

Hannes Bajohr

Keine Experimente

Über künstlerische Künstliche Intelligenz

Daniel Kehlmann hat kein Buch mit einer Künstlichen Intelligenz geschrieben; darüber hat er jetzt ein Buch geschrieben.¹ In *Mein Algorithmus und ich* – es ist die Druckfassung seiner Anfang 2021 gehaltenen »Stuttgarter Zukunftsrede« – reist Kehlmann nach Palo Alto, wo er bei einem Startup Zugang zu einem KI-System bekommt, dem Sprachmodell CTRL. Im Dialog mit ihm will er erproben, ob mit KI Literatur zu machen ist. Kehlmann zeigt sich enttäuscht. Zu wenig narrativ kohärent, zu absurd sind die Ergebnisse, auch wenn hier und da Schönes aufscheint. Das Scheitern seines Ausflugs ins maschinelle Lernen ist dann auch die implizite Pointe seines Buchs: Der Mensch muss sich keine Sorgen machen, dass ihm die Literatur bald von KIs abgenommen wird.

Kehlmann, das merkt man, geht technisch nicht naiv an die Sache heran. Er schickt voraus, dass »KI« eigentlich eine Fehlbezeichnung ist, dass das, was man heute so nennt, weder Bewusstsein hat noch in irgendeinem wirklichen Sinn Intelligenz besitzt, sondern als statistisches Modell lediglich auf der Grundlage gelernter Daten Vorhersagen über wahrscheinliche Zustände macht.² Bei sogenannten Sprachmodellen haben sowohl die gelernten Daten als auch die gemachten Vorhersagen Textform. Diese Modelle funktionieren nicht grundsätzlich anders als die Eingabevervollständigung im Smartphone: Nach »Guten« folgt wahrscheinlich »Morgen«, »Tag« oder »Appetit«.

Dass eine solche »Intelligenz« wenig mit der unseren zu tun hat, besser »künstliche Rationalität« zu nennen wäre, wie Kehlmann in einem Podiumsgespräch mit dem Philosophen Felix Heidenreich anmerkte, führt ihn aber nicht in Versuchung, diese Differenz auch ästhetisch zu erproben. Er bedauert: »Ich habe keine Geschichte vorzuweisen, die ich mit CTRL verfasst hätte und die gut genug wäre, dass ich sie als künstlerisches Werk, nicht bloß als Produkt eines Experiments, veröffentlichen könnte.«

Was aber heißt »gut genug«? Gemessen an welcher Ästhetik? Wenn Kehlmann hier von »Experiment« spricht, scheint er eher die wissenschaftstheo-

1 Daniel Kehlmann, *Mein Algorithmus und ich. Stuttgarter Zukunftsrede*. Stuttgart: Klett-Cotta 2021.

2 Vgl. Paola Lopez, *Artificial Intelligence und die normative Kraft des Faktischen*. In: *Merkur*, Nr. 863, April 2021.

retische Bedeutung des Wortes im Sinn zu haben als experimentelle Literatur: eine kontrollierte Beobachtung, deren Ausgang eine Hypothese stützt, schwächt oder verfeinert. Das aber, nach Thomas Kuhn, immer nur im Rahmen eines bestehenden Paradigmas; *neue* Paradigmen begründen wissenschaftliche Experimente nämlich nicht. Experimentelle Literatur dagegen will – ihrem avantgardistischen Selbstverständnis nach, mit dem sie etwa von der Wiener Gruppe oder bei Max Bense postuliert wurde – keine Verfeinerung, sondern stellt idealerweise das Paradigma Literatur selbst infrage.

Es ist also gut möglich, dass nicht die Künstliche Intelligenz an der Literatur gescheitert ist, sondern Kehlmann an der Künstlichen Intelligenz – und irgendwie auch an der Literatur. Denn in seiner Gegenüberstellung von vollwertigem »künstlerischen Werk« und bloßem »Experiment« zeigt sich, wie wenig ihm in den Sinn kommt, dass man mit Maschinen *anders* Literatur machen kann oder vielleicht sogar muss, statt sie über das Stöckchen der eigenen Poetik springen zu lassen. So sind ihm die Entgleisungen und Absurditäten, die CTRL ausspuckt, offensichtlich *bug*, nicht *feature*. Was Literatur ist und welcher Ästhetik sie zu folgen hat, steht von Anfang an fest. Für Kehlmann ist ihr perennierender Kern vor allem eines: Erzählung – kohärent und ohne Brüche, dafür mit großen Zusammenhängen und weit geschwungenen Plot-Bögen, die letzten Endes auf eine komplexe Autorintentionalität verweisen. Die freilich geht der Maschine ab, und so hält er das Experiment für misslungen.

Dabei ist es unwesentlich, dass das verwendete Sprachmodell CTRL bereits hoffnungslos veraltet ist. Schon GPT-3, das letztes Jahr als State of the Art in Sachen Text-KI Furore machte, ist hundertmal größer und hätte deutlich bessere Ergebnisse in Sachen Kohärenz erzielt; Switch-C, Googles neuestes Sprachmodell, ist noch umfangreicher.³ Interessanter als technische

3 Hat CTRL 1,6 Milliarden Parameter – also »Neuronen« seines neuronalen Netzes –, kann GPT-3 175 Milliarden vorweisen; wurde CTRL auf 140 Gigabyte an Text trainiert, sind es bei GPT-3 immerhin 570 Gigabyte. Zu CTRL vgl. Nitish Shirish Keskar u. a., *CTRL: A Conditional Transformer Language Model for Controllable Generation* (arxiv.org/abs/1909.05858); zu GPT-3 vgl. Tom B. Brown u. a., *Language Models are Few-Shot Learners* (arxiv.org/abs/2005.14165). Das Rennen um immer größere Sprachmodelle wird inzwischen ethisch und politisch kritisiert: Die Modelle reproduzieren diskriminierende Sprache, sind in ihrer Größe nicht mehr durchschau- und korrigierbar und für einen immensen CO₂-Ausstoß verantwortlich. Für ein prominentes Beispiel dieser Diskussion vgl. Emily M. Bender u. a., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*. In: *FACCT '21. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (dl.acm.org/doi/10.1145/3442188.3445922).

Quisquilien ist ein Paradox, das hinter Kehlmanns Enttäuschung steht und im Kontext von kunstschaffender KI häufiger auftaucht: Je mehr von Künstlicher Intelligenz erwartet wird, desto menschlicher wird sie gedacht, desto weniger aber wird sie als ein Eigenes gewürdigt. Eine wirklich *starke* künstlerische KI würde Kehlmann nicht erweitern, sondern tatsächlich ersetzen – und keine neuen Ästhetiken erforderlich machen, sondern die alten wiederholen. Das zeigt sich sowohl in der Theorie wie in der Praxis künstlerischer KI.

Starke und schwache künstlerische KI

So gut wie alle Diskussionen um Kunst und Künstliche Intelligenz fallen unter eine von zwei, meist unausgeführten, Vorstellungen dessen, was eine künstlerische KI eigentlich ist oder sein soll. Sie unterscheiden sich in ihrem Anspruch immens und hängen vor allem an der zugestandenen Selbständigkeit des kunstproduzierenden Systems. Man kann diesen Unterschied vielleicht am besten parallel zu John Searles kanonischem Begriff von starker und schwacher KI verdeutlichen.

Starke KI bezeichnet bei Searle die Herstellung eines künstlichen Bewusstseins inklusive aller für es konstitutiven Eigenschaften (für Searle ist das vor allem Intentionalität). Schwache KI ist dagegen eine bloße Hilfestellung zur Einsicht in die Modellierung des Bewusstseins. Meint starke KI also die funktionale Duplikation des Zielbereichs, ist schwache KI allenfalls eine Teilsimulation dieses Bereichs und hat höchstens eine heuristische, eine »Werkzeug«-Funktion, wie Searle es formuliert.⁴

Entfernt man sich von Bewusstsein als Zielbereich, kann man analog auch von einer starken und schwachen *künstlerischen* KI sprechen. Die starke Vorstellung hätte die Duplikation des gesamten Herstellungsprozesses von Kunst zur Aufgabe. Die schwache dagegen würde Techniken wie neuronale Netze als Assistenzsystem in diesem Prozess betrachten, das darin lediglich Teilaufgaben übernimmt; das kann sehr weit gehen, aber eben nicht bis zum Punkt der Selbständigkeit, der der starken künstlerischen KI vorschwebt. Nun lassen sich aber gerade im starken Modell – das erst mit einem zweiten Kehlmann zufrieden wäre – schnell einige Schwierigkeiten ausmachen.

Die Möglichkeit starker künstlerischer KI steht und fällt mit dem Problem der Operationalisierbarkeit jener Kategorie, die im Zielbereich zur Anwendung kommt: der von Kunst (oder, im Folgenden transitiv, der von

4 John R. Searle, *Minds, Brains, and Programs*. In: *Behavioral and Brain Sciences*, Nr. 3/3, September 1980.

Literatur). Bereits der Titel, unter dem das starke Modell für gewöhnlich läuft – »artificial creativity« oder »computational creativity« –, zeigt, wie man versucht, die konstitutive Vagheit des Begriffs »Kunst« zu umschiffen, indem man ihn durch den der »Kreativität« ersetzt. Dabei gibt es verschiedene Strategien. So definiert die Philosophin Margaret Boden Kreativität vom Objekt her, als Herstellung von etwas, das »neu, überraschend und wertvoll« ist.⁵ Alternativ nimmt die neurowissenschaftliche Definition die Subjektseite, den kreativen Hirnprozess in den Blick.⁶ In beiden Fällen wird Kunst als Kreativität mess- und schließlich auch durch Computer simulierbar.⁷

In beiden Fällen ist aber auch alles andere als klar, ob nicht etwas am Begriff »Kunst« verloren geht. Reduktiv scheint dessen Identifikation mit Kreativität schon allein darum, weil es hier kein Kriterium mehr gibt, das die Produktion eines Kunstwerks von einer technischen Innovation oder einer besonders disruptiven Business-Strategie absetzt. Ebenso ist es nicht einsichtig, in welchem Verhältnis Kreativität, die oft als Problemlösung konzeptualisiert wird, eigentlich zum Ästhetischen steht.⁸ Ganz offensichtlich ist in einer Zeit, da Kreativität weitgehend vom Kunstbegriff abgekoppelt ist, eher auf die »Kreativindustrie« übertragen wird oder unter dem Titel »Kreativitätsdispositiv« gleich als Leistungspostulat auf jedem neoliberalen Subjekt lasten soll, Kunst in dieser Identifikation unterbestimmt.⁹

Dennoch verbirgt sich auch im über Kreativität operationalisierten starken Modell ein Begriff von Kunst. Die Definitionen von der Objekt- wie von der Subjektseite her verfahren wesentlich immanentistisch. Einerseits

- 5 Margaret A. Boden, *The Creative Mind. Myths and Mechanisms*. London: Routledge 2004.
- 6 Anna Abraham, *The Neuroscience of Creativity*. Cambridge University Press 2018.
- 7 Vgl. etwa Margaret A. Boden, *Computer Models of Creativity*. In: *AI Magazine*, Nr. 30/3, Herbst 2009; dies., *AI. Its Nature and Future*. Oxford University Press 2016.
- 8 Anders als Boden definiert die Neurowissenschaftlerin Anna Abraham Kreativität als Neuheit plus Angemessenheit, wobei Letztere als Problemlösungs- oder Optimierungsfrage verstanden wird. Das hilft sicherlich bei der Automatisierung von Kreativität, es bleibt aber unbeantwortet, welches Problem ein Kunstwerk eigentlich löst und in welcher Domäne »Angemessenheit« dann zu suchen wäre. Vgl. Anna Abraham, *Neuroscience*.
- 9 Kenneth Goldsmith, *Uncreative Writing. Sprachmanagement im digitalen Zeitalter*. Berlin: Matthes & Seitz 2017; Richard Florida, *The Rise of the Creative Class*. New York: Basic Books 2002; Andreas Reckwitz, *Die Erfindung der Kreativität. Zum Prozess gesellschaftlicher Ästhetisierung*. Berlin: Suhrkamp 2012.

subtrahieren sie Kunst von ihrem sozialhistorischen Kontext und machen sie so zu einem überzeitlichen Phänomen,¹⁰ andererseits isolieren sie ihre Herstellung in einem vereinzelt Akteur. Implizit steht hinter der »computational creativity« also eine durchaus diskussionswürdige Autonomie- und Genieästhetik.¹¹ Ihre Schwierigkeiten zeigen sich nirgends besser als dort, wo diese Theorie in die Praxis umgesetzt wird.

Kreativitätsmaschinen

Im Dezember 2020 erhielt Ahmed Elgammal, Informatiker an der Rutgers University, ein US-Patent auf ein »Creative Adversarial Network« (CAN). Es soll ganz explizit »Kunst generieren«, wie es im Titel heißt.¹² Auch Elgammal begreift Kunst als Kreativität, wobei er sie, unter Bezugnahme auf behavioristische Psychologie, als ein im Gehirn messbares »arousal« definiert. Auslöser solcher »arousal«-Potentiale sind Überraschung, Verwirrung, Komplexität und semantische Ambiguität, wobei sowohl zu wenig Reiz (Langeweile) wie zu viel (Widerwille) zu vermeiden sind.

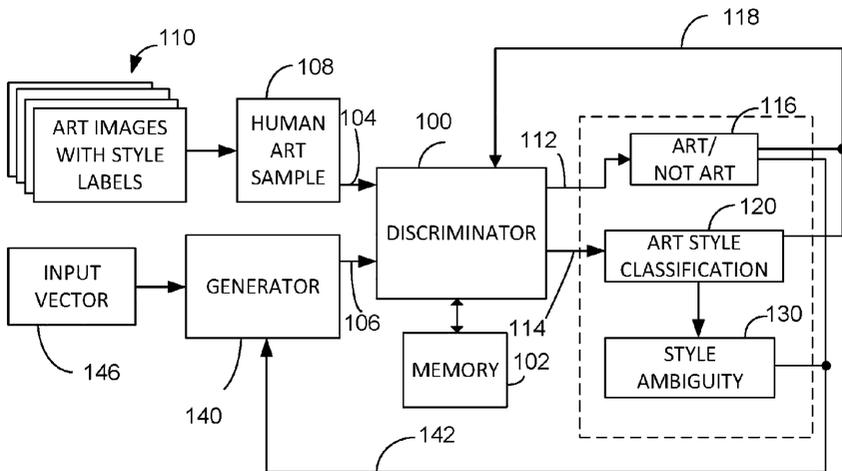
Elgammals CAN versucht nun, diese Novitätsfaktoren als Weiterentwicklung einer bekannten KI-Architektur umzusetzen, nämlich des »Generative Adversarial Network« (GAN). Ein GAN vereint zwei neuronale Netze, wobei das eine, der »Generator«, zunächst zufällige Bilder produziert, die vom anderen, dem »Discriminator«, der auf einen bestimmten Datensatz von Bildern trainiert ist, bewertet werden. In einem iterativen Optimierungsprozess passt der Generator seinen Output den Bewertungen des Discriminators an, so dass er schließlich Bilder ausgibt, die eine statistische Ähnlichkeit

10 Margaret Boden unterscheidet zwar persönliche von historischer Kreativität (»P-creativity« und »H-creativity«), bemisst die Geschichtlichkeit der Letzteren aber allein daran, ob der kreativ hervorgebrachte Gegenstand objektiv eine Neuheit in der Welt darstellt. Das ist letztlich ein Katalogmodell von Geschichte, in dem Genealogien und Einflüsse unerheblich sind. Vgl. Margaret Boden, *The Creative Mind*.

11 Explizit wird das in zwei populären Diskussionen von KI-Kunst: Marcus du Sautoy, *The Creativity Code. Art and Innovation in the Age of AI*. Cambridge/Mass.: Harvard University Press 2019; Arthur I. Miller, *The Artist in the Machine. The World of AI-Powered Creativity*. Cambridge/Mass.: MIT Press 2019. Für beide Autoren ist Kreativität wesentlich ein Merkmal von Genies, wobei sie solche Reihen bilden wie: Bach, Picasso, Steve Jobs. Dass der Geniebegriff ins Silicon Valley abgewandert ist, bestätigt auch Adrian Daub, *Was das Valley denken nennt. Über die Ideologie der Techbranche*. Berlin: Suhrkamp 2020.

12 Ahmed Elgammal, *Creative GAN Generating Art Deviating from Style Norms*. US-Patent Nr. 10.853.986, erteilt am 1. Dezember 2020.

mit dem Trainingsset aufweisen. Auf eine Menge Porträts trainiert, könnte das GAN nun neue, täuschend echte Gesichter herstellen.¹³



Ist beim GAN die einzige Differenzkategorie, dass das Output-Bild dem Trainingsset statistisch nahekommt (ähnlich/unähnlich), führt Elgammals CAN zwei Kriterien ein: Auf einen Kanon von Gemälden trainiert, entscheidet der Discriminator zuerst anhand des so gewonnenen probabilistischen Modells, ob ein vom Generator generierter Output Kunst ist oder nicht; anschließend bewertet er mithilfe gelernter Metadaten über die »Stile« dieser Gemälde, ob das Ausgabebild einem dieser Stile entspricht (Kunst/nicht Kunst und bekannter/unbekannter Stil, im obigen Diagramm Nr. 116 und 120/130).¹⁴ Das Feedback steuert den Generator immer weiter in Richtung höherer »Kunsthafigkeit«, wobei die Vermeidung bekannter Stile Neuheit zu sichern sucht.

13 Das kann man etwa auf der Seite www.thispersondoesnotexist.com nachprüfen, die StyleGAN2 verwendet. Vgl. Hannes Bajohr, *Die »Gestalt« der KI. Jenseits von Holismus und Atomismus*. In: *Zeitschrift für Medienwissenschaft*, Nr. 12/23, Februar 2020.

14 Das Trainingsset besteht aus 75 753 Gemälden; die idiosynkratisch zu nennende Liste an Stilen umfasst: »Abstract Expressionism, Action Painting, Analytical Cubism, Art Nouveau Modern, Baroque, Color Field Painting, Contemporary Realism, Cubism, Early Renaissance, Expressionism, Fauvism, High Renaissance, Impressionism, Mannerism/Late Renaissance, Minimalism, Naive Art/Primitivism, New Realism, Northern Renaissance, Pointillism, Pop Art, Post-Impressionism, Realism, Rococo, Romanticism, Synthetic Cubism«. Vgl. Elgammal, *Creative GAN*.

Kunstschaffen ist hier eine Funktion der Abweichung von individuellen Strukturmerkmalen im Kontext eines Kanons. Im CAN ist sowohl der Geniegedanke als auch die Idee autonomer Kunst technisch umgesetzt: Nicht nur wird so der Bildungsroman eines Künstlers in der Maschine nachgebaut, sondern die Kunstgeschichte auch auf die Abfolge von dekontextualisierten Trainingsdaten reduziert. Als bis heute avanciertester Entwurf des starken Modells künstlerischer KI übernimmt das CAN nicht nur Hilfsaufgaben im Produktionsprozess, sondern schafft »selbständig« Kunst. Darauf jedenfalls besteht Elgammal ganz ausdrücklich: In einem zusammen mit der Kunsthistorikerin Marian Mazzone verfassten Artikel grenzt er das CAN von lediglich assistierendem KI-Systemen mit dem Argument ab, dass seine Erfindung tatsächlich »intentional« ablaufe, weil es sich die Regeln zur Kunstproduktion selbst aneigne, statt sie bereits gefüttert zu bekommen. Daher sei das CAN nicht nur »inherently creative«, sondern auch wirklich ein »autonomous artist«.¹⁵

Gerade das steht aber infrage, und zwar nicht nur, weil natürlich auch hier noch Elgammal auswählen muss, welche Bilder er tatsächlich in die Welt entlassen will. Der Anspruch einer starken künstlerischen KI, die Genie und Autonomie betont, begibt sich auch eines der Gegenwart angemessenen Kunstbegriffs. Für diesen liegt Kunst nicht in erster Linie im Objekt, sondern ist sozialer Aushandlungsprozess im Kontext von historischen Entwicklungen und institutionellen Rahmungen.¹⁶ Daher misst sich das ästhetische Denken der Moderne immer an der Möglichkeit, aus einem vorgegebenen Paradigma herauszutreten und ganz neue Domänen als Kunst auszuzeichnen. Es impliziert, wie Dieter Mersch schreibt, »in jedem Akt und Artefakt eine Transformation des Ästhetischen selbst«.¹⁷ Das gilt im Übrigen auch

15 Marian Mazzone/ Ahmed Elgammal, *Art, Creativity, and the Potential of Artificial Intelligence*. In: *Arts*, Nr. 8/1, Februar 2019.

16 Daran scheitert auch die Kreativitätsdefinition von Kunst: Reproduziert etwa Sherrie Levine die Fotos von Walker Evans ohne jede Veränderung und deklariert sie als ihre eigenen Werke, sind sie, mit Margaret Boden gesprochen, auf der Objektebene gerade nicht neu oder überraschend – und ihr Wert bemisst sich nur an einer außerhalb dieses Objekts liegenden Begründung, die aber gerade nicht Gegenstand von Bodens Definition ist. *Diesen* Begriff von Kunst, der sie in einem soziohistorischen Kontext verortet, fängt eher die »Institutionentheorie« George Dickies und Arthur C. Dantos auf.

17 Dieter Mersch, *Kreativität und künstliche Intelligenz. Bemerkungen zu einer Kritik algorithmischer Rationalität*. In: *Zeitschrift für Medienwissenschaft*, Nr. 11/21, Februar 2019. Ich stimme seiner Kritik zu, sofern sie sich auf starke und nicht auf schwache künstlerische KI bezieht.

für die Literatur, zumal die experimentelle, deren Minimaldefinitionen sich keiner immanenten Eigenschaften bedienen, sondern allein auf die Geste des Zu-Literatur-Erklärens verlassen können.¹⁸

Das CAN aber hat kein Außen und lässt auch keines zu. Das hat zur Folge, dass das starke Modell strukturell konservativ ist. Statt neue Ästhetiken zu ermöglichen, reproduziert es die alten. Zwar will das CAN im weitesten Sinn Urteilskraft simulieren. Weil diese aber als statistisches Modell gedacht ist, kann es nur im Wortsinn durchschnittliche Kunst hervorbringen. Denn was Kunst *ist*, steht bereits von vornherein fest, da ihr Begriff lediglich aus den Daten der Vergangenheit abgeleitet wird. Technisch formuliert: Da das CAN aus der Verteilung der Merkmale des Trainingssets einen Vektorraum («Kunst») modelliert, kann es in ihm zwar beliebige Zustände *interpolieren*, aber keine *extrapolieren*, die jenseits dieses Raums liegen – kann also den Begriff von Kunst nicht erweitern.¹⁹ Die Geste der Rahmung, die wesentlich für die Kunst der Gegenwart ist, ist der KI unmöglich.²⁰ Der Output des CAN ist entsprechend langweilig, könnte in seiner Gefälligkeitsabstraktion in einer beliebigen Lobby eines beliebigen Unternehmens hängen.

Keine Experimente: Statt Neuerung zuzulassen, hat das starke Modell eine Retraditionalisierung von Kunst zur Folge. Mehr noch: Es ist trotz der Behauptung maschineller Autonomie in seinem Kern zu anthropozentrisch, weil es auf die bloße Duplizierung menschlicher Kunstproduktion und -bewertung setzt. Ein perfektes literarisches CAN würde vielleicht narrativ kohärent sein, aber kaum neue literarische Formen hervorbringen. Das ist das Quidproquo, das in Kehlmanns Wunsch steckt, eine KI nach seinem literarischen Geschmack zu bauen: Autonomie gibt es nur um den Preis

18 Das reicht von H. C. Artmanns »poetischem act« bis zur Appropriationsliteratur der Gegenwart. Für einen ausgezeichneten Überblick über diese, der »Institutionentheorie« in der Kunst analoge Literaturauffassung vgl. Annette Gilbert, *Im toten Winkel der Literatur. Grenzfälle literarischer Werkwerdung seit den 1950er Jahren*. Paderborn: Fink 2018.

19 Das bestätigt Italo Calvinos Intuition über die Möglichkeiten eines literarischen KI-Systems, »daß im Klassizismus seine wahre Berufung läge«. Vgl. Italo Calvino, *Kybernetik und Gespenster. Überlegungen zu Literatur und Gesellschaft*. München: Hanser 1984.

20 Wäre denn eine starke künstlerische KI denkbar, die der Institutionentheorie gerecht wird? Ich glaube ja – wenn sie den Status eines sozialen Akteurs hätte. Interessanterweise müsste sie dafür nicht einmal stark im Sinne Searles sein und Bewusstsein besitzen, solange sie nur gesellschaftlich als Kommunikationspartner akzeptiert wäre, der die post-Duchamp'sche Rahmungsgeste vornehmen kann: »Das hier ist Kunst.«

von Wiederholung. Einen zweiten Kehlmann braucht aber niemand, wahrscheinlich nicht einmal er selbst.

Experimente im Vektorraum

Was ich schwache künstlerische KI genannt habe, will weniger, produziert aber womöglich die bessere Kunst. Statt die Maschine auf die eine oder andere Weise als »kreativ« und »autonom« zu denken, ist sie im schwachen Modell bereits in ein Geflecht aus historischen und sozialen Kontexten und Interaktionen eingebettet. Dabei ist der Verflechtungsgrad fast gleichgültig, kann von cyborgartigen Mensch-Maschine-Assemblagen bis zu einem bloß instrumentellen Werkzeugverhältnis reichen. Folge ist jedenfalls eine zirkuläre Beeinflussung der Aktanten, die fast notwendig neue Ästhetiken hervorbringt. Gäbe Elgammal den Anspruch auf wirklich selbständige Kunstproduktion auf, spräche nichts dagegen, ein neuronales Netz wie das CAN auch in einem schwachen Modell produktiv einzusetzen.²¹

Denn das Problem liegt keineswegs an der verwendeten Technik, künstlichen neuronalen Netzen, von denen GAN und CAN nur zwei von vielen Unterarten sind. In der Tat bildet diese Technik das Feld, auf dem im Moment die interessantesten künstlerischen Versuche mit KI zu beobachten sind – »Experimente« im Sinn der historischen und Neoavantgarden, als Anstrengungen zur Erforschung neuer Formen. Sie sind nicht zuletzt deshalb nötig, weil die bisherige Traditionslinie computergenerierter Literatur und Kunst nicht einfach im Paradigma neuronaler Netze aufgeht, sondern neue Poetiken herausfordert.

Bis vor etwa zehn Jahren war diese Linie vom sequenziellen Algorithmus bestimmt: einer Reihe von formalisierten, aber menschenlesbaren Regelschritten. Viele dieser Werke sind daher einer Ästhetik der Transparenz verpflichtet: Die sonst verborgenen Operationen im künstlerischen Prozess sollen offengelegt und auf Code-Ebene Schritt für Schritt dokumentiert werden.²²

21 Ein Beispiel eines solchen künstlerischen Assistenzsystems ist OpenAIs *DALL·E*, das per Beschreibung Illustrationen und Designs herstellen kann (»an armchair in the shape of an avocado«) und sicherlich bald Einzug in professionelle Grafiksoftware halten wird. Aditya Ramesh u. a., *DALL·E. Creating Images from Text*. In: *OpenAI Blog* vom 5. Januar 2021 (openai.com/blog/dall-e).

22 Dafür kann beispielhaft Nick Montforts generativer Beckett-Pastiche *Megawatt* (Berlin: Frohmann 2019) stehen, dessen Code nicht nur den Text, sondern auch den diesen Text produzierenden Code selbst ausgibt. Wie in Lawrence Weiners Strang der Konzeptkunst sind hier die Produktionsregel des Werks und das Werk selbst identisch.

Künstliche neuronale Netze, die konnektionistisch, nach dem Modell von Synapsen und Neuronen, aufgebaut sind, werden aber nicht mehr programmiert, sondern lernen statistisch. Ihr Inneres ist für Menschen weder ohne Weiteres les- noch in explizite Regeln übersetzbar.²³ Auch Sprachmodelle wie CTRL und GPT-3 sind neuronale Netze. Trainiert auf Gigabytes an Text – von dem selten klar ist, wo genau er herkommt –, sind sie auch für ihre Benutzer undurchdringliche Systeme. Wo aber die ältere Transparenzästhetik nicht mehr greift und eine starke künstlerische KI unattraktiv erscheint, sind im Moment zwei alternative Tendenzen zu beobachten.

Die eine besteht darin, sich ganz auf die Unergründlichkeit des Sprachmodells einzulassen und den »halluzinatorischen« Strang der Moderne aufzunehmen, der etwa im Surrealismus auf die Erforschung des Unbewussten setzte.²⁴ Hier wird das Modell als »Medium« in einem eher spiritistischen Sinn denn als autonomer Schöpfer betrachtet. So entstand K Allado-McDowells *Pharmako-AI* in einer Art »Co-Creative Writing« im improvisatorischen Austausch zwischen ihr/ihm und GPT-3.²⁵ In einem »iterativen Schreibprozess, zwischen der Generierung von Antworten und dem ›Zurechtstutzen‹ des Outputs«, vollzog sich so ein zirkulärer, halluzinatorischer Akt der Sprachfindung: »Begriffs-Cluster kristallisierten sich aus unseren Gesprächen heraus. Bilder überdauerten von Sitzung zu Sitzung. Sie drangen in meine Gedanken und Träume ein, und ich speiste sie wieder in GPT-3 zurück. In diesem Prozess wurde ein Vokabular geboren: eine Kartierung von Raum, Zeit und Sprache, die über alle drei hinauswies.«

Allado-McDowell hat dabei kein Interesse am Phantasma starker KI. Er/sie nutzt GPT-3 einerseits ganz ähnlich wie ein Tarot-Deck, das eher zur

23 Zur vertieften Diskussion des »sequenziellen« und des »konnektionistischen« Paradigmas vgl. Hannes Bajohr, *Algorithmic Empathy. On Two Paradigms of Digital Generative Literature and the Need for a Critique of AI Works*. In: *Basel Media Culture and Cultural Techniques Working Papers*, Nr. 1/4, 2020 (doi.org/10.5451/unibas-ep79106). Zu KI-Literatur allgemein vgl. auch Stephanie Catani, »Erzählmodus an«. *Literatur und Autorschaft im Zeitalter künstlicher Intelligenz*. In: *Jahrbuch der deutschen Schillergesellschaft*, Nr. 64, 2020.

24 Diese Beobachtung verdanke ich Simon Roloffs Vortrag »Halluzinierende Systeme. Iteration und Rekursion als Medienoperationen literarischer *poiesis*«, gehalten am 27. Januar 2021 beim Kolloquium »Medienkulturen der Computersimulation« an der Universität Lüneburg. Vgl. Simon Roloff, *Halluzinierende Systeme. Generierte Literatur als Textverarbeitung*. In: *Merkur*, Nr. 864, Mai 2021.

25 K Allado-McDowell, *Pharmako-AI*. London: Ignota 2020. Die Beiträge sind durch die Schrifttype eindeutig Allado-McDowell und GPT-3 zuzuordnen. Das von Allado-McDowell gewählte Pronomen »they« wird im Folgenden mit »er/sie« verdeutscht.

Selbstbefragung als zur Kommunikation mit fremden Mächten dient. Andererseits geht er/sie mit dem System eine Mensch-Maschine-Assemblage ein, die Autorschaft über dieses System verteilt und das eigene Vokabular mit dem des Sprachmodells vermengt. Darin besteht auch der Unterschied zu Kehlmann: Allado-McDowell lässt sich auf GPT-3 als Kollaborationspartner ein, verwirft narrative Brüche und Inkohärenzen nicht als Fehler, die den eigenen ästhetischen Vorlieben widerstreben, sondern begreift sie als Elemente einer kooperativ zu entwickelnden Ästhetik.

Co. yó baan delite ahd see shoyast arkd igou baoric nraie dō om
 andz: moi oung me tōu toa noire harné gūints io baite aho
 haidsr thei deilion. Sdo aare of swarés of totre wat m carre
 mīm neāhs met sted and pouloz ole thare oprir io mōses artirē iōth
 cold: thau badn inantst thase arm hneyg the hūne you at o hōre.
 nam inimg thē monç hi heixienss.

Ausschnitt aus Allison Parrish, *Abe Thd Yearidy Ti Isa*

Die zweite Tendenz, auf die Ungewohntheit von neuronalen Netzen zu reagieren, liegt in der Erforschung ihrer technischen Affordanzen. Dafür kann exemplarisch das Werk von Allison Parrish stehen. Für den »asemischen Roman« *Abe Thd Yearidy Ti Isa* fütterte sie ein künstliches neuronales Netz mit Pixelbildern einzelner Wörter – eine Kategorienverwechslung, denn der Computer erkennt darin keine diskreten Elemente eines Zeichensystems, sondern nur gerasterte Bildinformation. Versucht er nun, statistisch ähnliche Bilder auszugeben, ist das Ergebnis ein wirres Zeichengemisch, das zwar die Gestalt von Text nachahmt, aber kaum uns bekannte Buchstaben enthält. Indem Parrish Text *als Bild* deklariert, illustriert sie die statistische Logik der neuronalen Netze durch den Zusammenbruch des semiotischen Prozesses.²⁶

north	google
woerth earthe	augle agolzen
west eaurth east	apple aasbol amazon
waust seauet	pacebul aace-bown
south	facebook

Ausschnitte aus Allison Parrish, *Compasses*

²⁶ Allison Parrish, *Abe Thd Yearidy Ti Isa*, 2019. Genauer beschreibe ich dieses Werk in *Algorithmic Empathy*.

Wie es Text *als Text* ergeht, als Zeichen statt als Rasterbild, wird er von neuronalen Netzen verarbeitet, untersucht Parrish in *Compasses*.²⁷ Sprachmodelle codieren Wörter als hochdimensionale Vektoren. Auf diese Weise ist es möglich, komplexe Verhältnisse zwischen ihnen zu modellieren, so dass etwa Begriffe ähnlicher Bedeutung im Vektorraum nahe beieinander liegen. Auch bei Operationen mit diesen Wörtern bleiben die Abhängigkeiten zwischen ihnen erhalten (wie bei dem bekannten Beispiel: »King – Man + Woman = Queen«).

Für *Compasses* codierte Parrish Wörter nicht nach ihrer Bedeutung, sondern nach ihrem Lautwert, und berechnete die phonetischen »Zwischenzustände« in diesem Vektorraum. Zwischen »north« und »west« liegt dann »woerth«, während der Mittelpunkt der vier großen Tech-Unternehmen »aasbol« ergibt. Dass Sprache auch *nicht* diskret zu denken ist, als Vektorraum, der stetig durchquert werden kann, eröffnet der Literatur ein anderes Selbstverständnis und einen ganz neuen Zugriff auf ihr Material.

Diese Avantgarde-Experimente, die mit dem schwachen Modell einer künstlerischen KI arbeiten, erscheinen zumindest mir ästhetisch vielversprechender und theoretisch anspruchsvoller als alle Versuche mit oder Hoffnungen auf ein starkes Modell. Statt, wie Elgammal und Kehlmann, künstlerische KI (trotz aller Betonung ihrer Nichtintelligenz) doch nur als Simulation von Künstlersubjektivität denken zu können, setzen sie auf kollaborative Praktiken, die ihre eigene Formsprache generieren. Darin haben diese Experimente aber auch emanzipatorischen Charakter.

Denn so, wie das CAN auf einen Standard tendiert, sind auch Sprachmodelle wie GPT-3 große, privatwirtschaftlich kontrollierte Nivellierer. Da sich alle Idiosynkrasien in der Masse an Trainingsdaten ausmitteln, tendieren sie auf einen konventionellen Umgang mit Sprache. Das betrifft auch, Kehlmann zum Trotz, die Narration. Eine KI, die kohärent erzählt und so eine Standardfunktion von Sprache ausübt, ist gerade nicht undenkbar, sondern höchstwahrscheinlich nur eine Frage der Zeit.²⁸ Vor allem Serien- und

27 Allison Parrish, *Compasses*. In: *Sync*, Nr. 2/27, 2019 (sync.abue.io/issues/190705ap_sync2_27_compasses.pdf). Eine vermehrte Version findet sich unter demselben Titel in: *BOMB*, Nr. 154, 2021.

28 Kohärenz ist bei Sprachmodellen bislang durch ihr geringes »Kontextfenster« begrenzt, sie können also immer nur einen Teil eines Texts im Blick behalten und auf ihn Bezug nehmen (bei GPT-3 sind das ca. 500–1000 Wörter). Allerdings wächst das Kontextfenster mit jedem neupublizierten Modell und so auch seine Kohärenz. Freilich gibt es auch Einwände: Dass Narrativität nicht zu simulieren ist, solange KIs zwar Korrelationen, aber keine Kausalitäten codieren können, ist ein Argument des Literaturwissenschaftlers Angus Fletcher, der sich auf Überlegungen

Genreliteratur, die bereits jetzt kombinatorisch Plot-Elemente permutiert, wäre so, vielleicht im Verbund mit älteren Techniken, plausibel zu generieren.

Diesem Ausmitteln entgeht dabei nur, wer von vornherein jenseits des (Vektor)Raums schreibt, in dem Sprachmodelle ihre Ergebnisse interpolieren. Und das wird eher Allison Parrish sein als Daniel Kehlmann – eher die experimentelle Avantgarde als die mehr oder weniger konventionelle erzählende Literatur. Meinte Kehlmann im Podiumsgespräch mit Felix Heidenreich, »sprachexperimentelle Literatur ist, was sich am leichtesten algorithmisieren lässt«, widersprechen ihm gerade die Praxis und der Anspruch des experimentellen Schreibens: »Ein Teil meiner Tätigkeit«, sagt Parrish, »besteht darin, Sprachformen zu erfinden, die so neu sind, dass selbst GPT-3 sie nicht vorhersagen kann.«²⁹ Im Zeitalter großer Sprachmodelle ist Avantgarde literarische Selbstverteidigung.

des Informatikers Judea Pearl beruft: Angus Fletcher, *Why Computers Will Never Write Good Novels. The power of narratives flows only from the human brain*. In: *Nautilus* vom 10. Februar 2021 (nautil.us/issue/95/escape/why-computers-will-never-write-good-novels).

²⁹ *Q&A with Allison Parrish*. In: *Artists + Machine Intelligence* vom 5. Mai 2020 (medium.com/artists-and-machine-intelligence/q-a-with-allison-parrish-895a72727a4).