Dieter Mersch

Eine monströse, unnatürliche Intelligenz

Zur Kritik synthetischer Rationalität

Meiner

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über ⟨https://portal.dnb.de⟩ abrufbar.

ISBN 978-3-7873-5095-7
PDF 978-3-7873-5096-4
EPUB 978-3-7873-5097-1

Kontaktadresse nach EU-Produktsicherheitsverordnung:
Felix Meiner Verlag GmbH, Richardstraße 47, 22081 Hamburg
info@meiner.de

© Felix Meiner Verlag Hamburg 2026. Alle Rechte vorbehalten. Der Verlag behält sich die Verwertung der urheberrechtlich geschützten Inhalte dieses Werkes für Zwecke des Text- und Data-Minings (§ 44b UrhG) vor. Jegliche unbefugte Nutzung ist hiermit ausgeschlossen.
Umschlaggestaltung: Stefan Adamick. Satz: satz&sonders, Dülmen.
Druck: Stückle, Ettenheim. Printed in Germany

Inhalt

Einleitung 9

1. Kybernetik, Mathematik, Dekonstruktion 24

1.1 Die Landschaften der Kybernetik als mathematische Karte 31
1.2 Die Theoriebasis der Norbert Wiener'schen Kybernetik . . 36
1.3 Die Mathematik im Hintergrund 40
1.4 Einige dekonstruktive Bemerkungen zur kybernetischen Mathematik 46
1.5 Übergang zur Kybernetik zweiter Ordnung 52
1.6 Zur logischen Form der Kybernetik zweiter Ordnung 61
1.7 Eine dekonstruktive Kritik der Kybernetik zweiter Ordnung 69

2. Das ›Rechenbare‹ und das ›Unrechenbare‹ 75

2.1 Logik, Reflexivität und Paradoxie 80
2.2 Die Grenzen des Algorithmischen und das ›Unrechenbare‹ 89
2.3 Das ›Orakel‹ mathematischer Intuition 97
2.4 Synthesis, Einbildungskraft und die Randomisierungen des Indeterminativen 104
2.5 Zur ›Unrechenbarkeit‹ des Schöpferischen 108

3. Können Maschinen denken? 118

3.1 Der Turing-Test und seine Logik 124
3.2 Zur medialen Struktur des Imitationsspiels 127
3.3 Turings ›intelligente‹ Maschinen 133
3.4 Präjudizierte Maschinenhaftigkeit des Tests 139
3.5 Unentscheidbarkeit versus Ununterschiedenheit: einige Einwände gegen Turing 144

4. Inkommensurabilitäten 156

4.1 Computer versus Gehirn 159
4.2 Computer, Denken und Bewusstsein 163

4.3 Heterologien zwischen Computation und Denken. Eine heuristische Skizze 171
4.4 Zur Frage der Beziehung zwischen Syntax und Semantik 177
4.5 Denken und ›Denken Denken‹ 185
4.6 Zwischen *genitivus subiectivus* und *obiectivus*: die Frage des Bewusstseins 193
4.7 Large Language Models (LLMs) und das Problem des ›Sinns des Sinns‹ 199
4.8 Stochastische Papageien 207

5. Die ›Welt‹ der Sehmaschinen 212

5.1 Trevor Paglens *Sight Machine* 215
5.2 Eine nichtmenschliche Art des Sehens 219
5.3 *Sight Machine* als ästhetische *Epistēmē* 228
5.4 Verfehlungen im Technischen 233

6. Von den ›Enden‹ der Kunst 238

6.1 Kunst und die Grundlagen von *Machine-Learning*-Programmen 240
6.2 KI-Generatoren auf der Basis von Bild- und Gesichtserkennungsprogrammen 247
6.3 KI-Generatoren auf der Basis von *Generative Adversarial Networks* 252
6.4 Promptgenerierte Text-Bild-Generatoren 259
6.5 Diskussion um die ›kreativen‹ Generatoren 263
6.6 Revisionen von Kunst und Kreativität 268
6.7 Resümee 275

Fazit 281

Literaturverzeichnis 293

Anmerkungen 311

Es gab allerdings Probleme, die man nicht so leicht vergessen konnte. Zum Beispiel, daß die Maschinen niemals etwas wirklich richtig machten. Und die Tatsache, daß es keine echten Veränderungen mehr gab. Keine neuen Erfindungen, keine neuen Ideen.

Rudy Rucker, *Gödel, Zappa, Rock'n'Roll*

Einleitung

Nach der Niederlage des koreanischen Großmeisters im Go-Spiel Lee Sedol, der damals als einer der besten Spieler der Welt galt, gegen das von Google *DeepMind* entwickelte Computer-Programm *AlphaGo* am 9. März 2016 in Seoul bezeichnete dieser die künstliche Intelligenz (KI) als eine »monströse, unnatürliche Intelligenz«. Das lag vor allem an Zug 37 in der zweiten Partie. Es handelte sich um einen äußerst ungewöhnlichen und unerwarteten Zug, dem viele Kommentatoren etwas Magisches, ja geradezu Unheimliches bescheinigten. Zunächst dachte man an einen Fehler, weil er kaum einzuordnen war, doch erkannte der chinesische Go-Meister Fan Hui in ihm nachfolgend das *Momentum* einer authentischen KI-Kreativität: »Es ist kein menschlicher Zug. Ich habe noch nie einen Menschen diesen Zug spielen sehen. […] So schön.«[1]

Offensichtlich unterschied sich die Art der von der KI durchgeführten Aktionen eklatant von menschlichen Spielzügen, denn die im Hintergrund der KI arbeitende Mathematik basiert auf einer ›dihairetischen Suche‹ auf der Grundlage eines binären Entscheidungsbaumes, der Regelableitungen und Evaluationen durch Feedback sowie die aus der Statistik bekannte zufallsbasierte Monte-Carlo-Methode unter Einschluss spezifischer Optimierungsverfahren miteinander kombinierte. Zwar lässt sich im menschlichen Maßstab ihre einzigartige ästhetische Qualität bewerten, doch haben wir es im Grunde mit einer Inkommensurabilität zu tun, insbesondere auch deshalb, weil das Spiel Go mit seinen 10^{170} Möglichkeiten wesentlich komplexer als das Schachspiel funktioniert und im hohen Maße Intuition verlangt, während das Programm mit mathematischer Präzision und Unerbittlichkeit seine Steine setzt.

Der spontane Befund Lee Sedols, der im Wettbewerb zwischen Mensch und Maschine einmal mehr das irreführende Narrativ von einer weiteren Demütigung der Menschheit, einer Art vierter Kränkung nach der Kopernikanischen Wende, der Darwin'schen

Evolutionsthese und dem Freud'schen Unbewussten provozierte, führt zur Frage nach dem Verhältnis von Computer, Mathematik, Denken, Rationalität und humaner Kreativität. Die Absicht der folgenden Überlegungen ist, das Verhältnis näher auszuloten und es dorthin zurückzustellen, wo es hingehört: In eine Reihe wissenschaftlicher Mystifikationen und Übertreibungen, mit denen künstliche Intelligenzen belehnt werden, um ihre tendenzielle Überlegenheit gegenüber menschlichen Intelligenzen unter Beweis zu stellen. Stattdessen wird es darum gehen, die mit ihnen verbundenen Ansprüche synthetischer Rationalität durch Rückgang auf ihre implizite Mathematik zu trivialisieren und ihnen ihre Grenzen aufzuzeigen. Zwar erlebt die KI-Forschung nach etlichen Rückschlägen und Wintereinbrüchen[2] durch die Erfolge der *Large Language Models* (LLMs) und der multimodalen Text-zu-Bildgeneratoren gegenwärtig eine beispiellose Hausse, gleichwohl scheint deren erneute Krise nur eine Frage der Zeit zu sein. Die Maschinen sind mit keinem Geheimnis behaftet, wie es sich Science-Fiction-Dystopien ausmalen, als ob sie plötzlich Bewusstsein erlangten und ihre Erfinder zu unterdrücken beginnen könnten. Auch türmt sich ihre Intelligenz nicht in schwindelnde Höhen, die den Menschen klein und beschämt zurücklassen; vielmehr sind sie spezialisiert, in ihrer Arbeitsweise beschränkt, voller Fehler und mit Fallen behaftet sowie im Ganzen *andersartig*, sodass es an der Zeit ist, ihre Eigenart und Leistungsfähigkeit einer nüchternen Analyse zu unterziehen. Die These der folgenden Untersuchungen bezieht sich auf diese Andersartigkeit, die zuweilen monströse Züge annimmt und vor allem im Sozialen mit Bezug auf Öffentlichkeit und Vertrauen[3] sowie in Ansehung von Kunst, Wissenschaft und dem kulturellen Sinn überhaupt[4] schwere Schäden anrichtet, die jedoch in Summe weder dämonisiert noch überhöht oder als schlichtweg unverständlich abgetan werden dürfen.

Eine eingehende philosophische Auseinandersetzung mit den jüngsten Entwicklungen der sogenannten KI-Revolution steht jedoch erst am Anfang.[5] Vorranggegangen waren in den 1980er und 90er Jahren bereits umfangreiche Kritiken symbolischer KI-Modelle, vor allem mit Blick auf eine verlässliche Wissensproduktion, der Darstellbarkeit semantischer Zusammenhänge oder ihrer Fähigkeit zu verstehen, sinnvolle Interaktionen zu führen und auf

überzeugende Weise Denken und Bewusstsein zu simulieren – Überlegungen, die angesichts von *Brute Force* und statistischer, auf Mustererkennung in großen Datenmengen basierender Verfahren alt geworden zu sein scheinen, auch wenn sie – aus systematische Gründen, wie wir zeigen werden – kaum an Aktualität eingebüßt haben. Weiterhin steht das ungeheure Versprechen einer möglichen *General Artificial Intelligence* (GAI) am Horizont; träumt die Tech-Industrie von einer fulminanten Erweiterung menschlicher Intelligenz; wird alles versucht, den Maschinen ›Selbstbewusstsein‹ einzuimpfen oder ihnen zumindest den Status selbständiger Akteure zuzusprechen – allerdings sind die Methoden andere, weniger durchsichtig und so, dass in allem, was die Systeme tun oder hervorbringen, schon das menschliche Wissen, die Sprache oder kulturelle Produktionen eingegangen sind, denn die Quelle der Trainingsdaten ist die Gesamtheit des Internets und der darin kursierenden Informationen unter Einschluss aller Fehl- und Desinformation. Tatsächlich fällt die Wette, die menschliche Intelligenz binnen des nächsten Jahrzehnts zu schlagen – mit *ChatGPT*, *Llama*, *Gemini* oder *DeepSeek*, *GROK* oder das europäische *TEUKEN*, um nur die wichtigsten derzeitigen Player zu nennen – nicht weniger vollmundig aus als vorher, wird doch nichts weniger versucht als die menschliche Vernunft, Urteilskraft und schöpferische Kreativität und damit auch die Grundfeste philosophischer Reflexion durch Systeme algorithmischer Berechenbarkeit zu ersetzen.

Die kritischen Interventionen der beiden Dezennien vor der Millenniumswende hatten dagegen noch prinzipielle epistemologische Bedenken erhoben – man denke an Hilary Putnams Zurückweisung radikaler Simulationen durch »Gehirne im Tank«[6] oder an Donald Davidsons Einwände gegen den Turing-Test,[7] an Hubert Dreyfus' phänomenologische Rejektion eines reinen, körperlosen Denkens[8] oder an John R. Searles Gedankenexperiment vom »Chinesischen Zimmer«.[9] Zu erwähnen sind außerdem die vielfältigen Referenzen auf die einschränkenden Resultate der Gödel-Theoreme vor allem bei Roger Penrose[10] und Douglas R. Hofstadters Repliken[11] sowie viele andere Bemühungen um Klärung. Im Gegensatz dazu positionieren sich heutige Kritiken in erster Linie im Feld von Ethik und Ästhetik.[12] Viele Einlassungen üben sich darüber hinaus in interessierten Apologien oder postulieren

eine Kontinuität zwischen hermeneutischen und informatischen Operationen, wie sie umgekehrt eine vorsichtige Relativierung übertriebener Geltungsansprüche unter Anerkennung dessen vornehmen, was als Digital Humanities oder ›digitale Philologie‹ firmiert: Es scheint, als seien besonders die Geisteswissenschaften dem diskreten Charme des Diskreten und seiner mathematischen Totalisierung sowie den schier grenzenlosen Anwendungsmöglichkeiten von KI in Ökonomie, Medizin oder Physik, Biologie und Neurowissenschaften, aber auch in Politik, Rechtsprechung und Textinterpretation bis hin zur Objekt- und Gesichtserkennung restlos erlegen.[13]

In einem bemerkenswerten Essay hat Daniel-Martin Feige allerdings eine *Kritik der Digitalisierung* vorgelegt, die eine Philosophie des Digitalen mit dem kritischen Impetus der Frankfurter Schule zusammenschließt, indem er den spezifischen Vernunfttypus computationaler Systeme aus einer ebenso technisch-instrumentellen wie ökonomischen Rationalität herzuleiten versucht.[14] Insbesondere geht es ihm darum, aufzuzeigen, wie sehr diese Logik, die er mit Stichworten wie »Blackboxing«, »Datafizierung« und »Alterisierung« unterlegt, unserer Grammantik des Sehens, Hörens, Denkens und Handelns verändert, indem sie diese im gleichen Atemzug reduziert wie entstellt, um ihre Gegenstände, Ziele und Inhalte auf eine grundlegende Weise »fungibel« zu machen.[15] »Digitalisierung« bedeute dabei vor allem die »produktive Radikalisierung einer sich verselbständigenden Logik instrumenteller Vernunft«.[16] So treffend die Diagnose auch sein mag, so sehr bleibt sie dennoch auf eine charakteristische Weise blind. Denn sie liest den Prozess der ›Digitalisierung‹ – ein unzureichender Ausdruck, den wir durch den Begriff eines ›computationalen Technizismus‹ ersetzen werden – hauptsächlich als Weiterführung einer ökonomischen und im Besonderen kapitalistischen Organisation, die sich die Wirklichkeit im Raster rein quantitativer Verwertbarkeit gefügig macht. Die Kritik setzt damit nicht auf die Geltung des Digitalen selbst, sondern auf seine Verbindung mit Macht bzw. der Ermächtigung und Aneignung des Realen durch seine ›Datafizierung‹. Das ist sicher richtig und fruchtbar, doch verweist demgegenüber die Geltung des Computationalen auf eine bestimmte Form mathematischer Rationalität, die nicht

mit der instrumentellen oder ökonomischen kurzgeschlossen werden darf. Ihre spezifisch *algorithmische* Architektur kann auf diese Weise nicht in den Blick kommen. Vielmehr bleibt die »Kritik der Digitalisierung« bei der Zerlegung und Rekonstitution der Welt aus der Struktur ihrer digitalen Diskretisierung stehen, das heißt bei ihrer *Codierung* und zielt nicht auf die mathematischen Methoden im Einzelnen, die sie eher summarisch einer statistischen Mustererkennung zuweist. Wie die mathematische Rationalität sich aber zu den klassischen Vernunftbegriffen im Allgemeinen und der instrumentellen Vernunft im Besonderen verhält, bleibt darin offen. Damit treffen die Analysen aber nicht in den eigentlichen Kern jener Theoriemaschinen, die die Diskurse und Modelle computationaler Technologien regeln, mithin auch jedes Ensemble von Prämissen und Verfahren, die die künstlichen Intelligenzen determinieren, um sie mit jenen ungeheuren Ansprüchen zu belehnen, die von ihrer universellen Anwendbarkeit bis zur kognitiven Übertrumpfung menschlichen Vermögens reichen.

Viele philosophische Kritiken besitzen typischerweise ihren Mangel darin, dass sie aus dem Begriff zu argumentieren scheinen, statt von der Sache her, sodass ihnen oftmals etwas Tautologisches anhaftet. So wird etwa eine Inkompatibilität zwischen humanem und maschinellem Denken und Handeln dadurch postuliert, dass, wie es auch bei Feige heißt, nur »rationalen Wesen« sinnvoll ein Denken zugesprochen werden könne, soweit sie ihre Überzeugungen durch Gründe belegen, wie auch gleichfalls »nur ein Lebewesen, das durch den Begriff der Vernunft formal bestimmt ist, […] ein handelndes Wesen sein (kann)«.[17] Nicht nur bleiben beide Aussagen anfechtbar – man denke an indeterministische Formen künstlerischer Performances einerseits oder an die Askesen des Nichthandelns andererseits –, vielmehr erweist sich so, was Denken und Handeln mit Bezug auf den Menschen bedeuten, bereits an Vernunft gebunden und auf ein Apriori an Vernünftigkeit zugeschnitten, wie ebenso die Begriffe einer theoretischen und praktischen Vernunft dadurch prävaliert werden, dass sie einseitig auf Sprache, Urteil und Begründung geeicht werden. Zudem bleibt das spezifische Verhältnis zwischen diesen Rationalitätsformen und der mathematischen, die sich doch nicht minder auf Beweise und Gründe stützt, offen. So können die Zuschreibungen zu einem

gewissen Grad auch für Computer und künstliche Intelligenzen gelten, soweit ihre Programme durch mathematische Termini und formale Sprachen definiert sind, die logisch und folglich selbst schon im Rahmen klassischer Vernunftbegriffe operieren. Sie können nicht vom Menschen getrennt werden, so wenig wie sich Mensch und Maschine vollkommen fremd gegenüberstehen, vielmehr sind die Maschinen zuallererst aus technischen Rationalitäten geboren, die wenigstens Ableitungen aus Vernunftbegriffen darstellen, ja mit ihnen sogar teilweise verwandt sind. Die Behauptung ›computationaler Technologien‹ besteht ja gerade darin, dass diese auf jene reduzierbar und die Komplexitäten humanen Denkens durch mathematische Modelle vollständig simulierbar seien. Die menschliche theoretische und praktische Vernunft wäre so nicht prinzipiell anders als die rechnende, vielmehr deren Fortsetzung und Radikalisierung, sodass zuletzt das dogmatische Gebot, dass ›nur rationale Wesen‹ angemessen denken oder handeln könnten, witzlos erscheint.

Mehr noch: Denken bedeutet vielerlei, denn auch die Kunst ›denkt‹,[18] wie keineswegs klar ist, ob es nicht auch ein emotionales oder körperliches, oder gleichfalls aisthetisches Denken gibt, sodass der Terminus ›Rationalität‹ nicht eindeutig ist, so wenig wie für seine Bestimmung kritische Urteilskraft ausschlaggebend ist. Dass Maschinen nicht denken können, dass sie nicht über eine genuine Kreativität verfügen, dass ihnen im emphatischen Sinne kein eigenständiges Handeln zukommt, kann deshalb nicht von außen dadurch angezeigt werden, dass sie in ein Netz von Grundbegriffen – wie ›Grund‹, ›Überzeugung‹, ›Absicht‹ oder ›Verstehen‹ – eingespannt werden, durch welches sich Denken angeblich erst erfüllt, um es umgekehrt technischen Artefakten abzusprechen. Als ob deren Sinn schon festgelegt sei und die Bedeutung von ›Gedanken haben‹ oder ›Handlungen ausführen‹ und Ähnliches durch sie paraphrasierbar wäre. Vielmehr bedarf es einer *Grenzziehung von innen*, indem die mathematische Rationalität und ihre Technologien an den Rand ihrer eigenen Möglichkeiten geführt werden und so ihre Andersartigkeit gegenüber humanen Rationalitätsformen preisgeben. So sehr also ein Beweis aussteht, dass den Ausdrücken ›Denken‹, ›Freiheit‹, ›Autonomie‹ oder auch ›Bewusstsein‹, ›diskursive Begründung‹ und ›Sinn‹ eine exklu-

sive Besonderheit zukommt und nicht ins Register maschineller Operationen passt, so sehr besteht auf der anderen Seite die Unangemessenheit dieser Argumentation darin, dass sie sich nicht gleichzeitig versichert, was Maschinen können oder kraft ihrer eigenen Möglichkeit zu leisten vermögen und worin sie ihre autochthonen Grenzen, ihre definitiven Demarkationen besitzen. Sie erwiesen sich dann selbst vielleicht als dem Menschen gegenüber mangelhaft und unterlegen, ja sogar – im vielfachen Wortsinn – *beschränkt und inkompatibel*.

Für die philosophische Kritik wäre deshalb zu belegen, dass eine bestimmte Form von Denkvermögen, von Rationalität und Urteilsvermögen aus logischen Gründen nicht im mathematischen Repertoire rekonstruierbar ist, sodass vielmehr zwischen dem Humanem und dem Computationalen ein fundamentaler Riss besteht. Die Fülle und Reichhaltigkeit des Humanen kann dann nicht zur Gänze im ›Rechenbaren‹ aufgehen – andernfalls bliebe die Behauptung seiner Andersartigkeit ein bloßes Postulat.[19] Man kann darüber hinaus schwerlich abtun, dass Mathematik und Logik zum ehernen Bestand dessen gehören, was einst dem *logos* zugedacht war, auch wenn sie nicht den höchsten Punkt im antiken *Quadrivum* ausmachten. Diesen Status erhielten sie erst im frühneuzeitlichen Wissenssystem. Damit verbunden ist eine weitere Schwierigkeit. Denn indem die philosophische Aufklärung des Digitalen sich zuvorderst der humanen Bedingungen vergewissern muss, um sie auf die technisch-mathematischen Verfahren zu applizieren, vermag sie lediglich wechselseitige Zuschreibungen miteinander zu vergleichen. Sie bleiben aber so lange steril, wie sie nicht durch historische Analysen flankiert werden. Diese dekuvrieren beide als Derivate einer sich fortschreibenden metaphysischen Tradition, die wiederum in der Gleichung von Mathematik und Vernunft ihr geheimes Telos fand. Dass in diesem Sinne die mathematische Rationalität zur Idealform der Vernunft überhaupt avancierte, wie sich durchweg im Anschluss an René Descartes und Gottfried Wilhelm Leibniz konstatieren lässt, impliziert, dass das, was anscheinend allein ›rationale Wesen‹ mit Bezug auf ihr Denken und Handeln auszeichnet, in besonderer Weise auch für die computationalen Technologien und ihre mathematischen Bedingungen zutreffen muss, soweit sie als selbständige Akteure

fungieren und Entscheidungen treffen, sodass die Beantwortung der Frage ihrer Übertragbarkeit oder Unterlegenheit sowie Nichtübertragbarkeit systematisch fehlgeht. Im Gegenteil: Unter den Prämissen einer immanenten Teleologie des Rationalen verhält sich das mathematische Denken gerade nicht reduktiv, wie auch die instrumentelle Vernunft nicht als ›halbierte‹ betrachtet werden darf, vielmehr prämiert die mathematische Vernunft in ihrer heutigen Totalisierung die technische, um alle anderen Formen von Rationalität in sich zu schließen und aus sich heraus zu entwickeln und zu begründen.

Wir halten darum die Geltungsfrage und den konträren Nachweis einer Nichtrückführbarkeit des humanen Denkens auf die Grundsätze von Technik, Logik und Mathematik für zwingend. Statt dabei auf gleichermaßen fruchtlose wie irreführende Selbstzuschreibungen zu rekurrieren, plädieren wir für die umgekehrte Vorgehensweise, nämlich für einen dezidierten Nachvollzug jener formalen Dispositionen, die den technisch-mathematischen Simulakren inhärent sind, und zwar sowohl im Hinblick auf ihre Potenziale als auch ihrer Einschränkungen und Barrieren, weil erst dann auffällig werden kann, wo sie versagen, worin ihre Unmöglichkeit besteht oder worin sie gegenüber der humanen Vernunft mangelhaft oder sogar monströs erscheinen. Den Ausgangspunkt unserer Überlegungen bilden deshalb die technischen und mathematischen Voraussetzungen der computationalen Maschinen und ihrer Modelle und Programme, um an ihnen aufzuzeigen, was sie nicht können, was sie zu menschlichen Vermögen in eine prinzipielle Distanz bringt, um gleichzeitig aufzuweisen, was an menschlicher Vernunft, an den Formen humanen Denkens und Handelns einer anderen als bloß syntaktischen Logik gehorcht und mathematischen Formalisierungen widersteht. Dazu gehört im Wesentlichen, am menschlichen Denken weniger seine Vernünftigkeit, das *sapere aude* der Aufklärung[20] zu betonen, als vielmehr – ebenfalls von Immanuel Kant her – das »Ich denke, [das] alle meine Vorstellungen begleiten können [muss]«, wie es in der »transzendentalen Deduktion« der *Kritik der reinen Vernunft* heißt:[21] Mithin das jederzeitig begleitende Vermögen einer Reduplizierung im Denken, denn – so unsere These – Denken bedeutet in erster Linie und immer schon ein ›Denken des Denkens‹, das

von sich her Reflexivität einschließt. Es ist diese reflexive Duplizierung, die anders ist und einer anderen Grammatik gehorcht, als sie aller mathematischen Formalisierung möglich ist, insofern die algorithmischen Modelle sie einzig durch eine sich unendlich aufstufende Rekursivität zu imitieren vermögen. Diese bleibt jedoch gegenüber der genannten Duplizität eines sich denkenden Denkens, wie es sich grammatisch im zweifachen Genetiv des *subiectivus* und *obiectivus* ausdrückt, systematisch inadäquat: Wir werden noch darauf zurückkommen.

Die Formel ›Denken Denken‹, auf die wir im Laufe unserer Untersuchungen immer wieder zu sprechen kommen werden (besonders in Kapitel 2 und 4), zeichnet somit das aus, was wir für das charakteristische Kennzeichen humanen Bewusstseins und Selbstbewusstseins halten, nämlich die Fähigkeit, sich gewiss zu sein, *dass* wir denken, *was* wir denken. Im Denken sind folglich – das wusste auch der Deutsche Idealismus – Was-sein (*quid*) und Dass-sein (*quod*) untrennbar miteinander verwoben. Eine Kernaussage unserer Überlegungen besteht dann darin, zu zeigen, dass diese Verschränkung im Rahmen computationaler Technologien weder logisch noch mathematisch oder technisch ›re-modellierbar‹ ist. Was Denken heißt – und das schließt seine transzendentale Semantizität genauso ein wie seine Kreativität –, erweist sich dann durch formale Systeme, seien sie symbolisch oder statistisch verfasst, als nicht rekonstruierbar. Künstliche Intelligenzen verhalten sich gegenüber menschlichen Intelligenzen als *alteritär*.

Das wird in sechs Schritten und sechs Kapiteln durchgeführt: *Erstens* mit Bezug auf kybernetische Feedbacksysteme und rekursives Lernen, *zweitens* in Ansehung des Turing-Tests und seiner Verwechslung zwischen Unentscheidbarkeit und Ununterschiedenheit, *drittens* durch eine Ausbuchstabierung interner Grenzen einer Mathematik des Denkens, *viertens* hinsichtlich der wesentlichen Inkommensurabilität zwischen humanem Denken – das immer schon das Denken des Denkens einschließt – und symbolisch oder statistisch operierenden KI-Systemen, *fünftens* im Hinblick auf die in ihnen eingeschlossene ›Welt‹ und schließlich *sechstens* mit Blick auf das Problem von Kreativität – vormals bei Kant der produktiven Einbildungskraft als dem Kern der Synthesis im Denkens. Die Usurpation von Kunst aus vermeintlich schöpfe-

rischen Bild-, Text- oder Soundgeneratoren, die den Lackmustest für eine mögliche maschinelle Kreativität bietet, erweist sich dabei als zufallsgesteuert, sodass wir es abermals mit einer Verwechslung zwischen Kreativität und Zufall zu tun bekommen. In der Tat beruht unser Hauptargument einer Inkommensurabilität zwischen Denken und Computationalität bzw. computationaler Maschinität auf diesen Verwechselungen. Wir werden sie in unterschiedlichen Varianten immer wieder durchspielen.

Im Einzelnen beschäftigt sich das *erste* Kapitel mit den theoretischen Grundlagen des Computationalen, insbesondere dessen Fundierung in einem mehrdimensionalen Theorieensemble aus radikalem Konstruktivismus, Informationstheorie, Systemtheorie und Kybernetik, wobei Letztere in Bezug auf die Ersteren als Oberbegriff fungiert. Uns geht es vor allem darum, zu zeigen, dass sämtliche dieser Theorieansätze auf bestimmten mathematischen Konzepten basieren, die ihr Maß, ihre Ausrichtung und Reichweite auf eine entscheidende Weise mitprägen. Eine maßgebliche Rolle spielt darin die Metamathematik der 1930er und 40er Jahre, wie sie mit Theorien der Entscheidbarkeit und Berechenbarkeit die Grundlagen des Computationalismus und damit der digitalen Technologien erst legten. In deren Zentrum stehen wiederum die Begriffe der Turingmaschine und der rekursiven Funktion, die zusammen mit der graphentheoretischen Beschreibung von Netzwerken die entscheidenden Modelle für die auch neurowissenschaftlich gestützte Isomorphie-These zwischen Gehirnen und Digitalcomputern bilden. Die fatale Gleichsetzung hat die Forschung seit über 70 Jahren irregeleitet. Was dabei die Kybernetik betrifft, so zielte diese in erster Linie sowohl auf eine Modellierung der zirkulären Logik dynamischer Systeme als auch auf die logische Anordnung von Metasystemen, wie sie scheinbar ebenfalls unsere Hirnstrukturen bestimmen. Allerdings bleiben auf diese Weise zentrale Dimensionen des humanen Denkens ausgeklammert, sodass ihre formale Rekonstruktion chronisch *unterhalb* ihrer emergenten Komplexitäten bleibt.

Das *zweite* Kapitel diskutiert dasselbe Problem vor dem Hintergrund der metamathematischen Theorien selbst, insbesondere der beiden Gödelsätze und ihrer Konsequenzen. Offensichtlich besitzen computationale Technologien ihre Beschränkung in dem, was

mathematisch als nicht berechenbar gilt, wozu im Besonderen die Berechnung ihrer eigenen Berechenbarkeitsweise bzw. die Frage zählt, ob mit Bezug auf ein präzise genug formuliertes Berechenbarkeitsproblem Maschinen in der Lage sind, in einer endlichen Zeitspanne befriedigende Lösungen zu erbringen. Wohl vermögen Turingmaschinen *prinzipiell* alles zu berechnen, was berechenbar *ist*; *wofür* es aber einen Algorithmus geben kann und wofür nicht, können sie hingegen nicht berechnen, denn es gibt keinen Algorithmus über alle möglichen Algorithmen, so wenig wie es eine ›Mathematik der Mathematik‹ geben kann. Die mathematischen Programme konfrontieren uns damit selbst mit ihrer eigenen Demarkation. Das gilt auch in Ansehung der Entwicklung kreativer Lösungsstrategien, die ihrerseits nicht wieder totalisierbar sind. Und doch besteht die Frage, ob wir nicht die gleichen Beschränkungen für das humane Denken annehmen sollten und müssen, sodass sich wiederum die Frage der Reichweite des Arguments stellt.

Wir leiten aus diesem Komplex die Skizze einer – im kantischen Sinne – ›Kritik algorithmischer Rationalität‹ ab, die zugleich eine immanente Kritik künstlicher bzw. synthetischer Intelligenz und Kreativität beinhaltet. Zu ihr gehört eine Zurückweisung der Isomorphie-These zwischen Gehirnen und Computer oder humanem Denken und Maschinen. Im Zentrum unserer Ausführungen steht vielmehr die Frage, was Denken heißt sowie das, was sich als Triangulation zwischen Sinn, Reflexion und Bewusstsein exponieren lässt und – so die weitere These – keiner synthetischen Vernunft zugänglich ist. Das Bild der Triangulation postuliert zwischen den Ausdrücken einen wechselseitigen Verweisungszusammenhang, der jedes Nachdenken über menschliches Denken und seine Bestimmung terminiert, sodass diese Ausdrücke nur im Zusammenhang, nicht isoliert betrachtet werden können. Es ergibt deshalb wenig Sinn, von einer Maschine zu behaupten, sie könne Reflexion ohne Bewusstsein oder Bewusstsein ohne Zugang zu komplexen, nichtlinearen Semantiken nachahmen. Unmittelbar damit verbunden ist dann die weiterführende Frage, ob Computer, seien sie formal intelligent oder nicht, jemals ein Bewusstsein bzw. Selbstbewusstsein und damit Denken in einer menschlichen Bedeutung erlangen können, was zugleich die alte Fragestellung

aus der Frühzeit der Computerisierung wiederholt, ob künstliche bzw. synthetische Systeme intentional sein können, ob sie im gehaltvollen Sinne etwas zu entscheiden oder zu wissen bzw. zu verstehen vermögen, ob sie also über Einsichten darüber verfügen, was sie selbst sind oder tun, oder ob sie sich gegebenenfalls selbst relativieren bzw. ihre Bezugnahmen revidieren und folglich auch zu gehaltvollen Selbstkorrekturen fähig sein können.

Das *dritte* Kapitel nimmt den Faden auf und spinnt ihn wiederum auf umgekehrte Weise fort, nämlich mit dem Vergleich zwischen synthetischer oder artifizieller Intelligenz einerseits und menschlichem Denken andererseits aufgrund des von Alan Turing vorgeschlagenen Testverfahrens. Es wird selbst noch einmal einer Geltungskritik unterzogen. Denn bei aller Heuristik und provisorischer Anordnung reduziert der Test die Frage von Unterschiedenheit (oder Ununterschiedenheit) von humanem und maschinellem Denken auf ein quasi-mathematisches *Entscheidungsproblem*. Er gleicht dann selbst einem – gewissermaßen analog konzipierten – Computerprogramm, das die Identität und Differenz zwischen Menschen und Maschinen entlang eines formalen Arrangements prüft, das wiederum in eine binäre Ja-Nein-Stellung mündet. Nicht nur reicht, wie Donald Davidson kritisch angemerkt hat, die regelrechte Produktion verstehbarer Sätze keineswegs aus, um Intelligenz unter Beweis zu stellen, sodass der Test nicht leistet, was er zu leisten vorgibt; vielmehr verwechselt seine Struktur selbst noch einmal ›Unentscheidbarkeit‹ mit ›Indifferenz‹ und damit Gleichheit, auch wenn Turing im Sinne einer zweifachen Negation, die immer nur bis zu einer gewissen Unbestimmtheit gelangt, einzig mit aller Vorsicht zuzugeben geneigt war, dass, wo der Test gelingt und eine Nichtentscheidbarkeit beider hervorbringt, wir lediglich *keine Gründe* besäßen, Computersystemen ein Denken *abzusprechen*. Dass wir ihnen dann notwendig Intelligenz zusprechen müssten, ist damit nicht gesagt. Turing, der zwar an das Denkvermögen von Computern glaubte, hielt sich hier bedeckt.

Computer sind – von ihrer Programmstruktur her – binäre Entscheidungsmaschinen; als Programme fußen sie auf Algorithmen, die grundsätzlich als Funktionen darstellbar sind. Das gilt auch für die statistisch verfahrenden Machine-Learning Programme

und die neuronalen Netze (*neural networks*). Aus Perspektive der grundlegenden Gödeltheoreme und dem Turing'schen Halteproblem aus betrachtet vermögen sie zwar eine Vielzahl von Berechenbarkeitsproblemen zu lösen, aber bei Weitem nicht alle. So heißt eine Turingmaschine – als mathematisches Verfahren – ›entscheidbar‹, wenn ihr Algorithmus nach einer endlichen Anzahl von Schritten mit einer Lösung zum Stehen kommt. Doch ist es unmöglich, in Bezug auf eine Turingmaschine und ein Berechenbarkeitsproblem auf deduktivem Wege zu entscheiden, *ob* sie anhalten wird oder nicht, das heißt ›entscheidbar‹ oder ›unentscheidbar‹ ist.[22] Dies wird im *vierten* Kapitel insbesondere anhand dessen beleuchtet, was wir als ›Denken Denken‹ ausgewiesen hatten. Es führt zu einer exemplarischen Liste von Reduktionismen oder Inkommensurabilitäten, die die Diskurse beherrschen, in deren Mitte sich neben der Duplizierung und der formalen Nachbildung durch eine logische Metasprachenhierarchie die Frage einer zureichenden Formalisierbarkeit des Semantischen stellt. Dies – und das ist das Überraschende – zwingt aus Konsistenzgründen zu einem Ausschluss von Selbstreferentialität. Reflexivität emergiert gerade nicht aus fortgesetzter Selbstanwendung; die These ist hier vielmehr, dass das computationale Denken ohne solche Ausschlüsse nicht funktionieren kann und daher die verschiedenen Modi humanen Denkens verfehlt. Die Inkonsistenz besteht dabei darin, dass die mathematische Rekonstruktion von ›Denken Denken‹ zu einer Struktur ohne Selbstreflexion führt. Sie birgt von sich her eine Disparität oder Unvergleichbarkeit zwischen menschlichem und mechanischem Denken. Die Behauptung ihrer Inkommensurabilität meint dann im Wortsinn etwas, was *keinem gemeinsamen Maß* genügt, was also sowohl *unvereinbar* als auch *unausmessbar* und darum *maßlos* ist. Was wir als spezifische ›Monstrosität‹ künstlicher Intelligenzen apostrophieren, fußt auf dieser Maßlosigkeit.

Aus ihr leiten sich dann als weitere Maßlosigkeiten die Nichtrückführbarkeit von Semantik auf formale Syntax sowie die Differenz zwischen schöpferischer Kreativität und Zufallssteuerung ab, die zuletzt paradigmatisch auf die Spur einer prinzipiellen Differenz zwischen Mechanität und Humanität führen. Sie gilt auch für maschinelles Lernen im Kontrast zu menschlichem Lernen. Beide sind von je anderer Art. Anhand des Beispiels von Identifika-

tions- oder Erkennungsprogrammen wie etwa den *Convolutional Neural Networks* (CNNs) stellt sich darüber hinaus die Frage, auf welche Art von ›Welt‹ sich diese berufen – nicht nur, welche Welten Maschinen sehen, wenn sie etwas sehen, sondern auch, was sie in einem phänomenologischen Sinne unter ›Welt‹ fassen und wie ihre Welten, im Unterschied zu erfahrbaren Wirklichkeiten, gebaut sind. Gegenstand des *fünften* Kapitels ist daher die Untersuchung, mit welchen durch die Trainingsdaten und ihrer Verarbeitungsweise präfigurierten Objekten Künstliche- Intelligenz-Systeme ausgestattet sind, was sie also unter Realität ›verstehen‹ und was somit für sie erkennbar ist und was nicht. Denn in Wahrheit bekommen wir es überall mit vorentschiedenen ›getaggten‹ Daten zu tun, die das Reale im Lacan'schen Sinne auf schlichte Objektontologien verkürzen, die zudem diskret angeordnet sind. Es werden uns primitive Semiotiken aus lauter wohlunterschiedenen Zeichen und nominalistische Zuordnungen präsentiert, die mit eindeutig identifizierbaren Indizes versehen sind, sodass sich ›Welt‹ und was als ›welthaltig‹ gilt, in eine Serie standardisierter Klischees verwandelt. Daraus gebiert sich, wie es wiederum Roland Barthes treffend formuliert hat, unweigerlich das »Monstrum Stereotypie«.[23] Maschinen stereotypisieren ›Welt‹, was in Sonderheit dort zum Problem avanciert, wo wir ihre Auskünfte mit Reliabilität belehnen. Es bedurfte bezeichnenderweise einer künstlerischen Aktion wie Trevor Paglens *Sightmachine* (2018), diese klischierten Räume der Maschinenwelten ihrerseits sichtbar zu machen, und zwar in Gestalt einer multimodalen Installation, die selbst nicht anders als monströs zu beschrieben ist. Denn was sie auslöste, war nicht nur Erschrecken und Erstaunen über die beispiellose Dummheit der Maschinen, sondern auch Gelächter, das erlaubte, ihnen den Ernst und die vermeintliche Erhabenheit ihrer erfolgreichen Massenanwendung zu nehmen.

Folgerichtig mündet das *sechste* Kapitel mit dem Titel *Von den Enden der Kunst* in eine Reihe von Überlegungen zum aufklärerischen Potenzial der Künste und dem Verhältnis zwischen Ästhetik, Denken, Reflexivität und maschineller Logik. Es wirft zugleich die tieferliegende Frage nach der relationalen Ordnung zwischen maschineller Generativität und Kunstproduktion auf. Dabei behandeln die ›Enden‹ der Kunst sowohl den sich aufdrängenden

Hegel'schen Topos vom *Ende der Kunst* angesichts ihrer drohenden Substitution und Entwertung durch generative Systeme als auch gleichzeitig deren Ende im Sinne ihres eigentlichen Zwecks und nicht zuletzt ihrer Positionierung am Ende und Ausblick der Erörterungen selbst, deren Kritik offensichtlich der künstlerischen Intervention und einem Lernen durch Kunst im Sinne einer prinzipiellen Unersetzbarkeit ihrer ästhetischen Reflexivität bedarf. Der Kunst fällt so ein notwendiger und zugleich zukünftiger Status zu. Ihre ästhetische Vernunft bezeugt darin gegenüber der monströsen synthetischen Rationalität der Computersysteme ihren einzigartigen humanistischen Wert.

Die Ausführungen schließen als Fazit mit einem Ausblick auf ein angenommenes künftiges Scheitern Künstlicher Intelligenz-Systeme. Diese sind mit Fehlern und Fehlschlägen behaftet, die sie auf lange Sicht ebenso ruinieren werden wie zuvor schon die beiden anderen großen ›Errungenschaften‹ der Digitalisierung: das Internet und die sogenannten Sozialen Medien. Vier augenfällige praktische Versagungen werden neben den theoretischen des Modellplatonismus und der Verwechslung zwischen mathematischen Systemen und ihrer Empirizität diskutiert: *Erstens* die im Rahmen von KI nicht wieder entscheidbare Differenz zwischen Wahrheit und *fake*. *Zweitens* das Problem der Qualität der Trainingsdaten, die aus einem diffusen Korpus gezogen werden und damit Unschärfen produzieren. *Drittens* das, was unter dem Begriff des ›Model Collapse‹ verhandelt wird, der dann entsteht, wenn KI-Systeme selbst auf KI-generierte Daten zurückgreifen und damit einem Element von Zirkularität unterworfen sind, das unweigerlich ›Demenzen‹ erzeugt. Schließlich sind wir *viertens* mit der Unlösbarkeit statistischer Halluzinationen konfrontiert, die aus systematischen Gründen bestenfalls abgemildert, nicht aber ausgeschlossen werden können. Wir wissen darum nie, *was* eine KI sagt: ob sie fabuliert oder Wissen reproduziert. Es bedarf der menschlichen Verifikation, und sie ihrerseits an Maschinen abzutreten, wäre fatal.