

Hannes Bajohr

Autorschaft und Künstliche Intelligenz

Computergenerierte Sprache ist fast so alt wie der Computer selbst. Nicht nur lässt sich das *natural language processing* (NLP, Nilsson 2010, 103–113) bis in die 1940er Jahre zurückverfolgen, wo es seit Warren Weavers Memorandum über *machine translation* vor allem im Bereich automatisierter Übersetzung der jungen Informatik ein vielversprechendes Forschungsfeld eröffnete (siehe für das gemeinhin als „Geburtsstunde“ der Maschinenübersetzung bezeichnete „Georgetown-IBM experiment“ (Gordin 2015, 213–240; Weaver 2003 [1949])). Die Frage nach der Autorschaft eines solchen Textes spielte auch im sehr viel breiteren Feld der Künstlichen Intelligenz seit Alan Turings „imitation game“ von 1950 (dem später so genannten „Turing-Test“) eine uneingestandene Hauptrolle: Um die Intelligenz von Maschinen behavioristisch operationalisierbar zu machen, so Turing, müsste ein Computer dann als intelligent gelten, wenn ein Mensch ihn in einem Gespräch selbst für einen Menschen hielte (Turing 1950). Da dieses Gespräch aus technischen Gründen textlich vermittelt sein musste, heißt das, in der Formulierung von Jay David Bolter: „artificial intelligence is the art of making texts“ (1991, 180). Insofern dieser Text auf Intelligenz verweisen soll, die zumindest im Verhaltensoutput intentional erscheinen muss, war Autorschaft also von Anbeginn an in aller KI-Forschung angelegt.

Dass dieses Potenzial bis in die jüngere Gegenwart brach lag, liegt daran, dass die Textproduktion von KI vor allem extraliterarisch angesiedelt war. Die User von Joseph Weizenbaums ELIZA-Chatbot etwa akzeptierten den Computer zwar wider besseres Wissen als intentionalen Kommunikationspartner (Weizenbaum 1966), aber nicht als Autor in der literaturtheoretischen Bedeutung des Wortes. Die Herstellung explizit literarischer Texte – und der dazugehörigen Auszeichnung einer spezifisch literarischen „Seinsweise des Diskurses“ sowie der „klassifikatorischen Funktion“ von Werkkohärenz und geistigem Eigentum (Foucault 2000, 210), für die der Begriff der Autorschaft normalerweise in seiner emphatischen Funktion reserviert ist – stand nicht im Interesse des Mainstreams der Entwicklung Künstlicher Intelligenz, sondern war eher in den künstlerisch-experimentalen Randbereichen des NLP angesiedelt. Im Folgenden sollen daher – auch weil die Grenze zwischen KI und bloßer Computerverwendung oft schwer zu ziehen ist – zunächst die historischen Debatten um Autorschaft computergenerierter literarischer Texte zusammengefasst werden, bevor systematisch die Autorschaft speziell Künstlicher Intelligenz Thema wird; in vielerlei Hinsicht ist sie lediglich eine Fortführung des älteren Diskurses, setzt aber in ihrer systematischen und ideologischen Ausgestaltung spezifische Akzente.

1 Historischer Abriss computergenerierter Literatur

Christopher Strachey, Kollege Turings an der Universität Manchester, gilt gemeinhin als Programmierer des ersten literarischen Experiments mit einem Computer. Seine 1952 auf dem Manchester Mark I ausgeführten und über einen Fernschreiber ausgedruckten „Love Letters“ funktionierten nach einem kombinatorischen Prinzip, so dass ein fester Satz an Wörtern in eine Reihe möglicher Wortschablonen für einen Liebesbrief eingefügt wurde (Link 2016; Wardrip-Fruin 2009). Gezeichnet waren diese Briefe stets mit „M.U.C.“ (Manchester University Computer). Bereits hier ist der Computer als Autor vorgestellt, aber eher im Sinne einer spielerischen Autorschaftsfiktion, die sich kaum von jener etwa eines Kater Murr unterscheidet. Strachey behauptete nie, der Computer übernehme dabei tatsächlich die Rolle eines Autors, sondern bestand im Gegenteil darauf, dass es sich hier um eine anthropomorphe Projektion handele. Die Frage sei nicht, ob Computer Briefe zu verfassen in der Lage sind, sondern ob man ein Programm zu diesem Zweck schreiben kann – „and this is really a very different sort of question because the writing of programs [...] is still an essentially human activity“ (Strachey 1954, 25–26). Autorschaft liegt hier eindeutig bei den Programmierenden, nicht beim Computer. Wie Turing aber hielt Strachey die Möglichkeit dieser Täuschung – und damit implizit auch die Annahme von Autorschaft – nicht nur für nicht ausgeschlossen, sondern für in der Zukunft auch wahrscheinlich (Strachey 1954, 31).

Die Informationsästhetik der Stuttgarter Schule um Max Bense verkomplizierte dieses Bild, indem sie die ontologischen Vorannahmen thematisierte, die solchen Aussagen über Autorschaft zugrunde liegen. Betonte Strachey, parallel zu Turings Vorschlag, die Erscheinung von Intentionalität als Voraussetzung von Autorschaft – und empfahl geschickte Täuschung per Programmierung –, artikulierte die Informationsästhetik einen Bruch in der Ontologie des Textes überhaupt. Bense unterschied zwischen „natürlicher“ und „künstlicher Poesie“ und sprach ersterer notwendig „eine Ichrelation und einen Weltaspekt“ zu, die als „ontologische Fortsetzung“ personaler oder sozialer Welterfahrung im Text fortwirkten. „Künstliche Poesie“ dagegen habe grundsätzlich keinen intentionalen und subjektrelativen, sondern nur einen „materiale[n] Ursprung“; sie bestehe aus kontextlosen festgelegten Textelementen, deren Kombination keine menschliche Sinnerfahrung ausdrücke, deren ästhetischer Gehalt aber unabhängig eines solchen humanen Maßstabs als statistische Unwahrscheinlichkeit von Zeichenfolgen gemessen werden kann (Bense 1962, 143). So gesehen wird die Frage nach Autorschaft zugunsten textimmanenter Kriterien abgewertet; computergenerierte Texte sind prinzipiell nicht-intentional, der Turing-Test demnach schlicht ein Kategorienfehler. Dass Bense von „simulierter

Poesie“ sprechen konnte (Bense 1998, 385), betont diese ontologische Differenz eher, statt sie aufzuheben und rückt seine Texttheorie in die Nähe des Diskurses um den „Tod des Autors“ (Barthes 2000).

Der Fokus auf die quantitativen Eigenschaften von Text, etwa relative Worthäufigkeiten, äußert sich auch in den „Stochastischen Texten“, die Benses Schüler Theo Lutz 1959 produzierte und die als erste computergenerierte Literatur im deutschen Sprachraum gelten. In ihnen wird eine Wortliste nach einem gewichteten Zufallsprozess kombiniert; dass das verwendete Vokabular Kafkas *Schloß* entstammte, widerspricht der Abwertung der Autorschaftsfrage nicht, sondern soll eine inhärente ästhetische Wertigkeit sicherstellen (Lutz 1959, 1960). Lutz allerdings – weniger texttheoretisch interessiert als Bense – scheint seine eigene Autorschaft vor allem auf das Programmskript bezogen und empathisch eingefordert zu haben, während er sich den Output nur sekundär zugeordnet hat (Bernhart 2020, 194). Wie bei Strachey ist Lutz' Arbeit im Bereich der NLP, nicht der KI anzusiedeln; explizit deklariert er sie als synthetisches Gegenstück einer „analytischen Philologie“ (so auch Bense 1998, 384; Lutz 1960, 11; 14). Ähnliches lässt sich auch von den ab 1964 generierten „Autopoemen“ Gerhard Stickels feststellen, die nach einem vergleichbaren Muster funktionieren. Und auch der Linguist Stickel hält die Annahme einer Autorschaft des Computers für eine unzulässige Anthropomorphisierung, „die nur durch die Unkenntnis seiner Arbeitsweise begründet ist“ (Stickel 1966, 123).

Nicht die Suche nach autonom agierenden, problemlösenden oder gar bewusstseinsäquivalenten Maschinen, sondern die Analyse syntaktischer Strukturen erscheint also als Startpunkt computergenerierter Literatur. Dieser Umstand deprivilegiert die Autorschaftsfrage ganz pragmatisch (im Fall von Lutz und Stickel) oder durch einen starken Textbegriff auf theoretischer Ebene (wie im Fall von Bense). Wo der linguistische oder informatische vermehrt einem künstlerischen Zugriff weicht, scheint es, als hätte in vielen Fällen das poetologische Interesse am „Tod“ des menschlichen Autors auch sein Wiederauferstehen als Computerautor verhindert (Schäfer 2017, 29) – jedenfalls wurden die generativen Texte der 1960er und 1970er Jahre eher im Kontext modernistischer Ästhetiken praktiziert, die in der Nachfolge John Cages und Marcel Duchamps mit aleatorischen Prozessen experimentierten, statt Autorschaft auf Maschinen zu projizieren. Bei den einschlägigen Werken dieser Zeit findet sich jedenfalls kaum etwas davon; im Falle von Alison Knowles' und James Tenneys „A House of Dust“ (1968) ist immerhin die Kollaboration zwischen Kunst (Knowles) und Ingenieurwesen (Tenney) auch im Autor:innenduo reflektiert (H. B. Higgins 2012). Doch wenn der wie Knowles zur Fluxus-Bewegung gehörende Dick Higgins für den Einsatz von „computers for the arts“ plädiert, so weiterhin unter der Maßgabe: „Computers are like most tools – deaf, blind and incredibly stupid.“ (D. Higgins 1970, 1)

Wo es zu einer bewussten Inszenierung von Autorschaft kommt, wird sie zunächst kaum sehr weit getrieben: Für die Sammlung von *Computer-Lyrik* (1968) zeichneten die Programmierer Manfred Krause und Götz F. Schaudt lediglich als Herausgeber verantwortlich, gestehen damit der generierenden Maschine paratextuell Autorschaft zu; im Vorwort aber wird diese Geste wieder zurückgenommen und es ist vom Computer als „nur ein[em] Werkzeug in unserer Hand“ die Rede (Krause und Schaudt 1969, 8). Der Dichter Karl Krolow, der den Band rezensiert, nimmt dennoch den Ball auf und spricht dem „Zuse Z 23, einjährig alter Computer-Dichter“ parodistisch-ernsthaft Autorschaft zu (Krolow 1968, 33), was die Inszenierung aber nur als durchschaute bestätigt. In der Tat ist die Diskrepanz zwischen Produktion und Rezeption von Computerliteratur auffällig: Der kritischen Rezeption dient sie als Projektionsfläche von Mensch-Maschine-Vergleichen – sei es als Folie einer irreduzibel menschlichen Fähigkeit zu ‚wirklicher‘ Literatur, wie Italo Calvino sie in der Verwurzelung jeder Geschichte im Mythos lokalisiert (Calvino 1984), sei es als bloßes Eichmaß des Handwerks, so wie Hans Magnus Enzensberger 1974 schreibt: „Wer nicht besser dichten kann als die Maschine, der täte besser daran, es bleiben zu lassen.“ (Enzensberger 2000, 52) Trotz dieser wiederholten Einreihung in moderne- und technikkritische Tropen (Rieger 2018; Schönthaler 2022) blieb die Diskussion um Autorschaftsfragen für die solche Literatur Produzierenden auffallend sekundär.

Das änderte sich in den 1980er Jahren mit der Veröffentlichung des Buchs *The Policeman's Bear is Half Constructed* (1984), bei der die Inszenierung von autonomer Maschinenautorschaft in ein bislang ungekanntes Extrem getrieben wird: Auf dem Umschlag angepriesen als „the first book ever written by a computer“, wird als Autor das Programm „Racter“ geführt (Racter 1984). Die aufwändige, mit Collagen versehene Buchausgabe steht nun nicht mehr für die als solche durchsichtige Inszenierung, sondern die ernsthafte Behauptung von Computerautorschaft. William Chamberlain und Thomas Etter, die Entwickler von Racter, setzten sich durch die Geheimhaltung des Programmcodes und der editorischen Eingriffstiefe einiger Kritik aus; der Nachweis von Autorschaft auf kausaler Ebene blieb aus, weshalb *The Policeman's Beard* heute als äußerst umstrittenes Schlüsselwerk generativer Literatur bezeichnet wird (Henrickson 2021). Aufschlussreich ist, dass der Klappentext Racter von KI absetzt: „Fundamentally different from artificial intelligence programming, which tries to replicate human thinking, Racter can write original work without promptings from a human operator.“ Autonome Maschinenautorschaft wird hier als das Paradigma von KI übersteigend deklariert, was nicht zuletzt unter dem Gesichtspunkt eines in den 1980er Jahren wieder an Fahrt aufnehmenden KI-Diskurses zu sehen ist.

In der Tat waren die 1980er eine Art Scheideweg: Die Entwicklung von text-basierter KI und generativem Schreiben verlief lange parallel. Weizenbaums ELIZA

hatte gezeigt, dass ein Computer den Turing-Test – kausale Autorschaft als anthropomorphisierende Zuschreibung – zumindest für kurze Zeit bestehen konnte. Literarisch aber nahm das Interesse an Textgeneratoren Ende der 1980er Jahre und in den 1990er Jahren ab. Stattdessen kam mit der Hypertextliteratur eine gänzlich andere Form computerzentrischen Schreibens auf, die nicht die Art der Produktion, sondern Modi der Rezeption privilegierte und das Lesen auf „Pfade, die sich verzweigen“ schickte (Borges 2000). Im Gegensatz zum frühen NLP, das fast gänzlich literaturtheoriefern war, kann die Hypertextliteratur geradezu als Schrift gewordener Poststrukturalismus gelten: Da Hypertextstrukturen multiple Lesepfade erlauben, ersetzen sie nicht nur den autoritären „Baum“ durch das anarchische „Rhizom“ (Deleuze und Guattari 1992, 11–42); sie entzögen, wie Florian Hartling entsprechend Roland Barthes' Autorschaftsmodell schreibt, auch Autor:innen ihre Macht, die nunmehr *scripteurs* von Materialkollektionen seien (Barthes 2000; Hartling 2009, 97–109), und werteten die Leser:innen als die eigentliche bedeutungsschaffende Instanz auf. Vorgeschlagen wurden neue Figuren wie der „Wreader“ (Landow 1997, 14) oder die „Browser“ (Guertin 2007; siehe dazu Meerhoff 2021, 97–109), die sich die eigenen, nichtlinearen Lesewege durch verlinkte Textabschnitte schlagen (dass gerade der Autorschaftsbegriff eben nicht verschwand, dazu Coover 2001 und Porombka 2001; kritisch auch Simanowski 2001). Der „Vernetzungsaspekt“ (Heibach 2001, 32) und die „Interaktivität, Intermedialität und Inszenierung“ (Simanowski 2001, 4) der Rezeption verstellten tendenziell den Blick für die Bedingungen der Produktion, und für längere Zeit war „Netzliteratur“ verwirrenderweise der Oberbegriff für jede computerbasierte und also auch generative Literatur, auch wenn sie mit dem Netz nichts zu tun hatte. Für die gegenwärtige Diskussion um Autorschaft von KI ist aber der seinerzeit neuartige Fokus auf Kollaboration wieder zentral: So wie verschiedene Schreiber kooperieren, so kann man auch das Verhältnis von Mensch und Maschine als kooperativ, als „Cyborg“-Autorschaft bezeichnen (Aarseth 1997, 132–134); gegenwärtige Ansätze nehmen vermehrt diesen Aspekt auf (Bajohr 2022; Henrickson 2018).

2 Kausale Autorschaft

Dieser Überblick über den Diskurs um maschinelle Autorschaft zeigt zunächst, wie NLP- und KI-Forschung sowie künstlerische Ansätze ganz verschiedene Interessen artikulieren: Setzt die KI-Forschung mit ihrer Suche nach Intelligenz oder Rationalität (wenn nicht gar Bewusstsein) Autorschaft als intentionale Kategorie weiterhin zumindest implizit voraus, ist NLP eher an den Prozessen der Textsynthese interessiert, die, depersonalisiert, als System der *langue* ganz ohne *parole* gedacht werden

können. Künstlerische Experimente können sich bei beiden Seiten bedienen und scheinen vielmehr von ästhetischen und poetologischen Positionen als von technischen Überlegungen getrieben zu sein, so dass der Autorschaftsdiskurs mit der poststrukturalistisch beeinflussten Hypertextliteratur schließlich eine völlig andere technische Grundlage erhielt.

Die Vernachlässigung des Produktionsaspekts im literaturtheoretischen Diskurs der 1980er und 1990er Jahre übersah freilich, dass in dieser Zeit eine grundsätzliche Neuausrichtung der KI-Forschung vor sich ging, die, mit Unterbrechungen, bis heute einflussreich ist. Konzentrierte man sich seit der Begründung der *artificial intelligence* als Forschungsfeld auf der Dartmouth-Konferenz 1956 vor allem auf symbolische KI, schwenkte man in den 1980ern auf subsymbolische KI um (Mitchell 2019). Der Unterschied ist einer ums Ganze: Bei symbolischer KI geht es um die explizite Formulierung von Regelschritten sowie dem Codieren von Weltwissen in sogenannte Expertensysteme (Alpaydın 2021, 60–62). Diese Systeme werden *sequenziell* programmiert, sind in ihrer Struktur menschenlesbar und ihre Lernfähigkeit ist auf die Erweiterung ihrer Wissensdatenbank beschränkt. Subsymbolische KI aber, deren Grundlagen in den 1960er Jahren gelegt, aber erst in den 1980ern wieder aufgenommen wurden, kommt ohne explizite Regeln aus und basiert auf stochastischem maschinellem Lernen (Rumelhart et al. 1986). Man kann es das *konnektionistische* Paradigma nennen, da im *deep learning* vor allem künstliche Neuronale Netze (KNNs) zum Zuge kommen, die eine Input-Output-Funktion über signalverstärkende und -hemmende „Neuronen“-strukturen modellieren, und die abstrakt dem Hirn nachempfunden sind (Alpaydın 2021, 105–142; zur Trennung sequenziell/konnektionistisch, siehe Bajohr 2021a).

Dennoch dauerte es bis in die 2010er Jahre, bis KI als *deep learning* auch breitenwirksam für literarische und künstlerische Zwecke zum Einsatz kam. Seitdem hat sich ihre Mächtigkeit im Vergleich zu klassischen, auf Kombinatorik oder Aleatorik basierenden Textgeneratoren gezeigt. Zudem wird im *unsupervised learning* die Idee „intelligenter“ Maschinen sehr viel überzeugender realisiert, so dass in ihnen nicht nur die Autonomie der Wissensanwendung, sondern auch der Wissensakkumulation angelegt ist. Da gegenwärtige NLP-Anwendungen wie maschinelle Übersetzung ebenfalls als KNNs implementiert sind und zudem zunehmend selbst als „KI“ bezeichnet werden, ist die frühere Distanz zwischen beiden Feldern geschrumpft. „Große Sprachmodelle“ wie ChatGPT, Bard oder Claude werden nun als „foundation models“ für umfassendere KI-Systeme gehandelt (Bommasani et al. 2021).

Erste explizit literarische Experimente mit KNNs fanden wahrscheinlich Anfang der 2000er statt (Funkhouser 2012 berichtet von einem Projekt, das 2006 Rumi-Verse generieren sollte). Doch erst die Popularisierung der schon seit den 1990er Jahren bekannten LSTM-RNNs („long short-term memory recurrent neural networks“) um 2015 brachte jenes breite Feld an literarischen Experimenten hervor,

das bis heute floriert (siehe auch → Generative Literatur). Da hierbei die Qualität der Ausgabe (im Sinne von annähernder Natürlichsprachigkeit) immer besser wurde, ist auch die Frage nach der Autorschaft solcher Systeme wieder vermehrt in den Blick gerückt. Interessant ist, dass die technische Substruktur eine unmittelbare Auswirkung auf die Reihe möglicher Autorschaftskandidat:innen hat; es macht einen Unterschied, ob sequenziell oder konnektionistisch geschrieben wird. *Kausale Autorschaft* ist daher heute, noch vor einer spezifisch literarischen Autorfunktion, zu einer zentralen Forschungsfrage geworden. Mit Blick auf die „kausale“, produktionsorientierte Seite der Texterstellung lässt sich dann ein Schema von *Distanzgraden* an Autorschaft konstruieren (ein erster Entwurf findet sich in Bajohr 2021c, eine genauere Ausarbeitung folgt in Bajohr 2024) und im Übergang von konventionellem zu sequenziellem zu konnektionistischem Schreiben eine Zunahme einer solchen kausalen Distanz beobachten.

Primäre Autorschaft wäre der immer noch geltende Standard konventionellen, „unvermittelten“ Schreibens, in dem ein:e Autor:in mehr oder minder direkt Text zu Papier oder auf einen Datenträger bringt. Die Instanzen, die zwischen Schreibenden und Text stehen, mögen zwar technischer Natur sein – von Federkiel über Schreibmaschine zu Wordprozessor –, doch ist weniger der determinierende Einfluss von Materialität und Medien auf den finalen Text ausschlaggebend (Kirschenbaum 2016; Stingelin 2012) als eher die Idee einer kausal als ungebrochen wahrgenommenen Verbindung von Autor:in und Text; der Schreibakt wird völlig auf der Seite der Person lokalisiert, die die Apparatur des Schreibens bedient, und alle anderen beteiligten Aktanten sind lediglich passiv-vermittelnd involviert. Dass diese Annahme sich keineswegs in allen Schreibformen durchhält, zeigen die weiteren Distanzschritte, die diesen Nexus durch einen Hiatus in seiner relativen Unmittelbarkeit aufbrechen.

Sekundäre Autorschaft wäre demnach eine Distanz erster Ordnung, die den Schreibakt verdoppelt. Streng genommen taucht sie in jeglicher Kalküldichtung auf (Cramer 2011) – sowohl in den kombinatorischen Dichtungsexperimente des Barocks oder der historischen Avantgarden wie im sequenziellen Paradigma: In all diesen Fällen besteht der Autor:innenbeitrag in der Formulierung einer Regelfolge, deren Ausführung das Werk produziert. Der Akt, der diese Regelfolge hervorbringt, ist auf kausaler Ebene nicht vom Schreiben primärer Texte zu unterscheiden, übernimmt aber eine vikariierende Funktion und fungiert als Modul in einer Operationskette, das nach Ausführung des ersten Textes selbst einen weiteren, den sekundären Text ausgibt. Autorschaft heißt hier also, einen Text zu schreiben, der einen Text schreibt. Ausschlaggebend ist, dass der primäre Text operativen Charakter hat: Hierunter fallen auch alle der oben genannten digitalen Literaturen (wobei der Hypertext gewissen Einschränkungen unterliegt) sowie die Transformationsregeln symbolischer KI wie bei ELIZA. Im Unterschied zur historischen Kalküldichtung hat

der Text des Codes hier tatsächlich eine eigene Kausalität und Performativität – er produziert den finalen Text im Wortsinne. Durch diese performative Beziehung stellt sich die Frage nach einer möglichen Computerautorschaft erst, die bei den Regelfolgen „analoger“ Kalküle kaum aufkommt.

Dennoch ist es nicht immer notwendig, dass der operative Text unmittelbar menschengemacht ist; er kann selbst Produkt eines früheren Herstellungsschritts sein. So schiebt sich ein zweiter Hiatus im *machine learning* des konnektionistischen Paradigmas in diese Operationskette und bringt eine *tertiäre Autorschaft* hervor. Der Informatiker François Chollet hat den hier wesentlichen Unterschied folgendermaßen beschrieben: Ist das sequenzielle Paradigma nach dem Schema „Regeln + Daten = Antworten“ aufgebaut, folgt das konnektionistische dem Schema „Daten + Antworten = Regeln“ (Chollet 2021, 4). Es werden also nicht mehr Regeln in einem Programm niedergelegt, dessen Anwendung (auf Daten) einen Output produziert, sondern es wird ein KNN auf einen großen Satz beispielhaften Outputs trainiert, der erst jene „Regeln“ macht, die schließlich zum finalen Text führen. Autor:innen bleibt damit allein der Lernalgorithmus zu programmieren (der aber meist schlicht von Dritten übernommen wird), der Datensatz für das Training zu definieren (aus denen das KNN selbstständig das Modell bildet) und die Parameter zu bestimmen (mittels derer das Modell schließlich den Output hervorbringt). Der Text, der im sekundären Modell den Output schreibt, ist damit selbst computergemacht.

Trotzdem ist Autorschaft auch im tertiären Modell noch im Maschinenraum des Digitalen angesiedelt: Die Auswahl des Trainingsalgorithmus und der „Hyperparameter“ sowie die Kuratierung des Trainingsdatensatzes unterliegen weiterhin menschlicher Kontrolle und Auswahl. Mit dem Aufkommen großer Sprachmodelle wie GPT-3 (2020 von OpenAI entwickelt und seitdem mehrmals aktualisiert, siehe Brown et al. 2020; zuletzt ChatGPT, siehe OpenAI 2022) ändert sich auch das, und man nun kann plausibel von *quartärer Autorschaft* sprechen. Als proprietäre Software – die überdies zu umfangreich ist, um von Usern je aufs Neue trainiert zu werden – sind Benutzer:innen auf die werksseitig getroffene Auswahl von Trainingsdaten eingeschränkt (auch wenn bei GPT-3 inzwischen ein eingeschränktes *finetuning* möglich ist). Mehr noch: Die einzige Interaktion mit diesen Modellen geschieht über ihr Interface, und zwar über natürlichsprachliche Aufforderungen. Statt also den Code zu schreiben, der ein Gedicht schreibt, oder ein Sprachmodell auf eine Vielzahl von Gedichten zu trainieren, das dann neue ausgibt, ist im quartären Modell allein die Eingabe ausschlaggebend, etwa: „Schreibe ein Gedicht im Stil von Wallace Stevens“ (Brown et al. 2020, 48–49); im Fall von *Text-to-Image AIs* wie Dall-E 2, ist die natürlichsprachliche Eingabe bereits die Beschreibung des im Output entstehenden Ausgabebildes (Ramesh et al. 2022). Quartäre Autorschaft heißt also: *Il n'y a rien hors du modèle*. Autorschaft ist eingekapselt in das Gehäuse einer kommerziellen Sprachtechnologie, deren genaues Funktionieren letztlich kaum mehr ergründlich ist.

3 Werkzeug oder Agent

Kausale Autorschaft, die die menschliche Komponente in Mensch-Maschine-Gefügen überblickt, ist ein notwendiger Einstieg in komplexere Beschreibungsmodelle von Autorschaft in Zeiten Künstlicher Intelligenz, weil zunächst alle beteiligten technischen Prozesse identifiziert sein müssen, um zu weitergehenden Konzeptionen vorzustoßen. Da primäre Autorschaft bisher als Folie eines „unmittelbaren“ Schreibens diente, erscheint die zunehmende Distanz auch mit einem zunehmenden Kontrollverlust einherzugehen: Konnten die Produktionsregeln im sequenziellen Paradigma noch exakt und deterministisch formuliert werden, funktioniert das *machine learning* des konnektionistischen über einen stochastischen Prozess, der in einer Wahrscheinlichkeitsverteilung über Textsequenzen resultiert. Ist es in ersterem ein Leichtes, beliebig viele baugleiche Ausgaben zu produzieren, ist letzteres auf die Reproduktion von statistisch ähnlichen, aber eben nicht identischen Outputs angelegt; statt Schablonen gibt es „Gestalten“ aus (Bajohr 2020). Hinzu kommt die Undurchsichtigkeit des Sprachmodells: Wird der Code der sekundären Autorschaft von Menschen geschrieben und ist daher auch von Menschen lesbar, ist das „Gewichtungsmo­dell“ des KNN, Resultat eines statistischen Prozesses, prinzipiell nicht in Regelschritte rückübersetzbar und daher auch kein Gegenstand möglicher hermeneutischer Lektüre, sondern eher natürliches Untersuchungsobjekt – eher Sterncluster als Manuskriptvariante. So kann es in ihm nicht mehr um das Verstehen seines Sinnanspruches, sondern nur noch um das Erklären seines Funktionierens gehen (Kirschenbaum 2021). Dieser Kontrollverlust lässt die Möglichkeit autonomer KI-Autorschaft neuerlich virulent werden. Bewegt man sich von kausaler Autorschaft zur Frage, wem jene Foucault'sche Autorfunktion zukommt, die Diskurse ordnet, Dissemination reguliert und Werkkohärenz hervorbringt (Foucault 2000), tun sich heute verschiedene Interpretationen auf. Die wesentliche Achse ist jene der *agency* des Systems, die vom bloßen Werkzeug bis zum autonomen Agenten reicht (Henrickson 2018): Ist KI am Ende selbst Autor oder doch nur eine bessere Schreibmaschine?

Wie im ersten Abschnitt gezeigt, wehren sich die klassische Informatik sowie der Großteil der digitalen Literatur gegen die Zuschreibung von Autorschaft an den Computer: „Der Programmierer hat die Regeln gewählt, nach denen der Computer schreibt. Ob jetzt der Programmierer ein Lob verdient, weil er gerade diese Bearbeitungsregeln ausgesucht hat, weiß ich nicht. [...] Jedenfalls sollte er nicht sagen, der Computer dichtet.“ (Weizenbaum 2001, 103) Bereits Turing bedachte diesen Einwand unter dem Titel der „Lady Lovelace's objection“ (Turing 1950, 450–451). Er spielte damit auf Ada Lovelace an, die für Charles Babbages Entwurf einer mechanischen Rechenmaschine die „Programmiersprache“ entwickelte und damit als erste Coderin gelten kann. In einem Kommentar zum Aufbau von

Babbages *analytical engine* spekulierte sie, dass die Maschine prinzipiell nichts Neues hervorbringen könne, weil alle ihre Produkte nur auf die Anweisungen eines programmierenden Menschen zurückgingen (Lovelace 2010, 44). Turing hielt dagegen, dass auch Kalküle Überraschung produzieren können, weil menschliche Antizipationsfähigkeit angesichts komplexer Code-Output-Relationen begrenzt sei („Esprit der Regel“, Bajohr 2022, 37). Mehr noch: Angesichts der Möglichkeit lernen der Maschinen sei Lovelaces Einwand ohnehin nichtig. Wäre eine Maschine in der Lage, auf Basis invarianter Regeln dynamisch neues Verhalten zu erlernen, müsste sie selbst seinen Erbauer:innen unvorhersehbar, geradezu zufällig erscheinen (Turing 1950, 459).

Da *machine learning* zumindest eine Variante solchen „Lernens“ implementiert, scheint Turings Replik heute stichhaltiger als im sequenziellen Paradigma. Eine Möglichkeit liegt dann in der Reduktion menschlicher Autorschaft auf die Handlung des Kuratierens, die in jedem Schreiben eine Rolle spielt – minimal bereits im Sprechakt der Imprimatur, bei automatischen Prozessen aber vorher bereits im Feedbackprozess, der anhand einer Ausgabe die Programmparameter ändert. Die Abnahme an Kontrolle erfordert die Zunahme editorischer Auswahl. Es liegt also nahe, Autorschaft wieder als Herausgeberschaft zu denken, wie es bereits im Diskurs um Hypertextliteratur der Fall war (Wirth 2001), oder gegenüber dem Schreiben den Akt des Lesens weiter aufzuwerten, da bei nichtintentional generierten Texten Sinnanspruch noch plausibler als im Fall primärer Autorschaft erst durch die Lektüre an den Text getragen werden kann (Catani 2020, 304).

Diese Interpretationslinie, die den Diskurs um den Tod des Autors auch auf KI-generierte Texte überträgt – ohne im Gegenzug einen starken KI-Autor zu postulieren – ließe sich fortsetzen. So wäre Julia Kristevas Begriff der „Intertextualität“ nicht nur als „Mosaik aus Zitationen“ von Bedeutung (Kristeva 1969, 85), sondern gerade im Medienverbund sekundären bis quartären Schreibens sinnvoll einzusetzen. Da der Intertext hier ganz wörtlich vorliegt, bekommt man es mit einer Vervielfältigung von Autorschaftsbeziehungen zu tun. Abgesehen davon, dass Programmiersprachen selbst Autor:innen haben, ist bereits die Textur des Codes intertextuell verfasst, da ihre Funktionalität hohe Rekurrenz aufweist und Programmierende sowohl eigene wie fremde Code-Abschnitte wiederverwenden. Das kann in Repositorien wie GitHub oder Foren wie Stackoverflow geschehen, wo entweder ganze Funktionsbibliotheken angeboten werden, die bestimmte Aufgaben übernehmen und modular in den eigenen Code eingesetzt werden können, oder wo auf einer Mikroebene Codesequenzen gemeinsam korrigiert oder verbessert werden (Flender 2021, 139). Hier wird nicht selten bewusst auf eine Autorschaftsauszeichnung verzichtet; Mark Marino vergleicht solche Abschnitte mit dem Einsatz einer Schraube – Ingenieur:innen würde dessen Erfinder:in auch nicht bei jeder Verwendung zitieren (Marino 2020, 50). Die Idee kollektiver Autorschaft,

die in der Forschung um die Netzliteratur der 1990er Jahre dominierte (Heibach 2003, 205–206), ist daher auch auf ausführbare Primärtext, also Programmcodes, anzuwenden.

Ist damit noch das Zusammenspiel menschlicher Autor:innen gemeint, gehen die Science and Technology Studies weiter und beziehen auch nichtmenschliche Akteure in das Schreibkollektiv ein, explizit etwa bei Katherine Hayles, die „multiple authorships“ auch in der Hardware des Computers (Hayles 2005, 105) und im Computer selbst als „cognizer“ vermutet (Hayles 2019). Diese Haltung wird vor allem im Kontext der Akteur-Netzwerk-Theorie artikuliert, die in der Suche nach Aktanten und Akteuren *agency* gerade nicht als „limited a priori to what ‚intentional‘, ‚meaningful‘ humans do“ (Latour 2005, 81), sondern in komplexen Operationsketten als über ein Gefüge verteilt versteht. Statt Autor:innen allein als menschlich zu denken, würden sie in einer ANT-inspirierten Autorschaftstheorie zu einem Prozess und „die empirische Person nur ein menschlicher Akteur in einem Akteur-Netzwerk [sein], welches man ‚Aktor‘ nennt“ (Schröter 2022, 626).

So populär die Figur der Verteilung auch ist, so ist sie doch nicht unproblematisch. Das zeigt das Beispiel großer Sprachmodelle: Systeme wie ChatGPT beziehen ihre Mächtigkeit aus dem Umfang der ihnen zugrundeliegenden Trainingsdaten, die inzwischen mehr oder minder das gesamte Internet umfassen. Da in diese System also die Sprache der Almende eingeht, bekäme „kollektive Autorschaft“ eine neue Bedeutung, die eher an ein kollektives Unbewusstes als Quelle ‚automatischen‘ Schreibens gemahnt (O’Gieblyn 2021). Dass in solchen Metaphern einer Verteilung von *agency*, seien sie nun eine Übertragung altbekannter Konzepte der Literaturtheorie, wie eben Intertextualität, oder aber von der ANT inspirierte, durchaus Nachteile liegen, zeigt Jasmin Meerhoff, wenn sie darauf hinweist, dass die hier implizierte „maximal verteilte, verstreute oder ganz zerstäubte Autorschaft“ ökonomische Abhängigkeiten und ethisch fragwürdige Appropriation ausblendet (Meerhoff 2021, 55). Die Herkunft der Trainingsdaten lässt sich aus dem fertig trainierten Sprachmodell nicht ablesen und ist nur nachzuverfolgen, wenn sie eigens offengelegt wurden (etwa in der Open-Source-Initiative EleutherAI, siehe Gao et al. 2020). Hier mag es notwendig sein, die Rede einer Maschinenagency zu vermeiden und stattdessen die tatsächlich menschlichen Autor:innen hinter der KI-Ausgabe zu würdigen: als Arbeiter:innen, aber auch als Träger:innen bestimmter Identitäten (ähnlich auch Piringer 2019). Diese Aufforderung schließt an Betrachtungen der Kunstsoziologie an, die an das infrastrukturelle „support personnel“ erinnert, das hinter den scheinbar autonomen Größen des Kunstsystems steht und deren Aktivitäten erst ermöglicht (Becker 1982, x; siehe auch Gilbert 2021, 66).

In der Tat: Betrachtet man die bisher getätigten, überschaubaren literarischen Experimente, dann stünde zu erwarten, es würde entweder die Kollektivität der Almende oder die Kollaboration mit der Maschine betont, was in beiden Fällen

eine Auflösung der Autorfunktion nahelegte. Bis auf wenige Ausnahmen (siehe Kaufmann 2022) scheinen Autor:innen aber weiterhin ihren Namen auch an in Kollaboration mit KI-Systemen entstandene Werke zu heften. Selbst ein Buch wie *Amor Cringe* des:der amerikanischen Dichters:in K Allado-McDowell erscheint unter dessen:deren Namen (Allado-McDowell 2022), auch wenn es explizit in Kooperation mit GPT-3 geschrieben wurde. Das heißt aber nicht notwendig die reaktionäre Reaktivierung überkommener Autorschaftsideen wie jener des Genies. Für die Dichterin Allison Parrish ist es weniger eine ontologische Frage als eine politische Aufgabe, Computer als Werkzeuge zu verstehen: „it will always be a mistake to attribute volition to the computer and not to the people who programmed it because attribution of volition is removing personal responsibility: the algorithm did it, not me.“ (zit. nach Miller 2019, 223) Die Reklamation von Autorschaft kann gerade als Abwehr sowohl gegen das Phantasma des technisch optimierten KI-Genies wie der absoluten Zerstäubung von Autorschaft in Anschlag gebracht werden, die politische, ökonomische und ethische Verantwortung nicht mehr nachvollziehen lässt. Statt gleichberechtigter Kollaboration wird hier die klar hierarchisierte Ko-Kreation vorgezogen (wie etwa in Johnston 2019; siehe auch Kaufmann 2022, 38).

Der Anspruch wirklich „starker“ KI-Autorschaft (zu diesem Begriff, siehe Bajohr 2021b), die jenseits von Kollaboration, Ko-Kreation und Verteilung zu autonomen Schöpfungen in der Lage ist, ist heutzutage vor allem in der Tech-Branche anzutreffen, deren Vertreter:innen Kunst als Katalysator für die Entwicklung künstlichen Bewusstseins verstehen (siehe auch Daub 2020; exemplarisch Elgammal und Mazzone 2020). Hier überleben Ideen von Autorschaft als Genie sehr viel unbefragter als im Kunst- und Literaturbereich, die mit der Dekonstruktion solcher Großbegriffe seit mehr als sechzig Jahren beschäftigt sind. Hier zeigt sich aber auch, dass ein romantischer Autor nicht notwendig menschlich gedacht sein muss und die ursprüngliche, aus dem Nichts Neues schaffende Originalität des romantischen Genies durchaus auch auf KI-Modelle übertragen werden kann, wie Carys J. Craig und Ian R. Kerr zeigen (2019). Nicht in der praktisch-literarischen Arbeit mit Sprachmodellen, sondern im Valley scheint die Grundideologie der KI-Forschung fortzuleben, die Bewusstsein und Intelligenz mit intentionaler Autorschaft kurzschließt, während die Literatur weiterhin im Modus der NLP die strukturierte Hervorbringung von Text verfolgt, der so verwendet werden kann oder eben auch nicht.

Sehr viel wahrscheinlicher als das Phantasma ‚künstlicher Autor:innen‘ wären – freilich kann das nur spekulativ gesagt werden – autorlose Texte, wie sie sich Foucault ausmalte: „Man kann sich eine Kultur vorstellen, in der Diskurse verbreitet oder rezipiert würden, ohne daß die Funktion Autor jemals erschiene.“ (Foucault 2000, 227) Ganz pragmatisch ist das in unserer Kultur zum Teil bereits Realität: Im Fall von automatisierten Informationssystemen, Interfaceausgaben und anderer Maschinen-

kommunikationen, denen wir tagtäglich ausgesetzt sind, stellt sich die Frage nach Autorschaft überhaupt nicht – weder würde man aus der Anzeige auf dem Bankautomaten auf eine Personalität der Maschine schließen, noch sich wesentlich den Kopf über deren ursprüngliche Autor:innen zerbrechen. Je mehr die Zirkulation von maschinell generierten Texten zunimmt, desto wahrscheinlicher wird es, sie als „post-artifiziell“ (Bajohr 2023) zu betrachten; die Frage nach Autorschaft verlöre sich dann in der „Namenlosigkeit des Gemurmels“ (Foucault 2000, 227).

Literaturverzeichnis

- Aarseth, Espen J. *Cybertext: Perspectives on Ergodic Literature*. Baltimore 1997.
- Allado-McDowell, K. *Amor Cringe*. New York 2022.
- Alpaydin, Ethem. *Machine Learning: Revised and Updated Edition*. Cambridge, Mass. 2021.
- Bajohr, Hannes. „Die ‚Gestalt‘ der KI: Jenseits von Holismus und Atomismus“. *Zeitschrift für Medienwissenschaft* 12.2 (2020): 168–181.
- Bajohr, Hannes. „Algorithmische Einfühlung: Über zwei Paradigmen digitaler generativer Literatur und die Notwendigkeit einer Kritik ästhetischer KI“. *Sprache im technischen Zeitalter* 59.4 (2021a): 471–497.
- Bajohr, Hannes. „Keine Experimente: Über künstlerische Künstliche Intelligenz“. *Merkur* 75.863 (2021b): 32–44.
- Bajohr, Hannes. „Künstliche Intelligenz und digitale Literatur“. *Digitale Literatur II*. Hrsg. von Hannes Bajohr und Annette Gilbert. München 2021c: 174–185.
- Bajohr, Hannes. *Schreibenlassen: Texte zur Literatur im Digitalen*. Berlin 2022.
- Bajohr, Hannes. „Artifizielle und postartifizielle Texte: Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben.“ *Sprache im technischen Zeitalter* 61.1 (2023): 37–61.
- Bajohr, Hannes. „Writing at a Distance: Notes on Authorship and Artificial Intelligence.“ *German Studies Review* 47.1 (2024, i.E.).
- Barthes, Roland. „Der Tod des Autors“. *Texte zur Theorie der Autorschaft*. Hrsg. von Fotis Jannidis, Gerhard Lauer, Matias Martinez und Simone Winko. Stuttgart 2000: 185–193.
- Becker, Howard S. *Art Worlds*. Berkeley 1982.
- Bense, Max. *Ästhetik und Texttheorie*. Hrsg. von Elisabeth Walther. Stuttgart 1998.
- Bense, Max. „Über natürliche und künstliche Poesie“. *Theorie der Texte. Eine Einführung in neuere Auffassungen und Methoden*, Köln 1962: 143–147.
- Bernhart, Toni. „Beiwerk als Werk: Stochastische Texte von Theo Lutz“. *editio* 34 (2020): 180–206.
- Bolter, Jay David. „Artificial Intelligence“. *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, NJ 1991: 171–193.
- Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx et al. „On the Opportunities and Risks of Foundation Models“. <http://arxiv.org/abs/2108.07258>, 2021.
- Borges, Jorge Luis. „Der Garten der Pfade, die sich verzweigen“. *Der Erzählungen erster Teil*, München 2000: 161–176.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal. „Language Models are Few-Shot Learners“. <http://arxiv.org/abs/2005.14165>, 2020.

- Calvino, Italo. „Kybernetik und Gespenster“. *Kybernetik und Gespenster: Überlegungen zu Literatur und Gesellschaft*, München 1984: 7–26.
- Catani, Stephanie. „Erzählmodus an: Literatur und Autorschaft im Zeitalter künstlicher Intelligenz“. *Jahrbuch der deutschen Schillergesellschaft*, Berlin 2020: 287–310.
- Chollet, François. *Deep Learning with Python*. 2. Auflage. Shelter Island 2021.
- Coover, Robert. „Goldene Zeitalter: Vergangenheit und Zukunft des literarischen Wortes in den digitalen Medien“. *Text+Kritik* 152 (2001): 22–30.
- Craig, Carys J. und Ian R. Kerr. „The Death of the AI Author“. *Ottawa Law Review* 52.1 (2019): 31–86.
- Cramer, Florian. *Exe.cut[up]able statements: Poetische Kalküle und Phantasmen des selbstausführenden Texts*. München 2011.
- Daub, Adrian. *What Tech Calls Thinking: An Inquiry into the Intellectual Bedrock of Silicon Valley*. New York 2020.
- Deleuze, Gilles und Félix Guattari. *Tausend Plateaus*. Berlin 1992.
- Elgammal, Ahmed und Marian Mazzone. „Artists, Artificial Intelligence and Machine-Based Creativity in Playform“. *Artnodes* 26 (2020): 1–8.
- Enzensberger, Hans Magnus. *Einladung zu einem Poesie-Automaten*. Frankfurt am Main 2000.
- Flender, Karl Wolfgang. „Do Conceptualists Dream of Electric Sheep? Algorithmische Interpretation des Unbewussten in Conceptual Writing und konzeptueller Codeliteratur“. *Digitale Literatur II*. Hrsg. von Hannes Bajohr und Annette Gilbert. München 2021: 134–144.
- Foucault, Michel. „Was ist ein Autor?“ *Texte zur Theorie der Autorschaft*. Hrsg. von Fotis Jannidis, Gerhard Lauer, Matias Martinez und Simone Winko. Stuttgart 2000: 198–229.
- Funkhouser, Christopher T. *New Directions in Digital Poetry*. New York 2012.
- Gao, Leo, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster u.a. „The Pile: An 800GB Dataset of Diverse Text for Language Modeling“. <http://arxiv.org/abs/2101.00027>, 2020.
- Gilbert, Annette. „Kollaterales Schreiben: Digitale Kollaboration im Zeitalter von Crowdfunding und Algotaylorismus“. *Digitale Literatur II*. Hrsg. von Hannes Bajohr und Annette Gilbert. München 2021: 62–74.
- Gordin, Michael D. *Scientific Babel: How Science Was Done Before and After Global English*. Chicago 2015.
- Guertin, Carolyn. „Wanderlust: The Kinesthetic Browser in Cyberfeminist Space“. <https://web.archive.org/web/20150921150355/http://www.performancestudies.ucla.edu/extensionsjournal/guertin.htm>. *Extensions: Online Journal of Embodiment and Technology* 3 (2007) (30. März 2023).
- Hartling, Florian. *Der digitale Autor: Autorschaft im Zeitalter des Internets*. Bielefeld 2009.
- Hayles, N. Katherine. *My Mother Was a Computer*. Chicago 2005.
- Hayles, N. Katherine. „Literary Texts as Cognitive Assemblages: The Case of Electronic Literature“. *Interface Critique* 2 (2019): 173–195.
- Heibach, Christiane. „Ins Universum der digitalen Literatur: Versuch einer Typologie“. *Text+Kritik* 152 (2001): 31–42.
- Heibach, Christiane. *Literatur im elektronischen Raum*. Frankfurt am Main 2003.
- Henrickson, Leah. „Tool vs. Agent: Attributing Agency to Natural Language Generation Systems“. *Digital Creativity* 29.2–3 (2018): 182–190.
- Henrickson, Leah. „Constructing the Other Half of The Policeman’s Beard“. <https://electronicbookreview.com/essay/constructing-the-other-half-of-the-policemans-beard>. *Electronic Book Review* 4.4.2021 (30. März 2023).
- Higgins, Dick. *Computers for the Arts*. Somerville 1970.

- Higgins, Hannah B. „An Introduction to Alison Knowles's The House of Dust“. *Mainframe Experimentalism: Early Computing and the Foundations of the Digital Arts*. Hrsg. von Hannah B. Higgins und Douglas Kahn. Berkeley 2012: 195–199.
- Johnston, David Jhave. *ReRites: Human + A.I. Poetry. Raw Output*. Montreal 2019.
- Kaufmann, Manuel. *Dreaming Data: Aspekte der Originalität und Autorschaft in der künstlichen Kreativität*. Zürich 2022.
- Kirschenbaum, Matthew G. *Track Changes: A Literary History of Word Processing*. Cambridge, Mass. 2016.
- Kirschenbaum, Matthew G. „Spec Acts: Reading Form in Recurrent Neural Networks“. *ELH* 88.2 (2021): 361–186.
- Krause, Manfred und Götz F. Schaudt, hrsg. *Computer-Lyrik*. 2. Auflage. Düsseldorf 1969.
- Kristeva, Julia. *Séméiotikè: Recherches pour une sémanalyse*. Paris 1969.
- Krolow, Karl. „Die Seele wimmert“. *Die Tat*, 13.01.1968: 33.
- Landow, George P. *Hypertext 2.0. The Convergence of Contemporary Critical Theory and Technology*. Baltimore 1997.
- Latour, Bruno. „Third Source of Uncertainty: Objects too Have Agency“. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford 2005: 63–87.
- Link, David. „There Must Be an Angel: On the Beginnings of the Arithmetic of Rays“. *Archaeology of Algorithmic Artefacts*. Minneapolis 2016: 55–78.
- Lovelace, Ada. „Translator's Notes to M. Menebrea's Memoir“. *Babbage's Calculating Engines: Being a Collection of Papers Relating to Them; Their History, and Construction*. Hrsg. von Henry P. Babbage. Cambridge 2010: 32–50.
- Lutz, Theo. „Stochastische Texte“. *Augenblick: Zeitschrift für aktuelle Philosophie, Ästhetik, Polemik* 4.1 (1959): 3–9.
- Lutz, Theo. „Über ein Programm zur Erzeugung stochastisch-logistischer Texte“. *Grundlagenstudien aus Kybernetik und Geisteswissenschaft* 1.1 (1960): 11–16.
- Marino, Mark C. *Critical Code Studies*. Cambridge, Mass. 2020.
- Meerhoff, Jasmin. „Verteilung und Zerstäubung: Zur Autorschaft computergestützter Literatur“. *Digitale Literatur II*. Hrsg. von Hannes Bajohr und Annette Gilbert. München 2021: 49–61.
- Miller, Arthur I. The Artist in the Machine *The Artist in the Machine: The World of AI-Powered Creativity*. Cambridge, Mass. 2019.
- Mitchell, Melanie. *Artificial Intelligence: A Guide for Thinking Humans*. New York 2019.
- Nilsson, Nils J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge 2010.
- O'Gieblyn, Meghan. *Babel: Could a machine have an unconscious?*. https://www.nplusonemag.com/issue-40/essays/babel-4_n+1_40.1 (2021) (30. März 2023).
- OpenAI. „ChatGPT: Optimizing Language Models for Dialogue“. <https://openai.com/blog/chatgpt/>. *OpenAI Blog* 2022 (30. März 2023).
- Piringer, Jörg. „elektrobarden“. *Transistor* 1.2 (2019): 78–83.
- Porombka, Stephan. *Hypertext: Zur Kritik eines digitalen Mythos*. München 2001.
- Racter. *The Policeman's Beard is Half Constructed*. New York 1984.
- Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen. „Hierarchical Text-Conditional Image Generation with CLIP Latents“. <https://arxiv.org/abs/2204.06125>, 2022.
- Rieger, Stefan. „Bin doch keine Maschine ...“. Zur Kulturgeschichte eines Topos“. *Machine Learning: Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*. Hrsg. von Christoph Engemann und Andreas Sudmann. Bielefeld 2018: 117–142.

- Rumelhart, David E., James L. McClelland und PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Bd. 1. Cambridge 1986.
- Schäfer, Jürgen. „Passing the Calvino Test? Writing Machines and Literary Ghosts“. *Digital Media and Textuality: From Creation to Archiving*. Hrsg. von Daniela Côrtes Maduro. Bielefeld 2017: 23–44.
- Schönthaler, Philipp. *Die Automatisierung des Schreibens und Gegenprogramme der Literatur*. Berlin 2022.
- Schröter, Jens. „Autorschaft aus dem Blickwinkel der Akteur-Netzwerk-Theorie“. *Grundthemen der Literaturwissenschaft: Autorschaft*. Hrsg. von Michael Wetzel. Berlin 2022.
- Simanowski, Roberto. „Autorschaften in digitalen Medien: Eine Einleitung“. *Text+Kritik* 152 (2001): 3–21.
- Stickel, Gerhard. „Computerdichtung“: Zur Erzeugung von Texten mit Hilfe von datenverarbeitenden Anlagen“. *Der Deutschunterricht* 18.2 (1966): 120–125.
- Stingelin, Martin. „UNSER SCHREIBZEUG ARBEITET MIT AN UNSEREN GEDANKEN. Die poetologische Reflexion der Schreibwerkzeuge bei Georg Christoph Lichtenberg und Friedrich Nietzsche“. *Schreiben als Kulturtechnik: Grundlagentexte*. Hrsg. von Sandro Zanetti. Berlin 2012: 83–104.
- Strachey, Christopher. „The ‚Thinking‘ Machine“. *Encounter* 3 (1954): 25–31.
- Turing, Alan M. „Computing Machinery and Intelligence“. *Mind* 59.236 (1950): 433–460.
- Wardrip-Fruin, Noah. „Christopher Strachey: The First Digital Artist?“ <https://grandtextauto.soe.ucsc.edu/2005/08/01/christopher-strachey-first-digital-artist>. *Grand Text Auto* 2009 (30. März 2023).
- Weaver, Warren. „Translation“ [1949]. *Readings in Machine Translation*. Hrsg. von Sergei Nirenburg, H. L. Somers und Yorick Wilks. Cambridge, Mass 2003: 13–17.
- Weizenbaum, Joseph. „ELIZA: A Computer Program for the Study of Natural Language Communication Between Man And Machine“. *Communications of the ACM* 9.1 (1966): 36–45.
- Weizenbaum, Joseph. „Kunst und Computer“. *Computermacht und Gesellschaft: Freie Reden*. Hrsg. von Gunna Wendt und Franz Klug. Frankfurt am Main 2001.
- Wirth, Uwe. „Der Tod des Autors als Geburt des Editors“. *Text+Kritik* 152 (2001): 54–64.