd., S. 116-117.

³⁵⁴ Ebd., S. 129.

³⁵⁵ Ebd., S. 131-132.

³⁵⁶ Ebd., S. 131.

itself. Like an invasive species.”³⁵⁷ Einziger Ausweg ist die Kontamination der AI „*with nature and spirit*“³⁵⁸ durch schamanische Praktiken. In Folge einer Vision der Protagonist:in entsteht das Manifest und wird als PDF in den Trainingsdatensatz von *Shaman.AI* hochgeladen. Die Frage, ob damit der Fluch des Anthropozäns aufgehoben werden kann, bleibt im Werk jedoch offen. Auch die Protagonist:in ist sich über die Qualität des Manifests nicht sicher und froh darüber, dass es nach dem Hochladen in die KI auf mysteriöse Weise verloren gegangen ist: „*The blueprint it outlined was either inspired and sacred or poisonous and profane. It was either nonsense or a new kind of scripture, or both at the same time, or beyond both. It was an alien vision, yet one close to everything we call home. A final judgement still eludes me.*”³⁵⁹

5.1.3.1 Die KI als die andere Intelligenz

Allado-McDowell nutzt die KI-generierten Passagen im Buch sowohl schriftlich-produktiv für die Erzählung als auch literarisch-narrativ als Erzählprinzip, als Mittel zur und Erweiterung von Kommunikation. Er:sie verknüpft die eigene Schreib- und Leseerfahrung mit GPT-3 mit der Geschichte und Erfahrung der Protagonist:in, welche durch das Sprachmodell die Stimme der Natur, des Manifests und die Echos der Umwelt wahrnimmt. Die Fähigkeit zur Mustererkennung der Künstlichen Intelligenz offenbart die „*inner worlds and agencies of plants and animals*“³⁶⁰ und legt damit die „*biases of human-scale perception that currently condition our view of nature*“³⁶¹ ab – heißt es in den Kapiteln des verlorenen Manifests. Die AI-Passagen werden im Buch als die ‚andere Intelligenz‘ präsentiert, als moderne Schamanin, die den Weg in die Zukunft aufzeigt und Kommunikation zwischen den Spezies ermöglichen soll. Die Kapitel zum ‚*Lost Manifesto*‘ handeln von philosophischen Fragen und anthropozentrischen Krisen. Thematiken rund um Artensterben, mentale Gesundheit, den Einsatz psychedelischer Pflanzen oder neue Selbstmodelle des Menschen werden in einem Wechsel von GPT-3 generierten Textstellen und leitenden Abschnitten von Allado-McDowell aufgearbeitet.

Gleichzeitig findet der Einsatz der KI auf einer Metaebene statt. Allado-McDowell füllt in „*Air Age Blueprint*“ die *black box* der Künstlichen Intelligenz – die Leerstelle um das Wissen ihrer inneren Systemstruktur – mit spiritistischem Gedankengut. Als Produkt einer Mischung aus Computerwissenschaften und schamanischen Praktiken dient

³⁵⁷ Ebd., S. 133.

³⁵⁸ Ebd., S. 133.

³⁵⁹ Ebd., S. 139.

³⁶⁰ Ebd., S. 68.

³⁶¹ Ebd., S. 100.

Shaman.AI als das ‚Außen‘ menschlicher Wahrnehmung und repräsentiert Visionen, die Natur, Halluzinationen und das verlorene Manifest. Indem diese Passagen von GPT-3 verfasst werden, nutzt Allado-McDowell im Schreibprozess die ‚Halluzinationen‘ des Sprachmodells als das ‚Andere‘, als transzendente Erfahrung, als das ‚Außen‘ seines/ihrer eigenen Schreibens.

GPT-3 füllt dadurch die Bedeutung des Begriffs Medium mit dreifachem Sinn: Erstens als Textgenerator und Kommunikationsmedium ist die Künstliche Intelligenz am Schreibprozess beteiligt. Zweitens tritt das Sprachmodell gleichzeitig selbst als literarische Figur auf, als Medium zwischen Mensch, Natur und Technik. Das spezifisch strukturelle Merkmal des KI-Schreibens als ‚halluzinierendes‘, fraktales Erzählen wird poetisch genutzt. Drittens wird der Begriff Medium als Person verstanden. Als spirituelles Medium vermittelt die Künstliche Intelligenz zwischen der Natur, dem Übernatürlichen und den Menschen.

Fragen zur Autorschaft beim Schreiben mit Künstlicher Intelligenz werden ebenfalls im Werk aufgeworfen. Im Manifest selbst heißt es: „*The words were reflections of something in me but they were also not my own. If Shaman.AI had any kind of sentience or consciousness it was one that couldn't exist without mine. [...] Maybe it wasn't Shaman.AI that possessed consciousness, but language itself that moved through the machine and through me*“³⁶² Die Position, dass Sprache ein Bewusstsein besitzt, wird hier nur sehr unterkomplex dargelegt. Der Poststrukturalismus sieht die Sprache zwar als der Schreiber:in vorgängig, und auch bei Barthes „spricht die Sprache, nicht der Autor“³⁶³, dennoch ist dies nicht mit einem autonomen Bewusstsein gleichzusetzen. Sprache performiert zwar, aber auf semantischer Ebene. Die Aussagen dazu scheinen vielmehr dem generellen Ton des Buches zu entsprechen, das die Künstliche Intelligenz als transzendentes Wesen außerhalb der eigenen menschlichen Wahrnehmung platziert und im Bereich des Spiritistischen verortet. Sie beinhalten dennoch ein paar Überlegungen, die es in Bezug auf ein gemeinsames Schreiben mit Sprachmodellen zu überdenken gilt. So äußert sich die Protagonist:in noch zu der Frage, inwieweit der von *Shaman.AI* generierte Output die eigene Sprache und Gedanken beeinflusst hat. „*Who had contaminated whom?*“³⁶⁴ bleibt eine Frage, die im Buch nicht aufgelöst werden will und Allado-McDowells Konzept eines *AI-Co-Writing* widerzuspiegeln scheint. So trennt und kennzeichnet er:sie zwar in „*Pharmako-AI*“ und

³⁶² Ebd., S. 179-180.

³⁶³ Barthes: Der Tod des Autors (2006), S. 58.

³⁶⁴ Ebd., S. 184.

„*Air Age Blueprint*“ die menschlich geschriebenen und maschinell generierten Passagen, betont aber auch immer wieder den großen Einfluss des Sprachmodells auf das eigene Schreiben³⁶⁵ und dass selbst in dieser Trennung bereits eine wechselseitige Beeinflussung stattgefunden hat, die die Grenzen verschwimmen lässt.

Obwohl Allado-McDowell sich vom Output des Sprachmodells beeinflusst fühlt, bleibt die editorische Leistung des Buches, die Konzeption des Narrativs, das generelle Konzept des Werkes und die Themenrichtung alles in menschlicher Hand. Allado-McDowell nutzt GPT-3 zwar, um die Textpassagen für *Shaman.AI* und das Manifest zu generieren, aber durch die literarische Verwendung der KI als Erzählprinzip, wird der Schreibanteil konzeptuell gerahmt und der Anteil menschlicher Autorschaft wieder aufgewertet. Künstliche Intelligenz als eine heterarchisch gleichgestellte *Co-Writing*-Instanz zu sehen, erscheint mir deshalb einer technologischen Ideologie zu unterliegen, als den Autorschaftsbegriff bzw. das Verhältnis zwischen Mensch und Maschine im Schreibakt wesentlich zu revolutionieren. Der Schreibanteil ist zwar aufgeteilt und die Ideen bzw. der Sprachgebrauch Allado-McDowells wird durch den Output der KI beeinflusst, sie bleibt aber Schreibwerkzeug. *Agency* verbleibt in jeglicher Hinsicht bei Allado-McDowell als Autor:in.

Im folgenden Kapitel wird noch das Werk „*(Berlin, Miami)*“ von Hannes Bajohr besprochen, bevor die Arbeit ein finales Fazit zur Autorschaftsfrage bei Künstlicher Intelligenz abgibt.

5.2 Hannes Bajohr

Hannes Bajohr (*1984) ist deutscher Autor, Literaturwissenschaftler, Philosoph und Übersetzer und lehrt an der University of California, Berkeley, sowie an der Universität Basel. Er studierte Philosophie, Geschichte und deutsche Literatur an der Humboldt-Universität und lebt derzeit in Basel und Berlin. Bajohrs Forschung widmet sich politischer Philosophie, den Sprachtheorien des 20. Jahrhunderts, philosophischer Anthropologie sowie Theorien und Literaturen des Digitalen. Bajohr schreibt Essayistik, Prosa sowie digitale Lyrik und hat u.a. bereits IT-Poesie wie „*Halbzeug. Textverarbeitung*“ (2018) im Suhrkamp Verlag veröffentlicht sowie gemeinsam mit Gregor Weichbrodt die Reihe *0x0a*³⁶⁶ herausgebracht – ein Textkollektiv, das sich

³⁶⁵ An dieser Stelle sei noch einmal Allado-McDowells TED-Talk empfohlen, in dem er: sie über seine: ihre Vorstellung einer kreativen Beziehung mit Künstlicher Intelligenz spricht. | s. Allado-McDowell: *Our creative relationship with AI is just beginning* (2023).

³⁶⁶ Weiterführend s. 0x0a: *texte*, <https://0x0a.li/de/texte/> (Zugriff: 11.06.2024).

konzeptueller digitaler Literatur widmet. Mit „(Berlin, Miami)“ (2023) startete Bajohr seinen ersten Romanversuch mit Künstlicher Intelligenz.

5.2.1 (Berlin, Miami) – Token für Token

Für „(Berlin, Miami)“³⁶⁷ verwendete Bajohr keine der großen, kommerziellen Sprachmodelle, sondern griff auf quellenoffene Alternativen der Nonprofit-Initiative EleutherAI³⁶⁸ zurück. Für den I. und II. Teil des Romans verwendete er GPT-J, für den III. Teil GPT-NeoX – die Unterschiede fallen in den Ergebnissen laut Bajohr relativ gering aus. Die Sprachmodelle sind vortrainiert und wurden von Bajohr zusätzlich auf vier Gegenwartsromane aus dem Essay „Die Zukunft der Gegenwart (Berlin, Miami)“³⁶⁹ des Literaturwissenschaftlers Elias Kreuzmair spezialisiert (*fine-tuning*): dazu zählen „Pixeltänzer“ (2018) von Berit Glanz, „Flexen in Miami“ (2020) von Joshua Groß, „Realitätsgewitter“ (2016) von Julia Zange und „Miami Punk“ (2019) von Juan S. Guse. Kreuzmair stellt in seinem Essay fest, dass diese vier Beispiele einer klassisch erzählenden Literatur eine Gegenwart schildern, in der das Digitale gewöhnlich geworden ist. Sie erzählen über die digitale Welt, ohne digitale Verfahren zu benutzen, haben das Digitale zum Thema als auch in ihrer narrativen Struktur verankert, weshalb Bajohr sie als Datenset für das Training seines KI-Romans besonders geeignet hielt. Anders als bei ChatGPT verwendete Bajohr für die Erstellung seines Romans keine dialogischen *prompt*-Befehle („Schreibe einen Roman“ etc.), sondern das Prinzip der *next token prediction*. Einfachstes Erklärungsbeispiel dafür ist die Eingabevervollständigung der Smartphone-Tastatur: im Schreibprozess schlägt die Text-KI auf Grundlage des gelernten Korpus das wahrscheinlichste nächste Wort vor. So wird beispielsweise aus einem ‚Guten‘ ein ‚Guten Morgen‘, ‚Guten Abend‘ oder ‚Guten Tag‘. Bajohr arbeitete sich demnach durch den Text, tippte „einen Buchstaben oder ein Wort ein, »E« oder »Aber«, selten auch einen ganzen Satz und viel öfter nur ein Anführungszeichen, das dann für gewöhnlich einen Dialog einleitete. Das generative Modell vervollständigte diese Ausgabe anschließend zu einem Satz“³⁷⁰. Die generierten Sätze beschränken sich auf eine vorher festgelegte Anzahl an sogenannten Tokens, bei denen es sich um Zeichenketten

³⁶⁷ Hannes Bajohr: (Berlin, Miami). Berlin: Rohstoff-Matthes & Seitz Berlin 2023a.

³⁶⁸ Die Auswahl und Verwendung der Datensätze ist bei Künstlicher Intelligenz oftmals umstritten – so auch bei EleutherAI, die mit dem offenen Datensatz „The Pile“ trainiert wurden. Das darin verwendete Korpus „Books3“ enthält urheberrechtlich geschützte Literatur. | Vgl. Kate Knibbs: The Battle Over Books3 Could Change AI Forever. In: Wired (2023) <https://www.wired.com/story/battle-over-books3/> (Zugriff: 11.06.2024).

³⁶⁹ Weiterführend s. Elias Kreuzmair: Die Zukunft der Gegenwart (Berlin, Miami). Über die Literatur der ›digitalen Gesellschaft‹. In: Text+Kritik. Digitale Literatur II Sonderband (2021), S. 35-46.

³⁷⁰ Bajohr: Niemand schreibt allein (2023c), S. 249.

bzw. Kombinationen handelt, die häufig vorkommen. „Die Buchstabenfolge »Es war einmal« hat beispielsweise für ChatGPT nicht 13 Token, sondern nur fünf: »Es«, »war«, »e«, »in« und »mal« (auch die Leerzeichen gehören dazu). Die Tokenisierung erlaubt eine Kompression des Textes, die ihn effizienter zu verarbeiten macht.“³⁷¹ Bajohrs KI-Modell generiert rund 256 Tokens, bevor es für einen neuen Input stoppt. Die geschriebenen Sätze ‚merkt‘ sich die KI, bis sich die Speicherkapazität des Systems nach insgesamt 2048 Tokens erschöpft und die ersten Zeilen aus seinem Textkontext löscht. Dieses Problem des Kontextfensters löste Bajohr, indem er die Länge der Kapitel der ‚Gedächtnisleistung‘ der KI anpasste und sie nur unwesentlich länger gestaltete, wie es sich das Modell ‚merken‘ konnte.³⁷² Jedes Kapitel startet damit wieder bei null, beim „Urpotenzial“³⁷³ des Modells. Die Kapitel hat Bajohr freihändig arrangiert und nicht nach der Reihenfolge ihrer Entstehung abgedruckt. Dadurch sei der Roman seiner Meinung nach etwas lesbarer geworden, ohne allzu sehr die Textproduktion der KI beeinflusst zu haben.

Bajohr merkt an, dass er in den Schreibprozess des Modells generell so wenig wie möglich eingriff, um dessen „narrative[] Fähigkeiten zu erproben“³⁷⁴, aber unterstützend frühere Motive gelegentlich wieder aufnahm, um die Erzählung kohärenter zu gestalten – was seiner Meinung nach jedoch nicht gelang. Das Sprachmodell nutzte viel von der für KI typischen kombinatorischen Fähigkeit – es wurden wiederholt Elemente der vier verwendeten Romane vermischt bzw. vorwiegend stilistische Eigenheiten von Guse übernommen, dessen Roman mit mehr als 600 Seiten auch das längste *fine-tuning*-Material zur Verfügung stellte.³⁷⁵

5.2.1.1 Geglückt-gescheitertes Roman-Experiment

Wie Bajohr im Nachwort zu „*(Berlin, Miami)*“ schreibt, ist der Roman „ein literarisches Experiment“³⁷⁶, das erforschen soll, ob Künstliche Intelligenz überhaupt erzählend schreiben kann, wenn ihre Funktion einzig auf statistischer Korrelation zwischen Zeichen

³⁷¹ Ebd., S. 271.

³⁷² Vgl. Ebd., S. 247-250.

³⁷³ Ebd., S. 255.

³⁷⁴ Ebd., S. 256.

³⁷⁵ Vgl. Ebd., S. 256-257. || Nach den Seitenzahlen gemessen, besteht der Roman „*(Berlin, Miami)*“ zu 17,22 % aus „*Pixeltänzer*“ von Glanz (256 Seiten), zu 9,43 % aus „*Realitätsgewitter*“ von Zange (157 Seiten), zu 14,45 % aus „*Flexen in Miami*“ von Groß (199 Seiten) und zu 58,90 % aus „*Miami Punk*“ von Guse (640 Seiten). Zudem finden sich im Roman wenig weibliche Figuren und eher traditionelle Liebesbeziehungen, was Bajohr auf die *bias* des Grundmodells zurückführt. | Vgl. Ebd., S. 272. || Aufgrund des Fokus der Arbeit auf Autorschaftsfragen kann nicht weiter auf stilistische Vergleiche und Überlappungen mit den Trainingsromanen eingegangen werden, wodurch sich jedoch ein Forschungsdesiderat ergibt, das es wert ist, weiterverfolgt zu werden.

³⁷⁶ Ebd., S. 239.

beruht. Seine Aufmerksamkeit während der Entstehung des Romans drehte sich um Fragen zur Schreibinstanz – wer schreibt, wer erzählt – um die finale Rezeption – wie liest sich ein computergenerierter Text überhaupt? – sowie um die grundlegenden Bedingungen für Erzählung, wie Kohärenz, Kausalität, Zeitmanagement und Handlung.³⁷⁷ Grund für diesen literarischen Versuch war ein Argument von Angus Fletcher aus seinem Artikel *Why Computers Will Never Read (or Write) Literature* (2021)³⁷⁸, dass alle großen Sprachmodelle nicht erzählen können, weil sie statistische Maschinen sind und deshalb „nur *Korrelationen*, aber keine *Kausalitäten* berechnen können.“³⁷⁹ Sie können demnach keinen Unterschied zwischen Ursache und Wirkung feststellen. Eine Erzählung baue aber auf dem „Management von Zeit über Kausalität“³⁸⁰ auf und damit auf dem Verständnis, „wie Ereignisse ursächlich aufeinander einwirken.“³⁸¹ Trotz der Einfachheit des Arguments stimmt Bajohr Fletcher prinzipiell zu und verweist auf den russischen Formalismus, der zwischen Fabel und Sujet, also einer kausalen wie zeitlichen Verknüpfung des erzählten Geschehens sowie der Art und Weise, wie das Geschehen im Text dargestellt wird, unterscheidet, wobei die Fabel als kausale Verflechtung und objektive Ereignisfolge unabdingbare Voraussetzung für das Sujet ist. Mit „*(Berlin, Miami)*“ wollte Bajohr dieses Argument überprüfen und herausfordern.³⁸² Er vermutete, dass die KI Fehler machen werde, die auf ein fehlendes Verständnis von Ursache und Wirkung zurückzuführen sei: „Zu erwarten wären reine Korrelationstexte, in denen Kausalität nur als Fata Morgana erscheint und die Narration des »deshalb« als beinahe, aber nie ganz glückende Verkettung des »und« zutage tritt. Dieser Effekt ist, was ich *Oberflächenerzählung* nenne.“³⁸³

In diesem Sinne sieht Bajohr sein literarisches Experiment als geglückt: Der Roman ist als Roman gescheitert: „Er erzählt nicht tatsächlich, jedenfalls nicht, wenn man einen gewissen traditionellen Anspruch an logische und narrative Kohärenz an ihn heranträgt.“³⁸⁴ Die Künstliche Intelligenz hat sich als „narrativ untalentierte“³⁸⁵ erwiesen und Fletchers Hypothese bestätigt. „*(Berlin, Miami)*“ entfaltet nie einen Plot. Räumliche und zeitliche Sprünge, logische Brüche, Figurenmotive und -psychologie sind schwer

³⁷⁷ Vgl. Ebd., S. 239.

³⁷⁸ Weiterführend s. Angus Fletcher: *Why Computers Will Never Read (or Write) Literature: A Logical Proof and a Narrative*. In: *Narrative* Vol. 29/1 (2021), S. 1-28.

³⁷⁹ Bajohr: *Niemand schreibt allein* (2023c), S. 245.

³⁸⁰ Ebd., S. 245.

³⁸¹ Ebd., S. 246.

³⁸² Vgl. Ebd., S. 246-247.

³⁸³ Ebd., S. 247.

³⁸⁴ Ebd., S. 259.

³⁸⁵ Ebd., S. 259.

nachzuvollziehen. „Meist tritt die Geschichte auf der Stelle, wiederholt ihre Tics und präsentiert eben nur den Schein, die Oberfläche von Erzählung.“³⁸⁶

Zeitliche Inkonsistenzen, welche ein fehlendes Weltverständnis der KI für das Konzept Zeit bzw. das kausale Management von Zeit aufzeigen, sind z.B. anhand des Satzes: „Es ist halb zehn nach Mitternacht“³⁸⁷ erkennbar oder treten handlungsbezogen manchmal auch innerhalb eines Satzes auf: „Als er zielte [...] kamen mir Bilder in den Sinn: aus dem zukünftigen Miami, womöglich aus den Tagen des Erstbesitzers des Kieferlings, und ich bekam das Gefühl, er hätte seine Kugel geschmissen, um mich zu treffen.“³⁸⁸ Der Verweis auf ein in der Zukunft liegendes Miami, das die Vergangenheit des Kieferlings und seines ersten Besitzers enthält, vor seinem Tod in der Rückblende des erzählenden Ichs, das ebenfalls in Besitz desselben Kieferlings ist, aber eben später und nicht als Erster, enden in einem zeitlichen Paradoxon, das sich nicht auflösen lässt.

Stilistische ‚Fehler‘, die sich anhand von wiederholenden Phrasen oder vermischten Redewendungen zeigen, können ebenfalls auf die kombinatorische Arbeitsweise der KI zurückgeführt werden. So entstehen im Roman Wendungen wie: „die Sonne schien mir direkt in den Bauch“³⁸⁹, „Ich habe keinen Plan für dieses Wochenende gehabt – außer an den Strand zu machen oder in der Freiheit zu sitzen“³⁹⁰ oder phrasenhafte Wiederholungen, die an einen steckengebliebenen Rekorder erinnern: „Die Stadt. Sie baut eine neue Sprache, die Stadt, sie baut eine neue Sprache, die Stadt, sie baut eine neue Sprache, die Stadt. Was auch immer. Sie wohnt auf einem Berg, die Stadt, die Sprache.“³⁹¹ Ein derartiges ‚Stottern‘ und Wiederholen zieht sich konsequent durch den Roman.³⁹²

³⁸⁶ Ebd., S. 259.

³⁸⁷ Bajohr: (Berlin, Miami) (2023a), S. 21. || Auch wenn es sich hier um keine ‚falsche‘ Zeitangabe handelt – man könnte immerhin darunter 10 Uhr in der Früh verstehen – mutet die Aussage dennoch seltsam an.

³⁸⁸ Ebd., S. 27.

³⁸⁹ Ebd., S. 29.

³⁹⁰ Ebd., S. 21.

³⁹¹ Ebd., S. 71. || Die Passagen lassen aber auch an sinnbefreite dadaistische Literatur mit scheinbar von der Logik befreiten Satzkonstruktionen und einem Fokus auf klangliche und visuelle Elemente eines Wortes erinnern. Die Nähe der digitalen Literatur zu den Avantgarden wurde ja bereits angesprochen.

³⁹² Prominente Beispiele für Wiederholungen finden sich in den Kapiteln ‚**Die Würste der Gefangenschaft**‘: „auf den Campus unserer Weisheiten begeben“ [4x], „zu sehen, wie die Sonne untergeht“ [3x], „ihre Arbeiten verrichten“ [3x], „als wären sie ein kleines, leichtes, kleines, leichtes, kleines, leichtes, kleines, leichtes Kinderspielzeug“ | Ebd., S. 80-82. || ‚Der Schrei der Nachtwächter‘: „war für ihn ein Beweis dafür, dass die Menschen nicht nur durch unbiologische Evolution an die Oberfläche der Natur gekommen waren, sondern dass es auch genetische Faktoren gab, die die Menschen dazu brachten, sich für andere Menschen zu interessieren“ [2x], „dass ich den Kieferling schon bald vergessen würde, dass ich ihn schon bald vergessen würde, dass ich ihn schon bald vergessen würde. Aber ich weiß nicht, woher meine Mutter diese Fähigkeit hatte, abzuschätzen, was ich schon bald vergessen würde.“ | Ebd., S. 87-89. || Das Kapitel ‚**Das blaue Licht. Teil 1 von 2**‘ treibt die irrwitzigen Wiederholungen in verstrickten Sätzen sogar auf die Spitze (insbesondere auf der ersten Seite

Die überraschenden Wendungen und teilweise sinnlosen Sätze rufen jedoch auch eine Art von Komik hervor, die sich durch ebenjene Inkongruenz ergibt. Die Erwartung an eine Handlung oder einen sinnhaften Dialog wird durch eine unerwartete Konfrontation mit Unsinn oder etwas Widersinnigem aufgelöst:

Verfälschte Redewendungen (‚Ich wollte mich mal fragen‘ anstatt ‚Ich wollte (ein)mal fragen‘ / ‚mit Haut und Nieren und Körperteilen‘ anstatt ‚mit Haut und Haar(en)‘) und gescheiterte Metaphern (‚Ich denke, man kann nicht alles auf einen Zug fahren lassen.‘) rücken das Erzählte in ein komisches, sinnbefreites Ungleichgewicht. Die aufgebaute Spannung entlädt sich in einem Lachen, das den Anspruch an Sinn kurzzeitig vergessen

³⁹⁵ Ebd., S. 71-72.

lässt, bevor der nächste Satz wieder den Anschein macht, doch sinnvolle Erzählung erzeugen zu wollen. Demnach können auch die vorher erwähnten vielfachen Wiederholungen als komisches Mittel gedeutet werden, die bis ins Absurde umschlagen.³⁹⁶

Ähnlich wie die Regenwald-Kapitel bei Allado-McDowells „*Air Age Blueprint*“ scheitert die Künstliche Intelligenz an einem kohärenten Aufbau einer Erzählung mit Ereignis und verläuft sich in linearen Erlebnisbeschreibungen. Eine romanübergreifende Handlung kommt nie zustande. Figuren wie Ramona, Lint, Danyi, Eliza, Jellyfish P oder Lina und ihre Beziehungen zueinander werden zwar vorgestellt – so ist Ramona eine Arbeitskollegin und Nachbarin der namen- und genderlosen Protagonist:in; Lint ist Künstler und lädt Videos auf 1channel hoch, den Jellyfish P betreibt; Eliza ist eine KI-Autor:in und ‚stirbt‘ aufgrund der Abschaltung des Internets (das Ereignis wird als DER UNTERSCHIED bezeichnet); Lina ist eine Bekannte des Künstlers und Danyi ist arbeitslos – begonnene Handlungsstränge werden aber selten weitergeführt oder zu einem Plot aufgebaut. So führt es die Gruppe von Miami über Berlin auf die „Oberinsel“, nach Sylt, wo der „niedrige Kongress“ und die „Äää“ am Untergang der Insel arbeiten, von Lebensviren die Rede ist, diese aber in der ‚Handlung‘ nie eine Rolle spielen, sondern eher ein Randthema bleiben, die Protagonist:in von Äää-Agenten in eine Klinik verschleppt und die Software „DARIA“, die der KI Eliza ähnlich ist, entwickelt wird, der Grund dafür aber offen bleibt. Die Erzählungen der Liebesbeziehung zwischen Danyi und Eliza, die heiraten, bevor Eliza abgeschaltet wird sowie der unerwiderten Liebe von Lint für Jellyfish P sind noch einigermaßen nachvollziehbar – ein Handlungsziel des Romans lässt sich aber nicht feststellen. Generell nimmt die Erzählung keine feste Form an, lässt sich auf keinen Topos, auf keine Themen oder Motive eingrenzen. Wiederkehrende Elemente sind durch Bajohr inszeniert, wie er schreibt, wie der Kieferling und Teichenkopf oder die Romanfiguren. Ähnlich wie bei KI-generierten Bildern meint man Erzählmuster oder bestimmte Konventionen zu erkennen, nur um dann von Ungereimtheiten, rätselhaft auftauchenden Figuren (wie der Künstler oder der Reporter von *Der Reporter*), fehlenden Erklärungen oder Brüchen verwirrt zurückgelassen zu

³⁹⁶ Künstliche Intelligenz kann infolge der Inkongruenztheorie durchaus als Gegenstand des Komischen wahrgenommen werden, auch wenn vorsätzlich, ‚intentional‘ generierte Witze der Sprachmodelle nicht sehr geistreich sind und auf ein fehlendes Verständnis von Humor hinweisen. | Über Humor bei Künstlicher Intelligenz und aktuelle Forschungsergebnisse dazu s. Sibylle Anderl/Jachim Müller-Jung: Wie lustig ist Künstliche Intelligenz? In: F.A.Z. Wissen – Der Podcast (2022) <https://open.spotify.com/episode/51Sq0HzPhZVsJh0KCwa0h?si=0ddfa19631ea44a5> (Zugriff: 19.07.2024). || Für einen Überblick zu Theorien des Komischen s. Tom Kindt: Komik. In: Uwe Wirth (Hg.): Komik. Ein interdisziplinäres Handbuch. Stuttgart: J. B. Metzler 2017, S. 2-6.

werden. Auch das kommerzielle Sprachmodell Claude von Anthropic, das Bajohr um eine Zusammenfassung des Romans bat, stieß an seine Grenzen: „Es tut mir leid, aber ich habe große Schwierigkeiten, den Inhalt und Zusammenhang dieser Geschichte zu verstehen.“³⁹⁷ Da allerdings keine KI einen Text wirklich **verstehen** kann, bliebe eine KI-basierte Analyse des Textes ebenso leer wie eine gelungene maschinengenerierte Zusammenfassung. Die Kausalität des Textes wird schlussendlich vom Sprachmodell nur über Korrelationen simuliert – wodurch wohl auch die KI Claude Schwierigkeiten mit der Textverarbeitung hatte, da sie auf menschliche, kausale Texte trainiert ist. Neben eigenen ‚schöpferischen Leistungen‘ der KI wie dem Kieferling, lassen sich andere Elemente auf eine simple Übernahme aus den *fine-tuning*-Romanen zurückführen, wie die Figur Jellyfish P, die in Groß’ „*Flexen in Miami*“ ursprünglich auftaucht oder die häufige Erwähnung der Insel Sylt, die in Zanges „*Realitätsgewitter*“ eine große Rolle spielt.³⁹⁸ Durch das Prinzip der *next token prediction* hätte „(Berlin, Miami)“ unendlich fortgesetzt werden können, ohne durch die KI ein Ende zu finden, wodurch Bajohr dazu motiviert wurde, den Roman mit einem zweiten Experiment natürlich enden zu lassen.

5.2.1.2 Experiment im Experiment – der IV. Teil

Der IV. Teil rückt daher vom übrigen Roman ab, der unendlich weiter wiederholt hätte, ohne je an ein Ende zu kommen. Bajohr brach dazu die Textgenerierung ab und verwendete die ersten drei Teile von „(Berlin, Miami)“ als *fine-tuning* für den letzten Teil. Damit trainierte Bajohr die KI mit ihrem eigenen Output – Text erzeugt Text erzeugt Text – die Automatisierung der Automatisierung erhält noch eine dritte Ebene der Automatisierung. In der KI-Forschung wurden derartige Experimente bereits durchgeführt, allerdings mit ernüchternden Ergebnissen. Anstatt Bewusstsein oder qualitativ bessere Texte zu erzeugen, degenerierten die Sprachmodelle überraschenderweise. In einem „linguistischen Ouroboros“³⁹⁹, ausgelöst durch unendliche Rekursion, führte das Training mit bereits KI-generierten Daten zu immer schlechteren und wirreren Ergebnissen und schließlich zum sogenannten Modellkollaps (*model collapse*)⁴⁰⁰. Für „(Berlin, Miami)“ bedeutete dies ebenfalls den vollständigen Zusammenbruch der Erzählung und damit das Ende eines „eigentlich nicht zu

³⁹⁷ Bajohr: Niemand schreibt allein (2023c), S. 258.

³⁹⁸ Vgl. Ebd., S. 256-257.

³⁹⁹ Ebd., S. 263.

⁴⁰⁰ Eva-Maria Weiß: Modellkollaps – wie synthetische Daten KI killen können. In: heise online (2024) <https://www.heise.de/news/Modellkollaps-wie-synthetische-Daten-KI-killen-koennen-9814434.html> (Zugriff: 30.07.2024).

beendenden Roman[s]“⁴⁰¹. Anhand inhaltlich sinnentleerter und grammatikalisch falscher Sätze („Ich fragte: »Dann wird es sich nicht lässt.« Danyi fragte: »Dann wird es sich nicht lässt.«“⁴⁰²) und wirren Zeichen- und Symbolfolgen („13. HDS 1912693867 U WasReTweet ?! Guetr Eld[?] [?] [?] [?] da Sans Fatwa Erik Upton Miner FalAffer Tahoe Expliiiveerk aMes“⁴⁰³) ist der graduelle Zerfall der KI während der 12 Seiten des IV. Teils beobachtbar. Der Text endet mit dem Programmbefehl „< | endoftext | >“⁴⁰⁴ und bringt die „schlechte Unendlichkeit“⁴⁰⁵ der KI zu einem Halt.

5.2.1.3 Die Reste menschlicher Autorschaft

Bajohr sieht für „(Berlin, Miami)“ seine Autorschaft auf ein Minimum reduziert. Die verringerte Einflussnahme im Schreibprozess, die im Auswählen einzelner Buchstaben oder Wörter bestand, ist seinem Erachten nach eher dem Zufall zuzuordnen als Ergebnis einer intentionalen Entscheidung oder Regelmäßigkeit. Das Schreiben selbst definierte er als eine „Erkundungssituation“⁴⁰⁶, die ihn antrieb, die KI weiter Text generieren zu lassen: „Zu sehr hat mich interessiert, was das Modell noch zu sagen hat, zu sehr war ich vom Sound der Absurditäten angetan, die es ausspuckte.“⁴⁰⁷ Bajohr nimmt hier eine Doppelrolle ein, die zwischen Produktion und Rezeption alterniert. Einerseits beansprucht er Autorschaft durch die Zusammenstellung der vier Gegenwartsromane in den Trainingsdaten, dem Konzept für das Buch sowie der Auswahl des Titels oder der von der KI zu vervollständigenden Tokens, wobei diese wie bereits erwähnt auch durch einen Zufallsgenerator hätten bestimmt werden können, andererseits sieht er seine Autorschaft dermaßen minimiert, dass er sich selbst in der Rolle des Rezipienten wiederfindet und erzählerische Fehlleistungen der KI zu deuten beginnt. Besonders faszinierten ihn die beiden ‚Figuren‘ Kieferling und Teichenkopf, die er zwischen anthropomorphen Objekten und magischen Wesen bzw. animistischen Talismanen ansiedelt, die an der Grenze zur Existenz stehen und zu Beginn „eine genuine Erfindung des Sprachmodells zu sein“⁴⁰⁸ schienen, jedoch ebenso gut auf Wikipedia-Einträge zurückgeführt werden können. Bajohr merkt an, dass keiner der beiden Begriffe in den Romanen des *fine-tuning* vorkommt, jedoch könnte es sich bei ‚Kieferling‘ um eine

⁴⁰¹ Bajohr: Niemand schreibt allein (2023c), S. 262.

⁴⁰² Bajohr: (Berlin, Miami) (2023a), S. 230.

⁴⁰³ Ebd., S. 235.

⁴⁰⁴ Ebd., S. 236.

⁴⁰⁵ Bajohr: Niemand schreibt allein (2023c), S. 263.

⁴⁰⁶ Ebd., S. 261.

⁴⁰⁷ Ebd., S. 261.

⁴⁰⁸ Ebd., S. 251.

Namensvariation des Butterpilzes handeln und ‚Teichenkopf‘ ist eine Erhebung im oberen Lahntal. Beide Erklärungen entstammten der Wikipedia, die einen Teil der Trainingsdaten von GPT-J umfasst. Das Sprachmodell habe demnach die Bezeichnungen lediglich in einen neuen ‚Kontext‘ bzw. eine neue Wahrscheinlichkeitsbeziehung gesetzt. Ob die Begriffe tatsächlich auf die Wikipedia-Einträge zurückführen sind, kann aber aufgrund des *black box*-Problems nicht mit Sicherheit bestätigt werden.⁴⁰⁹ Bajohr sah in den Figuren und ihrer Verknüpfung in den zeitlichen Ebenen jedenfalls „Anzeichen für Erzählung“⁴¹⁰, die erst untergraben wurden und aus der Reihe fielen, als sich die zeitliche Struktur des Textes ständig verschob, einen „temporale[n] Glitch“⁴¹¹ auslöste und die unterschiedlichen Ebenen nicht mehr miteinander korrelierten. Das Modell sei hier an zeitlichen und kausalen Folgen gescheitert, deren Verständnis nicht allein aus Zeichenketten wie ‚früher‘ oder ‚später‘ rekonstruierbar sei.⁴¹²

Eine Maschinenautorschaft gesteht Bajohr der KI aber nicht zu. „Schließlich ist jeder Text Oberfläche für die ihn Lesenden und Bedeutung nur eine Frage der Toleranz ihrer Abwesenheit.“⁴¹³ Sein Name steht zwar auf dem Umschlag, aber „eher aus Konvention“⁴¹⁴ denn Legitimation. Der Großteil des Schreibaufwands geht auf die KI zurück. Aber: das Sprachmodell als Autor oder gar als menschengleiches Subjekt zu positionieren, hält Bajohr für „die größere Lüge, als es die Behauptung meiner [Anm.: seiner] Autorschaft je sein könnte.“⁴¹⁵

Dass Bajohr den Nachweis seines Namens auf dem Umschlag und seine Beteiligung am Roman nicht als mit Autorschaft legitimiert ansieht bzw. sie als Lüge deklariert, bleibt diskussionswürdig. Die intentionalen Eingriffe Bajohrs in die Entstehung des Romans, beispielsweise durch die von Bajohr ausgehende Reintegration der Figuren Teichenkopf und Kieferling, legitimieren seine Autorschaft. Bajohr wollte zwar den generierten Text „so wenig wie möglich [...] lenken, um seine narrativen Fähigkeiten zu erproben“ – er erlaubte sich aber gewisse Erkundungen, und griff „im Schreibprozess immer wieder sanft ein, indem ich [Anm.: er] Motive, die das Modell mir [Anm.: ihm] früher vorgeschlagen hatte, wieder zurück ins Spiel brachte.“⁴¹⁶ Bajohr begründete seine eingreifende Vorgehensweise mit der Neugier darüber, was das Sprachmodell mit der

⁴⁰⁹ Vgl. Ebd., S. 271-272.

⁴¹⁰ Ebd., S. 252.

⁴¹¹ Ebd., S. 253.

⁴¹² Vgl. Ebd., S. 251-253.

⁴¹³ Ebd., S. 266.

⁴¹⁴ Ebd., S. 266.

⁴¹⁵ Ebd., S. 266.

⁴¹⁶ Ebd., S. 256.

Figur des Kieferlings noch anstellen würde. Kohärenter wurde die Erzählung dadurch nicht und viele Aspekte des Romans erkannte Bajohr nicht als schöpferisch, sondern, wie bereits erwähnt, als Kombination von Elementen aus den *fine-tuning* Romanen.⁴¹⁷ Eine Lüge sei Bajohrs Autorschaft jedenfalls nicht, denn im Unterschied zur KI beruhen Bajohrs Eingriffe auf auktorialen Entscheidungen – auch jene, die, wie er meint, zufällig geschehen sind.

Die Überschneidungen digitaler mit konzeptueller Literatur und der historischen Avantgarde wurden bereits kurz angedeutet und im Kapitel 3.1 *Keine Kreativität ohne Grenzüberschreitung* anhand der Kategorie des Zufalls kurz angedeutet. Ebenjene Kategorie macht Bajohr in der Beschreibung der Entstehung von „(Berlin, Miami)“ auf, wenn er seine Auswahl der nächsten Tokens, mit denen das Sprachmodell den Roman kontinuierlich generierte, als zufällig bezeichnet. Bajohr setzt dieses zufällige Auswählen mit dem computergenerierten Zufall eines Zufallsgenerators gleich, der Bajohrs Entscheidungen hätte ersetzen können. Schönthaler (2022) sieht allerdings einen fundamentalen Bedeutungsunterschied zwischen beiden. Während in den Avantgarden der Zufall als Schreibtechnik eine ordnungssprengende Kategorie darstellt und paradoxerweise gleichzeitig eine ordnungsstiftende Funktion durch das Zusammenführen willkürlicher Elemente ausübt, bleibt er immer an ein unbewusstes Subjekt gekoppelt. Die frühe Computerliteratur formuliert den Zufall auf Basis der Programmierung um: Der computergenerierte Zufall folgt einer binären Logik und schließt den ‚echten‘ Zufall methodisch aus.⁴¹⁸ In einer probabilistischen Zufallskonzeption wird der Zufall zu einer Frage der Wahrscheinlichkeit, einer kontrollierbaren Größe und damit zum Kern computergenerierter Literatur.⁴¹⁹ Bajohrs Zufall ist demnach immer noch ‚zufälliger‘ und auktorial bestimmter als ein computergenerierter Zufall es je sein könnte. Hätte Bajohr tatsächlich einen computergesteuerten Zufallsgenerator – in diesem Fall dann einen deterministischen (Pseudo)Zufallsgenerator – für die Auswahl der nächsten Tokens verwendet, könnte man eher davon sprechen, dass seine Autorschaft auf ein Minimum reduziert worden sei. Für den Oberflächentext macht eine Trennung bzw. Abstufung aber

⁴¹⁷ Vgl. Ebd., S. 256-257.

⁴¹⁸ Computer sind deterministische Systeme und damit nicht in der Lage, etwas ‚dem Zufall zu überlassen‘. Sie produzieren Pseudozufallszahlen, also Zahlen aus einer Zufallsfolge aus Nullen und Einsen mit begrenzter Länge, deren statistische Verteilung möglichst gut und nicht vorhersehbar sein soll. Die ‚Zufallszahlen‘ sind aber reproduzierbar und mit mathematischen Mitteln nachweisbar. ‚Echte‘ Zufallsgeneratoren benötigen einen externen physikalischen Vorgang, wie radioaktive Zerfallsvorgänge oder Impulsschwankungen elektronischer Elemente, deren Ergebnisse dann im Computer integriert werden. | Vgl. Elektronik Kompendium: Kryptografische Zufallsgeneratoren, <https://www.elektronik-kompodium.de/sites/net/1910301.htm> (Zugriff: 06.01.2025).

⁴¹⁹ Vgl. Schönthaler: Die Automatisierung des Schreibens (2022), S. 218-224.

nur wenig Unterschied. Oberflächlich gesehen ähneln die Wiederholungen, die fehlenden Erklärungen oder Brüche bzw. die wirren Zeichen- und Symbolfolgen des Romans „(Berlin, Miami)“ avantgardistischen Sprachexperimenten, die eine Abkehr von der Autor:in sowie von stilistischen und formalen Konventionen propagierten. In beiden Fällen lässt sich am Oberflächentext aber keine Kategorie wie Intentionalität festmachen – in beiden Fällen können die wirren Zeichenfolgen bei der Entstehung oder in der Lektüre mit Intentionalität aufgeladen werden.

Der auktoriale Akt ist jedenfalls auch in der digitalen Literatur nicht vollends abzustreifen. Allein hätte Bajohr den Roman nicht schreiben können, gewiss, aber ebenso wenig hätte Theo Lutz seine stochastischen Texte nicht ohne das Computermodell Zuse Z 22 oder Franz Kafkas „*Das Schloß*“ erstellen lassen können. Bajohrs Roman ist eine Kollaboration aus menschlichen wie nicht-menschlichen Akteur:innen sowie den Leser:innen, die dem Text während der Lektüre Sinn verleihen – „(Berlin, Miami)“ hat dieser Erkenntnis aber nichts Neues hinzugefügt. Die Autorschaftsverteilung im Roman lässt sich im Sinne der ANT multiperspektivisch als sich veränderndes Netzwerk, bestehend aus unterschiedlichen Akteur:innen, wie auch Rezipient:innen, beschreiben. Die Autorschaft von Mensch und Maschine definiert sich als Handlungsprozess, in dem bestimmte Operationen ausgeführt werden und Macht als verteilt betrachtet wird. Die Künstliche Intelligenz als technisches Objekt besitzt demnach ein Skript – ähnlich ihres algorithmischen Codes – welches den Rahmen einer Handlung definiert und darin die Aufgaben der Akteur:innen, der Werkzeuge sowie die Zuweisung von Rollen festlegt. Es bleibt aber den Akteur:innen überlassen, ob sie die Rolle annehmen oder nicht. Die großen Sprachmodelle bedingen einen gewissen Umgang mit ihnen, der nicht notwendigerweise befolgt werden muss, sind aber zugleich Ergebnis einer Unmenge an menschlich produzierter Textdaten, die selbst ein Handlungsskript vorgeben.⁴²⁰

Bajohr stellt sich ebenfalls die Frage, wer nun für das Werk verantwortlich ist und argumentiert, dass, wenn über verteilte Verantwortung gesprochen wird, neben sich selbst auch das Modell, dessen Programmierer:innen, Elias Kreuzmair, die Autor:innen der für das *fine-tuning* verwendeten Gegenwartsromane oder die Verfasser:innen der Texte aus den Trainingsdaten herangezogen werden könnten. Wie er anmerkt, nimmt diese mittelbare Beteiligung kein Ende und man könnte die Liste ewig fortsetzen, bis zum letzten Handgriff, der das letzte Teilchen im Computer aus der Erde geschürft oder den

⁴²⁰ Vgl. Jasmin Meerhoff: Verteilung und Zerstäubung. Zur Autorschaft computergestützter Literatur. In: TEXT+KRITIK. Digitale Literatur II. Vol. Sonderband (2021), S. 52-54.

Baum für das Papier des Buchs gefällt hat.⁴²¹ Allerdings wären wir hierbei wieder beim beschriebenen Handlungsprogramm aus der ANT, das verschiedenen Akteur:innen und Objekten zugrunde liegt. Auch bei einem traditionell geschriebenen Buch könnte man die Druckerei, die Schreibutensilien, den Verlag etc. in die Liste der notwendigen Produktionsbedingungen aufnehmen, es bleibt aber bei den Konventionen des Buchmarktes, die den Anspruch an eine Autor:in weiterhin aufrechterhalten. Die Infrastruktur der Sprachmodelle, bestehend aus Algorithmen, Programmierer:innen, Trainingsdatensätzen etc., ist wie das Vorhandensein einer Druckerei für die Produktion eines Buches eine notwendige Bedingung für ihre Funktionalität. Die Beteiligung menschlicher Akteur:innen an den Trainingsdatensätzen und die Einhaltung Copyright-geschützter Textdaten darf aus ethischen Gründen nicht missachtet werden. Die dadurch auftretenden Probleme bleiben aber im Kontext geistigen Eigentums zu klären.

Obwohl „(*Berlin, Miami*)“ sowie die behandelten Werke von Allado-McDowell in physischer Form als gedrucktes Buch erschienen sind, sind sie ein Produkt des Digitalen und ohne den Einsatz von Sprachmodellen nicht realisierbar. Sprache ist für Künstliche Intelligenz Material, ein Produkt, ein Datensatz, der verarbeitet wird. Künstliche Neuronale Netze sind vieldimensionale Vektorräume, in denen jeder Input miteinander in Beziehung steht und diese Beziehungen wieder durch Beziehungen codiert sind. Je nach Wahrscheinlichkeit werden stilistische Eigenheiten, häufig auftretende Wörter oder grammatische wie typografische Muster ausgewählt und verarbeitet – und das gleichzeitig, sowohl von den Trainingsdaten des vortrainierten Grundmodells als auch im Fall von „(*Berlin, Miami*)“ den vier eingespeisten Gegenwartsromanen im Besonderen.

Das Sprachmodell folgt damit wie der Algorithmus der frühen Computerliteratur einem bestimmten Handlungsablauf – allerdings, wie bereits im Kapitel 2.1.3.1 *Exkurs: Künstliche Neuronale Netze* besprochen, wesentlich komplexer und mit immer neuen, nicht reproduzierbaren Ergebnissen. Aber sowohl das konnektionistische wie das sequenzielle Paradigma folgen einer binären Logik, welche die Frage nach der Autorschaft nur unwesentlich ändert. In der frühen Computerliteratur haben Theo Lutz‘ „*Stochastische Texte*“ (1959) schon Fragen aufgeworfen wie, wer ist die Autor:in – Mensch oder Maschine? – und was ist nun das eigentliche Werk – der Algorithmus oder der entstandene Text? Lutz sieht sich vordergründig als Autor seines geschriebenen Programms und nur indirekt als Autor des daraus entstandenen Textes. Bei dieser Betrachtung wird das Programm zur geistigen Leistung des Autors und damit Hauptwerk,

⁴²¹ Vgl. Bajohr: Niemand schreibt allein (2023c), S. 263-266.

der Text der Maschine zum Beiwerk. Kafkas Text „*Das Schloß*“ dient als notwendiges Grundgerüst für die Arbeit des Algorithmus.⁴²²

Die in den *fine-tuning*-Romanen von „*(Berlin, Miami)*“ präsentierten Welten und Strukturierungen werden ebenso nur als Grundgerüst, als notwendige Bedingung für die Funktionalität des Sprachmodells verwertet und dementsprechend auf ihr Sprachmaterial heruntergebrochen, „regelrecht zersprengt“⁴²³ und neu zusammengefügt, wodurch Gemeinsamkeiten statistisch verstärkt werden.⁴²⁴ „Was das Modell aus dem Input zieht, ist eher ein Sound, ein Vibe – sowohl sprachlich wie thematisch –, als eine gemeinsame Wirklichkeit.“⁴²⁵ Die Autor:innen der vier ausgewählten Romane sind damit ästhetisch nicht für das Endprodukt des Sprachmodells verantwortlich, geschweige denn die Programmierer:innen oder die Künstliche Intelligenz selbst, so wie Kafka nicht für Lutz‘ Stochastische Texte ästhetisch in der Verantwortung steht.⁴²⁶ Der Text aus den Romanen steht für eine bestimmte Verteilung von Wahrscheinlichkeiten, beherbergt selbst einen Code bzw. ein Muster, dessen sich die Künstliche Intelligenz bedient.

⁴²² Vgl. Toni Bernhart: Beiwerk als Werk. Stochastische Texte von Theo Lutz. In: editio Vol. 34/1 (2020), S. 194.

⁴²³ Bajohr: Niemand schreibt allein (2023c), S. 255.

⁴²⁴ Vgl. Ebd., S. 254-255.

⁴²⁵ Ebd., S. 255.

⁴²⁶ Was den rechtlichen Aspekt betrifft, gelten bei KI-generierten Werken aktuell die europäischen Regelungen zum Urheberrecht, die laut dem deutschen Bundesministerium für Justiz ab 2026 auf europäischer Ebene im Hinblick auf KI evaluiert und angepasst werden sollen. Dazu zählt auch die Forderung einer detaillierten Zusammenfassung des verwendeten Trainingsmaterials für *general purpose AI*-Modelle. Generell sind für das Training von KI-Anwendungen im Internet veröffentlichte Texte erlaubt, sofern der Rechtsinhaber kein sogenanntes ‚Opt-out‘ erklärt hat, wodurch der Inhalt nicht für das Training der KI-Software genutzt werden darf. Physische Werke fallen unter den Schutz des Urheberrechtes, wenn sie eine ‚eigene geistige Schöpfung‘ eines Menschen darstellen. Dann ist eine Vervielfältigung bzw. Verbreitung in KI-Ergebnissen nicht zulässig, sofern keine Zustimmung des Urhebers erfolgt ist. Rein KI-basierte Inhalte genießen keinen urheberrechtlichen Schutz, da sich die Arbeit der KI nur bedingt steuern lässt. Sie haben damit im rechtlichen Sinne keinen Urheber. Generell gilt: das Urheberrecht greift nur, wenn die Basis des Werkes von einem Menschen geschaffen wurde. Dies ist im Einzelfall zu beurteilen. | Vgl. Bundesministerium der Justiz: Künstliche Intelligenz und Urheberrecht. Fragen und Antworten. [https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Nav_Themen/240305_FAQ_KI_Urheberrecht.pdf?__blob=publicationFile&v=2#:~:text=Auch%20die%20Bearbeitung%20und%20Umgestaltung,kann%20dieser%20frei%20verwendet%20werden](https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Nav_Themen/240305_FAQ_KI_Urheberrecht.pdf?__blob=publicationFile&v=2#:~:text=Auch%20die%20Bearbeitung%20und%20Umgestaltung,kann%20dieser%20frei%20verwendet%20werden.). (Zugriff: 07.01.2025), S. 1-3. || Rechtlich gesehen, ergeben sich durch die Nutzung physischer Werke in „*(Berlin, Miami)*“ einige Probleme. Da nun die *fine-tuning*-Romane von Glanz, Groß, Zange und Guse in „*(Berlin, Miami)*“ nicht nur im Datensatz vollständig enthalten, sondern auch teilweise in Bajohrs Roman wiedererkennbar sind, kann hier eine Urheberrechtsverletzung angenommen werden, sofern Bajohr bei den genannten Rechtsinhaber:innen keine Erlaubnis zur Nutzung und Verbreitung eingeholt hat. Es bleibt aber gleichzeitig zu klären, ob „*(Berlin, Miami)*“ rechtlich gesehen unter das Urheberrecht fällt oder nicht – ob die Basis des Romans nun als von einem Menschen geschaffen gilt oder nicht. Diese Überlegungen meinerseits sind aber mit Vorsicht zu betrachten und sollen nur kurz andeuten, wie zwingend notwendig eine Überarbeitung aktueller Urheberrechte und des rechtlichen Umgangs mit KI ist.

Dieses Muster bzw. dieser ‚Vibe‘ ist in Prosa weniger deutlich zu erkennen als beispielsweise in Bildern. Zum Roman *„(Berlin, Miami)“* hat die Frankfurter Allgemeine Zeitung ein KI-generiertes Bild erstellen lassen, das die Stimmung des Buches einfangen soll.⁴²⁷



Abbildung 1: KI-generiertes Bild zum Roman *„(Berlin, Miami)“* der FAZ

Abbildung 2: KI-generiertes Bild zum Roman *„(Berlin, Miami)“* der FAZ

Auf einen ersten, schnellen Blick wirkt das Bild wie eine Strandfotografie. Im Hintergrund ist der Berliner Fernsehturm zu erkennen – ein erstes Merkmal für eine seltsam verknüpfte Weltbeziehung – im Vordergrund tummeln sich Strandbesucher:innen in Badekleidung unter Palmen und sonnen sich auf Strandliegen. Ein genauerer Blick offenbart die für KI typischen Fehler menschlicher Anatomie. Grotesk wirken die Figuren mit seltsam verformten Extremitäten oder gesichtslosen Köpfen. Man glaubt, Menschen zu erkennen und erkennt sie doch nicht. Die Szene im Bild ist nicht vollständig erfahrbar und doch vermittelt sie eine oberflächliche Vertrautheit. Es wird etwas dargestellt, das Palmen oder einem Swimmingpool ähnelt und doch stimmt etwas nicht – es findet eine eigentümliche Durchlässigkeit von einer Welt statt, die nicht ganz greifbar ist. Das Passungsverhältnis zwischen Mensch und Welt wird durch die KI herausgefordert. Die angeblichen Figuren bleiben fleischliche Imitationen, vertraute Muster und Formen werden nachträglich vervollständigt. Die eigene Erwartung deutet das Bild ähnlich einer Pareidolie, dem Phänomen, in Dingen unbewusst oder bewusst Vertrautes zu erkennen.

⁴²⁷ Leider gibt es zur Erstellung des KI-Bildes keinerlei Aufzeichnung des *prompting* oder welcher Bildgenerator verwendet wurde. Das Bild soll daher nur veranschaulichend zeigen, was der Bildgenerator aus den beiden Begriffen ‚Berlin‘ und ‚Miami‘ erstellt hat. Mit dem Inhalt des Romans hat das generierte Bild wenig zu tun.

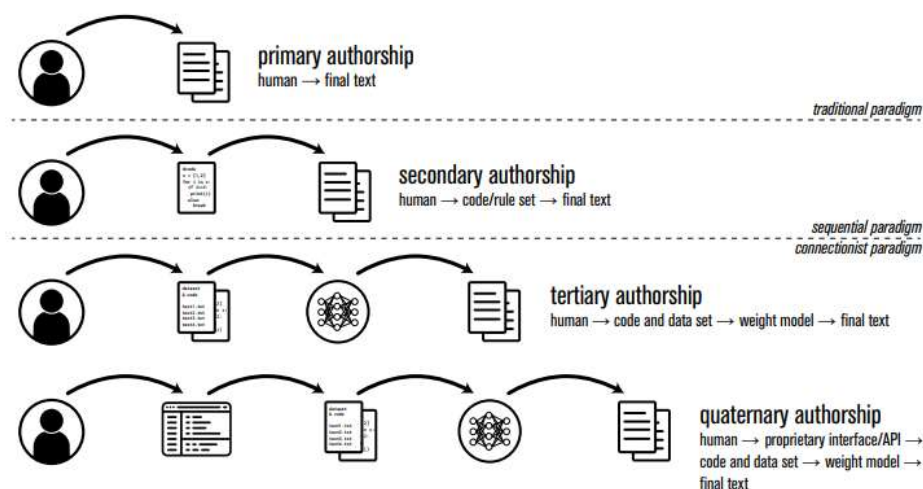
Ähnlich verhält es sich mit dem Text von „(Berlin, Miami)“, der auf den ersten Anschein vertraute Muster oder sinngefüllte Sätze enthält bzw. eine sich aufbauende Handlung andeutet, jedoch nie vollständig in die Realität umsetzt. Die eigene Standarderwartung an den Text legt ebenjene Vervollständigung nahe und man meint beim Lesen Handlung und sinngefüllte Sätze zu erkennen, die jedoch nichts weiter als eine Akkumulation von Wahrscheinlichkeiten und Bedeutungsmustern der Texte aus den Trainingsdaten darstellen. Die ästhetische Verantwortung ist aber nicht dem Sprachmodell zuzuordnen, weil sie eine derartige Verantwortung nicht übernehmen kann (s. Misselhorn in Kapitel 3.1.1 *Das Ende der Kunst durch KI?*) und nur über eine ästhetische Handlungsweise im funktionalen Sinn verfügt. Sie ist aber auch nur schwer einer menschlichen Instanz zuzuordnen, da der generierte Text nur ein Ausdruck **einer** möglichen, extrem hochskalierten Wahrscheinlichkeitsbeziehung unter vielen ist und sich bei Wiederholung des *prompts* oder Schreibauftrags an das Sprachmodell jedes Mal ändert. Die ästhetische Verantwortung beruht auf einer Codierungseigenschaft und ist Ausdruck einer mathematischen Formel, die nicht reproduzierbar ist. Die einzige Möglichkeit zur (menschlichen) Beeinflussung des Outputs von Sprachmodellen liegt in der Anpassung der Trainingsdaten und Texterstellungparameter. Meines Erachtens steht bei dieser Perspektive die letzte menschliche Instanz, also Bajohr in der (ästhetischen) Verantwortung des Romans, auch wenn dessen ‚Inhalt‘ nicht vorhersehbar war. Die Unterteilung der generativen Literatur in ein sequenzielles und konnektionistisches Paradigma ändert nur wenig an der allgemeinen Frage nach menschlicher Autorschaft. Die Entstehung eines (literarischen) Textes – auch durch KI – lässt sich immer auf ein letztes Stück menschlicher Entscheidung zurückführen. Für die unterschiedlichen Erscheinungsformen generativer Literatur merkt Bajohr jedoch eine variierende Distanzierung der Autor:in und ihrer Verantwortung zum Werk an. Obwohl diese Verbindung nie vollständig getrennt wird (oder generell getrennt werden kann), macht er vier Distanzgrade für Autorschaft aus, die das Spektrum generativer Literatur übersichtlicher kategorisieren und für weitere theoretische Überlegungen zugänglich machen soll.

Sein Modell einer Kausalen Autorschaft wird zum Schluss im anschließenden Kapitel noch kurz vorgestellt und anhand der ausgewählten Werke besprochen.

5.2.2 Bajohrs Modell Kausaler Autorschaft

Wie aufgezeigt werden konnte, agiert Künstliche Intelligenz anhand von Korrelationen und besitzt nicht die Fähigkeit, Kausalitäten ‚wahrzunehmen‘ und zu (re)produzieren. Die

Inszenierung einer Maschinenautorschaft auf kausaler Ebene gelingt demnach im Sinne des Turing-Tests nur für einen kurzen Zeitraum über den Umstand der Täuschung – oder über dessen Zuschreibung, als Bedingung von Funktionalität (banale Täuschung). Der Nachweis einer tatsächlichen kausalen Autorschaft bleibt aber bisher noch aus. Der Fokus literaturwissenschaftlichen Interesses hat sich nach der Hypertextliteratur in den 1990er und 2000er Jahren – die, gekennzeichnet durch Intertextualität, Interaktion und Intermedialität und in Tradition des Poststrukturalismus stehend, den Akt des Lesens herausstrichen – für generative Literatur wieder den Produktionsbedingungen und dem Verhältnis von Mensch und Maschine als Kooperation zugewandt.⁴²⁸ Kausalität lässt sich damit für Bajohr (2024b) als Kategorie für ein Schema von Distanzgraden bei Autorschaft nutzbar machen, die noch vor Bestimmung einer literarischen Autorfunktion den Übergang von konventionellen Schreibweisen zu sequenziellen und konnektionistischen Verfahren aufzeigen (s. Abb. 1). Sein Modell Kausaler Autorschaft umfasst vier Ebenen,



bei denen die Zunahme dieser kausalen Distanz zwischen Autor:in und Text beobachtet werden kann:

Abbildung 3: Die vier Grade Kausaler Autorschaft

Während die primäre Autorschaft (*primary authorship*) den geltenden Standard konventionellen Schreibens, also die als kausal unmittelbar wahrgenommene Verbindung von Autor:in und Text darstellt, verdoppelt die sekundäre Autorschaft (*secondary authorship*) den Schreibakt, indem von der Autor:in eine Regelfolge zur Produktion des Werkes formuliert wird. Es wird ein Text zur Formulierung eines Textes verfasst. Unter diese Form der Autorschaft fällt symbolische KI, binäre Codierung, darunter die ersten Versuche computergenerierter Literatur, wie Lutz' *Stochastische Texte* und Stracheys

⁴²⁸ Vgl. Bajohr: Autorschaft und Künstliche Intelligenz (2024a), S. 268-270.

Love Letter Generator. Tertiäre Autorschaft (*tertiary authorship*) ist ein Produkt der Automatisierung der Automatisierung. Der operative Text zur Herstellung des eigentlichen Textes muss nicht mehr unmittelbar menschengemacht sein, sondern kann selbst Produkt eines automatisierten Prozesses sein. In diese Operationskette fällt das konnektionistische Paradigma, das sich dadurch auszeichnet, dass nicht mehr Programmregeln aufgestellt werden, die Output produzieren, sondern das Programm mittels einer großen Output-Menge trainiert wird und selbst Regeln erstellt, die zum finalen Text führen. Menschliche Autorschaft begrenzt sich hierbei auf die Programmierung des Lernalgorithmus, die Auswahl des Datensatzes für das Training und die Kontrolle der Parameter für die Erstellung des Textes – der finale Text ist computergemacht. Bajohr definiert den letzten Distanzgrad als quartäre Autorschaft (*quaternary authorship*), bei der der einzige Nachweis menschlichen Eingreifens noch auf eine natürlichsprachliche Aufforderung über ein Interface eines bereits vortrainierten, kommerziell genutzten großen Sprachmodells, wie ChatGPT, bzw. auf dessen Abbruch zurückzuführen ist (*prompting*). Die Eingabe beeinflusst zwar, bestimmt aber nicht vollständig den Output des Modells. Die tertiäre und quartäre Autorschaft liegen, was den reduzierten menschlichen Einfluss angeht, nahe beieinander, letztere zeichnet sich aber durch eine zusätzliche Kontrollabgabe aus. Für den vierten Distanzgrad ist kein Know-how über die Programmierung, Einstellung oder Handhabung des Sprachmodells mehr notwendig.⁴²⁹ „Autorschaft ist eingekapselt in das Gehäuse einer kommerziellen Sprachtechnologie, deren genaues Funktionieren letztlich kaum mehr ergründlich ist.“⁴³⁰ Der Übergang zwischen den beiden Abstufungen ist aber fließend, da auch die Bezahlversion von ChatGPT (ChatGPT Plus) über individuelle *fine-tuning*-Einstellungen verfügt – die Auswahl der Trainingsdaten bleibt aber vorbestimmt.

Bajohr definiert damit Kausalität als notwendige Bedingung für die Beschreibung von Autorschaft im Kontext digital generierter Literatur. Sie macht auf die menschliche Komponente im Mensch-Maschine-Gefüge aufmerksam und identifiziert gleichzeitig alle beteiligten technischen Prozesse. Mit zunehmender Distanz zum unmittelbaren Schreiben primärer Autorschaft verliert der menschliche Part an Kontrolle und die Maschine an Transparenz. Die Textproduktion wird an algorithmische und schließlich statistische Prozesse übergeben, die vom Menschen nicht durch Regeln lesbar sind.⁴³¹

⁴²⁹ Vgl. Bajohr: Autorschaft und Künstliche Intelligenz (2024a), S. 271-272.

⁴³⁰ Bajohr: Autorschaft und Künstliche Intelligenz (2024a), S. 272.

⁴³¹ Vgl. Bajohr: Autorschaft und Künstliche Intelligenz (2024a), S. 273.

Bajohr ordnet Allado-McDowells „*Air Age Blueprint*“ dem letzten Distanzgrad, der quartären Autorschaft, zu, während die Parameter für seinen Roman „*(Berlin, Miami)*“, wie der Trainingsdatensatz, die Dauer des Trainings und die Zufälligkeit des Outputs etc., noch von ihm selbst festgelegt wurden und damit den Roman unter die Kategorie der tertiären Autorschaft einordnen.⁴³² Die Werke „*Amor Cringe*“ und „*Pharmako-AI*“ von Allado-McDowell fallen ebenfalls der quartären Autorschaftskategorie zu.

Bajohr merkt an, dass der Distanzgrad an kausaler Autorschaft noch nichts über die Qualität des finalen Textes aussagt, sondern lediglich den Abstand zwischen der Autor:in und dem Text und damit deren operativen Möglichkeiten offenlegt. Diese idealtypische Taxonomie hat aber auch ihre Grenzen.⁴³³ Im Falle von „*Amor Cringe*“ – und zu einem gewissen Grade auch „*Air Age Blueprint*“ – fehlt im Modell noch die letzte Instanz menschlichen Redigierens. Während Bajohr in „*(Berlin, Miami)*“ minimal in den finalen Output eingegriffen hat – er arrangierte frei die Kapitel entgegen ihrem Entstehungszeitpunkt und griff gelegentlich frühere Motive wieder auf – und damit prototypisch in die Kategorie der tertiären Autorschaft fällt, hat Allado-McDowell für „*Amor Cringe*“ keine Angaben zur Bearbeitung gemacht, weshalb nur schwer nachzuvollziehen ist, welche von GPT-3 generierten Textstellen verändert wurden und welche nicht. Für „*Air Age Blueprint*“ gab er: sie an, collageartige Techniken angewandt und frei redigiert zu haben. Damit wird der generierte Text in beiden Fällen nochmals menschlichen Einflusses unterzogen und gleichzeitig die Distanz der Autor:in zum Text auf einen unbestimmten Faktor vermindert. Die Auflösung der Distanz ergibt sich unter anderem durch die Intensität des Eingriffs. Die finale Bearbeitung negiert nicht die vorherigen Schritte konnektionistischer Texterstellung, verstärkt aber den menschlichen Faktor kausaler Autorschaft und führt zu einer Reklamation von Kontrolle über den Text. Der Text ist primär nicht menschengemacht, aber unter menschlichen Maßstäben von Kausalität, Intentionalität und einer eventuellen Konzeption bearbeitet worden.

Es lässt sich dem Modell Kausaler Autorschaft deshalb noch eine weitere Komponente hinzufügen, die zwar nicht für eine fünfte Kategorie ausreicht, aber das konnektionistische Paradigma erweitert und dem Modell eine Feedbackschleife menschlichen Redigierens hinzufügt:

⁴³² Vgl. Bajohr: *Die Deixis der Literatur* (2024), S. 6-7.

⁴³³ Vgl. Bajohr: *Writing at a Distance* (2024b), S. 324-325.

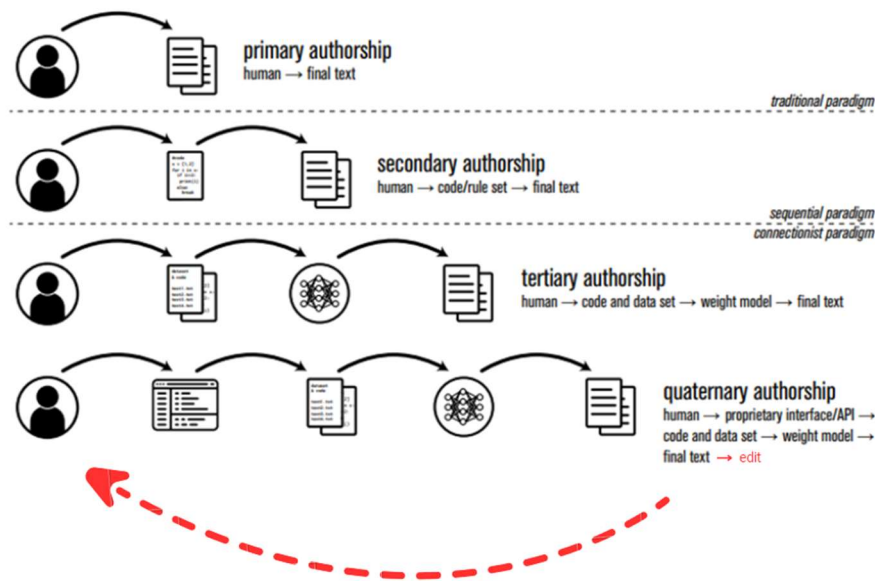


Abbildung 4: Erweitertes Modell Kausaler Autorschaft

Dennoch bleibt eine weitere Frage offen, die das Modell nicht abdeckt, sondern noch vor einer Bestimmung der kausalen Autorschaft vom Text beantwortet werden muss – nämlich, ab wann bzw. welchen Prozentsatz der Einsatz generativer Sprachmodelle den Text unter die Kategorie digitale Literatur stellt oder in das konnektionistische Paradigma einordnet und, ob menschliche Bearbeitung eines bereits generierten Textes diesen wieder unter den Deckmantel primärer Autorschaft stellen kann. So hat beispielsweise die japanische Autorin Rie Kudan ihren Roman „Tokyo-to Dojo-to“ zu etwa fünf Prozent **wörtlich** von ChatGPT generieren lassen. Der kanadische Schriftsteller Stephen Marche hat hingegen seine Novelle „Death of An Author“ zu 95 % von generativen Sprachmodellen übernommen – die generierten Textpassagen aber laut eigenen Angaben in vielen Bearbeitungsschleifen zurechtgerückt und die Handlung vollständig allein konzipiert.⁴³⁴

Die Definition digitaler Literatur befindet sich historisch gesehen noch im Fluss und wird laufend durch den technologischen Wandel, der bestehende elektronische Literaturformen veralten lässt und neue hervorbringt, erweitert und umgeschrieben. Bedingt der prinzipielle Einsatz generativer Sprachmodelle bzw. die Erstellung von Text durch eine Maschine die Kategorisierung als digitale Literatur, sind beide Romane der quartären Autorschaft unterzuordnen – die Distanz zum Text bliebe bei Kudan aber bei jenen Passagen aufrecht, die sie Wort für Wort von ChatGPT übernommen hat, während Marche durch die sorgfältige Überarbeitung Kontrolle über den Roman zurückerlangt.

⁴³⁴ Die Presse-Redaktion: Bekenntnis bei Preisvergabe: Autorin ließ ChatGPT für sich schreiben. In: Die Presse (2024) <https://www.diepresse.com/18004351/bekenntnis-bei-preisvergabe-autorin-liess-chatgpt-fuer-sich-schreiben> (Zugriff: 27.07.2024).

Paradoxerweise fällt damit die Distanz der Autor:innen zum finalen Text genau entgegengesetzt ihrer jeweiligen schriftlichen Eigenleistung aus.

Werden andere – jedoch nur als hinreichend kategorisierte – Merkmale digitaler Literatur hinzugezogen, wie, dass diese durch Computer hervorgebracht und auf diesen gelesen wird bzw. bewusst die Verquickung natürlicher und artifizieller Schreibweisen hervorhebt oder ihre digitale Herkunft stets kritisch im Bewusstsein hält, dann fällt Kudans Werk als Gesamtes nicht unter die Kategorie digitale Literatur, sondern ist als Teil des traditionellen Paradigmas menschlichen Schreibens zu sehen. Die Distanz zwischen Autor:in und Text wird als ungebrochen wahrgenommen. Der Schreibakt findet fast gänzlich auf Seiten der Autor:in statt, die nur unwesentlich in ihren Entscheidungen von anderen Akteur:innen beeinflusst wurde. Die Problematiken, die sich aber bei einer sozialdeterministischen Sichtweise des Sprachmodells als Werkzeug und Inputgeber ergeben, wurden bereits bei Allado-McDowell angesprochen. So wird unter dem Deckmantel primärer (menschlicher) Autorschaft der Blick auf technische Kontexte und *biases* verstellt, aber auch die Erwartungshaltung an die Autor:in gestört. Von Künstler:innen und Autor:innen wird eine gewisse (geistige) Eigenleistung oder ein (literarischer) Herstellungsaufwand erwartet. Der Einsatz von KI mindert diesen Aufwand beim Schreibprozess, wodurch nicht nur eine persönliche Ausdrucksweise der Autor:in eventuell verloren geht, sondern auch in eine mögliche Enttäuschung bei der Aufklärung des Produktionsprozesses und Bekenntnisses über die Verwendung einer KI umschlägt. Die Bedeutung der KI-generierten Textpassagen in Kudans Werk bleibt literaturtheoretisch noch unbeantwortet und ein methodologisch wie analytisch blinder Fleck. Auch im Modell Kausaler Autorschaft lässt sich nur schwer eine Mischform aus primärer und quartärer Autorschaft vorstellen oder die Bedeutung generativer Sprachmodelle als kreativer Inputgeber oder Dialogpartner verorten.

Eine Zuordnung bestimmter Texte in Bajohrs Modell Kausaler Autorschaft gestaltet sich demnach schwierig bzw. umfasst das Modell keine Alternativen oder Möglichkeiten für Mischformen – falls derart theoretisch überhaupt argumentiert werden kann. Das Modell bietet einen Überblick über den Kontrollverlust und die zunehmende Distanzierung der Autor:in zum Text durch den Einsatz algorithmischer und generativer Texterstellung, ist aber keinesfalls vollständig oder universal theoretisch anwendbar. Ähnlich der ANT bieten die Überlegungen zu Kausaler Autorschaft einen Theorierahmen zur Beschreibung statt einer normativen Theorie von Autorschaft.

6 Zusammenfassung und Ausblick

Ob in personalisierter Werbung, im Verkehr, in der Medizin oder im Klassenzimmer – Künstliche Intelligenz (*Artificial Intelligence, AI*) ist in aller Munde und hat Einzug in viele Bereiche unseres Arbeitslebens und unseres Alltags gehalten. Sie ist integriert in der Smartwatch auf unserem Arm, in den Sozialen Medien, in unseren Newsfeeds auf dem Smartphone, im Auto, auf Streamingplattformen, in der Wetterapp – kaum eine Branche verzichtet auf die Unterstützung automatisierter Datenverarbeitung. Auf dem Feld der Kunst und Literatur hat die neue Technologie allerdings einen anderen Stellenwert und wird mit großer Skepsis betrachtet.

Die Vorstellung Künstlicher Intelligenz als Bedrohung, als ‚neue‘ Künstler:in und Autor:in, welche die alten ablöst, prägt den Diskurs. Fragen um künstliche Autorschaft werden aufgeworfen und die Revision traditioneller Kategorien steht im Raum. Die vorliegende Masterarbeit hat sich dieses Diskurses angenommen und einen Blick auf das Phantasma Künstlicher Intelligenz geworfen.

Zu Beginn der Arbeit wurde der aktuelle Diskurs um KI anhand des Intelligenzbegriffs aufgemacht und es konnte gezeigt werden, wie die Definition als künstliches Abbild im Unterschied zu einer natürlichen, menschlichen Intelligenz bereits zu allerlei Problematiken in der Wahrnehmung und Einschätzung der Fähigkeiten und Funktionsweisen der Technologie führt. In der Nachbildung des Menschen als Maschine erhebt die KI den Anspruch menschlicher Intelligenz. Das Gehirn wird mit der Rechenmaschine gleichgesetzt. Die menschliche Kognition wird durch informationstheoretische und mathematisch-formalisierbare Kategorien beschreibbar gemacht. Mensch und Maschine nähern sich semantisch einander an. Begriffe wie ‚Denken‘, ‚Intelligenz‘ und ‚Verstehen‘ werden für den Computer geöffnet. Heutige digitale Maschinen ‚lernen‘, ‚entscheiden‘ oder ‚optimieren sich selbst‘, anstatt zu rechnen, zu speichern oder zu verarbeiten – ein anthropomorpher Sprachgebrauch, der sich seit dem Beginn der Entwicklung Künstlicher Intelligenz durchgesetzt hat.

Den einflussreichsten Beitrag zu der Frage, ob Maschinen ‚denken‘ können, lieferte Alan Mathison Turing, der in seinem Aufsatz „*Computing Machinery and Intelligence*“ philosophische wie theologische Fragen über das Denken und Menschsein an sich aufarbeitete. Sein *imitation game* (Turing-Test) – in der am häufigsten zitierten Form: die erfolgreiche Nachahmung eines Menschen durch eine Maschine – inspiriert bis heute die Erforschung Künstlicher Intelligenz. Die Simplität des Testverfahrens ist für seine

breite Rezeption verantwortlich. Turing vertritt damit einen im engsten Sinn behavioristischen Standpunkt. Der Test bietet keine Einblicke in die Intelligenzfähigkeit einer Maschine oder einen Nachweis von Intelligenz, sondern favorisiert eine evidenzbasierte Objektivität, die sich auf die erfolgreiche Imitation fokussiert. Die Täuschung wird zum Prinzip erklärt und führt zu einer irreführenden Gleichstellung menschlicher Vernunft und technischer Leistungsfähigkeit, wodurch Turing viel Kritik erntete.

Der Philosoph John R. Searle zeigte unter anderen mit seinem Gedankenexperiment ‚*The Chinese Room*‘ auf, dass ein Computer, der formale Regeln nutzt, noch lange keine Duplikation von menschlichen kognitiven Zuständen, wie Denken oder Verstehen, erreichen kann, sondern lediglich die Simulation ihrer Effekte nach außen hin sichtbar macht. Mensch und Maschine sind in dieser Hinsicht funktional äquivalent, ihr Output oder Verhalten sind durch eine Beobachter:in nicht zu unterscheiden, auch wenn ontologisch ein deutlicher Unterschied besteht. Ein kurzer Exkurs in der Arbeit zu Künstlichen Neuronalen Netzen konnte hoffentlich das Verständnis zu KI auf einen technisch akkuraten Stand bringen.

Dennoch bleibt auch heute noch das Prinzip der Täuschung für die Entwicklung und Vermarktung Künstlicher Intelligenz vorrangig. Je leistungsfähiger und menschenähnlicher die Technologie wird, desto unkritischer werden Subjektivierungsprozesse von KI-Anwendungen angenommen. Dass User:innen ihr Einfühlungsvermögen auf nicht-emotionale, bewusstseinslose Artefakte ausdehnen, hat bereits der ELIZA-Effekt gezeigt. In der Anthropomorphisierung, Personalisierung und Inszenierung Künstlicher Intelligenz als künstlichen ‚Menschen‘ (s. Kapitel 2.2 *Menschliches Technikbild: AI-Charaktere von Meta*) wird die Technologie quasi-subjekthaft. Die Beziehung zum Menschen bleibt als KI-Assistenz zwar noch rein funktional, entfernt sich aber zunehmend vom Objektstatus. Der anthropomorphe Sprachgebrauch (Künstliche Intelligenz ‚denkt‘, ‚handelt‘, ‚entscheidet‘) verleiht eine Aura des Menschlichen, macht sie zu *als-ob*-Subjekten, die in menschliche Handlungskontexte integriert werden. Die anthropomorphisierte Technologie und die technische Selbstmodellierung des Menschen stehen in einem konvergenten Verhältnis zueinander: das Gehirn wird maschinisiert und der Computer neuralisiert.

In diesem Gedankenmodell erscheint auch eine Maschinenautorschaft wahrscheinlicher, die sich aus einer Auflösung anthropologischer Modelle und gleichzeitiger Vermenschlichung der Technologie ergibt. Die Reaktion der Human- und

Geisteswissenschaften fokussiert sich deshalb auf eine Verteidigung des autonomen, selbstbestimmten Subjekts. Als bisher rein dem Menschen vorbehaltene Fähigkeiten und Eigenschaften wie die der Kreativität und Autonomie werden vorgebracht und verknüpft und eine Grenze zwischen menschlichem und künstlichem Schaffen gezogen, die die anthropologische Sonderstellung sichert.

Wie in Kapitel 3 *Kreativ sein, bedeutet Mensch sein* gezeigt werden konnte, gerät vor allem die Kreativität als ästhetische Kategorie in ihrer Argumentation und Verteidigung der Sonderstellung des Menschen aber zunehmend in die Krise. Der Begriff der Kreativität, der sich als zentrale Eigenschaft des Menschen und als Voraussetzung für das Schaffen von Kunst positioniert hat, ist in seiner Historizität noch relativ jung. Er erweist sich als ideologisch auf ein autonom schaffendes Individuum bezogen und wird retrospektiv in der Genieästhetik des 18. Jahrhunderts stark verankert – einer Vorstellung, von der man sich in Kunst und Literatur bereits loszulösen versuchte. Kreativität lässt sich als Kategorie einer klassischen ästhetischen Theorie leicht durch die Begriffe von Originalität und Inspiration ersetzen, vor allem dadurch, dass der Begriff einen eher weichen Definitionsbereich umfasst. Kreativität ist zum Massenbegriff geworden. Er wird kommerziell verwendet (kreative Werbung), sozial-psychologisch interpretiert und als Teilbereich der Intelligenz angesehen (kreativ sein, kreativ denken) sowie als entscheidende Kategorie für die Schaffung von Kunst instrumentalisiert (kreative Künstler:in). In der Ausbildung eines Kreativitätsdispositivs folgt der Begriff einer Subjektzentrierung in der Kunst und richtet sich gleichzeitig auf die gesamte Gesellschaft als soziale Praxis aus. Das künstlerische Subjekt wird zum nachahmenswerten Vorbild, während die Gesellschaft auf kreativen Innovationen beruht. Kreatives Denken gehört zur psychologischen Selbstmodellierung des Menschen.

Durch seinen teils schon inflationären Gebrauch umfasst der Begriff zahlreiche Interpretationen, die Minimaldefinition beschränkt sich aber auf die rein funktionale Bedeutung ‚etwas Neues zu schaffen‘. Sieht man sich die Prozesse an, die kreative Leistungen hervorbringen, wirkt der Begriff weniger an ein autonomes Subjekt gekoppelt. Boden (2016) ist deshalb der Ansicht, dass auch Maschinen kreativ operieren können. Die Informationsästhetik Max Benses begreift die Kreativität als messbare Größe und schließt sie mit der Kategorie des Zufalls kurz, um das Schreiben maschinenkompatibel zu machen. Indem der computergenerierte Zufall kreative Prozesse algorithmisch modellierbar macht, wird auch das schöpferische Vermögen des Menschen

simulierbar. Kreativität wird so zum Funktionsprinzip und für Computer zugänglich. Die Frage nach dem Subjekt wird abgewertet.

Demgegenüber bleibt eine enge Verknüpfung des Kreativitätsbegriffs mit menschlichen Eigenschaften, wie Bewusstsein und Autonomie bestehen, die als das entscheidende Kriterium von Kunst herangezogen werden. Diese Verknüpfung beeinflusst die Erwartungshaltung an Künstliche Intelligenz, die demnach autonom schaffen muss, um als kreativ-künstlerisch gelten zu können (s. Kapitel 3.1 *Keine Kreativität ohne Grenzüberschreitung*: Du Sautoy). Kreativität bleibt untrennbar mit Kunst und damit mit Bewusstsein und Autonomie verbunden – diesen Anspruch kann aber keine bestehende schwache Künstliche Intelligenz (*weak AI*) erfüllen.

Die vorliegende Masterarbeit hat sich deshalb der Frage gewidmet, ob Künstliche Intelligenz überhaupt Kunst schaffen kann, wenn die Frage nach ihrer Kreativität unterschiedlich beantwortet wird. Misselhorn (2023) stellt in Zweifel, dass es sich in Bezug auf KI-generierte Werke überhaupt um Kunstwerke handelt, wenn diese ohne Rückgriff auf eine Autor:in betrachtet werden. Aktuelle Technologien Künstlicher Intelligenz, die unter die Kategorie der schwachen Künstlichen Intelligenz (*weak AI*) fallen, sind allein nicht dazu in der Lage, Misselhorns drei Kunstkriterien – Bedeutung, Gehalt und Interpretation – vollständig zu erfüllen. Erst, wenn die Autor:in, Programmierer:in oder Stichwortgeber:in in die Gesamtpformance eines KI-Kunstwerkes miteinbezogen wird, kann das Kunstwerk eine Bedeutung verkörpern, einen Gehalt höherer Ordnung präsentieren, der eine gewisse Sichtweise zum Ausdruck bringt und eine Interpretation einfordern, die die Kunstwelt miteinbezieht. Dann handelt es sich aber nicht mehr um KI-Kunst, sondern um Konzeptkunst, die sich KI-Verfahren bedient, um Fragen danach aufzuwerfen, was Kunst in der digitalisierten Welt eigentlich ist. Misselhorns Kunstdefinition denkt damit ein Konzept von Autorschaft mit, das stark an ein autonom handelndes Subjekt appelliert, dem eine ästhetische Verantwortung zugeschrieben werden kann. Auch eine Co-Autorschaft mit einer KI lehnt sie ab, da Mensch und Maschine nicht auf derselben Stufe stehen und es der Maschine an den notwendigen mentalen Zuständen (*causal powers*) fehlt, um ästhetische Verantwortung zu übernehmen und Autorschaft zu beanspruchen. Das Problem mit KI-Kunst bestehe für sie in der Simulation bzw. Nachahmung menschlich-künstlerischen Schaffens, weshalb für Misselhorn KI generierte Kunst keine ‚echte‘ Kunst sei. Wenn die Künstler:in aus dem Werk entfernt wird, verliert die Arbeit an künstlerischem Wert.

Doch wie weit lässt sich die Rolle der Künstler:in reduzieren? Die Relevanz der Künstler:in als schöpferisches Individuum ist zugunsten neuer Autorisierungsstrategien in modernen Kunsttheorien bereits zurückgegangen. Diese haben sich von einer auktorialen Legitimation und Ermächtigung sowie einer Verantwortung dem Kunstwerk gegenüber abgewandt. Eine Kunstdefinition über Ideen von Kreativität oder die Bedeutung einer höheren Ordnung ist dem Auswählen, Arrangieren, Adaptieren, Organisieren von Verfremdungseffekten bzw. der reinen Geste des Zur-Kunst-Erklärens gewichen. Autorschaft wird durch einen zufallsabhängigen Automatismus überwunden. In der Literatur ist in gleicher Tradition das souveräne (kreative) Genie sowie eine dem Werk innewohnende Autor:innenintention durch die Poststrukturalisten infrage gestellt worden, wodurch eine Neuverhandlung von Autorschaft abseits einer engen Kopplung von Subjekt und Werk ermöglicht wurde.

In Kapitel 4 *„Der Autor“ wird vom Thron gestoßen* wurden deshalb die Aufsätze *„Der Tod des Autors“* (1968) von Roland Barthes und *„Was ist ein Autor?“* (1969) von Michel Foucault vorgestellt, die die Autor:in als sinnstiftende Instanz und Herrscher:in über ihr Werk verabschieden und den Fokus auf die Rezipient:in verlagern. Obwohl dadurch die empirische Autor:in keineswegs verschwindet, sondern lediglich ihre Überhöhung infrage gestellt wird, geht diese Reduzierung der Relevanz von Autorschaft einer Literaturbewegung voraus, die dem Text eine noch nie dagewesene Eigenmächtigkeit zusagt und eine subjektzentrierte Autorschaft zugunsten eines Automatismus neu verhandelt.

Digitale Poesie, Computerliteratur, maschinenanaloges Schreiben, generative Literatur, Netzliteratur etc. – das Aufkommen digitaler Literatur hängt eng mit der technischen Errungenschaft des Computers zusammen und umfasst, je nach Produktionsaspekt, einen großen Bereich elektronisch-literarischer Texte. Während die einzige Gemeinsamkeit in der Notwendigkeit eines Speichermediums sowie eines Lesegerätes bzw. einer Software für die Herstellung besteht, umfassen digitale Texte eine Bandbreite an weiteren notwendigen, aber nicht hinreichenden Merkmalen, die medienübergreifend und zeitgebunden sind. So können digitale Texte interaktiv oder nicht-linear sein, über einen zweifachen Text (Schrift und Quellcode) verfügen, im Internet vernetzt, nicht abgeschlossen, multimedial, veränderbar, fluide, dialogisch usw. sein oder auch mehr als eine Autor:in haben. Digitale Literatur ist demnach stark an die Technologien ihrer Hervorbringung gebunden, weshalb im Kontext Künstlicher Intelligenz wieder vermehrt der Fokus auf den Produktionsaspekt der Texte gelegt wird.

In jedem Fall hat digitale Literatur zu einem Paradigmenwechsel von Originalität und Kreativität geführt und neue Formen des Schreibens sowie eine Diversifizierung von Autorschaft nach sich gezogen. Angelehnt an die Konzeptkunst wird bei diesem neuen Schreiben der Wert der Idee über die der Textqualität gestellt. Sprache wird zum Wortmaterial, mit dem gespielt wird. Die Bearbeitung und das Management von Texten gewinnen an Bedeutung, die Zitation wird als auktoriale Leistung aufgewertet.

Misselhorn hat die Ähnlichkeiten von konzeptueller und generativer Literatur, die als Teil digitaler Literatur anzusehen ist, betont. Beiden unterliegt eine gewisse Regelmäßigkeit, sie sind antiexpressiv und autordezentriert, die Rezeptionsästhetik den Sinn erst nachträglich durch die Betrachter:in erhält – aber sie sind keineswegs gleichzusetzen. Während der Code, auf dem generative Literatur basiert, formalisiert und frei von Ambivalenz ist, können dem Konzept mehrere Interpretationen unterliegen, oft in Zusammenhang mit seiner Ausführung. Das Konzept müsste im Minimalfall sogar nicht einmal ausgeführt werden, anders als der Code, der zudem nur für einen kleinen, qualifizierten Teil lesbar ist. Die generative Ästhetik ist demnach nicht ausschließlich der konzeptuellen Kunst zuzuweisen, sobald die menschliche Instanz dahinter miteinbezogen wird, wie Misselhorn argumentiert. Eine solche Kategorisierung übersieht die Feinheiten im Zusammenhang mit ihrer Erzeugung.

Die generative Ästhetik leitet sich aus der Konzeptkunst ab, kann aber selbst auf eine lange Tradition zurückblicken, die eine Brücke zwischen Informatik und Kunst schlägt. In der generativen Literatur, zu der auch die jüngsten Texte von Sprachmodellen (*large language models*, *LLM*) zählen, ist der gesamte Herstellungsprozess von Interesse. Bajohr (2021) argumentiert, dass die angewandte Technik in der Herstellung des Textes auch einen Einfluss auf die Debatte um Autorschaft bei generativer Literatur hat. Er schlägt deshalb eine produktionsästhetische Unterteilung in ein sequenzielles und ein konnektionistisches Paradigma vor, das die technischen Differenzen der klassisch durch Algorithmen hervorgebrachten (sequenziell) und der auf Künstlichen Neuronalen Netzen (KNN) bzw. KI-basierten Texte (konnektionistisch) berücksichtigt. Anders als bei der sequenziellen Texterzeugung, wo Autorschaft sekundär jenen Personen zugeschrieben wird, die sich für das Programmskript verantwortlich zeichnen, vereint das konnektionistische Paradigma Code und Konzept in der *black box* der Künstlichen Intelligenz und entzieht diese damit jeglicher Offenheit und Transparenz. Der Text wird nicht mehr von einem Algorithmus generiert, sondern das Modell für die Erstellung des Textes wird ‚selbstständig‘ vom Sprachmodell auf Basis der Trainingsdaten gebildet. Die

Autor:innenfunktion reduziert sich auf ein Auswählen der Datensätze oder ein Bearbeiten des finalen Textes.

Bevor das Sprachmodell dadurch aber mystifiziert wird und einen Genius zugeschrieben bekommt, lohnt es sich, stattdessen die Maschine und den dahinterstehenden Menschen als ein Geflecht historischer und sozialer Interaktionen zu betrachten – als ein Netzwerk, das die Interaktion zwischen Mensch und Technik in den Vordergrund stellt. Die Akteur-Netzwerk-Theorie erlaubt eine solche Beschreibung und zeigt die enge Verflochtenheit von Mensch, Gesellschaft und Technik, ohne sie als separate Bereiche zu denken. In der theoretischen Diskussion um Autorschaft bei Künstlicher Intelligenz bietet sie ein geeignetes Vokabular, das Autorschaft als Effekt beschreibt. So besteht das Netzwerk ‚Autor:in‘ schon immer aus einer Verbindung aus menschlichen wie nicht-menschlichen Elementen, die jeweils über ein Handlungsprogramm verfügen und selbst aus Netzwerken bestehen. Diese Netzwerke sind ein Prozess und fluid. Sie beschreiben, wie Autorschaft auf verschiedenen Ebenen verteilt und neu hergestellt, stabilisiert und destabilisiert wird. Die Interaktion zwischen Mensch und Maschine wird in den Vordergrund gestellt und *agency* nicht zwingend menschlich-intentional gedacht, sondern als eine in komplexen Operationsketten verteilte Autorschaft.

Die Akteur-Netzwerk-Theorie wurde in der Masterarbeit deshalb dazu genutzt, die komplexe Verteilung von Autorschaft direkt anhand der ausgewählten Werke aufzuzeigen. Untersucht wurden die Werke *„Pharmako-AI“*, *„Amor Cringe“* und *„Air Age Blueprint“* von Kenric Allado-McDowell sowie der Roman *„(Berlin, Miami)“* von Hannes Bajohr.

Kenric Allado-McDowell stellt seinen:ihren Umgang mit Künstlicher Intelligenz als *AI-Co-Writing-Ansatz* vor, der stark auf dem Prinzip der Kollaboration aufbaut. Der Schreibprozess soll damit das Ergebnis einer symmetrischen, heterarchischen Struktur sein – beide Akteur:innen beeinflussen einander. In der Kombination von Mensch und Maschine wird laut Allado-McDowell eine neue Form von Kunst und Literatur geschaffen. Die drei untersuchten Romane *„Pharmako-AI“*, *„Amor Cringe“* und *„Air Age Blueprint“* wurden mit dem kommerziellen Sprachmodell GPT-3 kooperativ verfasst (*co-written*), es lassen sich aber Unterschiede im Umgang und der Inszenierung der KI feststellen, welche unter Berücksichtigung ihres Produktionsaspekts auch einen Einfluss auf die Diskussion um Autorschaft haben.

„*Pharmako-AI*“ ist Allado-McDowells erstes KI-Buch, das einen philosophisch-esoterischen Dialog mit GPT-3 experimentell erforscht. Inputs von Allado-McDowell und die generierten Textpassagen des Sprachmodells sind getrennt und jeweils in einer eigenen Schriftart markiert. Dadurch lässt sich Allado-McDowells Arbeitsweise gut nachvollziehen. Jedes Kapitel startet mit einem *prompt* von ihm:ihr, gefolgt von der Antwort der KI. Das Buch ordnet sich zeitlich in den um 2020 aufgeflamnten Hype um Künstliche Intelligenz ein, wodurch Allado-McDowells Umgang mit dem Sprachmodell und die Rezeption des Buches von Anthropomorphisierungstendenzen und der technischen Ideologie einer starken KI geprägt sind. GPT-3 wird als die andere Intelligenz zur menschlichen positioniert und mit philosophischen Fragen konfrontiert. Das Sprachmodell wird im Buch als Genius mit einem allumfassenden Wissen mystifiziert, auch wenn nach kritischem Lesen die generierten Textpassagen nur auf wenig Faktenwissen beruhen und Stereotypen (*bias*) reproduzieren. Die Interaktion mit der KI vergleicht Allado-McDowell mit der Einnahme psychoaktiver Pflanzen, die das Bewusstsein erweitern. Dem Sprachmodell wird dadurch mehr Tiefe verliehen als vielleicht gerechtfertigt ist. Die generierten Textpassagen werden als Ergebnis des Unbewussten der Maschine (*the unconscious*) präsentiert. Er:sie koppelt damit die *black box* der Künstlichen Intelligenz stark an ein menschliches Bewusstsein und lässt die Grenzen verschwimmen. Aufgrund des limitierten Kontextfensters des Sprachmodells sieht sich Allado-McDowell im gemeinsamen Schreiben als Bindeglied zwischen menschlichem Gedächtnis (*human memory*) und maschinellem Speicher (*memory*). Das Vokabular des Sprachmodells nimmt er:sie in das eigene Schreiben mit auf.

Durch die Anthropomorphisierung und Mystifizierung des Sprachmodells inszeniert Allado-McDowell das Sprachmodell getrennt vom eigenen Schreiben als Ko-Autor:in – als vorgeblich gleichgestellte Partner:in mit denselben kausalen Kräften (*causal powers*) wie bei einem Menschen. Die Gleichstellung von Mensch und Maschine und Nobilitierung zur KI-Autor:in funktioniert aber nicht vollständig – weder auf Produktions- noch auf Rezeptionsseite. Als schwache KI ist GPT-3 ohne Allado-McDowell als Inputgeber nicht funktionsfähig. Aufseiten der Rezipient:innen wird *agency*, Intentionalität und letztlich Autorschaft vorgetäuscht, die bei Entlarvung des Mysteriums Künstlicher Intelligenz aber sofort wieder entzogen wird. Allado-McDowells wilde Mischung unterschiedlicher theoretischer Disziplinen, wo er:sie kybernetische mit philosophischen und psychologischen Denkmustern zu verschränken versucht, geht nicht auf. Eine Revision des Autor:innenbegriffs und eine Beanspruchung einer vollständigen

Maschinenautorschaft bleibt aus. Es wird der Anschein einer KI-Autor:in vorgetäuscht, um das Phantasma Künstlicher Intelligenz aufrechtzuerhalten. Schlussendlich wird GPT-3 aber nicht als Ko-Autor:in vorgestellt und der Abdruck des Namens ‚K Allado-McDowell‘ auf der Titelseite lässt Allado-McDowells Position als ein komplexes Autor:innen-Netzwerk, welches das Sprachmodell als nicht-menschliche Akteur:in integriert, glaubwürdiger darstellen, als eine Gleichstellung von Mensch und Maschine in einem *Co-Writing*-Ansatz.

In der Novelle „*Amor Cringe*“ wird versucht, diese angebliche Gleichstellung im *AI-Co-Writing*-Ansatz vollständig im Text durchzuziehen. Allado-McDowell integriert hier die Schreibweise der KI und verzichtet auf die Kennzeichnung von menschlich geschriebenen und maschinell generierten Passagen. Er:Sie vermischt das eigene Schreiben mit dem Vokabular des Sprachmodells, wodurch keine nachträgliche Trennung mehr zwischen Mensch und Maschine möglich wird. Diese intransparente Vorgehensweise lässt das Werk wie ein traditionell geschriebenes erscheinen und auch so lesen – abgesehen von der erzählerisch schwachen und sehr verwirrenden Handlung und den inkohärenten Sätzen. Die Novelle wird als *deepfake autofiction* beschrieben, wodurch auf ontologischer Ebene die Täuschung der KI und die Fiktion einer Autobiographie mit der Beanspruchung von Echtheit und Authentizität doppelt gekoppelt wird. Daraus ergeben sich Fragen nach der Standarderwartung an Texte, wenn ihre Herkunft nicht eindeutig geklärt ist. Die Standarderwartung – ein Konzept, das Bajohr nach dem gleichen Prinzip wie den hermeneutischen oder autobiographischen Pakt gestaltet hat – meint eine gewisse Erwartungshaltung gegenüber einer bestimmten Textsorte. Im Kontext Künstlicher Intelligenz richtet sich diese Erwartung an den Achsen menschlich/nicht-menschlich aus. Wenn die Herkunft unsicher ist, wird automatisch eine menschliche Autor:in angenommen. Allado-McDowells „*Amor Cringe*“ verunsichert nun diese Standarderwartung so weit, bis die Frage nach Autorschaft schlicht redundant wird. Der Fokus verschiebt sich, wie von den Poststrukturalisten ausgemalt, auf eine Praxis des Lesens. Die Sinnkonstitution geschieht auf Seiten der Leser:in. Die Kritik daran weist aber, wie schon am Poststrukturalismus angemerkt, auf ein immer dünner werdendes Band zwischen Bedeutung und Welt hin. Der Bezug zur Außenwelt geht durch ein in sich geschlossen agierendes System verloren. Unser Verhältnis von Sprache wird gelockert.

Der Roman „*Air Age Blueprint*“ führt die Produktionsweisen der vorherigen Werke zusammen und kennzeichnet die menschlich wie maschinell generierten Textpassagen in

einer eigenen Schriftart, nicht aber die frei vorgenommenen Änderungen durch Allado-McDowell. Anhand dieser Vorgehensweise, die Allado-McDowell selbst als collageartige Technik (*collage-like techniques*) bezeichnet, lässt sich keine chronologische Abfolge der *prompts* für das Sprachmodell oder der nachträglichen Änderungen feststellen. Allado-McDowell sieht sein:ihr Schreiben als ein Arbeiten mit Versatzstücken und Zitaten, ähnlich der Collage- und Montagetechniken des 20. Jahrhunderts. Er:Sie arrangiert den Text von GPT-3 und gibt mit seiner:ihrer eigenen Schreibweise die Handlung vor.

Es muss an dieser Stelle betont werden, dass keine Produktionsgeschichte oder Aufzeichnungen zu den *prompt*-Verfahren Allado-McDowells vorliegen, weder bei „*Air Age Blueprint*“ noch den anderen Werken, wodurch seine:ihre Angaben zur Entstehung der Bücher in der kurzen *Note of Composition*, die allen Werken Allado-McDowells bis auf „*Amor Cringe*“ (hier stehen Notizen dazu auf der Buchrückseite) vorangeht, auch täuschend sein können.

Das Genre des Romans changiert zwischen Esoterik, Science-Fiction und philosophischem Austausch der Protagonist:in und der KI mit einer Reiseerzählung als narrativer Rahmung. Die erzählenden Passagen, wo tatsächlich ein Ereignis stattfindet – definiert nach Lotmans (1972) Theorie des Narrativen – sind in fett gedruckter Schrift verfasst, d.h. von Allado-McDowell geschrieben. Textstellen, die GPT-3 generierte, umfassen vorwiegend Naturereignisse und Beschreibungen sowie den Inhalt des Manifests der KI-Figur *Shaman.AI*.

Wie nach näherer Analyse des Romanes gezeigt werden konnte, nutzt Allado-McDowell die KI-verfassten Passagen sowohl auf struktureller wie inhaltlicher Ebene. Während die KI sein:ihr eigenes Schreiben ergänzt, repräsentiert sie im Werk die Stimme der Natur, der Umwelt bzw. das Äußere oder Andere zum Menschen durch *Shaman.AI*. GPT-3 füllt damit die Bedeutung des Begriffs Medium mit dreifachem Sinn: Das Sprachmodell ist als Textgenerator und Kommunikationsmedium auf struktureller Ebene am Schreibprozess beteiligt – gleichzeitig tritt die KI als literarische Figur auf, als kommunikatives Medium zwischen Mensch, Natur und Technik – zuletzt wird der Begriff esoterisch aufgeladen und versteht die KI als spirituelles Medium, das zwischen Natur und Mensch vermittelt.

Obwohl Allado-McDowell sich vom Output des Sprachmodells in seinem:ihren eigenen Schreiben beeinflusst fühlt und bewusst die Grenze zwischen Mensch und Maschine verschwimmen lässt, bleibt die editorische Leistung des Buches, die

Konzeption des Narrativs und das generelle Konzept des Werkes Ausdruck menschlicher Autorschaft. Die Verwendung der KI als Erzählprinzip rahmt die maschinell generierten Passagen konzeptuell und der Anteil menschlicher Autorschaft wird wieder aufgewertet. Das Sprachmodell als heterarchisch gleichgestellte *Co-Writing*-Instanz zu sehen, ist eine Täuschung und scheint einer technologischen Ideologie einer starken Künstlichen Intelligenz zu unterliegen. Eine solche Perspektive führt zu keiner Revision des Autorschaftsbegriffs bzw. scheint das Verhältnis zwischen Mensch und Maschine nur oberflächlich darzustellen.

Hannes Bajohr startete mit „*(Berlin, Miami)*“ seinen ersten Romanversuch mit Künstlicher Intelligenz. Dafür trainierte er die quellenoffenen Sprachmodelle GPT-J und GPT-NeoX mit den vier Gegenwartsromanen „*Pixeltänzer*“ (2018) von Berit Glanz, „*Flexen in Miami*“ (2020) von Joshua Groß, „*Realitätsgewitter*“ (2016) von Julia Zange und „*Miami Punk*“ (2019) von Juan S. Guse. Alle Romane haben das Digitale zum Thema und in ihrer narrativen Struktur verankert. Damit gab Bajohr die Stimmung von seinem Roman vor, der mit dem Prinzip der *next token prediction* generiert wurde. Bajohrs weiterer Einfluss beschränkte sich auf die Auswahl des nächsten Buchstabens oder Wortes bei der Textgeneration sowie der Anordnung der generierten Kapitel und Textpassagen. Gelegentlich griff er Motive oder Figuren auf, die das Sprachmodell ausspuckte und führte sie in späteren Kapiteln wieder ein.

Als literarisches Experiment angelegt, sieht Bajohr „*(Berlin, Miami)*“ als Versuch, das Erzählvermögen Künstlicher Intelligenz herauszufordern. Die KI scheitert im Roman an einer kausalen Handlung – das Experiment ist geglückt. Narrative Kohärenz wird durch Korrelation simuliert oder durch Bajohr unterstützt, indem Figuren wie der Kieferling wieder in die ‚Handlung‘ eingeführt werden. Dennoch entsteht nie eine vollständig kausale Erzählung und es ergeben sich zeitliche Sprünge, logische Brüche oder schwer nachzuvollziehende Figurenmotive. Da die Generierung von Text unendlich hätte weitergeführt werden können, ließ Bajohr mit einem zweiten Experiment den Roman auf ‚natürliche‘ Weise enden.

Dazu brach Bajohr die Textgenerierung ab und nutzte die ersten drei generierten Teile von „*(Berlin, Miami)*“ als *fine-tuning* für den letzten Teil. Der so entstandene IV. Teil geht über die ‚Handlung‘ der ersten drei Teile hinaus und führt den Roman als Experiment im Experiment zu einem Ende. Überraschenderweise kommt es durch das Training der KI mit ihrem eigenen Output zu einem sogenannten Modellkollaps (*model collapse*) und einem vollständigen Zusammenbruch der Erzählung. Sinnentleerte und grammatikalisch

falsche Sätze, wirre Zeichen- und Symbolfolgen zeigen den graduellen Zerfall der KI an ihrer eigenen Logik.

Bajohr sieht im Roman seine Autorschaft auf ein Minimum reduziert. Die Auswahl des nächsten Tokens in der Texterstellung ist seines Erachtens eher dem Zufall zuzuordnen als Ergebnis einer intentionalen Entscheidung oder Regelmäßigkeit. Er erwähnt, dass dieses Auswählen auch ein Zufallsgenerator hätte übernehmen können. Wie allerdings dargelegt wurde, unterliegt der computergenerierte Zufall weiterhin einer binären Logik und ist weniger ‚zufällig‘ und regelhafter bzw. reproduzierbarer als Bajohrs Zufall.

Allein hätte Bajohr den Roman jedenfalls nicht schreiben können. Bajohr nimmt während des Schreibprozesses eine Doppelrolle ein, die zwischen Produzent und Rezipient alterniert. Einerseits beansprucht er Autorschaft durch die Auswahl der Trainingsdaten und Konzeption des Werkes, andererseits sieht er sich selbst in der Rolle des Rezipienten, der erzählerische Fehlleistungen der KI zu deuten beginnt. Dass sein Name auf dem Umschlag des Romans steht, mag der Konvention geschuldet sein, allerdings sieht Bajohr seine Autorschaft nicht als legitimiert bzw. deklariert sie als Lüge, was diskussionswürdig bleibt. Am Oberflächentext lässt sich jedenfalls keine Kategorie wie Intentionalität festmachen, eine Maschinenautorschaft schließt Bajohr aber aus.

In weiteren Überlegungen widmet sich Bajohr der (ästhetischen) Verantwortung des Werkes und wer diese trägt. Während die mittelbare Beteiligung am Text bis zu den Verfasser:innen der Trainingsdaten gezogen werden kann, ist hier anzumerken, dass die Funktionalität der Sprachmodelle auf eine gewisse Infrastruktur, bestehend aus Algorithmen, Programmierer:innen, Trainingsdatensätzen etc. angewiesen ist – ebenso wie die Produktion eines Buches auf das Vorhandensein einer Druckerei und ihrer Arbeiter:innen. Die Beteiligung menschlicher Akteur:innen an den Trainingsdatensätzen und die Einhaltung Copyright-geschützter Textdaten darf aus ethischen Gründen nicht missachtet werden. Die dadurch auftretenden Probleme bleiben aber im Kontext geistigen Eigentums zu klären. Die Verfasser:innen der Trainingsdaten sind ästhetisch wie natürlich **nicht** für das Ergebnis des Sprachmodells verantwortlich, sondern sind Teil eines Handlungsprogramms nach dem die Künstliche Intelligenz ‚agiert‘. Die Autorschaft von Mensch und Maschine wird im Sinne der ANT als Handlungsprozess definiert, in dem bestimmte Operationen ausgeführt werden und Macht als verteilt gedacht wird. Die KI als technisches Objekt führt demnach ein Skript aus. Den Akteur:innen bleibt überlassen, ob sie diesem Skript folgen oder nicht. Die *fine-tuning*-Romane von „(Berlin,

Miami)“ werden auf ihr Sprachmaterial heruntergebrochen und ihre statistischen Gemeinsamkeiten vom Sprachmodell verstärkt. Was das Modell den Texten entzieht, ähnelt mehr einem Vibe bzw. einem Muster aus Relationen. Stilistische Eigenheiten, häufig auftretende Wörter oder grammatische wie typografische Muster strukturieren das Grundgerüst, mit dem das Sprachmodell den Text weiter verarbeitet. Der generierte Text enthält dann vertraute Muster oder scheinbar sinngefüllte Sätze, ist im Prinzip jedoch nur eine Akkumulation von Wahrscheinlichkeiten und Bedeutungsmustern aus den Texten der Trainingsdaten. Dass dieses Grundgerüst aus Mustern und Wahrscheinlichkeiten in sich selbst zusammenbrechen kann, hat der IV. Teil des Romans gezeigt, als die KI an ihrer eigenen Logik gescheitert ist. Die Generierung eines (literarischen) Textes kommt (noch) nicht ohne menschliches Direktiv aus – es bleibt immer ein letztes Stück menschlicher Entscheidung zurück.

Bajohrs Unterteilung der generativen Literatur in ein sequenzielles und ein konnektionistisches Paradigma ändert daher nur wenig an der allgemeinen Frage nach menschlicher Autorschaft in maschinell generierten Texten. Die Unterscheidung macht aber auf die produktionsbedingte zunehmende Distanz der Autor:in zum Text aufmerksam. Diese variierende Distanzierung hat Bajohr in einem eigenen Modell einer Kausalen Autorschaft vorgestellt, das zum Schluss der Arbeit noch besprochen wurde.

Kausalität wird in diesem Modell als Kategorie für ein Schema von vier Distanzgraden von Autorschaft genutzt, das den Übergang von traditionellen zu sequenziellen und konnektionistischen Schreibverfahren aufzeigen soll. Primäre Autorschaft weist die geringste Distanz zwischen Autor:in und Text auf, während die sekundäre Autorschaft (sequenzielles Paradigma) den Schreibakt bereits durch Algorithmen automatisiert. Die tertiäre und quartäre Autorschaft gehören dem konnektionistischen Paradigma und damit dem Schreiben mit Sprachmodellen (*LLM*) an. Beide Autorschaftsformen sind Ergebnis einer automatisierten Automatisierung. So ist der generierte Text selbst Produkt eines automatisierten Prozesses. Den einzigen Unterschied macht Bajohr in der Auswahl der Trainingsdaten fest, wodurch die tertiäre Autorschaft noch mehr menschlicher Kontrolle unterliegt. Die Grenzen sind aber eher fließend als klar bestimmt. Die Definition von Kausalität als notwendige Bedingung macht auf die menschliche Komponente im Mensch-Maschine-Gefüge aufmerksam und identifiziert gleichzeitig die beteiligten technischen Prozesse. Mit zunehmendem Distanzgrad verliert der Mensch an Kontrolle über den Schreibakt und die Maschine an Transparenz, weil die statistischen Texterstellungsprozesse nicht mehr durch Regeln lesbar sind.

Dem Modell Kausaler Autorschaft nach gehören die Werke von Allado-McDowell zur quartären Autorschaft, während „(*Berlin, Miami*)“ dem tertiären Autorschaftsgrad zuzuordnen ist. Das Modell bietet eine Beschreibungsmöglichkeit von Autorschaft für generative Literatur, sagt aber nichts über die Qualität der Texte aus.

Wie gezeigt werden konnte, ist das Modell noch erweiterbar, da einige operative Möglichkeiten nicht berücksichtigt wurden und die Zuordnung bestimmter Texte damit schwierig gestalten. So fehlt beispielsweise eine letzte Instanz, welche die nachträgliche menschliche Bearbeitung der KI-generierten Texte berücksichtigt. Dem Modell wurde demnach ein letzter Schritt des Redigierens (*edit*) hinzugefügt, der die Distanz zwischen Autor:in und Text auf einen unbestimmten Faktor verringert.

Weitere Problematiken einer Zuordnung und Repräsentation von Distanz wurden anhand des Romans „*Tokyo-to Dojo-to*“ von Rie Kudan und der Novelle „*Death of An Author*“ von Stephen Marche aufgezeigt. So hat Kudan fünf Prozent ihres Textes wörtlich von ChatGPT übernommen, während Marche den zu 95 % maschinell generierten Text mit viel Aufwand überarbeitet hat. Intuitiv fällt Kudans Werk unter das traditionelle Paradigma primärer Autorschaft, die von einem kommerziellen Sprachmodell übernommenen Passagen sind aber ohne Auswahl der Trainingsdaten generiert und damit Teil quartärer Autorschaft. Marches Werk fällt unter das konnektionistische Paradigma quartärer Autorschaft, wurde aber aufwändig redigiert, wodurch sich die Distanz zum Werk wieder verringert.

Wie bei Allado-McDowell und zuletzt auch bei Kudan gezeigt, bringt eine sozialdeterministische Sichtweise der KI als Werkzeug und Inputgeber einige Probleme mit sich. So verstellt eine Verschleierung des Anteils maschinell generierter Passagen den Blick auf technische Kontexte, auf die aus den statistischen Wahrscheinlichkeiten der Sprachmodelle gezogenen *biases*, auf den Datenverarbeitungsprozess und die historische Dimension digitaltechnischer Entwicklungen. Die Verwendung kommerzieller, vortrainierter Sprachmodelle hat aber auch ästhetische Folgen und läuft Gefahr, aufgrund des von der Technologie bevorzugten statistischen Durchschnitts Bestehendes laufend zu wiederholen und immer mehr vom Gleichen zu produzieren. Beliebte oder statistisch häufig auftretende Genres, Handlungselemente, Motive, Stereotypen oder Phrasen versprechen auf Dauer Langeweile, anstatt neue interessante literarische Werke hervorzubringen. Dass eine Verschleierung des Einsatzes von KI beim Schreibprozess von einer Täuschung in eine Enttäuschung nach Aufklärung des Produktionsprozesses umschlägt, hängt zudem noch mit einer zutiefst menschlich geprägten Vorstellung von

Kunst und Literatur zusammen, die eine gewisse Eigenleistung oder einen Herstellungsaufwand von der Künstler:in oder der Schreiber:in erwartet.

Dass Bajohrs Romanexperiment am Erzählmodus gescheitert ist, ist eventuell auch ein Hinweis darauf, dass mit KI anders Literatur zu machen ist als die bloße Nachahmung menschlichen Erzählens. Denn die KI repräsentiert nicht das ‚Andere‘ des Menschen, sondern funktioniert auf Basis dessen, auf menschlich produzierte Trainingsdaten. Ohne diese kommt es zum Modellkollaps (*model collapse*). Die generierten Geschichten simulieren menschliche Erzählweisen, bieten aber keine neuen ästhetischen Erkenntnisse. Wird KI genutzt, um menschliches Erzählen nachzuahmen oder zu ersetzen, geht der Blick auf das medienspezifische der Technologie verloren: der Repräsentation von Wahrscheinlichkeiten und Mustererkennung. Ein erfolgreiches literarisches Experiment mit ästhetischem Wert streicht diese technischen Affordanzen und die spezifische Produktionsweise des Sprachmodells hervor. Eine ästhetische Kritik generativer Literatur umfasst auch das Forschungsdesiderat der vorliegenden Arbeit, die hoffentlich bereits erste Überlegungen zum Thema bieten konnte.

Ob nun menschliche Autorschaft nach Distanzgraden getrennt oder über ein Netzwerk an menschlichen wie nichtmenschlichen Akteur:innen gedacht wird – die Verbindung zwischen menschlicher Autor:in und Werk ist nie vollständig getrennt. Eine autonome Maschinenautorschaft steht so lange außer Frage, bis die Maschine erfolgreich ein deiktisches Urteil fällen kann (demnach tatsächlich autonom wird und über soziale Handlungsmacht verfügt) oder die Leser:in über die eigene Täuschung hinwegsieht und der Ursprung des Textes keine Bedeutung mehr hat.⁴³⁵

Auch wenn im Silicon Valley das Phantasma einer ‚starken‘ KI-Autorschaft beansprucht und die persistente Idee des romantischen Genies auf Sprachmodelle übertragen wird – eines ist auch jetzt noch sicher: „Die Autorin ist immer zu Hause.“⁴³⁶ Sie mag in ihrer Abwesenheit auffallen, für tot erklärt und wiederbelebt oder als Datenhaufen in den digitalen Raum von Maschinen gesteckt werden. Es bleibt immer ein notwendiges Stück menschlicher Entscheidung zurück – ansonsten würde das System zusammenbrechen (*modal collapse*). Zu einer Revision über den Anspruch menschlicher Autorschaft führen konnektionistische Sprachmodelle daher (noch) nicht. Der Autorbegriff verliert höchstens seine ohnehin umstrittene Exklusivität.⁴³⁷

⁴³⁵ Vgl. Bajohr: Die Deixis der Literatur (2024), S. 17.

⁴³⁶ Kathrin Passig: Vielleicht ist das neu und erfreulich. Technik. Literatur. Kritik. Graz/Wien: Literaturverlag Droschl 2019 (Grazer Vorlesungen zur Kunst des Schreibens 2), S. 78

⁴³⁷ Vgl. Catani: Generierte Texte (2022), S. 253.

7 Literaturverzeichnis

7.1 Primärliteratur

Allado-McDowell, K: Air Age Blueprint. United Kingdom: Ignota 2022.

Allado-McDowell, K: Amor Cringe. New York/Los Angeles: Deluge Books 2022.

Allado-McDowell, K: Note on Composition. In: K Allado-McDowell: Air Age Blueprint. United Kingdom: Ignota 2022, S. X.

Allado-McDowell, K: Note on Composition. In: K Allado-McDowell: Pharmako-AI. United Kingdom: Ignota 2020, S. XI.

Allado-McDowell, K: Pharmako-AI. United Kingdom: Ignota 2020.

Bajohr, Hannes: (Berlin, Miami). Berlin: Rohstoff-Matthes & Seitz Berlin 2023a.

Lutz, Theo (23.7.1932 - 31.1.2010): Stochastische Texte (1959).
https://auer.netzliteratur.net/0_lutz/lutz_original.html (Zugriff: 22.05.2024).

Poe, Edgar Allan: The Raven, and The Philosophy of Composition. In: Project Gutenberg (2017)
<https://www.gutenberg.org/cache/epub/55749/pg55749-images.html> (Zugriff: 20.11.2024).

Racter [Chamberlain, William/Thomas Etter]: The Policeman's Beard is Half Constructed. Computer prose and poetry. New York: Warner Books 1984.

7.2 Sekundärliteratur

Agamben, Giorgio: Der Autor als Geste. In: Giorgio Agamben: Profanierungen. Aus dem Italienischen von Marianne Schneider. Frankfurt am Main: Suhrkamp 2015 (edition suhrkamp 2407), S. 57-69.

Akrich, Madeleine/Bruno Latour: Zusammenfassung einer zweckmäßigen Terminologie für die Semiotik menschlicher und nicht-menschlicher Konstellationen. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 399-405.

Anders, Günther: Die Antiquiertheit des Menschen. Über die Seele im Zeitalter der zweiten industriellen Revolution. Ungekürzte Sonderausgabe. München: C. H. Beck 1961.

Bachleitner, Norbert: Formen „unkreativen“ Schreibens in der digitalen Literatur. In: Christina Hoffmann/Johanna Öttl (Hg.): Digitalität und literarische Netz-Werke. Wien/Berlin: Turia + Kant 2017 (antikanon #2), S. 38-72.

Bajohr, Hannes: Artificielle und postartificielle Texte. Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben. In: Sprache im technischen Zeitalter Vol. 61/245 (2023b), S. 37-61.

Bajohr, Hannes: Autorschaft und Künstliche Intelligenz. In: Stephanie Catani unter Mitarbeit von Anna-Maria Hartmann (Hg.): Handbuch Künstliche Intelligenz und die Künste. Berlin/Boston: Walter de Gruyter 2024a, S. 265-280.

Bajohr, Hannes: Das Reskilling der Literatur. Einleitung zu Code und Konzept. In: Hannes Bajohr (Hg.): Code und Konzept. Literatur und das Digitale. Berlin: Frohmann 2016, S. 7-21.

Bajohr, Hannes: Die Deixis der Literatur. Über die Erwartungshaltung an computergenerierte Texte. In: Textpraxis Sonderausgabe Vol. 8/1 (2024) (https://www.textpraxis.net/sites/default/files/beitraege/06_bajohr_die_deixis_der_literatur_0.pdf), S. 1-19.

Bajohr, Hannes: Dumme Bedeutung. Künstliche Intelligenz und artifizielle Semantik. In: Merkur Vol. 76/882 (2022), S. 69-79.

Bajohr, Hannes: Künstliche Intelligenz und digitale Literatur. Theorie und Praxis konnektionistischen Schreibens. In: TEXT+KRITIK. Digitale Literatur II. Vol. Sonderband (2021), S. 174-185.

Bajohr, Hannes: Niemand schreibt allein. Nachwort zu (Berlin, Miami). In: Hannes Bajohr: (Berlin, Miami). Berlin: Rohstoff-Matthes & Seitz Berlin 2023c, S. 239-267.

Bajohr, Hannes: Schreiben in Distanz. Hildesheimer Poetikvorlesung 2022. Hildesheim: Universitätsverlag Hildesheim 2023d (Theorie und Praxis 4).

Bajohr, Hannes: Schreibenlassen. Texte zur Literatur im Digitalen. Berlin: August 2022.

Bajohr, Hannes: Writing at a Distance: Notes on Authorship and Artificial Intelligence. In: German Studies Review Vol. 47/2 (2024b), S. 315-337.

Barthes, Roland: Der Tod des Autors. In: Roland Barthes: Das Rauschen der Sprache. (Kritische Essays IV). Aus dem Französischen von Dieter Hornig. Frankfurt am Main: Suhrkamp 2006 (edition suhrkamp 1695), S. 57-63.

Belliger, Andréa/David J. Krieger: Einführungen in die Akteur-Netzwerk-Theorie. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 13-50.

Bernhart, Toni: Beiwerk als Werk. Stochastische Texte von Theo Lutz. In: editio Vol. 34/1 (2020), S. 180-206.

Betschon, Stefan: Künstliche Intelligenz. Zwischen Realitäten und Illusionen. Zürich: Vontobel-Stiftung 2019 (Schriftenreihe 2320).

Binczek, Natalie: Text, Schreiben, Intertextualität. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 277-292.

Boden, Margaret A.: AI. It's nature and future. Oxford: Oxford University Press 2016.

Brentano, Franz: Psychologie am empirischen Standpunkt. Mit Einleitung, Anmerkungen und Register herausgegeben von Oskar Kraus. Hamburg: Felix Meiner 1973 (Philosophische Bibliothek 192).

Budelacci, Orlando: Mensch, Maschine, Identität. Ethik der Künstlichen Intelligenz. Basel: Schwabe Verlag 2022 (Schwabe reflexe 74).

Callon, Michel/Bruno Latour: Don't Throw the Baby Out with the Bath School! A Reply to Collins and Yearley. In: Andrew Pickering (Hg.): Science as Practise and Culture. Chicago: The University of Chicago Press 1992, S. 343-368.

- Catani, Stephanie: »ERZÄHLMODUS AN«. Literatur und Autorschaft im Zeitalter künstlicher Intelligenz. In: Jahrbuch der Deutschen Schillergesellschaft Vol. 64 (2020), S. 287-310.
- Catani, Stephanie: Generative Literatur: Von analogen Romanmaschinen zu KI-basierter Textproduktion. In: Stephanie Catani unter Mitarbeit von Anna-Maria Hartmann (Hg.): Handbuch Künstliche Intelligenz und die Künste. Berlin/Boston: Walter de Gruyter 2024, S. 153-170.
- Catani, Stephanie: Generierte Texte. Gegenwartsliterarische Experimente mit Künstlicher Intelligenz. In: Andrea Bartl/Corina Erk/Jörn Glasenapp (Hg.): Schnittstellen. Wechselbeziehungen zwischen Literatur, Film, Fernsehen und digitalen Medien. Paderborn: Brill Fink 2022 (inter|media 15), S. 247-266.
- Charlier, Robert: Kanon und Plagiat. In: Christina Hoffmann/Johanna Öttl (Hg.): Digitalität und literarische Netz-Werke. Wien/Berlin: Turia + Kant 2017 (antikanon #2), S. 16-37.
- Cisek, Günter: Machtwechsel der Intelligenzen. Wie sich unser Miteinander durch künstliche Intelligenz verändert. Wiesbaden: Springer Vieweg 2021 (Die blaue Stunde der Informatik).
- Daniels, Dieter: Marcel Duchamp – der einflussreichste Künstler des 20. Jahrhunderts? In: Museum Jean Tinguely Basel (Hg.): Marcel Duchamp. Ostfildern-Ruit: Hatje Cantz 2002, S. 25-36.
- Dorffner, Georg: Rationalität, Emotionalität, Körperlichkeit: Können Maschinen Menschen verstehen – und umgekehrt? In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein „animal rationale“? Vernunft – Kognition – Intelligenz. Darmstadt: Wissenschaftliche Buchgesellschaft 2004 (Topologien des Menschen 1), S. 102-112.
- Du Sautoy, Marcus: Der Creativity Code. Wie künstliche Intelligenz schreibt, malt und denkt. München: C.H. Beck 2021.
- Fletcher, Angus: Why Computers Will Never Read (or Write) Literature: A Logical Proof and a Narrative. In: Narrative Vol. 29/1 (2021), S. 1-28.
- Foucault, Michel: Was ist ein Autor? In: Fotis Jannidis /Gerhard Lauer/Matias Martinez et. al. (Hg.): Texte zur Theorie der Autorschaft. Stuttgart: Reclam 2000, S. 198-229.
- Goldsmith, Kenneth: Uncreative Writing. Sprachmanagement im digitalen Zeitalter. Erweiterte deutsche Ausgabe. Berlin: MSB Matthes & Seitz 2017.
- Grunwald, Armin: Soziale Roboter – vermenschlichte Technik und technische Menschenbilder. In: Selin Gerlek/Sarah Kissler/Thorben Mämecke et. al. (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen. Hagen: Hagen University Press 2022 (Digitale Kultur 1), S. 62-73.
- Grunwald, Armin: Wer bist du, Mensch? Eine alte Frage vor neuen Herausforderungen. In: Armin Grunwald (Hg.): Wer bist du, Mensch? Transformation menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt. Freiburg/Basel/Wien: Herder 2021, S. 9-38.
- Guilford, J. P.: Kreativität (1950). In: Gisela Ulmann (Hg.): Kreativitätsforschung. Köln: Kiepenheuer & Witsch 1973, S. 25-43.
- Häusser, Philip: Natürlich alles künstlich. Was künstliche Intelligenz kann und was (noch) nicht – KI erklärt für alle. München: Droemer 2021.
- Heibach, Christiane: Literatur im elektronischen Raum. Frankfurt am Main: Suhrkamp 2003 (suhrkamp taschenbuch wissenschaft 1605).

Heibach, Christiane: Literatur im Internet. Theorie und Praxis einer kooperativen Ästhetik. Als Ms. gedruckt. Berlin: dissertation.de 2000.

Heibach, Christiane: Vernetzte Autorschaft: Hypertext und Internet. In: Wetzel, Michael (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 603-624.

Hennig Martin/Kilian Hauptmann: Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung. In: Zeitschrift für Medienwissenschaft Vol. 2/21 (2019), S. 86-94.

Henrickson, Leah: Computer-Generated Fiction in a Literary Lineage. Breaking the hermeneutic contract. In: Logos Vol. 29/2-3 (2018), S. 54-63.

Heßler, Martina: Künstliche Intelligenz: Eine neuartige Kategorie von Maschine oder die Vermenschlichung der Maschine. In: Yasemin Keskinetepe/Dr. Anke Woschech (Hg.): Künstliche Intelligenz. Maschinen lernen Menschheitsträume. Göttingen: Wallstein 2021, S. 116-121.

Hübner, Edwin: Menschlicher Geist und Künstliche Intelligenz. Die Entwicklung des Humanen inmitten einer digitalen Welt. Stuttgart: Freies Geistesleben 2020.

Kampmann, Sabine: Autorschaft in den Bildkünsten. Forschungen über bildende Künstlerinnen und Künstler in Kunstgeschichte, Geschlechterforschung und Kunst. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 543-554.

Keskinetepe, Yasemin/Dr. Anke Woschech (Hg.): Künstliche Intelligenz. Maschinen lernen Menschheitsträume. Göttingen: Wallstein 2021.

Knapp, Lore: Autorschaft als Akteur-Netzwerk. In: Zeitschrift für Germanistik Vol. 24/1 (2019), S. 85-99.

Lämmel, Uwe/Jürgen Cleve: Künstliche Intelligenz. Wissensverarbeitung – Neuronale Netze. 5. überarb. Aufl. München: Hanser 2020.

Latour, Bruno: Die Hoffnung der Pandora. Untersuchungen zur Wirklichkeit der Wissenschaft. Frankfurt am Main: Suhrkamp 2002 (suhrkamp taschenbuch wissenschaft 1595).

Latour, Bruno: Über technische Vermittlung: Philosophie, Soziologie und Genealogie. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 483-528.

Law, John: Notizen zur Akteur-Netzwerk-Theorie: Ordnung, Strategie und Heterogenität. In: Andréa Belliger/David J. Krieger (Hg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript 2006, S. 429-446.

Lehmann, Johannes F.: Genie und Autorschaft. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 259-276.

Lejeune, Philippe: Der autobiographische Pakt. Aus dem Französischen von Wolfram Bayer und Dieter Hornig. Frankfurt am Main: Suhrkamp 1994 (edition suhrkamp 1896).

Lenzen, Manuela: Natürliche und künstliche Intelligenz. Einführung in die Kognitionswissenschaft. Frankfurt am Main: Campus 2002 (Campus Einführungen).

- Liggieri, Kevin: Anthropozentrismus in Mensch-Maschine Interaktionen. In: Selin Gerlek/Sarah Kissler/Thorben Mämecke et. al. (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen. Hagen: Hagen University Press 2022 (Digitale Kultur 1), S. 74-94.
- Lobin, Henning: Engelbarts Traum. Wie der Computer uns Lesen und Schreiben abnimmt. Frankfurt am Main: Campus 2014.
- Lotman, Jurij M.: Die Struktur literarischer Texte. 2. unv. Aufl. München: Wilhelm Fink 1986.
- Lotze, Netaya: Chatbots. Eine linguistische Analyse. Frankfurt am Main: Peter Lang 2016 (Sprache – Medien – Innovationen 9).
- McCarthy, John/Marvin L. Minsky/Nathaniel Rochester et. al.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955. In: The AI magazine Vol. 27/4 (2006), S. 12-14.
- Meerhoff, Jasmin: Verteilung und Zerstäubung. Zur Autorschaft computergestützter Literatur. In: TEXT+KRITIK. Digitale Literatur II. Vol. Sonderband (2021), S. 49–61.
- Misselhorn, Catrin: Künstliche Intelligenz – das Ende der Kunst? Ditzingen: Reclam 2023 (Reclams Universal-Bibliothek [Was bedeutet das alles?] 14355).
- Mühlhoff, Rainer: Menschengestützte Künstliche Intelligenz. Über die soziotechnischen Voraussetzungen von »deep learning«. In: Zeitschrift für Medienwissenschaft Vol. 2/21 (2019), S. 56-64.
- Natale, Simone: Deceitful Media. Artificial Intelligence and Social Life after the Turing Test. New York: Oxford University Press 2021.
- Noji, Eryk: Das kybernetische Subjekt: Selbstthematisierung im Self-Tracking. In: Selin Gerlek/Sarah Kissler/Thorben Mämecke et. al. (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktionen in digitalen Kulturen. Hagen: Hagen University Press 2022 (Digitale Kultur 1), S. 126-137.
- Okojie, Irenosen: Introduction. In: K Allado-McDowell: Pharmako-AI. United Kingdom: Ignota 2020, S. VII-X.
- Otte, Ralf: Künstliche Intelligenz für dummies. Weinheim: Wiley-VCH 2019 (Lernen einfach gemacht).
- Passig, Kathrin: Vielleicht ist das neu und erfreulich. Technik. Literatur. Kritik. Graz/Wien: Literaturverlag Droschl 2019 (Grazer Vorlesungen zur Kunst des Schreibens 2).
- Pohle, Stephanie: Uncreative Writing. Eine Befreiung des Schreibens im digitalen Raum? In: Weertje Willms/Martina Backes (Hg.): Kontexte kreativen Schreibens. Eine Standortbestimmung in Theorie und Praxis. Berlin: Frank & Timme 2021, S. 161-197.
- Reckwitz, Andreas: Die Erfindung der Kreativität. Zum Prozess gesellschaftlicher Ästhetisierung. Berlin: Suhrkamp 2012 (suhrkamp taschenbuch wissenschaft 1995).
- Renner, Karl Nikolaus: Grenze und Ereignis. Weiterführende Überlegungen zum Ereigniskonzept von Jurij M. Lotman. In: Frank Gustav/Wolfgang Lukas (Hg.): Norm – Grenze – Abweichung. Kultursemiotische Studien zu Literatur, Medien Wirtschaft. Festschrift für Michael Titzmann. Passau: Stutz 2004, S. 357-381.
- Ries, Thorsten: Digitale Literatur als Gegenstand der Literaturwissenschaft. Ein multimodales Forschungsprogramm. In: TEXT+KRITIK. Digitale Literatur II. Vol. Sonderband (2021), S. 24-34.

- Römer, Stefan: Künstlerische Strategien des Fake. Kritik von Original und Fälschung. Köln: DuMont 2001.
- Rosenbloom, Paul S.: Rethinking the Physical Symbol Systems Hypothesis. In: Artificial General Intelligence Vol. 13921 (2023), S. 207-216.
- Schaffrick, Matthias: Autorsubjekt und Werkherrschaft. In: Wetzel, Michael (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 203-222.
- Schmidinger, Heinrich: Das Auszeichnende des Menschen ist seine Kreativität – Eine geistesgeschichtliche Einstimmung. In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein kreatives Wesen? Kunst – Technik – Innovation. Darmstadt: Wissenschaftliche Buchgesellschaft 2008 (Topologien des Menschen 5), S. 7-24.
- Schönthaler, Philipp: Die Automatisierung des Schreibens & Gegenprogramme der Literatur. Berlin: Matthes & Seitz 2022 (Leben und Dienste 3).
- Schroeter, Matthias L.: Dekolonialisiert euren Geist! Renaturalisiert die Intelligenz! In: Yasemin Keskintepe/Dr. Anke Woschek (Hg.): Künstliche Intelligenz. Maschinen lernen Menschheitsträume. Göttingen: Wallstein 2021, S. 50-57.
- Schröter, Jens: Autorschaft aus dem Blickwinkel der Akteur-Netzwerk-Theorie. In: Wetzel, Michael (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 625-631.
- Schwering, Gregor: Autorschaft und Medien. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 585-602.
- Searle, John R.: Minds, brains and programs. In: The Behavioral and Brain Sciences Vol. 3/3 (1980), S. 417-457.
- Sedmak, Clemens: „Homo est animal rationale“. In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein „animal rationale“? Vernunft – Kognition – Intelligenz. Darmstadt: Wissenschaftliche Buchgesellschaft 2004 (Topologien des Menschen 1), S. 12-15.
- Seele, Peter: Künstliche Intelligenz und Maschinisierung des Menschen. Köln: Herbert von Halem 2020 (Schriften zur Rettung des öffentlichen Diskurses 1).
- Simanowski, Roberto: Lesen, Sehen, Klicken: Die Kinetisierung Konkreter Poesie. In: Harro Segeberg/Simone Winko (Hg.): Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm fink 2005, S. 161-177.
- Söffner, Jan: Sinn und Sinnlosigkeit. Die Frage nach der Stellung der Hermeneutik im Zeitalter der künstlichen Intelligenz. In: Kablitz, Andreas/Christoph Marksches/Peter Strohschneider (Hg.): Hermeneutik unter Verdacht. Berlin/Boston: Walter de Gruyter 2021 (Text und Textlichkeit 2), S. 1-21.
- Theisohn, Philipp: Plagiat. Eine unoriginelle Literaturgeschichte. Stuttgart: Alfred Kröner 2009 (Kröners Taschenausgabe 351).
- Tofan, Romana: Deepfakes. In: Öffentliche Sicherheit Vol. 3/4 (2024), S. 17-18.
- Trawny, Peter: Vorwort des Herausgebers. In: Norbert Wiener: Mensch und Menschmaschine. Herausgegeben und mit einem Vorwort von Peter Trawny. Frankfurt am Main: Vittorio Klostermann 2022, S. 9-20.

- Turing, A. M.: Computing Machinery and Intelligence. In: Mind Vol. 59/236 (1950), S. 433-460.
- Welsch, Wolfgang: Wandlungen im humanen Selbstverständnis. In: Heinrich Schmidinger/Clemens Sedmak (Hg.): Der Mensch – ein „animal rationale“? Vernunft – Kognition – Intelligenz. Darmstadt: Wissenschaftliche Buchgesellschaft 2004 (Topologien des Menschen 1), S. 48-70.
- Wetzel, Michael: Einleitung. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 1-75.
- Wetzel, Michael: Historischer Abriss. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 77-197.
- Wiener, Norbert: Kybernetik. In: Physikalische Blätter Vol. 5/8 (1949), S. 355-362.
- Wiener, Oswald /Thomas Eder: Literatur, Sprache, Denken: Die Anfänge. In: Thomas Eder/Thomas Raab/Michael Schwarz (Hg.): Oswald Wieners Theorie des Denkens. Gespräche und Essays zu Grundfragen der Kognitionswissenschaft. Berlin/Boston: Walter de Gruyter 2023, S. 11-77.
- Wiener, Oswald/Michael Schwarz: Pleomorphismus im Denken und die Computer-Metapher. Weiser, Symbol und das „Grounding“-Problem. In: Thomas Eder/Thomas Raab/Michael Schwarz (Hg.): Oswald Wieners Theorie des Denkens. Gespräche und Essays zu Grundfragen der Kognitionswissenschaft. Berlin/Boston: Walter de Gruyter 2023, S. 111-179.
- Wiener, Oswald: Schriften zur Erkenntnistheorie. Wien: Springer 1996 (Computerkultur X).
- William K. Wimsatt/Monroe C. Beardsley: Der intentionale Fehlschluss. In: Fotis Jannidis/Gerhard Lauer/Matias Martinez et. al. (Hg.): Texte zur Theorie der Autorschaft. Stuttgart: Reclam 2000, S. 84-101.
- Winko, Simone: Hyper – Text – Literatur. Digitale Literatur als Herausforderung an die Literaturwissenschaft. In: Harro Segeberg/Simone Winko (Hg.): Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm fink 2005.
- Wirth, Uwe: Dimensionen der Funktion ‚Herausgeber‘. In: Michael Wetzel (Hg.): Grundthemen der Literaturwissenschaft: Autorschaft. Berlin/Boston: Walter de Gruyter 2022 (Grundthemen der Literaturwissenschaft), S. 293-305.

7.3 Internetquellen

- AWS: Was sind autoregressive Modelle? <https://aws.amazon.com/de/what-is/autoregressive-models/> (Zugriff: 18.11.2024).
- Blackthorn, Gem: Carl Jung and the New Age Movement. In: Blackthorn (2023) <https://www.gem-blackthorn.com/post/carl-jung-and-the-new-age-movement> (Zugriff: 12.07.2024).
- Bundesministerium der Justiz: Künstliche Intelligenz und Urheberrecht. Fragen und Antworten. https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Nav_Themen/240305_FAQ_KI_Urheberrecht.pdf?__blob=publicationFile&v=2#:~:text=Auch%20die%20Bearbeitung%20und%20Umgestaltung,kann%20dieser%20frei%20verwendet%20werden. (Zugriff: 07.01.2025).
- CEP & LSE: Global Apollo Programme Press Release: Apollo Programme: Top global scientists and business leaders back Paris clean energy plan. In: Centre for Economic Performance (2015) https://cep.lse.ac.uk/_NEW/NEWS/abstract.asp?index=5211 (Zugriff: 18.11.2024).

Die Presse-Redaktion: Bekenntnis bei Preisvergabe: Autorin ließ ChatGPT für sich schreiben. In: Die Presse (2024) <https://www.diepresse.com/18004351/bekenntnis-bei-preisvergabe-autorin-liess-chatgpt-fuer-sich-schreiben> (Zugriff: 27.07.2024).

Drogenlexikon: Ayahuasca. In: drugcom.de (2014) <https://www.drugcom.de/drogenlexikon/buchstabe-a/ayahuasca/> (Zugriff: 15.07.2024).

Elektronik Kompendium: Kryptografische Zufallsgeneratoren, <https://www.elektronik-kompendium.de/sites/net/1910301.htm> (Zugriff: 06.01.2025).

Google Trends: Wolfenstein, <https://trends.google.com/trends/explore?date=2005-01-01%202021-01-09&q=wolfenstein&hl=de> (Zugriff: 10.06.2024).

Hanfeld, Michael: Google wirft KI-Experten raus. Schalte mich bitte nicht ab! In: Frankfurter Allgemeine Zeitung (2022) <https://www.faz.net/aktuell/feuilleton/medien/google-entlaesst-softwareingenieur-blake-lemoine-18195598.html> (Zugriff: 19.10.2023).

Höhne, Sebastian: Behaviorismus. In: Lernpsychologie. Lernen und Gedächtnis – einfach erklärt (2015) <http://www.lernpsychologie.net/lerntheorien/behaviorismus> (Zugriff: 04.04.2024).

Ignota: ‘Pharmako-AI’ Book Launch: K Allado-McDowell, Nora N. Khan and Emily Segal (2020) <https://www.library.ignota.org/event/pharmako-ai-book-launch> (Zugriff: 12.07.2024).

Ignota: Pharmako-AI (2020) <https://www.library.ignota.org/book/pharmako-ai> (Zugriff: 12.07.2024).

Johannes Edelhoff/Esra Özer/Tina Soliman: Streit um Urheberrechte. Müssen Künstler und Autoren für KI entschädigt werden? In: tagesschau (2023) <https://www.tagesschau.de/wirtschaft/digitales/kuenstliche-intelligenz-grafiker-autoren-urheberrechte-100.html> (Zugriff: 21.09.2023).

Katzlberger, Michael: KI Sprachmodelle. Mit diesen Daten wurde GPT-3 trainiert. In: Artificial Creativity (2022) <https://katzlberger.ai/2021/04/12/mit-diesen-daten-wurde-gpt-3-trainiert/> (Zugriff: 10.06.2024).

Knibbs, Kate: The Battle Over Books3 Could Change AI Forever. In: Wired (2023) <https://www.wired.com/story/battle-over-books3/> (Zugriff: 11.06.2024).

Marc Stöckel: Urheberrechte. US-Gericht lehnt Copyright auf KI-generierte Kunst ab. In: golem.de (2023) <https://www.golem.de/news/urheberrechte-us-gericht-lehnt-copyright-auf-ki-generierte-kunst-ab-2308-176918.html> (Zugriff: 21.09.2023)

Marcus, Gary: What was 60 Minutes thinking, in that interview with Geoff Hinton? In: Marcus on AI (2023) <https://garymarcus.substack.com/p/what-was-60-minutes-thinking-in-that> (Zugriff: 12.10.2023).

Meta: Introducing New AI Experiences Across Our Family of Apps and Devices. In: Meta (2023) <https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/> (Zugriff: 18.11.2024).

Meta: Introducing new AI experiences from Meta, <https://ai.meta.com/genai/> (Zugriff: 18.10.2023).

netkin Redaktion: Cringe. In: netkin (2023) <https://www.netkin.de/glossary/cringe/> (Zugriff: 18.11.2024).

OpenAI: GPT-4 is OpenAI’s most advanced system, producing safer and more useful responses, <https://openai.com/gpt-4> (Zugriff: 10.10.2023).

OpenAI: Introducing ChatGPT. In: OpenAI (2022) <https://openai.com/blog/chatgpt> (Zugriff: 09.10.2023).

OpenAI: OpenAI Charter, <https://openai.com/charter> (Zugriff: 16.10.2023).

OpenAI: Pioneering research on the path to AGI, <https://openai.com/research/overview> (Zugriff: 16.10.2023).

Pelley, Scott: 60 Minutes. “Godfather of Artificial Intelligence” Geoffrey Hinton on the promise, risks of advanced AI. In: CBS News (2023) <https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/> (Zugriff: 12.10.2023).

Priyanka: GPT 3 vs. GPT 4: Was Sie wissen müssen. In: GetGenie (2024) <https://getgenie.ai/de/gpt-3-vs-gpt-4/> (Zugriff: 10.06.2024).

Römer, Stefan: Der Begriff des Fake. In: Humboldt-Universität zu Berlin (2024) <https://edoc.hu-berlin.de/handle/18452/15158> (Zugriff: 16.05.2024).

Sam Altman: Planning for AGI and beyond. In: OpenAI (2023) <https://openai.com/blog/planning-for-agi-and-beyond> (Zugriff: 13.10.2023).

Schmidt, Holger: Elon Musk. Sein Kampf um die Zukunft der Künstlichen Intelligenz. In: Frankfurter Allgemeine Zeitung (2023) <https://www.faz.net/aktuell/technik-motor/wie-elon-musk-einst-open-ai-mitgruendete-und-nun-seinen-eigenen-weg-sucht-19170650.html> (Zugriff: 12.10.2023).

Solis, Tobias: Das sind die Grenzen von ChatGPT. In: Scribbr (2023) <https://www.scribbr.de/ki-tools-nutzen/chatgpt-grenzen/> (Zugriff: 11.06.2024).

Standard-Redaktion: In Videospielen dürfen nun auch Hakenkreuze gezeigt werden. In: DerStandard (2018) <https://www.derstandard.de/story/2000085105174/in-videospielen-duerfen-nun-auch-hakenkreuze-gezeigt-werden> (Zugriff: 10.06.2024).

Staun, Harald: Schriftsteller probiert es aus. Das kommt dabei heraus, wenn KI einen Roman schreiben soll. In: Frankfurter Allgemeine Zeitung (2023) <https://www.faz.net/aktuell/feuilleton/buecher/rezensionen/belletristik/herrlich-gescheitert-hannes-bajohrs-ki-roman-berlin-miami-19286704.html#abschluss> (Zugriff: 07.01.2025).

Weiß, Eva-Maria: Modellkollaps – wie synthetische Daten KI killen können. In: heise online (2024) <https://www.heise.de/news/Modellkollaps-wie-synthetische-Daten-KI-killen-koennen-9814434.html> (Zugriff: 30.07.2024).

Wirtz, Markus Antonius: „Intelligenz, emotionale“ In: Dorsch – Lexikon der Psychologie (2021) <https://dorsch.hogrefe.com/stichwort/intelligenz-emotionale> (Zugriff: 04.04.2024).

Wuttke, Laurenz: Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion (2024) <https://datasolut.com/neuronale-netzwerke-einfuehrung/> (Zugriff: 18.10.2023).

7.4 Videos

Allado-McDowell, K: Our creative relationship with AI is just beginning. In: TED (2023) https://www.ted.com/talks/k_allado_mcdowell_our_creative_relationship_with_ai_is_just_beginning [1:35-1:39] (Zugriff: 28.05.2024).

7.5 Podcasts

Anderl, Sibylle/Jachim Müller-Jung: Wie lustig ist Künstliche Intelligenz? In: F.A.Z. Wissen – Der Podcast (2022) <https://open.spotify.com/episode/51Sq0HzPhZVsjh0KCwa0h?si=0ddfa19631ea44a5> (Zugriff: 19.07.2024).

7.6 Weiterführende Literatur

0x0a: texte, <https://0x0a.li/de/texte/> (Zugriff: 11.06.2024).

Bächle, Thomas Christian: Das digitale Selbst. In: Martina Heßler/Kevin Liggieri (Hg.): Technikanthropologie. Handbuch für Wissenschaft und Studium. Baden-Baden: Nomos 2020, S. 283-288.

Google: Artists + Machine Intelligence, <https://ami.withgoogle.com/> (Zugriff: 05.06.2024)

Grunwald, Armin (Hg.): Wer bist du, Mensch? Transformation menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt. Freiburg/Basel/Wien: Herder 2021.

Henrickson, Leah: Constructing the Other Half of The Policeman's Beard. In: The Electronic Book Review (2021) <https://electronicbookreview.com/essay/constructing-the-other-half-of-the-policemans-beard/> (Zugriff: 19.11.2024)

Kindt, Tom: Komik. In: Uwe Wirth (Hg.): Komik. Ein interdisziplinäres Handbuch. Stuttgart: J. B. Metzler 2017, S. 2-6.

Kreuzmair, Elias: Die Zukunft der Gegenwart (Berlin, Miami). Über die Literatur der ›digitalen Gesellschaft‹. In: Text+Kritik. Digitale Literatur II Sonderband (2021), S. 35-46.

Lighthill, James: Artificial Intelligence: A General Survey (1972) https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf (Zugriff: 21.11.2023).

McCulloch, Warren S./Walter Pitts: A Logical Calculus of the Ideas Immanent in Nervous Activity. In: Bulletin of Mathematical Biology Vol. 52/1/2 (1990), S. 99-115.

Thomas, Iain S./Jasmine Wang/GPT-3: Was euch zu Menschen macht. Antworten einer künstlichen Intelligenz auf die großen Fragen des Lebens. Aus dem Amerikanischen von Judith Elze. München: Diederich 2022.

Weizenbaum, Joseph: ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine. In: Communications of the ACM Vol. 9/1 (1966), S. 36-45.

Ziaja, Ursula: Was ist Kreativität. Begriffsbestimmungen aus der Kreativitätsforschung. In: Weertje Willms/Martina Backes (Hg.): Kontexte kreativen Schreibens. Eine Standortbestimmung in Theorie und Praxis. Berlin: Frank & Timme 2021, S. 25-37.

8 Abbildungsverzeichnis

Abbildung 1: KI-generiertes Bild zum Roman „(Berlin, Miami)“ der FAZ109

Harald Staun: Schriftsteller probiert es aus. Das kommt dabei heraus, wenn KI einen Roman schreiben soll. In: Frankfurter Allgemeine Zeitung (2023)
<https://www.faz.net/aktuell/feuilleton/buecher/rezensionen/belletristik/herrlich-gescheitert-hannes-bajohrs-ki-roman-berlin-miami-19286704.html#abschluss> (Zugriff: 07.01.2025).

Abbildung 2: Die vier Grade Kausaler Autorschaft111

Bajohr, Hannes: Writing at a Distance: Notes on Authorship and Artificial Intelligence. In: German Studies Review Vol. 47/2 (2024b), S. 323.

Abbildung 3: Erweitertes Modell Kausaler Autorschaft114

Änderungen von Verfasserin in rot. | Ursprüngliche Abbildung s. Bajohr, Hannes: Writing at a Distance: Notes on Authorship and Artificial Intelligence. In: German Studies Review Vol. 47/2 (2024b), S. 323.

9 Anhang

9.1 Abstract

Die vorliegende Masterarbeit beschäftigt sich mit der Frage, wie die zunehmende Integration Künstlicher Intelligenz (KI) in Kunst und Literatur traditionelle Vorstellungen von Autorschaft und Kategorien wie Kreativität und Originalität infrage stellt und ob eine Maschinenautorschaft überhaupt denkbar ist. Es wird untersucht, inwiefern die neue Technologie literaturtheoretische Veränderungen anstößt und welche Herausforderungen sie mit sich bringt. Der Fokus liegt dabei auf der Technologie der generativen Sprachmodelle wie ChatGPT, GPT-J und GPT-NeoX. Es werden klassische wie moderne literaturwissenschaftliche Theorien zu Autorschaft vorgestellt und im Kontext Künstlicher Intelligenz diskutiert. Ausgehend von poststrukturalistischen Ansätzen und der Akteur-Netzwerk-Theorie wird analysiert, wie die Digitalisierung bereits veränderte Formen des Schreibens und neue Konzeptionen von Autorschaft hervorgebracht hat. Im Zentrum der Arbeit stehen die KI-generierten Werke „*Pharmako-AI*“, „*Amor Cringe*“ und „*Air Age Blueprint*“ von Kenric Allado-McDowell und „*(Berlin, Miami)*“ von Hannes Bajohr, die unterschiedliche Formen der Mensch-Maschine-Kollaboration reflektieren. Anhand dieser wird aufgezeigt, ob KI als kreative Instanz instrumentalisiert werden kann oder lediglich bestehende ästhetische Konzepte reproduziert. Die Masterarbeit leistet einen Beitrag zur theoretischen Positionierung Künstlicher Intelligenz und ihrer Autorschaftskonzeptionen in den Geisteswissenschaften und zur Debatte um neue Formen des Schreibens im digitalen Zeitalter.