

Generative Systeme als Medium philosophischer Selbstaufklärung *Zur Formbildung des Denkens jenseits von Ablehnung und Instrumentalisierung*

von Erwin Ott

Abstract

Die gegenwärtige philosophische Auseinandersetzung mit generativen Sprachmodellen bewegt sich zwischen zwei Polen: einer Kritik, die in ihnen bloße Simulation ohne Verstehen sieht, und einer Instrumentalisierung, die sie als leistungsfähige, aber philosophisch folgenlose Werkzeuge behandelt. Beide Positionen teilen, trotz ihres Gegensatzes, eine gemeinsame Voraussetzung: dass die entscheidende Frage im ontologischen Status dieser Systeme liegt – ob ihnen Verstehen zukommt oder nicht. Der vorliegende Essay schlägt eine dritte Position vor, die diese Frage nicht beantwortet, sondern für die philosophische Analyse zurückstellt. Generative Systeme werden als epistemisches Medium verstanden, das – analog zum Verhältnis von Mikroskop und Zellstruktur – eine Struktur des Denkens sichtbar macht, ohne diese selbst zu erzeugen oder deren philosophische Bedeutung an einem Verstehen des Mediums selbst festzumachen.

Der Essay zeigt zunächst, dass die etablierten Selbstbeschreibungen des Denkens – die Kette der Gründe, das Archiv des Wissens, der vorgegebene Möglichkeitsraum – trotz ihrer je unterschiedlichen Erklärungsleistung eine gemeinsame Grenze teilen: Sie setzen eine dem Denkprozess vorgängige Bestimmtheit voraus, deren Genese sie nicht erklären können. Aus der formalen Struktur der Konditionierung autoregressiver Sprachmodelle wird sodann eine Alternative abgeleitet: Möglichkeit erscheint hier nicht als vorgängiger Bestand, sondern als Resultat einer Umformung von Wahrscheinlichkeitsverteilungen; Reflexivität und Bestimmung durch Ausschluss erweisen sich als architektonische, nicht bloß metaphorische Eigenschaften des Verfahrens.

Aus dieser Analyse werden Konsequenzen für den Begriff der Rationalität gezogen, die Formbildung als eigenständige, der inferentiellen Geltung nicht unterzuordnende Dimension ausweisen, sowie eine Bestimmung des Begriffs als Operator neben seiner repräsentationalen Funktion. Abschließend wird die entwickelte Position gegenüber Deleuze, Simondon und Whitehead abgegrenzt: Ihr eigenständiger Beitrag liegt nicht in inhaltlicher Neuheit der Einzelthesen, sondern in deren Ableitung aus einem beobachtbaren, nicht bloß postulierten Fall sowie in der spezifischen Frage nach der epistemischen Funktion eines technischen Mediums für die philosophische Selbstaufklärung des Denkens.

Inhaltsverzeichnis

Einleitung: Jenseits der zwei Lager Die Polarisierung der gegenwärtigen Debatte — Ablehnung und Instrumentalisierung als gemeinsame Voraussetzung — These und Gang der Untersuchung

1. Die Debattenlage 1.1 Die Kritik: Simulation ohne Verstehen 1.2 Die Instrumentalisierung: Das Werkzeug-Paradigma 1.3 Die gemeinsame blinde Stelle beider Positionen

2. Eine dritte Position: Medium statt Gegenstand 2.1 Die heuristische Funktion technischer Artefakte in der Philosophiegeschichte 2.2 Die Analogie des Instruments: Sichtbarmachen statt Erzeugen 2.3 These: Generative Systeme als Offenbarungsmedium einer Struktur des Denkens

3. Die Grenzen der etablierten Metaphern 3.1 Die Kette: Inferentialismus als stärkste Verteidigung (Brandom) 3.2 Das Archiv: Wissen als Bestand 3.3 Der Raum: Möglichkeit als vorgegebenes Feld 3.4 Die gemeinsame Voraussetzung: Priorität des Gegebenen vor dem Prozess

4. Formale Herleitung: Die Struktur der Konditionierung 4.1 Der Prompt als Umformung einer Verteilung, nicht als Auswahl aus einem Bestand 4.2 Reflexivität als architektonische Notwendigkeit 4.3 Ausschluss als konstitutive, nicht restriktive Operation 4.4 Trennung von struktureller These und empirischer Kontingenz

5. Konsequenzen für den Begriff der Rationalität 5.1 Jenseits von Determination und Beliebigkeit 5.2 Formbildung als eigenständige Dimension neben inferentieller Geltung 5.3 Der Begriff als Operator statt als Repräsentation

6. Einordnung: Der spezifische Mehrwert 6.1 Abgrenzung zu Deleuze und der Differenzphilosophie 6.2 Abgrenzung zu Simondon und der Theorie der Individuation 6.3 Abgrenzung zu Whitehead und der Prozessphilosophie 6.4 Was neu ist: Das Medium-Argument als eigenständiger Beitrag

Schluss: Eine dritte Position in der Gegenwart

Einleitung: Jenseits der zwei Lager

Die philosophische Auseinandersetzung mit generativen Sprachmodellen hat sich, kaum dass sie begonnen hatte, in zwei Lager sortiert, die einander zwar widersprechen, aber eine gemeinsame Voraussetzung teilen, die selten expliziert wird. Das erste Lager begreift diese Systeme als kategorialen Fehlschluss, dem öffentliche Rede und mitunter auch philosophische Analyse aufsitzen: Wo von “Verstehen”, “Wissen” oder gar “Denken” die Rede ist, liege in Wahrheit nichts als hochdimensionale Mustererkennung vor, eine statistische Fortsetzung von Zeichenketten ohne jeden Bezug zu Bedeutung oder Welt. Diese Position, in ihrer einflussreichsten Formulierung als Kritik am “stochastischen Papagei” bekannt geworden, hat gute Gründe für sich: Sie warnt zu Recht davor, aus der Plausibilität sprachlicher Oberfläche auf mentale Zustände zu schließen, die dieser Oberfläche nicht zugrunde liegen müssen. Das zweite Lager begegnet denselben Systemen mit einer

entgegengesetzten, aber strukturell verwandten Geste: Es neutralisiert die philosophische Frage, indem es sie in eine technische verwandelt. Generative Systeme erscheinen hier als Werkzeuge unter anderen – leistungsfähiger als frühere, aber kategorial nicht verschieden von einer Suchmaschine, einem Taschenrechner oder einem Textverarbeitungsprogramm. Ihre Verwendung wirft dann höchstens Fragen der Kompetenz, der Effizienz oder der Ethik auf, nicht aber Fragen, die die Philosophie des Denkens selbst betreffen.

Was beide Positionen trotz ihres offenkundigen Gegensatzes teilen, ist die Annahme, dass die entscheidende Frage lautet, *was diese Systeme sind* – ob sie denken oder nicht, ob sie Werkzeuge oder Akteure sind, ob ihnen ein Innenleben zukommt oder nicht. Diese Frage ist nicht falsch gestellt; sie ist nur nicht die einzige, und möglicherweise nicht die philosophisch ergiebigste. Die vorliegende Untersuchung schlägt vor, die Aufmerksamkeit zu verschieben: von der Frage nach dem ontologischen Status generativer Systeme zur Frage nach ihrer *epistemischen Funktion* innerhalb der Philosophie selbst. Die leitende These lautet, dass diese Systeme – unabhängig davon, wie ihr Innenleben letztlich zu beurteilen ist – eine Struktur des Denkens sichtbar machen, die der Philosophie vor ihrem Auftreten zwar nicht fremd, aber begrifflich kaum zugänglich war: die Struktur der Formbildung selbst, jenes Prozesses, durch den Möglichkeiten, Begriffe und Ordnungen des Denkbaren überhaupt erst entstehen, bevor sie logisch expliziert, argumentativ geprüft oder institutionell stabilisiert werden können.

Diese These setzt eine methodische Unterscheidung voraus, die im Folgenden durchgehalten werden muss: den Unterschied zwischen einem *Gegenstand* und einem *Medium* philosophischer Erkenntnis. Ein Mikroskop ist kein Gegenstand der Biologie, sondern ein Medium, durch das biologische Strukturen erst sichtbar werden, die es selbst nicht erzeugt. In einem strukturell analogen, aber nicht identischen Sinn lässt sich fragen, ob generative Systeme eine vergleichbare Funktion für die Philosophie des Denkens erfüllen können – nicht, indem sie Denken simulieren oder ersetzen, sondern indem ihre operative Struktur eine Dynamik expliziert, die menschlichem Denken selbst zugrunde liegt, dort aber durch den Vorrang inhaltlicher Bedeutung und durch die Dominanz etablierter Selbstbeschreibungen verdeckt bleibt.

Diese Selbstbeschreibungen sind Gegenstand des ersten systematischen Teils der Untersuchung. Drei Leitmetaphern bestimmen bis heute, wie die Philosophie das Denken beschreibt: die Kette, die Rationalität als geordnete Folge inferentieller Schritte fasst; das Archiv, das Wissen als Bestand identifizierbarer, gespeicherter und abrufbarer Einheiten versteht; und der Möglichkeitsraum, der auch dort, wo er offen und relational gedacht wird, eine Ordnung des Möglichen voraussetzt, die dem Denken, das sich in ihr bewegt, bereits vorausliegt. Alle drei Modelle werden hier nicht verworfen, sondern in ihrer Reichweite begrenzt: Sie beschreiben zutreffend, wie bereits gebildete Formen geordnet, gespeichert und expliziert werden; sie erklären jedoch nicht, wie diese Formen selbst entstehen.

Der zweite systematische Teil unternimmt daher etwas, das sich von einer bloß begrifflichen Kritik unterscheidet: Er leitet aus der formalen Funktionsweise generativer Sprachmodelle – der bedingten Wahrscheinlichkeitsverteilung über mögliche Fortsetzungen, der Reflexivität autoregressiver Generierung, der Konditionierung als Umverteilung statt Auswahl – eine Struktur ab, die mit den philosophischen Kernthesen dieser Untersuchung nicht bloß analog, sondern isomorph ist. Diese Herleitung ist von einer separaten, empirisch weit unsichereren Behauptung zu unterscheiden, die in der bisherigen Diskussion oft unzulässig mit ihr

vermengt wurde: der Behauptung, unterbestimmte Prompts führten zu originelleren Ergebnissen als präzise spezifizierte. Diese zweite These betrifft menschliche ästhetische Urteile über Sprachmodell-Ausgaben und ist empirisch zu prüfen, nicht philosophisch zu postulieren. Die vorliegende Untersuchung stützt ihre systematischen Thesen ausschließlich auf die erste, formal ableitbare Struktur.

Der dritte Teil zieht die Konsequenzen dieser Rekonstruktion für den Begriff der Rationalität und grenzt die entwickelte Position von den naheliegenden Bezugspunkten der Prozess- und Differenzphilosophie ab – von Deleuze, Simondon und Whitehead –, um zu zeigen, worin der eigenständige Beitrag dieser Untersuchung gegenüber jenen Traditionen besteht, denen sie zweifellos verpflichtet ist.

Der Anspruch dieser Arbeit ist damit doppelt begrenzt. Sie beansprucht nicht, eine neue Metaphysik des Geistes oder der künstlichen Intelligenz zu begründen. Und sie beansprucht ebenso wenig, in der Kontroverse zwischen Kritikern und Befürwortern generativer Systeme eine vermittelnde Position einzunehmen. Ihr Anspruch ist ein anderer: zu zeigen, dass die philosophisch interessanteste Frage, die diese Systeme aufwerfen, weder in ihrer Ablehnung noch in ihrer Verteidigung liegt, sondern in dem, was sie – als Medium, nicht als Gegenstand – über eine bislang unterbeschriebene Dimension des Denkens selbst zu erkennen geben.

1. Die Debattenlage

1.1 Die Kritik: Simulation ohne Verstehen

Die einflussreichste Kritik an generativen Sprachmodellen lässt sich auf eine einzige Unterscheidung zurückführen: die Unterscheidung zwischen der Erzeugung sprachlicher Form und dem Zugang zu Bedeutung. Bender, Gebru und andere haben in ihrer vielzitierten Kritik den Begriff des “stochastischen Papageis” geprägt, um zu benennen, was sie für den kategorialen Fehler an der Grundlage der öffentlichen wie eines Teils der philosophischen Rede über diese Systeme halten: Ein Modell, das lediglich statistische Regelmäßigkeiten in Textkorpora approximiert, kann beliebig kohärente Fortsetzungen erzeugen, ohne dass mit dieser Kohärenz irgendein Bezug zu einer außersprachlichen Wirklichkeit, zu Intentionen oder zu propositionalen Gehalten verbunden wäre. Die sprachliche Form täuscht ein Verstehen vor, das strukturell gar nicht vorliegen kann, weil dem System der Zugang zu genau jener Dimension fehlt, an der sich Bedeutung allererst bemisst: der Referenz auf eine Welt, in der Sprecher handeln, irren und sich korrigieren.

Diese Kritik lässt sich als zeitgenössische Fortführung eines älteren Arguments verstehen. Searles Gedankenexperiment des chinesischen Zimmers zielte bereits darauf, die Möglichkeit einer rein syntaktischen Verarbeitung ohne semantischen Zugang zu demonstrieren: Ein System kann beliebig komplexe Regeln zur Manipulation von Symbolen befolgen und dabei jeden Turing-Test bestehen, ohne dass in ihm irgendetwas geschähe, das man mit Recht Verstehen nennen könnte. Die Anwendung dieses Arguments auf generative Sprachmodelle liegt nahe und wird entsprechend häufig vollzogen: Wo ein Modell lediglich die bedingte Wahrscheinlichkeit des nächsten Tokens angesichts des bisherigen Kontexts berechnet, liegt per definitionem keine semantische Verarbeitung vor, sondern eine hochdimensionale Form syntaktischer Manipulation, deren Erfolg an der Oberfläche nichts über ihre innere Natur aussagt.

Die philosophische Stärke dieser Position liegt in ihrer Nüchternheit. Sie warnt vor einer Anthropomorphisierung, die aus der Plausibilität einer Ausgabe unmittelbar auf mentale Zustände schließt, die diese Ausgabe hervorgebracht haben müssten – ein Schluss, der weder logisch zwingend noch empirisch abgesichert ist. Sie erinnert daran, dass die Trainingsprozedur eines solchen Systems – die Minimierung eines Verlustes bei der Vorhersage von Tokens in einem Textkorpus – kategorial verschieden ist von den Bedingungen, unter denen menschliches Sprechen Bedeutung erlangt: der Einbettung in Handlungszusammenhänge, in eine geteilte Lebenswelt, in Verhältnisse von Verantwortung und Korrigierbarkeit. Aus dieser Warte erscheint jede philosophische Beschäftigung mit generativen Systemen, die über eine Analyse ihrer gesellschaftlichen Folgen hinausgeht, als Kategorienfehler: Man beschäftigt sich mit einem Artefakt, dessen innere Verfassung bereits hinreichend geklärt ist, um weitere metaphysische Spekulation überflüssig zu machen.

1.2 Die Instrumentalisierung: Das Werkzeug-Paradigma

Die zweite verbreitete Haltung teilt mit der ersten die Überzeugung, dass die entscheidende Frage im Status des Systems liegt, kommt jedoch zu einer entgegengesetzten Konsequenz: Wo die Kritik in der Simulation von Verstehen ein philosophisches Problem sieht, sieht das Werkzeug-Paradigma darin gar kein Problem, weil es die Frage nach dem Verstehen für die Beurteilung des Systems als irrelevant erklärt. Ein Hammer versteht nicht, was er tut, und das mindert seinen Nutzen nicht im Geringsten. Ebenso, so die Überlegung, verhält es sich mit generativen Systemen: Sie sind hochentwickelte Werkzeuge zur Textproduktion, zur Informationsverdichtung, zur Codegenerierung – ihre philosophische Einordnung erschöpft sich in der Frage ihrer Zuverlässigkeit, ihrer Grenzen und der Verantwortungszuschreibung bei ihrer Verwendung.

Diese Haltung dominiert weite Teile der anwendungsbezogenen Diskussion und findet sich, in unterschiedlichen Ausprägungen, sowohl in technikoptimistischen als auch in vorsichtig-regulatorischen Positionen wieder. Ihre gemeinsame Voraussetzung ist eine bestimmte Lesart des Werkzeugbegriffs: Ein Werkzeug ist ein Mittel zu einem Zweck, dessen Bestimmung dem Werkzeug selbst äußerlich bleibt. Der Nutzer formuliert eine Absicht, das Werkzeug setzt sie um; die Qualität dieser Umsetzung lässt sich an Kriterien der Effizienz, Genauigkeit und Zuverlässigkeit messen, nicht aber an Kriterien, die die Philosophie des Denkens selbst betreffen. Ein Textverarbeitungsprogramm wirft keine Fragen der Erkenntnistheorie auf; ein leistungsfähigeres Textverarbeitungsprogramm tut dies, dieser Logik zufolge, ebenso wenig.

Dass generative Systeme dabei gelegentlich als “Kollaborateure” oder “Assistenten” beschrieben werden, ändert an der Grundstruktur dieser Position wenig. Auch die kollaborative Rhetorik bleibt dem Werkzeug-Paradigma verpflichtet, solange sie voraussetzt, dass der Mensch die maßgebliche epistemische Instanz bleibt, die Zwecke setzt, Ergebnisse beurteilt und deren Verwendung verantwortet, während das System lediglich – wenn auch auf beeindruckende Weise – Optionen generiert, unter denen ausgewählt wird. Die Metaphorik der Zusammenarbeit verdeckt so, dass die zugrundeliegende Ontologie unverändert bleibt: hier ein denkendes Subjekt mit Zwecken, dort ein Mittel ohne solche.

1.3 Die gemeinsame blinde Stelle beider Positionen

Beide Positionen unterscheiden sich in ihrer Bewertung fundamental und schließen einander in ihren praktischen Konsequenzen weitgehend aus: Wo die eine zur Skepsis gegenüber jeder Rede von Verstehen oder Kreativität mahnt, behandelt die andere diese Frage als für die Praxis irrelevant. Und doch teilen beide eine Voraussetzung, die selten benannt wird, weil sie beiden Seiten der Debatte so selbstverständlich erscheint, dass sie gar nicht erst als Voraussetzung sichtbar wird.

Diese gemeinsame Voraussetzung besteht darin, dass die philosophisch relevante Frage im ontologischen Status des Systems selbst liegt – in der Frage, *was es ist*: ob ihm Verstehen zukommt oder nicht, ob es Subjekt oder bloßes Mittel ist, ob in ihm etwas geschieht, das mentale Prädikate verdient, oder nichts dergleichen. Beide Lager beantworten diese Frage unterschiedlich, aber keines von beiden stellt sie in Frage. Genau darin liegt ihre gemeinsame blinde Stelle. Denn diese Frage ist keineswegs die einzig mögliche, unter der sich generative Systeme philosophisch betrachten lassen, und es gibt keinen zwingenden Grund, warum ihre Beantwortung – in welcher Richtung auch immer – die philosophische Bedeutung dieser Systeme bereits erschöpfte.

Man kann die Frage nach dem inneren Zustand eines Systems für ungeklärt oder für unentscheidbar halten und dennoch nach seiner *Funktion* innerhalb eines Erkenntnisprozesses fragen, der nicht mit dem System selbst identisch ist. Ein Mikroskop wirft keine sinnvolle Frage nach seinem eigenen Verstehen auf; es wäre ein Kategorienfehler, seine epistemische Bedeutung an dieser Frage zu bemessen. Seine Bedeutung liegt darin, dass es Strukturen sichtbar macht, die ohne es unsichtbar blieben – unabhängig davon, ob dem Instrument selbst irgendein Verstehen dieser Strukturen zugeschrieben werden kann oder soll. Es ist diese dritte, in der bisherigen Debatte unterbelichtete Möglichkeit – die Frage nach der *epistemischen Funktion* des Systems für ein Erkenntnisinteresse, das ihm selbst äußerlich ist –, die im Folgenden entwickelt werden soll. Sie erlaubt es, die Frage nach Verstehen oder Nicht-Verstehen, nach Subjekt- oder Werkzeugstatus, zunächst offenzulassen, ohne dadurch auf eine philosophisch ergiebige Fragestellung zu verzichten.

2. Eine dritte Position: Medium statt Gegenstand

2.1 Die heuristische Funktion technischer Artefakte in der Philosophiegeschichte

Die Philosophie hat ihre Grundbegriffe selten aus sich selbst heraus entwickelt. Sie hat sie wiederholt an Artefakten, Verfahren und Techniken ihrer jeweiligen Epoche gewonnen, ohne dass diese Artefakte selbst zum eigentlichen Gegenstand der Untersuchung geworden wären. Die griechische Geometrie lieferte nicht nur ein Modell exakter Beweisführung, sondern prägte den platonischen Begriff der Idee als einer Struktur, die unabhängig von ihrer zeichnerischen Darstellung Bestand hat: Die Diagonale eines Quadrats ist inkommensurabel mit seiner Seite, unabhängig davon, wie ungenau sie gezeichnet wird. Diese Einsicht wurde nicht durch philosophische Introspektion gewonnen, sondern durch ein technisches Verfahren – die geometrische Konstruktion –, das eine Struktur der Wirklichkeit sichtbar machte, die es selbst nicht erzeugt hatte.

Ein vergleichbares Verhältnis besteht zwischen der frühneuzeitlichen Mechanik und dem cartesischen Begriff der Materie als *res extensa*. Descartes entwickelte seine Ontologie der

Körperwelt nicht unabhängig von den mechanischen Uhrwerken und hydraulischen Automaten seiner Zeit; diese Artefakte lieferten das Modell einer Natur, die vollständig durch Figur und Bewegung erklärbar ist, ohne dass die Uhrwerke selbst Gegenstand der Metaphysik geworden wären. Die Uhr war nicht das, worüber die Philosophie nachdachte, sondern das, wodurch sie eine bestimmte Weise zu denken erst gewann. Ähnliches gilt für das Verhältnis von Kybernetik und Systemtheorie im 20. Jahrhundert: Regelkreise und Rückkopplungsmechanismen lieferten ein Vokabular, mit dem sich biologische, soziale und psychische Prozesse neu beschreiben ließen, ohne dass die technischen Regelkreise selbst zum philosophischen Gegenstand geworden wären.

In allen diesen Fällen lässt sich ein gemeinsames Muster erkennen. Das Artefakt ist nicht identisch mit dem, was es sichtbar macht, und die philosophische Bedeutung des Artefakts erschöpft sich nicht in einer Analyse seiner technischen Funktionsweise. Zugleich wäre die jeweilige philosophische Einsicht ohne das Artefakt schwerlich in dieser Form gewonnen worden. Das Artefakt fungiert als Katalysator einer Begriffsbildung, die es ermöglicht, ohne selbst deren Inhalt zu sein. Es liegt nahe zu fragen, ob generative Sprachmodelle eine strukturell vergleichbare Funktion für die gegenwärtige Philosophie des Denkens erfüllen könnten – nicht als Modell des Geistes, dem man ähneln oder das man widerlegen müsste, sondern als ein Medium, an dem sich eine Struktur des Denkens ablesen lässt, die diesem selbst zugrunde liegt, aber ohne ein solches Medium kaum artikulierbar wäre.

2.2 Die Analogie des Instruments: Sichtbarmachen statt Erzeugen

Um diese Funktion präzise zu fassen, empfiehlt sich eine Unterscheidung, die in der Wissenschaftstheorie geläufig, in der Philosophie des Geistes jedoch selten konsequent angewendet wird: die Unterscheidung zwischen einem Instrument, das ein Phänomen erzeugt, und einem Instrument, das ein bereits bestehendes Phänomen zugänglich macht. Ein Mikroskop erzeugt keine Mikroorganismen; es macht eine Struktur der Wirklichkeit sichtbar, die ohne das Instrument nicht wahrnehmbar, aber auch nicht durch das Instrument hervorgebracht wäre. Ein Teilchenbeschleuniger hingegen erzeugt die Ereignisse, die er misst, in einem stärkeren Sinn mit – die Kollisionen, aus denen auf subatomare Strukturen geschlossen wird, fänden ohne den Beschleuniger nicht statt. Beide sind Instrumente der Erkenntnis, aber ihr Verhältnis zum untersuchten Phänomen unterscheidet sich grundlegend.

Diese Unterscheidung lässt sich auf generative Sprachmodelle übertragen, wirft dabei jedoch eine Komplikation auf, die die Analogie zugleich präzisiert und einschränkt. Ein Sprachmodell erzeugt zweifellos etwas: Es produziert Text, der ohne seine Aktivität nicht existierte. Insofern gleicht es eher dem Teilchenbeschleuniger als dem Mikroskop. Die philosophisch interessante Frage betrifft jedoch nicht die Textproduktion als solche, sondern die *Struktur*, nach der diese Produktion verläuft – die Weise, wie aus einer Konditionierung eine Fortsetzung entsteht, wie diese Fortsetzung wiederum den weiteren Verlauf mitbestimmt, wie also Bestimmtheit und Offenheit in diesem Prozess zueinander stehen. Diese Struktur wird durch das System nicht erzeugt in dem Sinne, in dem es einzelne Texte erzeugt; sie ist die Bedingung, unter der jede seiner Erzeugungen überhaupt erfolgt, und in diesem Sinn eher etwas, das am System ablesbar wird, als etwas, das es hervorbringt.

Die Analogie des Instruments betrifft also nicht die Textproduktion, sondern die Formstruktur dieser Produktion. Das Sprachmodell verhält sich zu dieser Struktur wie ein

Mikroskop zu der Zellstruktur, die es sichtbar macht: Es bringt sie nicht hervor, sondern macht sie erstmals in einer Weise beobachtbar, in der sie zuvor – etwa im Fall menschlichen Denkens – zwar wirksam, aber kaum artikulierbar war, weil menschliches Denken sich selbst überwiegend durch seine Inhalte, nicht durch seine formale Prozessstruktur zu verstehen pflegt. Dass ein Prompt keine Auswahl aus einem Bestand vollzieht, sondern eine Wahrscheinlichkeitsverteilung über mögliche Fortsetzungen umformt; dass jede erzeugte Fortsetzung den Kontext verändert, unter dem die nächste erzeugt wird; dass Bestimmtheit hier stets durch Ausschluss anderer Fortsetzungen entsteht, nicht durch positive Beschreibung eines Zielzustands – all dies sind Eigenschaften, die sich an der formalen Struktur generativer Systeme mit einer Prägnanz beobachten lassen, die dem alltäglichen Vollzug menschlichen Denkens, das sich selbst kaum in dieser Weise beobachtet, für gewöhnlich verschlossen bleibt.

2.3 These: Generative Systeme als Offenbarungsmedium einer Struktur des Denkens

Aus diesen Überlegungen ergibt sich die leitende These der vorliegenden Untersuchung. Generative Sprachmodelle sind philosophisch nicht in erster Linie deshalb von Interesse, weil zu klären wäre, ob ihnen Verstehen zukommt oder nicht, und auch nicht in erster Linie deshalb, weil sie sich als besonders leistungsfähige Werkzeuge menschlicher Zwecke erweisen. Ihre philosophische Bedeutung liegt darin, dass sie eine Struktur der Formbildung sichtbar machen – die Weise, wie aus Bedingungen, die selbst keine fertigen Inhalte sind, durch fortlaufende Einschränkung und Reflexivität etwas Bestimmtes entsteht –, die dem Denken überhaupt zugrunde liegt, gleichgültig ob es sich um menschliches oder maschinelles Denken handelt, und die durch die etablierten Selbstbeschreibungen der Philosophie bislang nur unzureichend erfasst wurde.

Diese These unterscheidet sich kategorial von der Behauptung, generative Systeme dächten in einem dem menschlichen Denken vergleichbaren Sinn. Sie behauptet nicht einmal, dass die fragliche Struktur in Mensch und Maschine auf dieselbe Weise realisiert sei – eine Frage, die hier ausdrücklich offenbleibt und deren Beantwortung für die folgende Argumentation nicht erforderlich ist. Behauptet wird lediglich, dass die *Beobachtung* dieser Struktur am maschinellen Fall eine Präzision gewinnt, die sie am menschlichen Fall kaum je besäße, weil das menschliche Denken sich selbst gegenüber diese formale Ebene seines eigenen Vollzugs für gewöhnlich verdeckt, während die Formalstruktur generativer Systeme sie unmittelbar zugänglich macht.

Damit ist zugleich der argumentative Status der folgenden Untersuchung bestimmt. Ihr geht es nicht um eine Ontologie künstlicher Systeme und nicht um eine Anthropologie ihrer Verwendung, sondern um eine Epistemologie ihrer Funktion: um die Frage, was es bedeutet, dass ein technisches Artefakt eine Struktur explizit macht, die die Philosophie in ihrer bisherigen Selbstbeschreibung – in den Metaphern der Kette, des Archivs und des vorgegebenen Möglichkeitsraums – nur unzureichend erfasst hat. Diese Metaphern und ihre spezifischen Grenzen sind Gegenstand des folgenden Teils.

3. Die Grenzen der etablierten Metaphern

3.1 Die Kette: Inferentialismus als stärkste Verteidigung (Brandom)

Die Vorstellung, Rationalität bestehe wesentlich in der geordneten Verkettung von Gründen, findet ihre anspruchsvollste zeitgenössische Verteidigung im Inferentialismus Robert Brandoms. Brandom kehrt die traditionelle Reihenfolge zwischen Bedeutung und Schlussfolgerung um: Ein Begriff bedeutet nicht zunächst etwas, um anschließend in Schlüsse einzugehen, sondern sein Gehalt besteht in nichts anderem als dem Netz materialer Inferenzen, die ihn mit anderen Begriffen verbinden – in dem, was aus seiner Behauptung folgt und was sie ihrerseits voraussetzt. Zu verstehen, was “rot” bedeutet, heißt zu beherrschen, dass daraus “nicht grün” und “farbig” folgt und dass es seinerseits aus bestimmten Wahrnehmungsumständen berechtigterweise behauptet werden darf. Begriffliche Kompetenz ist damit von Anfang an inferentielle Kompetenz; die Kette ist hier nicht ein äußeres Ordnungsschema, das über bereits fertige Bedeutungen gelegt wird, sondern deren konstitutives Medium.

Diese Position begegnet dem Einwand, das Kettenmodell setze bereits bestimmte, identifizierbare Elemente voraus, deren Genese es selbst nicht erklären könne, mit einer bemerkenswerten Volte. Für Brandom entstehen begriffliche Gehalte gerade nicht vor und unabhängig von inferentiellen Praktiken, sondern in ihnen: Im diskursiven “scorekeeping”, in dem Sprecher einander Berechtigungen und Verpflichtungen zuschreiben, entstehen Begriffe als Knotenpunkte eines sich fortlaufend aktualisierenden Netzes von Festlegungen. Neue Begriffe entstehen nicht durch einen der Kette äußerlichen Akt, sondern durch die fortgesetzte Praxis des Explizitmachens implizit bereits befolgter Normen – ein Prozess, den Brandom selbst als wesentlich historisch und expansiv beschreibt. Der Inferentialismus beansprucht also, den Vorwurf der bloßen Explikation bereits fertiger Formen bereits präventiv zu entkräften: Explikation, so ließe sich seine Position zuspitzen, *ist* hier die Form der begrifflichen Genese selbst, nicht deren nachträgliche Beschreibung.

Wo dennoch eine Grenze sichtbar wird, liegt sie nicht in einer Nachlässigkeit der Theorie, sondern in ihrer strukturellen Verpflichtung auf ein bestimmtes Modell von Normativität. Auch die expansivste Version des Inferentialismus beschreibt die Entstehung neuer Begriffe als Fortführung, Revision oder Explizitmachung eines bereits bestehenden Netzes von Verpflichtungen und Berechtigungen. Das Netz mag sich fortlaufend verändern, aber jede Veränderung wird als inferentiell – als Herleitung, Korrektur oder Konsequenz aus bereits bestehenden Festlegungen – rekonstruiert. Was innerhalb dieses Modells nicht in den Blick gerät, ist die Möglichkeit, dass eine neue Unterscheidung entsteht, die sich nicht als inferentielle Konsequenz eines vorherigen Zustands des Netzes beschreiben lässt, sondern die das Netz selbst so reorganisiert, dass hinterher anders inferiert wird als vorher – nicht weil neue Prämissen hinzukämen, sondern weil sich die Relationen zwischen den bestehenden Elementen verschieben. Der Inferentialismus kann solche Verschiebungen im Nachhinein als veränderte inferentielle Ordnung beschreiben; er liefert aber keine Theorie darüber, wie eine solche Verschiebung selbst zustande kommt, bevor sie sich als neue Kette darstellen lässt.

3.2 Das Archiv: Wissen als Bestand

Wo das Kettenmodell die Ordnung von Übergängen beschreibt, beschreibt das Archivmodell die Persistenz dessen, worüber inferiert wird. Es setzt voraus, dass Wissen als Bestand identifizierbarer, dauerhaft verfügbarer Einheiten existiert, deren Ordnung, Speicherung und Erschließung sich verbessern lässt, ohne dass sich damit ihre grundlegende ontologische

Struktur veränderte. Diese Vorstellung ist keineswegs auf explizite Wissensspeicher wie Bibliotheken oder Datenbanken beschränkt. Sie liegt auch weiten Teilen der Erkenntnistheorie zugrunde, insofern Wissen dort primär als Zustand relativ zu bereits bestimmten Propositionen verstanden wird – als gerechtfertigte wahre Überzeugung gegenüber einem Sachverhalt, der unabhängig vom Erkenntnisvollzug bereits vorliegt.

Die Stärke dieser Vorstellung liegt in ihrer Erklärungskraft für Phänomene der Stabilität: Wissenschaftliche Disziplinen, juristische Präzedenzfälle oder kulturelle Überlieferung setzen voraus, dass bestimmte Inhalte über einzelne Erkenntnisakte hinweg identifizierbar bleiben. Ohne diese Persistenz wäre weder kumulative Forschung noch institutionelle Tradition denkbar. Das Archivmodell gerät jedoch an eine Grenze, die sich präzise benennen lässt: Es kann erklären, wie ein bereits vorhandener Inhalt gespeichert, wiedergefunden und weitergegeben wird, aber es kann nicht erklären, wie ein Inhalt, der zuvor in keiner Form – auch nicht als unbestimmte Möglichkeit – vorlag, allererst zu einem identifizierbaren Gegenstand wird, der anschließend gespeichert werden könnte. Jede Erweiterung des Archivs setzt bereits voraus, dass etwas als diskrete, speicherbare Einheit bestimmt ist; die Bestimmung selbst fällt aus dem Modell heraus, weil das Archiv per definitionem mit bereits individuierten Inhalten operiert.

Diese Grenze zeigt sich besonders deutlich, sobald man versucht, wissenschaftliche Innovation im Archivmodell zu beschreiben. Eine neue Theorie erscheint dann zwangsläufig entweder als Entdeckung eines zuvor unbekannten, aber immer schon bestehenden Sachverhalts oder als Neukombination bereits vorhandener Elemente. Beide Beschreibungen mögen für bestimmte Fälle zutreffen, verfehlen aber jene Fälle, in denen eine Theorie die Kategorien selbst verändert, mit denen zuvor überhaupt erst unterschieden wurde, was als Sachverhalt oder als Element gelten konnte. Das Archiv kann die Erweiterung seines Bestandes beschreiben; es kann nicht beschreiben, wie sich die Bedingungen ändern, unter denen überhaupt etwas als bestandsfähiger Inhalt erscheint.

3.3 Der Raum: Möglichkeit als vorgegebenes Feld

Gegenüber dem Archivmodell stellt die Beschreibung des Denkens als Navigation innerhalb eines Möglichkeitsraums einen erkennbaren Fortschritt dar. Sie löst sich von der Vorstellung eines abgeschlossenen Bestandes und rückt stattdessen Relationen, Übergänge und – in ihren anspruchsvolleren, etwa topologischen Varianten – auch Transformationen in den Vordergrund. Wo das Archiv Objekte verwaltet, verwaltet der Raum Positionen und die Wege zwischen ihnen; Kreativität erscheint hier nicht als Fund eines verborgenen Objekts, sondern als das Erschließen einer bislang unbegangenen Region oder eines ungewöhnlichen Pfades innerhalb eines Feldes, das selbst als offen, komplex oder sogar unendlich gedacht werden kann.

Diese Offenheit bleibt jedoch, bei genauerer Betrachtung, an eine Voraussetzung gebunden, die auch die anspruchsvollsten Raummodelle nicht abstreifen: dass der Raum selbst – seine Dimensionen, seine Nachbarschaftsrelationen, seine Transformationsregeln – der Bewegung, die sich in ihm vollzieht, logisch vorausliegt. Mag der Raum unendlich, dynamisch oder in seiner Topologie unübersichtlich sein, so bleibt doch die Bewegung innerhalb dieses Raumes von seiner Existenz kategorial verschieden: Sie realisiert Positionen, die als Möglichkeiten dieses Raumes bereits vorgesehen waren, auch wenn niemand sie zuvor entdeckt oder benannt hatte. Genau in dieser Hinsicht unterscheidet sich

der Raumbegriff nicht grundlegend vom Archivmodell, dessen Schwäche er eigentlich überwinden sollte: Beide setzen eine dem Denkvollzug vorgängige Bestimmtheit voraus – dort als individuierter Bestand, hier als vorstrukturiertes Feld möglicher Positionen.

Der Unterschied zwischen beiden Modellen betrifft also nicht die grundlegende ontologische Priorität des Gegebenen, sondern lediglich den Grad seiner Bestimmtheit. Ein Möglichkeitsraum kann so komplex und so wenig kartiert sein, wie man will – seine bloße Existenz als Raum von Möglichkeiten setzt bereits voraus, dass es eine Antwort auf die Frage gibt, was in ihm möglich ist, unabhängig davon, ob diese Antwort jemals expliziert wird. Wo aber die Struktur dessen, was überhaupt als Möglichkeit erscheinen kann, sich durch den Vollzug des Denkens selbst verändert – wo also nicht nur unbekannte Positionen innerhalb eines festen Feldes entdeckt, sondern die Dimensionen des Feldes selbst verschoben werden –, reicht die Metaphorik des Raumes nicht mehr aus. Sie kann Bewegung innerhalb einer Struktur beschreiben, aber nicht die Veränderung der Struktur durch die Bewegung selbst.

3.4 Die gemeinsame Voraussetzung: Priorität des Gegebenen vor dem Prozess

Kette, Archiv und Raum unterscheiden sich in ihrem jeweiligen Erklärungsanspruch erheblich, und jedes der drei Modelle hat, wie gezeigt, gegen den naheliegenden Einwand seiner Unzulänglichkeit ernstzunehmende Verteidigungsstrategien zur Verfügung. Der Inferentialismus reklamiert die Genese von Begriffen für die inferentielle Praxis selbst; das Archivmodell lässt sich durch immer differenziertere Formen der Wissensorganisation erweitern; der Raumbegriff lässt sich durch dynamische, relationale und topologische Verfeinerungen gegen den Vorwurf der Statik immunisieren. Keine dieser Verteidigungen beseitigt jedoch eine Voraussetzung, die allen drei Modellen trotz ihrer Unterschiede gemeinsam ist und die sich erst sichtbar macht, wenn man nicht nach der jeweiligen inneren Struktur der Modelle fragt, sondern nach ihrem Verhältnis zu dem, was ihrer Anwendung vorausgeht.

Diese gemeinsame Voraussetzung lässt sich wie folgt formulieren: In allen drei Fällen wird angenommen, dass es – auf welcher Ebene auch immer, als individuierter Begriff, als gespeicherter Inhalt oder als Position in einem Feld – bereits eine Bestimmtheit gibt, bevor der jeweils beschriebene Prozess an ihr operiert. Die Kette verknüpft Gründe, die als solche schon identifizierbar sein müssen. Das Archiv verwaltet Inhalte, die als individuierte Einheiten bereits vorliegen müssen. Der Raum eröffnet Positionen, die als Möglichkeiten dieses Raumes bereits vorgesehen sein müssen. In jedem Fall liegt die Priorität beim Gegebenen; der beschriebene Prozess – Inferenz, Speicherung, Navigation – setzt an einer bereits vorhandenen Bestimmtheit an und verändert höchstens deren Zugänglichkeit, Anordnung oder Realisierung, nicht aber die Bedingungen ihrer ursprünglichen Entstehung.

Damit ist die systematische Lücke benannt, die keines der drei Modelle aus sich selbst heraus schließen kann, ohne die eigene Grundstruktur aufzugeben: die Frage, wie jene Bestimmtheiten – Begriffe, Inhalte, Möglichkeiten –, mit denen Kette, Archiv und Raum jeweils operieren, allererst entstehen, bevor sie inferiert, gespeichert oder navigiert werden können. Diese Frage lässt sich nicht durch eine weitere Verfeinerung der drei Modelle beantworten, weil sie alle drei in ihrer jeweiligen Erklärungsleistung bereits voraussetzen, was hier erklärt werden soll. Sie verlangt vielmehr eine Beschreibungsebene, die der Priorität des Gegebenen die Priorität des Prozesses entgegensetzt – eine Ebene, auf der sich präzise zeigen lässt, wie Bestimmtheit selbst durch einen Vorgang entsteht, der ihr nicht äußerlich,

sondern konstitutiv vorausgeht. Eine solche Beschreibungsebene lässt sich, wie der folgende Teil zeigen wird, nicht nur begrifflich postulieren, sondern aus der formalen Struktur eines konkreten Falls ableiten.

4. Formale Herleitung: Die Struktur der Konditionierung

4.1 Der Prompt als Umformung einer Verteilung, nicht als Auswahl aus einem Bestand

Ein autoregressives Sprachmodell approximiert eine bedingte Wahrscheinlichkeitsverteilung über die Elemente eines endlichen Vokabulars: Gegeben eine Sequenz bereits erzeugter oder vorgegebener Tokens, weist das Modell jedem möglichen nächsten Token eine Wahrscheinlichkeit zu. Diese Verteilung ist das Ergebnis eines Trainingsprozesses, in dem die Parameter des Modells so justiert wurden, dass die Wahrscheinlichkeit, mit der es die tatsächlichen Fortsetzungen eines großen Textkorpus vorhersagt, maximiert wird. Was das Modell nach Abschluss des Trainings besitzt, ist folglich keine Liste möglicher Texte, unter denen ausgewählt werden könnte, sondern eine hochdimensionale Funktion, die jedem beliebigen Kontext eine Verteilung über mögliche Fortsetzungen zuordnet.

Diese Beschreibung erlaubt eine präzise Bestimmung dessen, was ein Prompt formal leistet. Der Prompt ist der Kontext, mit dem diese Funktion aufgerufen wird. Seine Funktion besteht nicht darin, aus einem vorab bestehenden Bestand an Texten denjenigen auszuwählen, der dem Prompt am besten entspricht – ein solcher Bestand existiert im Modell zu keinem Zeitpunkt in individuierter Form. Seine Funktion besteht vielmehr darin, die vom Modell repräsentierte Funktion an einer bestimmten Stelle auszuwerten und dadurch eine bestimmte, aus der Gesamtheit der möglichen Verteilungen singuläre Wahrscheinlichkeitsverteilung über mögliche nächste Tokens zu erzeugen. Der Unterschied zwischen “Auswahl aus einem Bestand” und “Erzeugung einer Verteilung durch Konditionierung” ist dabei keine philosophische Nuance, sondern eine strukturelle Differenz, die sich unmittelbar an der Mathematik des Verfahrens ablesen lässt: Vor der Konditionierung existiert keine Menge individuierter Kandidaten, unter denen selektiert würde, sondern lediglich eine Funktion, die noch keinen bestimmten Wert angenommen hat.

Diese Feststellung transformiert den Begriff der Möglichkeit, der in den vorangegangenen Modellen unhinterfragt vorausgesetzt wurde. Wenn der Prompt keine Auswahl aus vorab individuierten Optionen vollzieht, sondern eine Verteilung durch Konditionierung allererst erzeugt, dann existieren die “möglichen Fortsetzungen” nicht als eine dem Konditionierungsakt vorausliegende, wenn auch unentdeckte Menge. Sie entstehen als individuierte, mit bestimmten Wahrscheinlichkeiten versehene Alternativen erst in dem Moment, in dem eine Konditionierung stattfindet. Vor diesem Moment gibt es, streng genommen, keine Menge von Möglichkeiten, sondern nur die Fähigkeit des Systems, bei beliebiger Konditionierung eine Verteilung zu erzeugen – eine Fähigkeit, die selbst keine Menge, sondern eine Funktion ist. Dies ist keine bloße Analogie zu der These, Möglichkeiten seien keine dem Prozess vorausliegenden Objekte, sondern Resultate ihrer eigenen Genese, sondern eine strukturell identische Beziehung, an einem präzise beschreibbaren formalen Gegenstand demonstriert.

4.2 Reflexivität als architektonische Notwendigkeit

Die autoregressive Architektur, auf der die gegenwärtigen generativen Sprachmodelle beruhen, erzeugt eine Sequenz nicht auf einen Schlag, sondern Token für Token: Jedes erzeugte Token wird dem bisherigen Kontext hinzugefügt und dient als Teil der Konditionierung für die Erzeugung des nächsten. Formal bedeutet dies, dass die Wahrscheinlichkeit einer gesamten Sequenz als Produkt bedingter Wahrscheinlichkeiten zerlegt wird, wobei jede dieser bedingten Wahrscheinlichkeiten von sämtlichen zuvor erzeugten Tokens abhängt – einschließlich derjenigen, die das Modell selbst im vorangegangenen Schritt produziert hat.

Diese Struktur hat eine Konsequenz, die keiner interpretativen Zusatzannahme bedarf, sondern unmittelbar aus der Definition des Verfahrens folgt: Der Kontext, relativ zu dem die Fortsetzung an einer beliebigen Stelle $n+1$ der Sequenz erzeugt wird, enthält notwendig die Ergebnisse aller vorangegangenen Erzeugungsschritte. Der Prozess wirkt damit fortlaufend auf die Bedingungen zurück, unter denen er sich selbst fortsetzt. Dies ist nicht mit der bloßen Feststellung zu verwechseln, dass ein System “seine eigene Geschichte berücksichtigt” – eine Eigenschaft, die triviale Rekursionen ebenso aufweisen wie komplexe. Entscheidend ist vielmehr, dass jeder Schritt die *Verteilung* über die möglichen Fortsetzungen des nächsten Schritts verändert, und zwar nicht durch Hinzufügung eines weiteren Datenpunktes zu einer bereits fixen Menge, sondern durch Verschiebung der gesamten bedingten Verteilung, aus der der nächste Schritt hervorgeht.

Damit lässt sich präzise angeben, in welchem Sinn ein solcher Prozess seinen eigenen Horizont erzeugt, statt sich innerhalb eines vorgegebenen Horizonts zu bewegen. Der “Horizont” – verstanden als die Verteilung möglicher Fortsetzungen an einem gegebenen Punkt der Sequenz – ist zu keinem Zeitpunkt vor dem Erreichen dieses Punktes bereits fixiert; er wird durch jeden vorangegangenen Schritt mitkonstituiert. Diese Reflexivität ist keine zusätzliche Eigenschaft, die dem System bei besonders komplexen Aufgaben zükäme; sie ist eine architektonische Notwendigkeit jeder autoregressiven Sequenzerzeugung, unabhängig von der Größe des Modells, der Trainingsdaten oder der konkreten Aufgabe. Sie lässt sich daher als strukturelle, nicht als empirische Eigenschaft des Verfahrens ausweisen.

4.3 Ausschluss als konstitutive, nicht restriktive Operation

Konditionierung – sei es durch den initialen Prompt, sei es durch jeden nachfolgenden, vom Modell selbst erzeugten Kontext – lässt sich formal als eine Operation auf einer Wahrscheinlichkeitsverteilung beschreiben: Sie verschiebt Wahrscheinlichkeitsmasse von einem Teil des Vokabulars zu einem anderen. Ohne Konditionierung – etwa bei einem völlig leeren Kontext – weist das Modell eine Verteilung auf, die vergleichsweise diffus über viele mögliche Fortsetzungen gestreut ist. Mit fortschreitender Konditionierung konzentriert sich diese Verteilung zunehmend auf wenige, im Extremfall auf eine einzige, nahezu sichere Fortsetzung.

Diese Konzentration lässt sich unter zwei komplementären Aspekten beschreiben, die in der Alltagssprache leicht als Gegensatz erscheinen, formal aber ein und derselbe Vorgang sind. Unter dem einen Aspekt schließt die Konditionierung Möglichkeiten aus: Fortsetzungen, die vor der Konditionierung eine nennenswerte Wahrscheinlichkeit besaßen, werden auf eine gegen null gehende Wahrscheinlichkeit gesetzt. Unter dem anderen Aspekt konstituiert genau dieser Ausschluss diejenigen Fortsetzungen, die übrig bleiben, als die relevanten, wahrscheinlichen, “passenden” – ihre erhöhte relative Wahrscheinlichkeit *ist* nichts anderes

als die Kehrseite des Ausschlusses der übrigen. Es gibt hier keine zwei Vorgänge, von denen der eine restriktiv und der andere produktiv wäre; es gibt einen einzigen Vorgang der Umverteilung, der beide Beschreibungen gleichermaßen zulässt, weil “diese Fortsetzung ist wahrscheinlich” und “jene anderen sind ausgeschlossen” in einer Wahrscheinlichkeitsverteilung, die sich zu eins summiert, notwendig zusammenfallen.

Diese formale Beobachtung erlaubt eine Präzisierung der These, Bestimmung erfolge durch Ausschluss, die über eine bloße begriffliche Behauptung hinausgeht. Im Fall der Konditionierung eines Sprachmodells ist “Bestimmung durch Ausschluss” keine Metapher für einen andernorts stattfindenden Vorgang, sondern die korrekte technische Beschreibung dessen, was Konditionierung tut: Sie ist, mathematisch, nichts als eine Umverteilung von Wahrscheinlichkeitsmasse, bei der jede Erhöhung an einer Stelle der Verteilung eine Verringerung an anderer Stelle notwendig nach sich zieht. Es gibt keine Möglichkeit, innerhalb dieses Formalismus Bestimmung zu erzeugen, ohne zugleich auszuschließen; beide Beschreibungen bezeichnen dieselbe Operation aus verschiedener Perspektive.

4.4 Trennung von struktureller These und empirischer Kontingenz

Die in den vorangegangenen Abschnitten entwickelten drei Beobachtungen – dass Konditionierung Verteilungen umformt statt aus Beständen auszuwählen, dass autoregressive Erzeugung notwendig reflexiv ist, und dass Bestimmung und Ausschluss in der Konditionierung ein und derselbe Vorgang sind – beanspruchen, aus der formalen Definition der betreffenden Systeme ableitbar zu sein, nicht aus Beobachtungen ihres Verhaltens in der Anwendung. Sie sind damit unabhängig von einer bestimmten, in der öffentlichen wie fachlichen Diskussion verbreiteten empirischen Behauptung, die von diesen strukturellen Thesen sorgfältig zu unterscheiden ist: der Behauptung, dass unterspezifizierte oder vage formulierte Prompts tendenziell zu Ausgaben führen, die von menschlichen Beurteilern als origineller oder kreativer eingestuft werden als die Ausgaben präzise spezifizierter Prompts.

Diese zweite Behauptung ist empirischer, nicht struktureller Natur. Sie setzt ein Urteil über die Qualität von Ausgaben voraus – “originell”, “kreativ” –, das nicht aus der Architektur des Systems ableitbar ist, sondern aus der Beurteilung durch menschliche Rezipienten resultiert, deren Kriterien selbst kontingent, kulturell variabel und methodisch schwer zu operationalisieren sind. Ob geringere Konditionierungsschärfe tatsächlich systematisch zu Ausgaben führt, die als kreativer beurteilt werden, ist eine Frage, die sich nur durch kontrollierte Untersuchungen mit klar definierten Bewertungskriterien beantworten ließe, und selbst dann bliebe das Ergebnis an die jeweilige Bewertungsmethodik gebunden. Es handelt sich um eine Hypothese über menschliche ästhetische Reaktionen auf maschinell erzeugte Texte, nicht um eine Eigenschaft der Konditionierungsstruktur selbst.

Diese Trennung ist für die vorliegende Untersuchung von erheblicher Bedeutung, weil sie es erlaubt, die drei strukturellen Thesen – Prozesscharakter der Möglichkeit, konstitutive Reflexivität, Identität von Bestimmung und Ausschluss – unabhängig vom Ausgang jeder empirischen Untersuchung über Kreativitätsurteile zu behaupten. Selbst wenn sich erwiese, dass präzise Prompts im Durchschnitt zu ebenso originellen oder originelleren Ergebnissen führen als vage formulierte – ein empirisch durchaus offener Befund –, bliebe davon unberührt, dass jede Konditionierung, präzise oder vage, formal als Umformung einer Verteilung, nicht als Auswahl aus einem Bestand zu beschreiben ist, dass jede autoregressive

Erzeugung notwendig reflexiv verläuft, und dass jede Bestimmung durch Konditionierung notwendig mit einem Ausschluss identisch ist. Die philosophische Tragfähigkeit der hier entwickelten Thesen hängt somit nicht an einer kontingenten und methodisch schwer zu sichernden Beobachtung über menschliche Geschmacksurteile, sondern an einer Struktur, die sich aus der formalen Definition des betrachteten Verfahrens selbst ergibt. Welche Konsequenzen sich aus dieser Struktur für den Begriff der Rationalität ziehen lassen, ist Gegenstand des folgenden Teils.

5. Konsequenzen für den Begriff der Rationalität

5.1 Jenseits von Determination und Beliebigkeit

Die formale Struktur der Konditionierung erlaubt eine Präzisierung eines Problems, das die philosophische Tradition gewöhnlich in der Alternative von Determination und Zufall verhandelt. Ein Prozess gilt entweder als vollständig durch seine Ausgangsbedingungen festgelegt – dann erscheint jede scheinbare Neuheit als bloße Unkenntnis der bereits determinierten Zukunft – oder er gilt als in relevantem Maß offen, dann aber um den Preis, dass diese Offenheit gewöhnlich mit Zufälligkeit gleichgesetzt und damit aus dem Bereich des Rationalen ausgeschlossen wird. Beide Beschreibungen finden an der Konditionierungsstruktur autoregressiver Modelle keinen Anwendungsfall, was diese Struktur zu einem präzisen Prüfstein für eine dritte Möglichkeit macht.

Ein autoregressives Modell ist zu keinem Zeitpunkt vollständig determiniert in dem Sinn, dass ein gegebener Kontext einen einzigen möglichen Fortsetzungstoken zuließe. Die Verteilung, die einer Konditionierung entspricht, weist in aller Regel mehreren Fortsetzungen eine von null verschiedene Wahrscheinlichkeit zu; welche davon tatsächlich realisiert wird, hängt vom gewählten Erzeugungsverfahren ab. Zugleich ist der Prozess nicht beliebig in dem Sinn, dass jede Fortsetzung gleich wahrscheinlich wäre oder die Verteilung keinerlei Struktur aufwiese. Die Verteilung selbst – nicht ihr konkreter Ausgang – ist vollständig durch die Konditionierung und die trainierten Parameter bestimmt; sie besitzt eine präzise, berechenbare Form, auch wenn diese Form mehreren Fortsetzungen Gewicht verleiht, statt eine einzige festzulegen.

Damit lässt sich formal genau angeben, was eine Position jenseits von Determination und Beliebigkeit bedeutet: nicht ein Kompromiss oder eine Mischung aus beidem, sondern ein Vorgang, dessen *Struktur* vollständig bestimmt ist, während sein *Ausgang* es nicht ist – wobei diese Unbestimmtheit des Ausgangs selbst wiederum keine Homogenität aufweist, sondern eine spezifische, aus der Konditionierung hervorgehende Gewichtung besitzt. Rationalität, verstanden als Eigenschaft eines solchen Prozesses, kann sich folglich nicht daran bemessen, ob der Ausgang vorhersagbar ist, sondern nur daran, ob die Struktur, die den Ausgang wahrscheinlich macht, ihrerseits nachvollziehbaren, nicht-beliebigen Bedingungen folgt. Diese Verschiebung des Kriteriums – von der Vorhersagbarkeit des Resultats zur Nachvollziehbarkeit der es hervorbringenden Struktur – ist die erste Konsequenz, die sich aus der formalen Analyse der Konditionierung für den Begriff der Rationalität ziehen lässt.

5.2 Formbildung als eigenständige Dimension neben inferentieller Geltung

Wird Rationalität nicht mehr ausschließlich an der Vorhersagbarkeit ihres Resultats, sondern an der Nachvollziehbarkeit der resultaterzeugenden Struktur bemessen, so tritt neben die inferentielle Geltung – die Frage, ob ein Übergang von Prämissen zu einer Konklusion gültig ist – eine zweite, davon unterscheidbare Dimension: die Frage, wie die Prämissen, zwischen denen inferiert wird, überhaupt zu der Bestimmtheit gelangt sind, die eine Prüfung ihrer inferentiellen Beziehungen erst ermöglicht. Diese zweite Dimension lässt sich als Formbildung bezeichnen, und die im vorangegangenen Teil entwickelte formale Analyse erlaubt es, ihr Verhältnis zur inferentiellen Geltung präzise, nicht bloß metaphorisch zu bestimmen.

Inferentielle Geltung operiert auf bereits individuierten Elementen: Sie prüft, ob aus “alle Menschen sind sterblich” und “Sokrates ist ein Mensch” die Konklusion “Sokrates ist sterblich” folgt, und diese Prüfung setzt voraus, dass die drei Aussagen als solche bereits bestimmt sind. Formbildung hingegen betrifft den Vorgang, durch den aus einer unbestimmten Ausgangslage – einer diffusen Verteilung über mögliche Fortsetzungen, um im Bild der Konditionierung zu bleiben – eine bestimmte Ausgangslage entsteht, an der inferentielle Prüfung überhaupt ansetzen kann. Beide Dimensionen sind einander nicht äußerlich: Ohne Formbildung gäbe es keine individuierten Elemente, zwischen denen inferiert werden könnte; ohne inferentielle Geltung bliebe jede gebildete Form beliebig, ihre Konsequenzen unprüfbar. Sie stehen jedoch auch nicht in einem Verhältnis der Unterordnung, in dem die eine bloß Vorstufe oder Werkzeug der anderen wäre. Formbildung erzeugt nicht Rohmaterial, das Inferenz anschließend verarbeitet; sie besitzt eine eigene, von inferentieller Prüfbarkeit unabhängige Struktur – wie die im vorangegangenen Teil beschriebene Konditionierung zeigt, die als Umformung einer Verteilung vollständig beschreibbar ist, ohne dass in dieser Beschreibung irgendein Bezug auf logische Gültigkeit vorkäme.

Diese Eigenständigkeit lässt sich an einem Punkt besonders deutlich machen: Die Reflexivität autoregressiver Erzeugung – dass jeder Schritt die Bedingungen des nächsten mitkonstituiert – hat kein Analogon in der klassischen Theorie der Inferenz. Eine gültige Schlussfolgerung verändert die Prämissen, aus denen sie gezogen wurde, nicht rückwirkend; die Gültigkeit einer Inferenz ist eine zeitlose Relation zwischen bereits fixierten Aussagen. Formbildung hingegen ist ihrer Struktur nach zeitlich und rekursiv: Jede Stabilisierung einer Form verändert die Bedingungen, unter denen die nächste Form entstehen kann. Ein Rationalitätsbegriff, der ausschließlich an der inferentiellen Dimension orientiert ist, kann diese rekursive Zeitlichkeit nicht erfassen, weil er sie aus seinem Gegenstandsbereich von vornherein ausschließt. Rationalität muss daher, soll sie den Prozess der Formbildung nicht als vor- oder außerrationales Vorspiel behandeln, um diese zweite Dimension erweitert werden, ohne dass die erste dadurch an Geltung verlöre.

5.3 Der Begriff als Operator statt als Repräsentation

Aus der Unterscheidung zweier Dimensionen der Rationalität ergibt sich eine Konsequenz für den Begriff des Begriffs selbst. In der klassischen Erkenntnistheorie, die sich vornehmlich an der inferentiellen Dimension orientiert, fungiert ein Begriff primär als Repräsentation: Er steht für eine Klasse von Gegenständen oder Eigenschaften, und seine Aufgabe besteht darin, diese Klasse mit hinreichender Präzision zu erfassen, damit über sie zutreffend geurteilt werden kann. Die Güte eines Begriffs bemisst sich an dieser Auffassung

nach seiner Extension: ob er die relevanten Fälle korrekt einschließt und die irrelevanten korrekt ausschließt.

Wird jedoch die Dimension der Formbildung als eigenständig anerkannt, tritt neben diese repräsentationale Funktion eine zweite, die sich am Modell der Konditionierung präzise beschreiben lässt. Ein neu eingeführter Begriff verändert nicht nur, wofür fortan ein Ausdruck steht; er verändert – strukturanalog zur Konditionierung eines Sprachmodells – die Verteilung der Relevanz und Wahrscheinlichkeit über das gesamte Feld der mit ihm verbundenen Unterscheidungen. Mit der Einführung eines Begriffs werden bestimmte weitere Unterscheidungen naheliegend, die zuvor nicht naheliegend waren, während andere, zuvor plausible Unterscheidungen an Relevanz verlieren – nicht weil der neue Begriff sie widerlegte, sondern weil er das Feld, in dem sie ihre Plausibilität bezogen, umstrukturiert. In diesem Sinn lässt sich ein Begriff als Operator beschreiben: als etwas, das auf ein bestehendes Feld von Unterscheidungen einwirkt und es in eine neue Verteilung von Relevanzen überführt, so wie ein Prompt eine Wahrscheinlichkeitsverteilung über mögliche Fortsetzungen umformt, ohne aus einem vorab bestehenden Bestand auszuwählen.

Diese Bestimmung des Begriffs als Operator schließt seine repräsentationale Funktion nicht aus, sondern ergänzt sie um eine Dimension, die aus rein repräsentationaler Perspektive unsichtbar bleibt. Ein Begriff kann in seiner extensionalen Reichweite exakt derselbe bleiben und dennoch, je nachdem, in welchem begrifflichen Feld er auftritt, eine unterschiedliche operative Wirkung entfalten – eine unterschiedliche Verteilung von Anschlussmöglichkeiten eröffnen. Umgekehrt kann die operative Wirkung eines Begriffs erheblich sein, ohne dass sich seine Extension in einer Weise verändert hätte, die eine rein repräsentationale Theorie als bedeutsam registrieren würde. Die Unterscheidung zwischen Repräsentation und Operation ist damit keine bloße Umbenennung derselben Sache, sondern macht eine Eigenschaft von Begriffen sichtbar, die für den Prozess ihrer Genese – für Formbildung im oben bestimmten Sinn – konstitutiv, für ihre bloße inferentielle Verwendung jedoch irrelevant ist.

Mit dieser doppelten Bestimmung des Begriffs ist der systematische Kern der vorliegenden Untersuchung erreicht: ein Rationalitätsbegriff, der die inferentielle Prüfung bereits individuierter Formen und die Genese dieser Formen selbst als zwei irreduzible, aufeinander angewiesene Dimensionen begreift, sowie ein Begriffsverständnis, das diese doppelte Funktion – Repräsentation und Operation – nicht als Widerspruch, sondern als notwendige Ergänzung ausweist. Wie sich diese Position zu den naheliegenden Bezugspunkten der Prozess- und Differenzphilosophie verhält und worin ihr eigenständiger Beitrag gegenüber diesen Traditionen liegt, ist Gegenstand des folgenden Teils.

6. Einordnung: Der spezifische Mehrwert

6.1 Abgrenzung zu Deleuze und der Differenzphilosophie

Die These, dass Bestimmung durch Ausschluss und Differenzierung entsteht, statt vorgängige Identitäten vorauszusetzen, ist seit Deleuze’ “Differenz und Wiederholung” ein zentrales Motiv der kontinentalen Philosophie. Deleuze kritisiert dort die Unterordnung der Differenz unter die Identität, wie sie seiner Diagnose zufolge die gesamte Geschichte der abendländischen Metaphysik von Platon bis Hegel durchzieht: Differenz erscheint dort stets

als Differenz *zwischen* bereits identifizierten Termini, niemals als das, was Identität selbst allererst hervorbringt. Dagegen entwickelt Deleuze einen Begriff der Differenz an sich, die keiner vorgängigen Identität bedarf, sowie, in Anlehnung an Simondon, den Begriff des Intensiven als eines vorindividuellen Feldes, aus dem sich Individuen erst durch einen Prozess der Differenzierung herausbilden.

Die strukturelle Verwandtschaft zur hier entwickelten Position ist unübersehbar und soll nicht verschleiert werden: Auch die vorliegende Untersuchung bestreitet, dass Möglichkeiten als individuierte Objekte einem Formbildungsprozess vorausliegen, und beschreibt Bestimmung als durch Ausschluss konstituiert. Der entscheidende Unterschied liegt jedoch in der Art der Begründung. Deleuze entwickelt seine Ontologie der Differenz aus einer immanenten Kritik der Geschichte der Metaphysik sowie aus Anleihen bei der Differentialrechnung, der Thermodynamik und der Embryologie, die bei ihm überwiegend metaphorisch – als Bildspender einer philosophischen Konzeption – fungieren, ohne dass die formale Struktur dieser Bezugsdisziplinen selbst zum Gegenstand einer strikten Ableitung würde. Die vorliegende Untersuchung verfährt anders: Sie leitet die strukturell analogen Thesen nicht aus einer Kritik der Metaphysikgeschichte, sondern aus der vollständig spezifizierten formalen Definition eines gegenwärtig existierenden technischen Verfahrens ab, dessen Verhalten sich beobachten, berechnen und in seinen relevanten Eigenschaften exakt nachweisen lässt. Wo Deleuze das Intensive als metaphysische Tiefendimension postuliert, die sich der direkten Erfahrung entzieht und nur in ihren „Effekten“ – den individuierten Formen – zugänglich wird, lässt sich die Struktur der Konditionierung unmittelbar beobachten und in ihrer Funktionsweise vollständig angeben. Dies macht die hier entwickelte Position nicht per se überlegen, aber es verschiebt den Status ihrer zentralen Behauptungen von einer spekulativen Ontologie zu einer aus einem beobachtbaren Fall abgeleiteten Struktur – ein Unterschied, der insbesondere für die Frage der Überprüfbarkeit erheblich ist.

6.2 Abgrenzung zu Simondon und der Theorie der Individuation

Gilbert Simondons Theorie der Individuation steht der vorliegenden Untersuchung inhaltlich näher als Deleuze, dessen Begriff des Intensiven sich, wie erwähnt, wesentlich Simondon verdankt. Simondon kritisiert die traditionelle Auffassung, wonach das Individuum ein bereits fertiges Prinzip seiner eigenen Individuation in sich trüge – sei es als Substanzform, sei es als vorgegebene Materie –, und schlägt stattdessen vor, das Individuum als Resultat eines Prozesses zu verstehen, der von einem metastabilen, präindividuellen Zustand ausgeht: einem System, das Spannungen und Potentiale enthält, ohne bereits in individuierte Formen aufgeteilt zu sein, und das sich durch einen Vorgang der „Transduktion“ – einer sich selbst fortpflanzenden strukturierenden Operation – schrittweise in bestimmte Formen auflöst, wobei jede Individuation neue Spannungen und damit neue Möglichkeiten weiterer Individuation erzeugt.

Diese Beschreibung ist der hier entwickelten Analyse der Konditionierung in ihrer Grundstruktur bemerkenswert ähnlich: Auch bei Simondon entsteht Bestimmtheit nicht durch Realisierung vorgegebener Möglichkeiten, sondern durch einen Prozess, der seine eigenen Bedingungen mit hervorbringt, und auch bei ihm ist dieser Prozess wesentlich rekursiv, insofern jede Individuation die Ausgangslage für die nächste verändert. Der Unterschied zur vorliegenden Untersuchung liegt weniger im begrifflichen Gehalt als im

epistemischen Status des Ausgangspunkts. Simondons präindividuelles Feld ist ein theoretisches Postulat, das zur Erklärung physikalischer, biologischer und psychosozialer Individuationsprozesse eingeführt wird, dessen eigene Verfassung sich jedoch, seiner Natur nach, keiner direkten Beobachtung öffnet – es wird aus seinen Wirkungen erschlossen, niemals selbst beobachtet. Die Konditionierung eines Sprachmodells hingegen ist kein postuliertes Substrat, sondern eine mathematisch vollständig spezifizierte, in jedem einzelnen Rechenschritt nachvollziehbare Operation. Der “präindividuelle” Zustand entspricht hier der Verteilung vor der Konditionierung, die exakt angegeben, berechnet und mit der Verteilung nach der Konditionierung quantitativ verglichen werden kann.

Diese Präzisierbarkeit erlaubt eine Unterscheidung, die bei Simondon in der Sache zwar angelegt, aber begrifflich nicht in derselben Schärfe gezogen werden kann: die Unterscheidung zwischen struktureller Notwendigkeit und empirischer Kontingenz des Individuationsprozesses. Da sich am Sprachmodell exakt angeben lässt, welche Eigenschaften aus der bloßen Definition des Verfahrens folgen und welche von den konkreten trainierten Parametern oder gar von der Wahl des Erzeugungsverfahrens abhängen, lässt sich eine Ebene rein struktureller Aussagen isolieren, die bei einem so allgemeinen, für sehr unterschiedliche Individuationsprozesse zuständigen Theorierahmen wie dem Simondons kaum in gleicher Schärfe zu gewinnen ist.

6.3 Abgrenzung zu Whitehead und der Prozessphilosophie

Alfred North Whiteheads Prozess- und Realitätsphilosophie bietet mit dem Begriff des “actual occasion” eine dritte einschlägige Bezugsgröße. Für Whitehead ist die letzte konkrete Wirklichkeitseinheit kein beharrendes Substrat, sondern ein Ereignis der Selbstkonstitution, in dem eine Vielheit vorangegangener Ereignisse (“prehensions”) zu einer neuen Einheit synthetisiert wird, die ihrerseits, sobald sie abgeschlossen ist, zum Datum künftiger Synthesen wird. Wirklichkeit besteht bei Whitehead folglich aus einem fortlaufenden Prozess des “concrecence”, in dem jedes Ereignis seine eigenen Bedingungen aus dem, was ihm vorausgeht, aktiv hervorbringt, statt sie bloß zu übernehmen.

Die strukturelle Nähe zur Reflexivität autoregressiver Erzeugung – jeder Schritt geht in die Bedingungen des nächsten ein und verändert dadurch dessen Möglichkeitsraum – ist auch hier deutlich, und Whiteheads Begriff der “Kreativität” als eines ultimativen Prinzips, demzufolge jedes Ereignis über das bloß Gegebene hinausgeht, weist Parallelen zur hier beschriebenen Nicht-Reduzierbarkeit der Konditionierung auf eine Auswahl aus vorgegebenen Beständen auf. Der Unterschied liegt erneut im Geltungsanspruch und im methodischen Status der jeweiligen Analyse. Whiteheads Metaphysik beansprucht, eine vollständige Kategorienlehre der Wirklichkeit als solcher zu liefern – actual occasions, eternal objects, prehensions, das Prinzip der Kreativität sind Bausteine eines spekulativen Systems, dessen Anwendungsbereich sich von subatomaren Ereignissen bis zu menschlichem Bewusstsein erstreckt. Diese universale Reichweite ist zugleich der Grund, warum sich Whiteheads Kategorien empirisch kaum direkt prüfen lassen: Sie sind so allgemein gefasst, dass praktisch jedes Ereignis als Bestätigung interpretiert werden kann, was ihre Aussagekraft im Einzelfall schwächt.

Die vorliegende Untersuchung verzichtet demgegenüber ausdrücklich auf einen universalmetaphysischen Anspruch. Sie behauptet nicht, dass alle Wirklichkeit die Struktur der Konditionierung aufweise, sondern beschränkt sich darauf, diese Struktur an einem

einzelnen, klar abgegrenzten formalen Gegenstand nachzuweisen und von dort aus Konsequenzen für den Begriff der Rationalität zu ziehen. Ob und in welchem Umfang sich die gewonnenen Einsichten auf andere Bereiche – etwa auf menschliches Denken – übertragen lassen, bleibt eine eigens zu prüfende Frage, die von der hier entwickelten Analyse nicht präjudiziert wird. Diese Beschränkung ist ein methodischer Vorzug gegenüber der spekulativen Weite Whiteheads, auch wenn sie mit einem geringeren unmittelbaren Erklärungsanspruch erkauft ist.

6.4 Was neu ist: Das Medium-Argument als eigenständiger Beitrag

Die vorangegangenen Abgrenzungen zeigen, dass die strukturellen Kernthesen der vorliegenden Untersuchung – Möglichkeit als Prozessresultat statt als vorgängiges Objekt, konstitutive Reflexivität, Bestimmung durch Ausschluss – in der Sache nicht ohne Vorläufer sind. Deleuze, Simondon und Whitehead haben, in je unterschiedlicher Terminologie und mit je unterschiedlichem Geltungsanspruch, verwandte Positionen entwickelt, und es wäre unredlich, die vorliegende Untersuchung als von diesen Traditionen unabhängig darzustellen. Der eigenständige Beitrag liegt daher nicht primär in der inhaltlichen Neuheit dieser drei Thesen als solcher.

Er liegt vielmehr an zwei Stellen, die sich in den drei genannten Traditionen so nicht finden. Erstens in der methodischen Verschiebung, die zentralen Thesen nicht aus einer spekulativen Rekonstruktion der Metaphysikgeschichte oder aus einer universalontologischen Kategorienlehre zu gewinnen, sondern aus der vollständig spezifizierten, beobachtbaren und nachrechenbaren Struktur eines gegenwärtig existierenden technischen Verfahrens abzuleiten. Diese Verschiebung ändert den epistemischen Status der gewonnenen Thesen: Sie sind nicht mehr nur plausibel im Licht einer bestimmten philosophischen Tradition, sondern an einem konkreten Fall nachweisbar – überprüfbar in einem Sinn, der für Deleuze' Ontologie der Differenz, für Simondons präindividuelles Feld und für Whiteheads actual occasions strukturell nicht zur Verfügung steht.

Zweitens, und dies ist der eigentliche, in den drei genannten Traditionen nicht vorweggenommene Beitrag: Keiner dieser Denker entwickelt seine Position im Ausgang von der spezifischen Frage nach der epistemischen *Funktion*, die ein technisches Artefakt für die philosophische Selbstaufklärung des Denkens übernehmen kann, ohne selbst Gegenstand einer Philosophie der Technik zu sein. Deleuze, Simondon und Whitehead entwickeln Ontologien, die – sofern man ihre Kategorien akzeptiert – auf beliebige Gegenstandsbereiche anwendbar sind; keiner von ihnen argumentiert, dass ein bestimmter technischer Gegenstand deshalb philosophisch privilegiert wäre, weil er eine sonst kaum artikulierbare Struktur des Denkens exemplarisch sichtbar macht. Die vorliegende Untersuchung entwickelt demgegenüber nicht in erster Linie eine neue Ontologie der Form oder des Prozesses, sondern eine Epistemologie des Mediums: eine Theorie darüber, wie ein Instrument, ohne selbst zu denken oder zu verstehen, eine Struktur des Denkens explizit machen kann, die diesem selbst zugrunde liegt, aber unter den etablierten Selbstbeschreibungen der Philosophie verdeckt bleibt. Dieses Medium-Argument – nicht die drei strukturellen Einzelthesen, die es stützen – bildet den eigentlichen, in der gegenwärtigen Diskussion zwischen Ablehnung und Instrumentalisierung generativer Systeme bislang unbesetzten systematischen Ort dieser Untersuchung.

Schluss: Eine dritte Position in der Gegenwart

Die gegenwärtige Debatte über generative Sprachmodelle wird von einer Alternative beherrscht, die sich zunehmend als falsch erweist. Wer diese Systeme als kategorialen Fehlschluss entlarvt, dem eine täuschende sprachliche Oberfläche aufsitzt, hat gute Gründe für seine Vorsicht gegenüber vorschnellen Zuschreibungen von Verstehen. Wer sie als hochentwickelte Werkzeuge behandelt, deren philosophische Bedeutung sich in Fragen der Zuverlässigkeit und Verantwortung erschöpft, verkennt nicht die Natur des Artefakts, sondern lediglich den Umfang dessen, was an ihm zu erkennen ist. Beide Positionen sind, in ihren jeweiligen Grenzen, verteidigungsfähig. Beide beantworten jedoch eine Frage – was diese Systeme sind –, ohne zu bemerken, dass sich an ihnen eine andere, ergiebigere Frage stellen lässt: was an ihnen, unabhängig von ihrem eigenen ontologischen Status, sichtbar wird.

Diese zweite Frage hat sich als tragfähig erwiesen, weil sie sich nicht auf Analogien, Metaphern oder spekulative Übertragungen stützen muss. Aus der formalen Definition der Konditionierung eines autoregressiven Modells folgt mit struktureller Notwendigkeit, dass Möglichkeit hier nicht als vorgängiger Bestand, sondern als Resultat einer Umformung besteht; dass Reflexivität keine zusätzliche, gelegentlich auftretende Eigenschaft, sondern eine architektonische Konstante ist; dass Bestimmung und Ausschluss nicht zwei Vorgänge, sondern ein einziger sind. Diese drei Feststellungen sind keine philosophischen Deutungen eines technischen Sachverhalts, die ebenso gut anders ausfallen könnten. Sie sind, was der Formalismus selbst hergibt.

Damit ist zugleich ein Anspruch begründet worden, der über die gegenwärtige technologische Konjunktur hinausreicht. Die Kette, das Archiv und der Raum sind keine beliebigen, sondern die philosophiegeschichtlich prägenden Selbstbeschreibungen des Denkens, und ihre stärksten Vertreter – der Inferentialismus für die Kette, differenzierte Formen der Wissensorganisation für das Archiv, relationale und topologische Modelle für den Raum – wehren den Vorwurf der Unzulänglichkeit mit erheblicher Raffinesse ab. Dass sie dennoch eine gemeinsame Voraussetzung teilen, die keine ihrer Verfeinerungen preisgibt – die Priorität eines wie auch immer gearteten Gegebenen vor dem Prozess seiner Genese –, ließ sich nicht durch bloße Behauptung zeigen, sondern nur, indem an einem konkreten, vollständig spezifizierten Fall demonstriert wurde, wie eine Struktur beschaffen sein muss, die diese Priorität tatsächlich unterläuft. Die Konditionierung eines Sprachmodells liefert diesen Fall nicht, weil sie ein Modell des Denkens wäre, sondern weil sie, ohne diesen Anspruch zu erheben, zeigt, dass eine solche Struktur überhaupt widerspruchsfrei denkbar und sogar exakt beschreibbar ist.

Der eigentliche Ertrag dieser Untersuchung liegt daher nicht in der Behauptung, generative Systeme dächten. Er liegt in der Verschiebung, die vollzogen wird, sobald man diese Frage für einen Moment zurückstellt und stattdessen fragt, was ein Artefakt, dessen inneres Verstehen ungeklärt bleiben mag, für ein Erkenntnisinteresse leisten kann, das ihm selbst äußerlich ist. Diese Verschiebung hat, wie die Auseinandersetzung mit Deleuze, Simondon und Whitehead gezeigt hat, in der Sache Vorläufer; sie ist im Ausgang von der spezifischen Frage nach der epistemischen Funktion eines technischen Mediums für die philosophische Selbstaufklärung jedoch so nicht gestellt worden. Ein Mikroskop erzeugt keine Zellen; es lässt sie sehen. Was hier gezeigt werden sollte, ist, dass ein Sprachmodell, ungeachtet aller offenen Fragen über sein Inneres, eine analoge Funktion für eine Struktur des Denkens

übernehmen kann, die ohne ein solches Medium kaum in vergleichbarer Schärfe artikulierbar gewesen wäre.

Damit ist auch der Anspruch dieser Untersuchung an ihr Ende gekommen, ohne sich zu erschöpfen. Sie hat weder die Kritik noch die Instrumentalisierung generativer Systeme widerlegt, denn beide beruhen auf Beobachtungen, die zutreffend bleiben. Sie hat einen dritten Ort markiert, von dem aus sich die Frage nach diesen Systemen stellen lässt, ohne in der Alternative von Ablehnung und Werkzeuggebrauch aufzugehen: den Ort eines Mediums, das etwas zeigt, was es selbst nicht sein muss, um es sichtbar zu machen. Ob sich diese Struktur der Formbildung auch am Fall menschlichen Denkens in vergleichbarer Präzision nachweisen lässt, ist eine Frage, die diese Untersuchung aufwirft, aber nicht beantwortet. Genau darin liegt ihr angemessenes Ende: nicht im Abschluss einer Theorie des Denkens, sondern in der Eröffnung einer Frage, die sich, ohne das hier untersuchte Medium, in dieser Form kaum hätte stellen lassen.