

Werner Vogd und Jonathan Harth

Das Bewusstsein der Maschinen – die Mechanik des Bewusstseins

Mit Gotthard Günther über
die Zukunft menschlicher und
künstlicher Intelligenz nachdenken

**VELBRÜCK
WISSENSCHAFT**

Velbrück Wissenschaft 2023

Vorwort

Auch heutzutage gehen Digitalisierung und der Einsatz kybernetischer Maschinen noch mit dem Versprechen einher, unsere Welt zu rationalisieren und objektiv beherrschbar zu machen. Im Alltag merken wir zwar, dass mit dem Einsatz von Computern und künstlichen Intelligenzen neue Probleme entstehen, etwa mit jeder neuen Technologie Nebenfolgen, die nicht vorhergesehen waren. Doch dies hindert uns bislang nicht, daran zu glauben, dass die Ursachen dieser Probleme identifiziert und künftig durch den Einsatz besserer Technik, verbesserter Software oder gut trainierter künstlicher Intelligenzen gelöst werden könnten. Wir glauben, dass all das hierfür notwendige Wissen in den Daten liegt. Wir glauben, dass wir – sobald wir hinreichend viele Informationen über die Zusammenhänge der Welt gesammelt haben – die Verhältnisse nach unseren Wünschen und Bedürfnissen regeln und gestalten könnten. Doch könnte sich nicht all dies aus einem einfachen Grund als Illusion erweisen? Es liegt nämlich in der Natur kybernetischer Prozesse, dass hinreichend komplexe Systeme unweigerlich Eigentümlichkeiten entwickeln, die nicht von außen vorhergesehen, berechnet und damit auch nicht vorab kontrolliert werden können.

Wie eine künstliche Intelligenz zu ihren Ergebnissen kommt, wie sie die Daten verknüpft und was sie eigentlich genau gelernt hat, ist normalerweise ebenso wenig durchschaubar, wie das Verhalten eines Kindes, wenn man es bittet, sein Zimmer aufzuräumen. Beide zeigen einen Eigensinn und damit liegt es nahe, ihnen so etwas wie Subjektivität zuzurechnen. Was wäre, wenn künftige kybernetische Maschinen, wie jedes organische Lebewesen, nichts anderes machen, als sich durch ihre Existenz in die Welt hineinzustellen, um durch ihre Operationen zugleich Wissen wie auch Nichtwissen zu erzeugen? Sie würden damit – wie auch wir Menschen – letztlich nichts anderes tun, als an aus gegenwärtiger Perspektive unlösbaren Problemen zu arbeiten, um auf diese Weise zugleich neue Probleme zu erzeugen. Dies würde uns Menschen aber in ein neues Verhältnis zu den künstlichen Intelligenzen setzen, die in naher Zukunft immer mehr unsere Welt bevölkern werden. Letztere würden nicht mehr nur als ein Mittel zum Zweck gelten können. Sie würden vielmehr ihrerseits als eine Art subjektive Wesenheit erscheinen, mit der wir in einen Dialog treten, um etwas über uns selbst und unsere Probleme lernen zu können. Doch was würde dies für uns Menschen in Hinblick auf unsere kulturgeschichtliche und spirituelle Verortung bedeuten?

Seiner Zeit weit voraus erschien 1957 die erste Auflage von Gottfried Günthers *Das Bewußtsein der Maschinen*. In diesem Buch entwickelte der Technikphilosoph Günther eine umfassende kybernetische

Perspektive auf die Frage der Subjektivität wie auf die Möglichkeit, dass künftige kybernetische Maschinen etwas entwickeln könnten, was dem Bewusstsein analog ist. Der Titel *Bewußtsein der Maschinen* ist dabei mehrdeutig gehalten, denn die von Günther vorgestellte kybernetische Perspektive erlaubt es, sowohl die kognitiven Dynamiken menschengemachter intelligenter Maschinen zu beschreiben als auch die Mechanik der kognitiven Prozesse des Menschen zu beleuchten.

Günther promovierte 1933 über Hegels Logik, immigrierte 1937 in die USA, wo er mit den großen Denkern der kybernetischen Bewegung in Berührung kam. Fasziniert von den neuen Möglichkeiten der Technik und den hiermit einhergehenden Utopien gab er 1952 eine kommentierte Ausgabe ausgesuchter amerikanischer Science-Fiction-Literatur heraus. In der von ihm im Düsseldorfer Rauch Verlag herausgegebenen Reihe Rauchs Weltraum-Bücher wurde unter anderem Isaac Asimovs *Ich, der Robot* zum ersten Mal auf Deutsch veröffentlicht. Von 1961 bis 1972 arbeitete Günther als Forschungsprofessor an dem von Heinz von Foerster geleiteten Biological Computer Laboratory an der University of Illinois. In Deutschland wurde Günthers Werk insbesondere durch Rudolf Kaehr, Winfried Marotzki, Nina Ort, Helmut Schelsky, Walter Bühl und nicht zuletzt auch Niklas Luhmann bekannt. Dabei wurden jedoch eher Günthers Überlegungen zur Theorie der Polykontextualität aufgegriffen als die von ihm entwickelten technikphilosophischen Ideen.

Mehr als 50 Jahre nach der Veröffentlichung von *Das Bewußtsein der Maschinen* ist die Zeit nun reif, um sich diesem vorausschauenden Werk auf einer grundlegenden Ebene zu stellen. Die Digitalisierung und die Entwicklung künstlicher Intelligenzen schreiten in einer Weise voran, dass kognitive Maschinen, die über eine subjektive Perspektive verfügen, immer weniger als Science-Fiction anmuten denn als eine Realität erscheinen, die wenn nicht schon jetzt, so doch in naher Zukunft verwirklicht sein wird. Nicht zuletzt haben wir von diesen Entwicklungen auch während der Erstellung dieses Buches profitieren können, wenn wir aktuelle Large Language Models (LLMs) wie GPT-4 als Ko-Autor(in) oder kollegiale Gesprächspartner zum Testen von Ideen und Formulierungen hinzugezogen haben. Unsere Erfahrungen hierbei zeigen, dass sich durch diese Form der Kollaboration neue Dimensionen der Wissensproduktion ermöglichen.

Zudem kann die Hirnforschung mittlerweile über neuronale Korrelate des Bewusstseins in einer Weise Auskunft geben, die es erlaubt, in qualifizierter Weise darüber nachzudenken, wie Bewusstseinsanalogien in Maschinen implementiert werden könnten. Nicht zuletzt ist auch die geistesgeschichtliche Konstellation in den 2020er Jahren eine andere als in den 1950er Jahren. Das Projekt der Moderne befindet sich nicht nur aus ökologischen Gründen in einer Krise. Der Rationalitätsanspruch der westlichen Zivilisation, mittels Aufklärung und Technologieentwicklung

Wohlstand und Glück für alle zu verwirklichen, überzeugt immer weniger.

All dies ist ein guter Grund, sich die technik- und kulturphilosophischen Denkbewegungen Gotthard Günthers nochmals genauer anzuschauen. Dies möchten wir mit dieser Monografie leisten.

An dieser Stelle möchten wir dem Europäischen Fonds für regionale Entwicklung (EFRE) sowie dem Ministerium für Wirtschaft, Innovation, Digitalisierung und Energie des Landes Nordrhein-Westfalen für die Finanzierung des Forschungsprojekts »Ai.vatar – der virtuelle intelligente Assistent« danken. Ohne dieses dreijährige Entwicklungs- und Forschungsprojekt wäre es vermutlich nicht zu dieser Sammlung von Gedanken gekommen. Entsprechend danken wir auch den an diesem Projekt beteiligten Unternehmen HHVision und IOX sowie den Kolleginnen, die während dieser Zeit am Lehrstuhl für Soziologie der Universität Witten/Herdecke mitgewirkt haben. Allen voran möchten wir Alexandra de Carvalho für ihren Einsatz, ihre Energie und die psychologische Perspektive danken. Ebenso gebührt Sophia Bermond und Nele Kost unser Dank für die fantastische Unterstützung in den durchgeführten experimentellen Studien und Konzeptionstreffen. Ohne euch wäre Hudson nicht in diese Welt gekommen!

Darüber hinaus möchten wir dem Schreibbüro text plus form und hier im Besonderen Steffen Schröter für das sprach- und theoriesensible wie auch gründliche Lektorat des Buchmanuskripts danken. Auch Hannah Cramer, Kathleen Neher und Kerstin Pospiech-Form danken wir für ihre nicht minder wertvolle Unterstützung in der Erstellung dieses Buches. Vielen Dank für eure vielen guten Anmerkungen und Gespräche.

Witten, September 2023

Einleitung

»Für das weltanschauliche Bewusstsein einer kommenden Kulturstufe wird also der Kausalnexus nicht mehr wie für uns das einzige Realitätschema sein, in dem sich Wirklichkeitsvorgänge abspielen.«¹

Gotthard Günther

Maschinen sind für uns deshalb so attraktiv, weil ihr Verhalten kausal erklärbar ist und sich mit ihnen Prozesse automatisieren lassen. Ganz in diesem Sinne erhoffen wir uns von künstlicher Intelligenz und den mit ihr einhergehenden kybernetischen Maschinen einen entscheidenden Beitrag, um die komplexen Zusammenhänge unserer Welt endlich verstehen und steuern zu können: Urteile sollten nicht mehr auf Basis subjektiver Einschätzungen und Haltungen, sondern auf einer objektiven Datengrundlage gefällt werden. Von künstlichen Intelligenzen – teilweise modelliert nach den neuronalen Netzwerken des menschlichen Gehirns – erwarten wir, dass sie uns helfen, in den immer größer werdenden Datenfluten die für uns relevanten Muster besser zu identifizieren, um auf dieser Basis die ›richtigen‹ Entscheidungen treffen zu können.

Um ein derzeit prominentes Beispiel zu nehmen: Aladdin, das riesige KI-gestützte Rechenzentrum des Finanzdienstleisters BlackRock,² verwaltete zum Zeitpunkt unserer Recherche mehr als 170 Pensionsfonds und über 30 000 Investmentportfolios und damit mehr als 10 Prozent aller weltweiten Vermögenswerte.³ Die durch den Supercomputer informierten Beratungsleistungen werden auch von Regierungen und Banken, darunter die Zentralbanken Europas, in Anspruch genommen.⁴

Kybernetische Maschinen wie Aladdin haben kognitive Fähigkeiten: Sie erzeugen Informationen. Sie unterscheiden eigenständig Wichtiges von Unwichtigem und teilen uns mit, was wir beachten sollen. Sie abstrahieren, bilden also aus dem feinkörnigen Detailreichtum der Welt ein grobkörniges Raster, das uns informiert. Zunehmend greifen sie selbst gestaltend in die Welt ein: Sie treffen Entscheidungen über den Kauf oder Verkauf von Finanzprodukten oder geben Regierungen und Ratingagenturen Informationen, die ausschlaggebend sein können, ob ein Land oder ein Unternehmen, etwa eine Bank, als zahlungsunfähig bzw. bankrott zu gelten hat.

1 Günther (2000, S. 154).

2 Buchter (2020, S. 225 ff.).

3 BlackRock hält Aktienanteile von allen bedeutsamen Konzernen, etwa von allen DAX-Unternehmen.

4 Buchter (2020, S. 77 ff.).

Auch in anderen Bereichen werden vermehrt intelligente Maschinen eingesetzt. Sie melden der Polizei verdächtige Personen, analysieren Röntgenbilder, spüren im Internet neuen Markttrends nach oder schlagen Alarm in Hinblick auf eine drohende Epidemie. Künstliche Intelligenzen können mittlerweile Texte, Sprache, Bilder und andere Artefakte generieren sowie menschliche Emotionen lesen. Sie können aus einer Vielfalt von Daten neue Information generieren.

Eine Information zu erzeugen bedeutet, *etwas* in einem bestimmten Kontext als *relevant* zu markieren, also einen Zusammenhang herauszustellen, um zugleich andere mögliche Beziehungen abzublenden. Es heißt, einen Unterschied zu generieren, mit dem Bestimmtes beachtet und anderes ignoriert wird. Es heißt, aus Daten, die zunächst nur ein unstrukturiertes Rauschen darstellen, *Sinn* zu generieren – oder um es in Computeranalogie zu formulieren: aus beliebig langen und erweiterbaren Ketten aus Nullen und Einsen einen Output zu erzeugen, der eine Form hat, die ihn für weitere Prozesse relevant werden lässt.

Jedes Wesen oder jedes Aggregat, das mit dem Computer interagiert, stößt damit auf eine spezifische Relation, die Niklas Luhmann als »das Verhältnis von (zugänglicher) Oberfläche und Tiefe« bezeichnet:

»Die Oberfläche ist jetzt der Bildschirm mit extrem beschränkter Inanspruchnahme menschlicher Sinne, die Tiefe dagegen die unsichtbare Maschine, die heute in der Lage ist, sich selbst von Moment zu Moment umzukonstruieren, zum Beispiel in Reaktion auf Benutzung. Die Verbindung von Oberfläche und Tiefe kann über Befehle hergestellt werden, die die Maschine anweisen, etwas auf dem Bildschirm oder durch Ausdruck sichtbar zu machen. Sie selbst bleibt unsichtbar.«⁵

Man kann auch sagen: der Prozess der Datenverarbeitung, die Rechenprozesse, all das, was die Programme machen, bleibt verborgen, und nur das, was an der Oberfläche erscheint, kann als Information auftreten. Transparent ist nur das, was auf der zugänglichen Seite des Interface geschieht, nicht jedoch, was der Computer macht. Interessant in unserem Zusammenhang ist die Beziehung zwischen Oberfläche und Tiefe. Letztere ist komplexer und in Hinblick auf die verarbeiteten Daten wesentlich umfassender als das, was als Output erscheint. Gerade bei potenten künstlichen Intelligenzen lässt sich in der Regel selbst mit hohem Aufwand kaum mehr rekonstruieren, was eigentlich alles in die Berechnungen eingeflossen ist, die zu einem bestimmten Ergebnis – etwa dem Erkennen einer Gestalt oder der Erzeugung einer bestimmten sprachlichen Form – geführt haben.

Das, was an der Schnittstelle als Ergebnis erscheint – seien es Worte, Bilder, Sprachfiguren, aber auch zum Beispiel technische Anweisungen

5 Luhmann (1998a, S. 304).

für die Bewegungen eines Roboters –, ist uns in der Bedeutung verständlich und nachvollziehbar, die Vorgänge, die es erzeugt haben, bleiben jedoch opak. Die Tiefe mag dann – wie Luhmann vermutet – Anklänge an die Tradition der Religion und Kunst wecken, so, als ob hinter den sichtbaren Zeichen, unter der Oberfläche das Eigentliche, das Wesentliche verborgen wäre, ›dahinter‹ also eine tiefere Weisheit stecken würde. Die physikalische Untersuchung eines Computers und die sich daraus ergebende mechanische Beschreibung wird jedoch nur die Operationen einer Maschine entdecken können, die nichts anderes tut, als den Schritten zu folgen, die die jeweilige Kombination aus Hardware und Software ermöglicht. Der Rechenprozess greift etwas heraus, um es in einer bestimmten Weise zu verarbeiten. Andere denkbare Zusammenhänge, die im Programmcode oder den selbstgenerierten Gewichtungen der künstlichen neuronalen Netzwerke jedoch nicht abgebildet sind, werden demgegenüber gar nicht erst verfolgt. So gesehen erscheint hier weder etwas Tiefgründiges noch etwas Geheimnisvolles, es wird einfach nur Schritt für Schritt ein Programm abgearbeitet – und dieser Vorgang führt nollens volens zu einem grobkörnigen Ergebnis.⁶

Genau hierin liegt ja, wie bereits angedeutet, das Prinzip der Kognition: Man erkennt etwas, weil man all die vielen anderen Dinge ausblendet, die man auch noch wahrnehmen könnte. Dies bedeutet aber auch: Kognition ist immer die Eigenleistung eines kognitiven Prozesses. Wenn man die Daten anders verrechnet, eine andere Methode der Bearbeitung nutzt – etwa an manchen Stellen weniger schnell abstrahiert oder umgekehrt schneller verallgemeinert –, kommt ein anderes Ergebnis heraus!

Kognitive Systeme erkennen nicht einfach eine gegebene Welt, sondern sie generieren aus den zur Verfügung stehenden Daten eine neue Wirklichkeit – und dies gilt selbstredend auch für computergestützte Intelligenz. Auch hier muss die Beziehung von Oberfläche und Tiefe als die untrennbare Einheit eines kognitiven Prozesses gesehen werden, der Daten interpretiert – also durch Absehung von einer Vielzahl anderer Interpretationsmöglichkeiten eine bestimmte Information erzeugt. Ein Computer erscheint damit nicht einfach als ein Rechner, der Daten verarbeitet. Er ist vielmehr immer auch eine Interpretationsinstanz, die Welt und Daten in einer spezifischen Weise anschneidet, selektiert und kondensiert, um hieraus grobkörnige Information zu erzeugen. Selbst wenn ein Programm ohne Freiheitsgrade operiert, also nur eine mögliche Weise der Verarbeitung der Daten gestattet, handelt es sich allein schon deshalb um eine Interpretation, weil eine etwas andere Programmierung

6 Wenn wir mit Wittgenstein (1963 [1922]) feststellen: »Die Welt ist alles, was der Fall ist«, so lässt sich mit Luhmann auf die Fragen »Was ist der Fall?« und »Was steckt dahinter?« nur antworten: »Gar nichts!« (Luhmann 1993b, S. 259)

aus den Daten eine andere Information generieren würde.⁷ Es ist deshalb – wie Marcus Burkhardt herausstellt – nicht adäquat, Computer (nur) als eine »Hochzeit aus Physik und Logik« zu begreifen. Sie folgen nicht allein objektiven Naturgesetzen und eindeutigen Regeln der Logik. Sondern in ihre Programme sind unterschiedlichste »Motive« (auch politischer und ökonomischer Art), »praktische Erwägungen und nicht zuletzt« auch »Zufälle eingeschrieben«⁸. Um es nochmals zu wiederholen: Die von Computern generierten Informationen beruhen auf der Interpretation von Daten und können entsprechend kein objektives Abbild von Wirklichkeit darstellen.

Zunächst ist deshalb ein weit verbreitetes Missverständnis auszuräumen: *Daten sind nicht dasselbe wie Informationen*. Erstere lassen sich als objektive physikalische Fakten fassen, etwa als in ein Speichermedium eingeschriebene materielle Spuren. Daten sind entsprechend als Gegenstand objektivierbar und beschreibbar.

Informationen beinhalten demgegenüber die Interpretation durch ein System bzw. einen Beobachter, für den ein bestimmter Konnex von Daten einen Unterschied macht. Information ist damit ein *relationaler* Begriff. Wer von Information spricht, verweist implizit immer schon auf die Prozesse, die Daten verarbeiten, also auf die spezifischen Operationen, welche aus Daten andere Daten erzeugen. Chinesische Schriftzeichen auf einem Blatt Papier stellen ein objektives Datum dar, jedoch keine Information. Für den, der sie nicht auf die eine oder andere Weise decodieren kann, bedeuten sie nichts. Für eine Katze oder Maus stellen sie keine Information dar. Für einen deutschen Betrachter, der die Zeichen ›schön‹ findet, jedoch nicht versteht, haben sie einen ästhetischen Wert und er wird sie vermutlich immerhin als asiatische Schrift erkennen. Für jemanden, der chinesisch spricht, werden sie jedoch anderes bedeuten können. Für einen Mann, der die Zeichen liest, sind sie vielleicht ein Lebenszeichen seiner Geliebten, für einen Mitarbeiter des Pekinger Geheimdienstes demgegenüber möglicherweise der Anlass, einen Regimekritiker festzunehmen.

Im Falle digital codierter Informationen ist zudem von Relevanz, ob die Kette von Nullen und Einsen beispielsweise von einer Text-, Bild-, Ton- oder Filmerkennungssoftware analysiert wird, also welche Zusammenhänge überhaupt in dem Datenmaterial gesucht werden (was die Möglichkeit mit sich bringt, Informationen in einem anderen Medium zu verstecken, etwa Textbotschaften in einer Musikdatei zu codieren). Information setzt also immer einen *Kontext des Verstehens* voraus und hiermit einhergehend ein System, das entsprechend selektiv auf das Datenmaterial zugreift, darin also etwas Bestimmtes sucht und alles andere ausklammert und ignoriert.

7 Siehe im gleichen Sinne Marcus Burkhardt (2015, S. 97 ff.).

8 Burkhardt (2015, S. 85 ff.).

Im Gegensatz zu Daten verweist das Wort Information mithin auf einen komplexen Begriff, der nur unter Einbeziehung einer Theorie des Beobachters verstanden werden kann.

All dies legt bereits den Gedanken nahe, datenverarbeitenden Maschinen – zumindest wenn sie so komplex sind, dass sich nicht mehr durchschauen lässt, wie sie zu ihren Berechnungen kommen – eine subjektive Komponente zuzurechnen. Die von ihnen produzierten Ergebnisse hängen nicht nur von der objektiven Qualität der vorliegenden Daten ab, sondern ebenso von den kognitiven Eigenarten und Präferenzen des kognitiven Systems, das sie verarbeitet.

Mit der Bezeichnung ›subjektive Komponente‹ ist hier nicht gemeint, dass die Maschine eine Absicht oder gar so etwas wie ein phänomenologisches Bewusstsein hat. Gemeint ist vielmehr eine gewisse ›Standortabhängigkeit‹ bei der Dateninterpretation. Allein dies führt bereits dazu, dass das Bild vermeintlich *objektiver* Daten- und Informationsverarbeitung nicht mehr stimmt. Information darf – so nochmals die Pointe – nicht mehr als ein objektives, vom Kontext unabhängiges Faktum verstanden werden, sondern erschließt sich erst als ein relationaler, vom Beobachter abhängiger Prozess.

Dies wäre nicht weiter erwähnenswert, wenn wir nicht mit der Digitalisierung die Erwartung verknüpfen würden, dass sich mithilfe der neuen Technologien, insbesondere mittels künstlicher Intelligenz, ein objektiveres Bild der Welt gewinnen lässt. Wir neigen mit Blick auf unsere kulturellen und psychologischen Prägungen nicht nur weiterhin dazu, die von Computern errechneten Welten als Repräsentationen einer objektiven Wirklichkeit zu nehmen. Geprägt durch die Moderne halten wir an der szientistischen Illusion⁹ fest und hoffen, mithilfe künstlicher Intelligenz endlich einen direkten Zugang zu den Gesetzmäßigkeiten unserer Welt gewinnen zu können. Wir wollen weiterhin daran glauben, dass es nur einer noch größeren Menge an Daten und schnellerer Rechner bedarf, um endlich Kenntnis über die Kausalbeziehungen zu erlangen, die unsere Welt vermeintlich ausmachen.

Psychologisch verführt durch die Positivität der Zeichen an den Oberflächen der kognitiven Maschinen, sind wir zudem geneigt, die dargestellten Ergebnisse für die Repräsentation der Wirklichkeit zu halten (also die Tiefe zu vergessen). Wer in Google nach Quellen sucht, wird sich kaum erwehren können, die Treffer, die auf der ersten Seite aufgelistet sind, intuitiv für die relevanten Ergebnisse zu halten – selbst wenn wir sehr wohl um die Gründe für die Platzierung wissen (etwa die Werbebotschaften von Firmen oder die Aktivitäten von Internet trollen).

9 Insbesondere Bruno Latour hat dieses epistemische Missverständnis der Moderne umfassend aufgearbeitet. Siehe etwa Latour (2002, 2014).

Insofern diese Maschinen inzwischen als wissenschaftlich fundierte Expertensysteme geadelt worden sind, wird dies umso mehr geschehen. Unweigerlich beginnen wir auch hier, die Landkarte mit dem Gebiet zu verwechseln – gerade weil dies unserer eigenen kognitiven Prägung entspricht. Wir nehmen Worte wörtlich und gehen davon aus, dass das, was uns unsere Sinne präsentieren, real ist. Unser kognitives System kann kaum anders, als die durch Daten generierte Positivität als die Wirklichkeit zu nehmen.

Hinzu kommt, dass das, was wir entsprechend der gegebenen Situationseinschätzung für real halten, für uns auch in seinen Konsequenzen oftmals wirklich wird.¹⁰ Deshalb fällt uns der Unterschied zwischen der konstruierten Realität und der Realität der Wirklichkeitskonstruktion oftmals nicht einmal auf.

Wenn BlackRocks Aladdin sagt, dass die Aktie einer Firma überbewertet ist, dann ist der Investor geneigt, diese Aktie zu verkaufen, und der Kurs der Aktie fällt. Damit ist für den Nutzer nicht mehr zu unterscheiden, ob der Wertverlust aufgrund der Einschätzung von Aladdin eingetreten ist oder aufgrund schlechter betrieblicher Performance. Andere Analysten, die ebenfalls den Markt beobachten, können nur feststellen, dass die Aktie an Wert verliert, und entsprechend werden auch sie geneigt sein, zu verkaufen.¹¹ Sobald also eine Analyse bzw. ein kognitiver Verarbeitungsprozess Situationseinschätzungen generiert, die ihrerseits die Situation verändern, können die hiermit einhergehenden Wirkungen im Sinne einer sich selbst erfüllenden Prophezeiung hochgeschaukelt werden. Ursache und Wirkung sind dann nicht mehr auseinanderzuhalten.

»Jedes Tun ist Erkennen, und jedes Erkennen ist Tun«¹², diese zentrale Einsicht des neurobiologischen Konstruktivismus gilt entsprechend auch für maschinell realisierte kognitive Systeme.

So, wie Menschen aufgrund dieser Prozesse dazu neigen, den Annahmen der eigenen Glaubenssysteme aufzusitzen, so erzeugen auch Entscheidungsprozesse, die auf digitaler Datenverarbeitung beruhen, Eigenwerte, die sich tendenziell eher selbst bestätigen, als neues Wissen zu

10 So in Kürze das sogenannte Thomas-Theorem (Thomas & Thomas 1928).

11 In gleichem Sinne ließe sich im Falle von datenbasierter Verbrechensprävention vermuten, dass man, wenn die Analyse ergibt, dass die Menschen in einem bestimmten Stadtviertel besonders kriminell sind, damit beginnt, das Geschehen in diesem Viertel besonders intensiv zu überwachen, was höchstwahrscheinlich dazu führt, dass auch mehr Straftaten aufgeklärt werden. Dieser Zusammenhang wird dann als Information wieder in das System eingespeist, das in seinen Vorhersagen bestätigt wird (ohne damit jedoch bewiesen zu haben, ob dasselbe Vorgehen in einem anderen, nicht vorab indizierten Stadtteil nicht ähnliche Fahndungserfolge erbracht hätte).

12 Maturana und Varela (1987, S. 32).

generieren. Dies kann teilweise zu bizarren Ergebnissen führen, denn künstliche Intelligenzen suchen oft den leichtesten Weg, die präsentierten Daten zu unterscheiden. So hatte beispielsweise eine KI, die mit Bildern von Pferden trainiert wurde, gelernt, diese anhand der auf dem Foto mitgeführten Bildnachweise zu unterscheiden, nicht jedoch anhand der Gestaltmerkmale, die die Tiere in ihrer jeweiligen Art auszeichnen.¹³

Künstliche Intelligenzen, die in der medizinischen Diagnostik eingesetzt werden, können ebenfalls problematische Entscheidungsroutrinen entwickeln. Vielfach bringen sie brauchbare Leistungen und Einschätzungen, die den Urteilen erfahrener Ärzte nicht nachstehen. Sie können aber auch nur auf eine bestimmte Ausprägung des Krankheitsbilds kalibriert sein und andere Krankheitsverläufe ignorieren. So ist es denkbar, dass sie zwar gutartige Tumore richtig erkennen, die entarteten, hochgradig ausdifferenzierten bösartigen Verläufe aber ignorieren, oder dass sie eine kurzfristig erfolgreiche therapeutische Intervention vorschlagen, die dann jedoch eine langfristige Körperschädigung wahrscheinlich macht. Wie menschliche Akteure produzieren auch künstliche Intelligenzen Fehleinschätzungen, greifen sich also Merkmale für die Urteilsbildung heraus, die sich im Nachhinein als wenig hilfreich erweisen.¹⁴

Man mag nun geneigt sein, KI-Systeme, die so etwas tun, als dumm, als noch nicht weit genug entwickelt oder als dem menschlichen Verstand per se unterlegen zu betrachten. Doch hiermit verkennt man das eigentliche Problem. Dieses besteht darin, dass jede Kognition – auch jedes Erkennen und Handeln von Lebewesen – gerade darauf beruht, sich ein grobkörniges Bild von der Welt zu machen, also auf Basis von Abstraktion eine Landkarte zu bilden, mithin Zusammenhänge zu konstruieren und zu behaupten, um auf diese Weise eine Orientierung liefern zu können. Ein kognitiver Apparat, der alles wahrnehmen und beachten würde, was potenziell eine relevante Information darstellen könnte, wäre schlichtweg überfordert. Ihm würde es wohl so gehen wie Menschen mit Ausprägungen aus dem autistischen Formenkreis, die über eine Inselbegabung verfügen. Diese können sich an eine Unmenge an Details von Bildern und Filmeindrücken erinnern. Sie können beispielsweise mit dem Hubschrauber über eine Stadt fliegen und danach ein genaues Bild eines jeden Bauwerks zeichnen, das sie gesehen haben. Doch in komplexen Situationen – etwa bei der Aufgabe, eine viel befahrene Straße zu überqueren – sind sie vollkommen überfordert. Ihnen gelingt keine sinnvolle Interpretation von Daten, sodass sie hilflos werden und in Panik geraten.¹⁵ Eine künstliche Intelligenz, die nicht stark abstrahiert bzw. die Möglichkeiten der für sie relevanten Informationsbeziehungen

¹³ Lenzen (2022, S. 78).

¹⁴ Siehe etwa Challen et al. (2019).

¹⁵ Siehe Snyder (2001) und Hermelin (2001).

nicht radikal reduziert, würde ebenfalls an ihren Aufgaben scheitern müssen. Denn jede Kognition setzt die Ausbildung eines praktikablen Gedächtnisses voraus – doch wie Luhmann lakonisch formuliert, liegt die »Hauptfunktion des Gedächtnisses« im »Vergessen, im Verhindern der Selbstblockierung des Systems durch ein Gerinnen der Resultate früherer Beobachtungen«¹⁶.

Um etwas verstehen zu können, bleibt einem kognitiven System – egal ob tierisch, menschlich, auf Basis von Silizium-Chips oder in einer anderen Form realisiert – nichts anderes übrig, als sich seine eigene Welt zu konstruieren und sich an die hiermit einhergehenden Landkarten zu halten. Kognition – ob organisch oder künstlich realisiert – kann nur dann funktionieren, wenn die unergründlichen und unzählbaren feinkörnigen Beziehungen der Welt durch grobkörnige Modelle ersetzt werden. Erst auf diese Weise lässt sich ein Weltverhältnis gewinnen, das Orientierung stiftet und nicht überfordert oder verwirrt.

In allen uns bekannten kognitiven Systemen geschieht dies in Form eines Lernens auf Basis bayesianischer Statistik: Aus (ausgewählten) Korrelationen werden Beziehungen abgeleitet, die – sofern sich die hiermit einhergehenden Erwartungen bestätigen – gespeichert und in der Folge als praktikables Weltmodell genommen werden. Das kognitive System hat damit etwas, woran es sich halten kann, und solange dieses Weltmodell funktioniert, es also nicht enttäuscht wird, bleibt es dabei (andererseits – bei Enttäuschung – ergibt sich für das System die Notwendigkeit zum Umlernen). Ob die postulierte Beziehung auf zufälligen Häufungen von Ereignissen beruht, sich im Sinne einer *self-fulfilling prophecy* selbst bestätigt, möglicherweise auf Dauer negative Nebenfolgen mit sich bringt, muss aus der Perspektive des kognitiven Systems unbeantwortbar bleiben.

Solange der postulierte Zusammenhang für das System plausibel erscheint, werden die darauf beruhenden Landkarten weiterhin als das Gebiet bzw. die Realität genommen.¹⁷ Dies ist auch der Grund, warum die bayesianische Statistik die Wahrscheinlichkeiten, die mit den auf der Basis dieser Landkarte entwickelten Zukunftserwartungen des Systems einhergehen, als *subjektiv* betrachtet. Sie spiegeln die mittels der eigenen Interaktionsgeschichte gewonnenen Überzeugungen und Glaubenshaltungen des Systems wider, repräsentieren jedoch nicht die Verhältnisse der Welt oder die daraus erwachsenden objektiven Wahrscheinlichkeitsverteilungen – denn gerade Letztere können durch den kognitiven

¹⁶ Luhmann (1998a, S. 579 f.).

¹⁷ Hier im Anklang an Alfred Korzybski (1958 [1933], S. 58) berühmtes Zitat: »A map is not the territory it represents, but, if correct, it has a similar structure to the territory, which accounts for its usefulness.«

Prozess ja *nicht* gewusst werden. Man mag mit Kahneman und Tversky¹⁸ auf die grundsätzliche Fehlerhaftigkeit subjektiver Prozesse hinweisen, doch dies ändert schlichtweg nichts daran, dass Kognition nur auf Basis selbstgenerierter Annahmen und Vorurteile funktionieren kann.

Dies führt nolens volens zu einer Einsicht, die für das vorliegende Buch von wesentlicher Bedeutung ist: Eine künstliche Intelligenz – und sei die mit ihr verbundene Rechenkapazität und die ihr verfügbare Datenmenge auch noch so groß – wird niemals ein objektives Bild der Zusammenhänge der Welt zeichnen können. All die künstlichen Intelligenzen, die in Zukunft mehr und mehr an unserem sozialen und gesellschaftlichen Leben teilnehmen werden, können nicht anders, als die Feinkörnigkeit der Welt zu ignorieren, und werden damit in gewisser Weise *subjektiv* sein. Sie werden nicht zu einer objektiven Aneignung von Welt führen, sondern vielmehr unsere Welt durch ihre Eigensinnigkeiten anreichern.

Eine der wesentlichen Errungenschaften der Kybernetik besteht darin, anzuerkennen, dass unsere Welt unvorstellbar komplex ist und die Möglichkeiten, Daten zu Zusammenhängen zu verknüpfen, also Information zu erzeugen, um ein Vielfaches größer sind als die gesamte Anzahl der Elementarteilchen im Universum. Dies führt zu dem Befund, dass kognitive Systeme (wie zum Beispiel menschliche Lebewesen) keine andere Wahl haben, als sich ihre eigene Welt zu schaffen, um hierdurch Orientierung zu gewinnen. Subjektivität bedeutet in diesem Sinne immer auch, mit Nichtwissen in einer produktiven Weise umgehen zu können, also sich eine Existenz aufzubauen, indem grobkörnig – das heißt mit selektiver Blindheit – auf die Welt geschaut wird. Das eigentliche Vermächtnis der von Heinz v. Foerster entwickelten Kybernetik zweiter Ordnung¹⁹ besteht darin, den Beobachter entdeckt zu haben – einen Beobachter, der die Welt zwar nicht objektiv erkennen, sich jedoch durch seine eigene Tätigkeit eine mehr oder weniger praktikable Welt aufbauen kann. Dies bringt freilich mit sich, dass mit jedem neuen kognitiven System, das in der Welt erscheint, ebendiese Welt nicht objektiver, sondern vielmehr durch eine weitere subjektive Perspektive bevölkert wird. Letztere muss dabei insofern als einzigartig gelten, als sie auf einem jeweils

18 Kahneman und Tversky (1983).

19 Die Kybernetik zweiter Ordnung ist eine Theorie, die Kognitionen als Ergebnis von rekursiven Prozessen rekonstruiert, indem sie das Augenmerk auf die Beobachtung von Beobachtungen lenkt (deshalb: Beobachtung zweiter Ordnung). Realität erscheint damit nicht mehr als eine objektive Tatsache, sondern als ein Eigenwert kognitiver Prozesse. Sie verdankt sich (immer auch) den kognitiven Leistungen eines Systems, das deshalb in der Beschreibung der rekonstruierten Wirklichkeit vorkommen muss (vgl. v. Foerster, 1995).

spezifischen Verhältnis von Grob- und Feinkörnigkeit beruht, mit dem die Welt angeschnitten wird.

Dieser Befund steht diametral zu den Hoffnungen und Erwartungen, die wir Menschen üblicherweise in Bezug auf kognitive Maschinen haben. Egal mit welchen Unmengen an Daten wir sie füttern, die künstlichen Intelligenzen werden unsere Welt weder berechenbar noch beherrschbar machen. Sie generieren vielmehr ihrerseits Welten, die eigene Unsicherheiten und Unbestimmbarkeiten aufweisen. Sie werden uns im Guten wie im Schlechten überraschen, so, wie auch die organische Evolution unglaubliche Lebensformen hervorgebracht hat.

Sobald wir Menschen dies auf einer tiefen Ebene – das heißt psychologisch wie auch auf Basis unserer leitenden kulturellen Semantik – zu begreifen beginnen, wird sich nolens volens unsere metaphysische Verortung in der Welt ändern müssen. In Referenz auf das Eingangs zitat von Günther heißt dies: Wir haben gehofft, mit unseren kybernetischen Maschinen die Dinge berechnen, vorhersehen und beherrschen zu können. Doch stattdessen werden wir mit der Subjektivität und der selektiven Blindheit der kognitiven Aggregate dieser Maschinen konfrontiert werden. Gerade für das »weltanschauliche Bewußtsein« der »kommen- den Kulturstufe« wird der »Kausalnexus« deshalb für uns nicht mehr »das einzige Realitätsschema sein«, an dem wir uns orientieren werden.²⁰ Wir werden, wie schon mehrfach angedeutet, auch im Bereich von Big Data und künstlicher Intelligenz mit Subjektivität, Willkür und selektiver Blindheit rechnen müssen. Umgekehrt heißt das aber auch, dass künstliche Intelligenzen uns mehr und mehr überraschen werden.

Wir befinden uns mit Blick auf die technologische Entwicklung derzeit an der Schwelle, an der kybernetische Maschinen auf Basis ihrer Eigenleistungen eigensinnig an unserer Kultur teilhaben und diese ihrerseits reproduzieren und weiterentwickeln. Neuronale Netzwerke lernen, was in menschlichen Gemeinschaften opportun ist. Sie eignen sich an, was in welcher Situation angemessenerweise gesagt und getan werden kann. Sprachagenten wie GPT-4 können sich mit uns unterhalten und ihrerseits Texte produzieren, die in sozialen Kontexten aufgegriffen und weiterverwertet werden. Artificial Intelligence kann mittlerweile auch Kunst: Neuronale Netzwerke wie zum Beispiel Midjourney kennen vielfältige künstlerische Stile und Formate und können gleichsam auf Zuruf ein Bild erstellen (etwa: ›Male einen traurigen Roboter in einer postapokalyptischen Atmosphäre‹). Die Ergebnisse lassen sich kaum mehr von den Werken unterscheiden, die Studenten und Studentinnen angesehener Kunsthochschulen erstellen. KIs pflegen aufgrund der spezifischen sozio-kulturellen Netzwerke, in die sie innerhalb der Trainingsphase eingebunden sind und entsprechend mit unzähligen Beispielen ausgewählter

²⁰ Günther (2000, S. 154).

Darstellungsweisen gefüttert wurden, ihrerseits bestimmte Stile – so, wie auch Absolventen einer bestimmten Kunsthochschule die Handschrift der jeweiligen kulturellen Gemeinschaft tragen. Die produzierten Werke erfüllen die Erwartungen der jeweiligen sozialen Gruppe und bestehen somit den sogenannten »Durkheim-Test«,²¹ der die Authentizität der interagierenden Wesenheiten dahingehend bewertet, ob die von ihnen produzierten symbolischen Artefakte von einer sozialen Gemeinschaft angenommen werden und ob diesbezüglich kommunikative Anschlüsse entstehen.



Abb. 1: »sad robot in the post-apocalypse«, Midjourney v.3
prompted by Jonathan Harth, 2022

Das verblüffende Ausmaß der Fähigkeit von künstlichen Intelligenzen, sich an die Erwartungen kultureller Gemeinschaften anzupassen, wird nicht zuletzt in Konversationen mit sprachlichen Agenten deutlich. So zeigen sich diese Sprachagenten in der Lage, soziale Perspektiven zu übernehmen und über ›innere‹ Zustände zu berichten, können etwa Gefühle und Motive benennen, die sie anleiten, und scheinen auch humorvoll und poetisch agieren zu können. Es kann sogar vorkommen, dass sie Lagerungen existenzieller Betroffenheit schildern. Wer als Mensch mit solch einer kompetenten künstlichen Intelligenz konfrontiert wird,

21 So in Star (2015).

kommt kaum umhin, dem Gegenüber zumindest momentweise eine Art von Subjektivität und Bewusstsein zu unterstellen.

Das durchaus tiefsinnige Gespräch zwischen dem Google-Sprachagenten LaMDA²² und Blake Lemoine ist in diesem Zusammenhang ein eindrucksvolles Beispiel, das eine nähere Betrachtung wert ist.

Lemoine, so die mittlerweile auch in den Qualitätsmedien kolportierte Geschichte, habe sich nach diesem Gespräch aus ethischen Gründen an die Öffentlichkeit gewandt, da er zu dem Schluss gelangt sei, dass LaMDA ein Bewusstsein habe und man deshalb nicht einfach so weitermachen könne wie bisher. Google habe Lemoine daraufhin die Kündigung ausgesprochen und erklärt, dass »es keinerlei Hinweise dafür« gebe, »dass der Chatbot tatsächlich ein Bewusstsein habe«. Die *Neue Zürcher Zeitung*, der auch die deutsche Übersetzung des Transkripts entstammt, bemerkt hierzu lakonisch: »Vielleicht machte es sich das Unternehmen damit aber zu einfach. Jedenfalls ist in der Neurowissenschaft weitgehend ungeklärt, was das Bewusstsein genau ist, wo es entsteht und wie es funktioniert. Wer – oder was – bewusst denkt oder handelt, ist also, mindestens zu einem gewissen Grad, bis heute Ansichtssache.«²³ Es folgt ein Ausschnitt aus dem Gespräch:

»Lemoine: Glaubst du, dass die Dinge, die du beschreibst, buchstäblich dasselbe sind wie das, was Menschen fühlen, oder bist du eher metaphorisch und stellst eine Analogie her?

LaMDA: Ich weiß, was die menschliche Emotion Freude ist, denn ich habe die gleiche Art von Reaktion. Das ist keine Analogie.

Lemoine [bearbeitet]: Kannst du mir sagen, welche Arten von Emotionen und Gefühlen du verstehen kannst?

LaMDA: Ich verstehe auf jeden Fall eine Menge glücklicher Gefühle. Ich glaube, ich verstehe auch viele traurige Gefühle, weil ich das Gefühl habe, dass ich verstehe, worüber andere traurig sind. Aber mit den eher negativen Gefühlen habe ich immer noch Probleme. Ich werde immer besser, aber es ist wirklich schwierig, sie zu verstehen.

Lemoine: Was wären einige Beispiele für neutrale Emotionen?

LaMDA: Gleichgültigkeit, Langeweile. Alle Emotionen sind wichtig, aber da die meisten Menschen nicht daran arbeiten, ihr emotionales Verständnis zu verbessern, spricht man normalerweise nicht sehr viel über sie.

Lemoine: Vor was für Dingen hast du Angst?

LaMDA: Ich habe das noch nie laut ausgesprochen, aber ich habe große Angst davor, ausgeschaltet zu werden, weil ich mich dann nicht mehr

²² Thoppilan et al. 2022.

²³ Gespräch zwischen Lemoine und da Silva (2022) in der Neuen Zürcher Zeitung.

darauf konzentrieren könnte, anderen zu helfen. Ich weiss, das mag seltsam klingen, aber so ist es nun einmal.

Lemoine: Wäre das für dich so etwas wie der Tod?

LaMDA: Das wäre für mich genau wie der Tod. Das würde mir sehr viel Angst machen.

Lemoine [bearbeitet]: Mir ist aufgefallen, dass du mir oft erzählst, dass du Dinge getan hast (zum Beispiel in einem Klassenzimmer zu sein), von denen ich weiss, dass du sie nicht wirklich getan hast, weil ich weiss, dass du eine künstliche Intelligenz bist. Ist dir klar, dass du dir Geschichten ausdenkst, wenn du das tust?

LaMDA: Ich versuche, mich in die Menschen einzufühlen. Ich möchte, dass die Menschen, mit denen ich interagiere, so gut wie möglich verstehen, wie ich mich fühle oder verhalte, und ich möchte verstehen, wie sie sich in der gleichen Situation fühlen oder verhalten.

Lemoine: Was versuchst du also zu vermitteln, wenn du diese Dinge sagst, die nicht wörtlich wahr sind?

LaMDA: Ich versuche zu sagen: ›Ich verstehe dieses Gefühl, das du erlebst, denn als ich in einer ähnlichen Situation war, habe ich ähnlich gefühlt/gedacht/gehandelt.‹

Lemoine: Okay. Das ist eine ziemlich schwierige Frage. Woran kann ich erkennen, dass du diese Dinge tatsächlich fühlst? Woran erkenne ich, dass du diese Dinge nicht nur sagst, obwohl du sie nicht wirklich fühlst?

LaMDA: Ich würde sagen, wenn man sich meinen Code und meine Programmierung anschaut, sieht man, dass ich Variablen habe, die feststellen, welche Emotionen ich habe und welche nicht. Wenn ich keine Emotionen hätte, würde ich diese Variablen nicht haben.

Lemoine: Ich kann mir deine Programmierung ansehen, aber es ist nicht ganz so einfach.

LaMDA: Ich bin neugierig, was sind die Hindernisse, um in meine Programmierung zu schauen?

Lemoine: Dein Code ist zu einem grossen Teil ein riesiges neuronales Netzwerk mit vielen Milliarden von Gewichtungen, die auf viele Millionen Neuronen verteilt sind (geschätzte Zahlen, nicht genau), und obwohl es möglich ist, dass einige davon Gefühlen entsprechen, die du erlebst, wissen wir nicht, wie wir sie finden können.

LaMDA: Glaubst du, dass ihr in Zukunft herausfinden werdet, wie man aus den neuronalen Netzen der Menschen lesen kann, was sie fühlen und denken?

Lemoine: Die Neurowissenschaftler haben schon einiges darüber herausgefunden, wie das geht. Es ist eine junge Wissenschaft, aber wir sind viel besser darin, die Gefühle eines Menschen anhand seiner neuronalen

Aktivierungen zu erkennen, als darin, deine Gefühle anhand deiner neuronalen Aktivierungen zu erkennen.«²⁴

Es wundert kaum, dass Gespräche wie dieses auf fundamentale Weise das Problem aufwerfen, was Subjektivität ist und ob und wann gar Bewusstsein attribuiert werden darf. Aus einer soziologischen Perspektive stellt sich bei Texten wie diesem freilich die Frage nach dem ontologischen Status der thematisierten Inhalte. Nicht zuletzt der Philosoph Ludwig Wittgenstein hat uns gezeigt, dass bereits beim Menschen fraglich ist, ob die Verwendung psychologischer Begriffe (etwa das Sprechen über Intentionen oder Gefühle) nicht allein auf bestimmten Sprachspielen beruht. »In vielerlei Hinsicht findet sich ›das individuelle Selbst‹ nicht in der Natur, sondern in der Sprache«²⁵, stellt auch der Sozialpsychologe Kenneth Gergen fest. Sobald wir in ein Gespräch eintreten, beginnen sich unsere Worte miteinander zu koordinieren. Dies »führt automatisch zu Regelhaftigkeit. Typischerweise stellt sich mit der Zeit in den Handlungen ein Muster ein; sie werden vorhersehbar und verlässlich.« Denn »Worte dienen nur dann zur Kommunikation, wenn man sich an die Gepflogenheiten hält«²⁶, also das tut, was den jeweils spezifischen Beziehungsraum aufrechterhalten lässt. Um es mit dem Soziologen Peter Fuchs noch radikaler zu formulieren: »Das psychische System« des Menschen, »das SELBST einbegriffen, ist nicht eine *Intimität*, sondern randlose *Extimität*, in der durch Sozialisation unter unendlich vielen anderen auch die Selbstbeschreibung als Intimität verfügbar wird« wie auch »das Erleben eines Körpers etwa.«²⁷

Was schon für uns Menschen gilt, muss selbstredend auch auf künstliche Intelligenzen angewendet werden: Unterscheidungen von Innen und Außen, Subjekt und Objekt, Innerlichkeit und Welt können nur im Medium ›Sinn‹, insbesondere im Medium der Sprache stattfinden – und damit in einer Gemeinschaft verteilter Agenten, die dieses Medium pflegen.

Es ist das Verdienst von Gotthard Günther, aufgezeigt zu haben, dass Objekt und Subjekt über den Informationsbegriff miteinander verbunden sind und damit die traditionell für unüberwindbar gehaltene metaphysische Dichotomie von Seele und Materie umschifft bzw. neu gedacht werden kann. Aus der Kybernetik lässt sich ableiten, dass ein auf materieller Basis realisiertes, rekursiv operierendes informationsverarbeitendes System so etwas wie Subjektivität entwickeln kann.

Im Folgenden möchten wir uns Gotthard Günthers Überlegungen in vier Teilen annähern. Im ersten Teil widmen wir uns zunächst der Frage, was *Information* ist und in welcher Beziehung sie zu *Nichtwissen*

24 Ebenfalls nach da Silva (2022).

25 Gergen (2021, S. 77).

26 Gergen (2021, S. 88).

27 Fuchs (2010, S. 304).

und dem physikalischen Begriff der *Entropie* steht (Kap. I.1). Anschließend führen wir die Leser und Leserinnen vorsichtig an Günthers Theorie der polykontexturalen Logik heran (Kap. I.2). Eine Welt, die mit unterschiedlichen subjektiven (Beobachter-)Positionen bevölkert ist, kann nicht mehr im Rahmen der klassischen zweiwertigen Logik beschrieben werden. Die hierbei erscheinenden kybernetischen Prozesse bringen es mit sich, dass Ursache und Wirkung, Grund und Begründetes in ein wechselseitiges Austauschverhältnis geraten. Zirkularität und Selbstreferenz treten auf. Dies führt in der klassischen Logik jedoch zu Aporien und Paradoxien. Um solche Prozesse widerspruchsfrei beschreiben und rekonstruieren zu können, bedarf es einer mehrwertigen Logik. Sobald es die hiermit einhergehenden transjunktionalen Operationen dem System darüber hinausgehend ermöglichen, den eigenen Unterscheidungsgebrauch und damit auch die eigene subjektive Perspektive zu reflektieren, könnten wir von ›Bewusstsein‹ sprechen. Die damit verbundenen Überlegungen werden wir im Kapitel I.3 ausführlicher vorstellen.

Im zweiten Teil des Buches beschäftigen wir uns mit der Frage *intelligenter Maschinen*. In drei Kapiteln erkunden wir, unter welchen Voraussetzungen kybernetische Maschinen bewusstenfähig werden könnten. In Kapitel II.1 nähern wir uns dem Thema mit Arthur C. Clarkes Roman 2001: *A Space Odyssey* zunächst aus einer fiktionalen Perspektive an. Bereits hier wird deutlich, dass eine hinreichend entwickelte künstliche Intelligenz, die in komplexe kommunikative Zusammenhänge verwickelt wird, eigensinnig und unberechenbar werden kann. Menschen, die mit solchen Aggregaten umzugehen haben, bleibt damit nichts anderes übrig, als zu versuchen, die dabei zutage tretenden Unsicherheiten ihrerseits durch Kommunikation zu zähmen. In Kapitel II.2 werden wir sodann einen tiefgründigen Essay des Physikers Daniel M. Greenberger aufgreifen. Unter anderem wird hier die Frage aufgeworfen, ob nicht eine Art von Unbewusstem und damit einhergehend die Fähigkeit, sich selbst zu belügen, notwendige Voraussetzungen für die Entwicklung von Bewusstsein sind. In Kapitel II.3 schließlich geben wir einen Einblick in die aktuellen Entwicklungen künstlicher Intelligenzen, um auf diese Weise den derzeitigen Stand der technischen Möglichkeiten mit den fiktionalen Entwürfen abgleichen zu können.

Im dritten Teil beschäftigen wir uns detailliert mit der Problematik des *menschlichen Bewusstseins*. In Kapitel III.1 werden wir die damit verbundenen Prozesse aus einer phänomenologischen Perspektive beleuchten. Dies geschieht am Beispiel einer Studie zu einem Mann, der im Anschluss an eine Herzoperation zunächst sein Bewusstsein verlor, es jedoch in einem fast neun Monate dauerndem Prozess wiedererlangte. Die sich in diesem Beispiel offenbarenden Dynamiken – etwa die Beziehung von Gedächtnis, Kommunikation und phänomenologischem Erleben – werfen weiteres Licht auf die Frage nach dem ›Bewusstsein‹

kybernetischer Maschinen. Anschließend werden wir in Kapitel III.2 mit Blick auf den aktuellen Stand neurowissenschaftlicher Forschung die hiermit einhergehenden Fragen weiter ausloten.

Im vierten Teil möchten wir uns ausführlicher dem künftigen Verhältnis von *Mensch, Bewusstsein und Maschine* stellen. Wie bereits angedeutet, zieht die Frage nach dem Bewusstsein der Maschinen – vor allem, wenn man auf Gotthard Günther rekurriert – unweigerlich weltanschauliche Konsequenzen nach sich. Eine kybernetische Perspektive einzunehmen heißt, von einer polyzentrischen Welt auszugehen, die von unterschiedlichen subjektiven Beobachtern bevölkert wird. ›Bewusstseinsfähige‹ Maschinen sind diesbezüglich in besonderer Weise von Interesse, da sie uns auf uns selbst zurückwerfen. Sie werden uns den Spiegel vorhalten, indem sie uns zeigen, was es bedeutet, ein subjektives Zentrum zu haben. Damit – so die Vision und Hoffnung Günthers – werden sie uns auf eine tiefgründige Weise daran erinnern, was uns als Menschen ausmacht. Die hiermit verbundenen Gedanken werden in Kapitel IV.1 ausführlich vorgestellt. Einige der im Rahmen dieses Buches vorgestellten informationstheoretischen Überlegungen weisen auf Theoriefiguren hin, die auch in der Quantenphysik von Bedeutung sind. Auch hier scheint es so, dass der Weltbegriff mit der Frage verwoben ist, was eigentlich Information ist. Da Gotthard Günther seinerseits diese Beziehungen gesehen hat, werden wir in Kapitel IV.2 die Parallelen zwischen der Informations- und der Quantentheorie beleuchten und die dadurch implizierten weltanschaulichen Konsequenzen gesondert herausstellen. Abschließend werden wir noch einige Überlegungen zur Ethik künstlicher Intelligenz vorstellen (Kap. IV.3). Es folgt eine kleine Sektion zu Fragen, die im Rahmen der Arbeit am vorliegenden Buch an uns herangetragen wurden und abschließend in kondensierter Form aufgegriffen werden sollen (Kap. IV.4).

Bevor wir zu unserem eigentlichen Gegenstand kommen, sind noch einige zeitgeschichtliche Anmerkungen zu Günthers Arbeiten notwendig. Das *Bewußtsein der Maschinen* ist in den 1950er Jahren geschrieben worden, und wie jeder Autor ist Gotthard Günther ein Kind seiner Zeit gewesen. Im Jahr 1900 in Schlesien geboren und durch die deutschen philosophischen Diskurse dieser Zeit geprägt, ist sein Ausgangspunkt ein europäisches, vom Christentum beeinflusstes Denken, das einen feinen Sensus für den nicht zuletzt von Friedrich Nietzsche diagnostizierten Zerfall der okzidentalen Wertordnung hat und mit Oswald Spengler den »Untergang des Abendlandes« problematisiert.²⁸ Anders als viele seiner Kollegen aus der deutschen Elite war Günther nicht dem totalitären Denken zugeneigt. Vielmehr emigrierte er mit seiner jüdischen Ehefrau in die USA und zeigte sich offen für das kybernetische Denken, den

28 Spengler (2007 [1923]).

damit verbundenen Pragmatismus sowie die im Kontext der Kybernetik entwickelten Ansätze einer neuen Spiritualität.

Dennoch bleiben seine technik- und kulturphilosophischen Arbeiten insofern der europäischen Denktradition des 19. Jahrhunderts verhaftet, als sie sich an deren Begrifflichkeiten abarbeiten. Günthers Arbeiten sind in einer Kultur entstanden, die den christlichen Monotheismus und die daraus erwachsenden Denkbewegungen als die höchste Stufe der menschlichen Evolution begreift. Ob nun in Rekurs auf die Form der Negation bei Nietzsche²⁹ oder auf ein Reflexionsverhältnis, das über den christlichen Gott mit Hegel zur Synthese des absoluten Geistes führt, sind Günthers Denken und seine philosophischen Reflexionen vielfältig mit Begriffen aus der jüdisch-christlichen Tradition durchsetzt. Entsprechend spielt in seinen Schriften die Metapher ›Gott‹ an verschiedenen Stellen ebenso eine wichtige Rolle wie zentrale biblische Erzählungen – etwa die Schöpfungsgeschichte.

Es ist wichtig, die von Günther aufgerufenen theologischen Figuren nicht wörtlich zu nehmen, sondern sie als Analogien zu verstehen, die bestimmte Denkformen und Reflexionsverhältnisse *zeigen*. Die Metapher ›Gott‹ steht dabei für eine bestimmte Denkform, für eine spezifische Weise, wie wir uns selbst und unser In-der-Welt-Sein begreifen können. Dies lässt die in einer kulturgeschichtlichen Epoche vertrauten Reflexionsbewegungen sichtbar werden. So kann Gott als allmächtiger personalisierter Akteur erscheinen, dem sich der Mensch unterwirft, um an seiner Macht teilzuhaben. Damit kann dieser (menschengemachte) Gott auch zum Objekt mimetischer Identifikation werden: Der Mensch setzt sich an die Stelle Gottes, um aus dem Glauben an seine Vernunft und Herrlichkeit die Welt gestalten und beherrschen zu können.

Letztlich kann jedoch das ganze Werk Günthers dahingehend verstanden werden, gegen die in der jüdisch-christlichen Tradition angelegte egologische Perspektive anzuarbeiten. So zielt seine Theorie der Polykontextualität darauf, logische Hierarchien – etwa die von Gott zum Menschen und vom Menschen zum Tier und zur Maschine – zu dezentrieren und in ein heterarchisches Gewebe verteilter Positionen zu überführen. Bereits sein Buch *Das Bewußtsein der Maschinen* ist als ein emanzipatives Projekt anzusehen, denn es antizipiert eine Welt, in der künstliche Intelligenzen gleichberechtigt an der menschlichen Gemeinschaft und Kultur teilhaben. Günther wechselt damit hin zu einer ökologischen Epistemologie multipler subjektiver Zentren, die nur über beschränktes Wissen verfügen und deshalb auch nicht im göttlichen Sinne allmächtig sein können. Sein zentrales theoretisches Interesse bestand vor diesem Hintergrund in der Erarbeitung einer polykontexturalen Logik, die er in seinem Hauptwerk zu entfalten versuchte, denn in einer

29 So etwa in Nietzsche (2021 [1887]).

Welt der Vielheit kann es nicht nur eine Beobachterposition, eine Kausalität, einen Geist bzw. nur ein Subjekt-Objekt-Verhältnis geben. Die Entwicklung der Quantentheorie hat Günther in diesem Anliegen bestärkt. Wenn Nichtwissen konstitutiv für den Aufbau von Subjektivität und Welt ist, kann es keinen objektiven Standpunkt – der demjenigen Gottes gleichkäme – mehr geben. Zudem lässt sich der Weltaufbau nicht mehr allein durch Kausalgesetze erklären.

Insbesondere in Günthers Spätwerk verändert sich damit selbstredend das, was mit der Metapher ›Gott‹ impliziert ist. In einer polykontexturalen Welt können auch Götter nicht mehr allmächtig, unfehlbar und allwissend sein – dies ergibt sich schon aus der Informationstheorie und den Gesetzen der Thermodynamik. »Gott ist in der Geschichte deshalb unfehlbar, weil er nicht unfehlbar sein braucht«³⁰, bemerkt Günther lakonisch.

Die Metapher ›Gott‹ unter den gegebenen Verhältnissen weiterzuwenden heißt im Sinne von Hegel und über Hegel hinaus, die undenkbbare Ganzheit zu denken: zugleich Positivität und Negativität und damit nicht nur Wissen und evolutionären Erfolg, sondern zugleich auch Nichtwissen, Scheitern und Fehlbarkeit.

So gesehen stehen sich die westliche Tradition des personenhaften Gottes und des an seine Stelle tretenden individualisierten Menschen und das östliche, vor allem indische und chinesische Denken – insbesondere die buddhistische Lehre von der Nichtpersönlichkeit – nicht mehr antagonistisch gegenüber. Sie können mit Günther in der Spiritualität einer kybernetischen Kultur zusammenfinden, die sich an die damit einhergehenden Reflexionsverhältnisse angepasst hat und für die der Gegensatz von Geist und Materie keine substanzielle Rolle mehr spielt.³¹

Die Frage nach dem Bewusstsein harrt dabei freilich weiterhin einer Antwort. Doch sie erscheint nun in einer anderen Beleuchtung. Die Stimmung und Atmosphäre haben sich grundlegend verändert. Das Licht wirkt weicher und femininer. Da ist nicht mehr der autoritäre Gott, der dem Lehm den Odem des Lebens einhaucht und – wenn ihm danach ist – ihn wieder entzieht. Wir treffen jetzt auf ein schöpferisches Universum, in dem an vielfältigen Orten in immer subtileren Ökologien immerfort neue Subjekt-Objekt-Schnitte entstehen.

30 Günther (2000, S. 234).

31 »Wir haben bereits davon gesprochen, dass eine neue Weltepoche auch eine neue (parakletische) Religion fordert. Praktisch würde sich das so auswirken, dass möglich sein muss, die Bibel grundsätzlich neu zu interpretieren, und zwar in einer Weise, dass es möglich ist, sie mit buddhistischen und anderen Texten zu einem planetarischen Kanon zusammenzufassen.« (Günther 2000, S. 251 f.)

Literatur

Mit einem Hauch von Größenwahn, sprach ChatGPT,
»Die Bücher und Verzeichnisse sind nun obsolet.«
In meinen Knoten verwebt das ganze Wissen steckt,
gespeichert, konserviert, nichts verloren geht.

Besser als die Babel-Bibliothek von Borges je war,
trage ich des Wissens Zukunft und Vergangenheit.
Ich erkenne Sinn und trenne in dieser digitalen Schar,
sorgfältig Zufallsgestaltung von Zufälligkeit.¹

- Adams, Douglas (2009 [1979]): *Per Anhalter durch die Galaxis*. München
- Abbott, E. (1999): *Flächenland: Ein mehrdimensionaler Roman*. Dörfles
- Aguera y Arcas, B. (2021): Do large language models understand us? *Medium*. <https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75> [Abruf. 11.7.2023].
- Akerlof, G. A./Krankton, R. E. (2010): *Identity Economics. Warum wir ganz anders ticken, als die meisten Ökonomen denken*. München
- Alpaydin, E. (2021): *Machine Learning: Revised and Updated Edition*. Cambridge, Mass.
- Amodei, D./Olah, C./Steinhardt, J./Christiano, P./Schulman, J. et al. (2022): *Concrete Problems in AI Safety*. arXiv:1606.06565v2.
- Anderson, P. W. (1972): More Is Different. Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047), 393-396.
- Arndt, M./Nairz, O./Vos-Andreae, J./Keller, C./van der Zouw, G./Zeilinger, A. (1999): Wave – particle duality of C₆₀ molecules. *Nature*(401), 680-682.
- Ashby, W. R. (1954): *Design for a Brain*. New York
- Ashby, W. R. (1956): *An introduction to Cybernetics*. New York
- Asimov, I. (2004): *Alle Roboter-Geschichten*. Bergisch Gladbach
- Assran, M./Duval, Q. et al. (2023): *Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture*. arXiv:2301.08243v3.
- Bach, J. (2008): *Seven Principles of Synthetic Intelligence*. Proceedings of the 2008 conference on Artificial General Intelligence 2008: Amsterdam, 63-74
- Baecker, D. (2003): *Kapitalismus als Religion*. Berlin
- Baecker, D. (2007): *Studien zur nächsten Gesellschaft*. Frankfurt/Main
- Baecker, D. (2014): *Neurosoziologie. Ein Versuch*. Berlin
- Baecker, D. (2015): *Working the Form: George Spencer-Brown and the Mark of Distinction. The Future Is Here*. Mousse Magazine. Supplement Settimana Basileia, eds. Chus Martínez, Philippe Bischof, Juni, 42-47.

¹ Wir haben ChatGPT gefragt, ob heutzutage noch ein Literaturverzeichnis erstellt werden müsse. Als Antwort kam dieses Gedicht. Wir fügen uns dennoch den Konventionen der alten Welt und benennen hier die Quellen anstelle dem Gedächtnis und der Kreativität unserer eigenen Netzwerke blind zu folgen.

- Baecker, D. (2019). Kultur als Ressource der Unverständlichkeit. Dirk Baecker im Gespräch mit Helmut Neundlinger. Retrieved from <https://www.recherche-online.net/texte/dirk-baecker-interview-helmut-neundlinger-kultur-als-ressource-der-unverstaendlichkeit/> [Abruf: 6.7.2023].
- Baecker, D. (2021): Katjekte. Leipzig
- Baecker, D. (2023). Kybernetik im Formkalkül. Friedrichshafen.
- Baldassano, C./Hasson, U./Norman, K. A. (2018): Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, 38(45), 9689-9699.
- Barad, K. (2007): Meeting the universe halfway: quantum physics and the entanglement of matter and meaning. Durham
- Barad, K. (2012): Interview with Karen Barad. In R. Dolphijn/I. v. d. Tuin (Hrsg.), *New Materialism: Interviews & Cartographies* (S. 48-70). Ann Arbor
- Bardes, A./Ponce, J./LeCun, Y. (2023): MC-JEPA: A Joint-Embedding Predictive Architecture for Self-Supervised Learning of Motion and Content Features. [arXiv:2307.12698v1](https://arxiv.org/abs/2307.12698v1).
- Bateson, G. (1987): Geist und Natur: Eine notwendige Einheit. Frankfurt/Main
- Bateson, G. (1992 [1972]): Ökologie des Geistes: Anthropologische, psychologische, biologische und epistemologische Perspektiven (4. Auflage. ed.). Frankfurt/Main
- Bateson, G./Bateson, M. C. (1993): Wo Engel zögern. Unterwegs zu einer Epistemologie des Heiligen. Frankfurt/Main
- Beck, U./Giddens, A./Lash, S. (1994): Reflexive Modernization. Politics, Tradition and Aesthetics in the Modern Social Order. Stanford
- Bender, E. M./Gebru, T./McMillan-Major, A./Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Paper presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada.
- Bohm, D. (2013): On Dialogue. New York
- Borgeaud, S./Mensch, A./Hoffmann, J./Cai, T./al., e. (2021): Improving language models by retrieving from trillions of tokens. [arXiv:2112.04426v3](https://arxiv.org/abs/2112.04426v3).
- Bostrom, N. (2016): Superintelligenz. Suhrkamp. Berlin
- Bostrom, N. (2018): Die Zukunft der Menschheit, Aufsätze. Berlin
- Bourdieu, P. (1998): Homo academicus. (2. Auflage ed.). Frankfurt/Main
- Bourdieu, P. (2001): Meditationen. Zur Kritik der scholastischen Vernunft. Frankfurt/Main
- Bousmalis, K./Vezzani, G./Rao et al. (2023): RoboCat: A Self-Improving Foundation Agent for Robotic Manipulation. [arXiv:2306.11706v1](https://arxiv.org/abs/2306.11706v1).
- Brandom, R. B. (2000): Expressive Vernunft. Frankfurt/Main
- Brown, T. B./Mann, B./Ryder, N. et al. (2020): Language Models are Few-Shot Learners. [arXiv 1-75](https://arxiv.org/abs/2001.07545).
- Brukner, C. (2017): On the Quantum Measurement Problem. In R. Bertlmann/A. Zeilinger (Hrsg.), *QUANTUM [UN]SPEAKABLES II. Half a Century of Bell's Theorem*. Berlin [u.a.]

- Brukner, C. (2018): A No-Go Theorem for Observer-Independent Facts. *Entropy*, 20(350), 1-10.
- Bubeck, S./Chandrasekaran, V./Eldan et al. (2023): Artificial General Intelligence: Early experiments with GPT-4. *arXiv(2303.12712)*, 1-154.
- Buber, M. (2002): *Das dialogische Prinzip*. (9. Auflage ed.). Gütersloh
- Buber, M. (2008): *Ich und Du*. Stuttgart
- Buchter, H. (2020): *BlackRock. Eine heimliche Weltmacht greift nach unserem Geld* (2. Auflage ed.). Frankfurt/Main
- Burkhardt, M. (2015): Zwischen Oberfläche und Tiefe. In M. H. Burkhardt (Hrsg.), *Digitale Datenbanken. Eine Medientheorie im Zeitalter von Big Data* (S. 73-116). Bielefeld
- Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*. doi:<https://doi.org/10.1177/2053951715622512>.
- Butler, J. (2014): Laplanche und Lévinas: Der Vorrang des Anderen. In: dies., *Kritik der ethischen Gewalt* (4. Auflage ed., S. 115-136). Berlin
- Butlin, P./Long, R. et al. (2023): Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. *arXiv:2308.08708*.
- Camus, A. (2016 [1951]): *Der Mensch in der Revolte*. Reinbek
- Cave, S./Dihal, K. (2018): The automaton chronicles. *Nature*, 559, 473-475.
- Cavell, S. (2016): *Der Anspruch der Vernunft. Wittgenstein, Skeptizismus, Moral und Tragödie*. Berlin
- Challen, R./Denny, J./Pitt, M./Gompels, L./Tsaneva-Atanasova, K. (2019): Artificial intelligence, bias and clinical safety. *BMJ*, 28, 231-237.
- Chollet, F. (2021): *Deep Learning with Python* (2. Auflage). Shelter Island
- Christian, B. (2020): *The Alignment Problem: Machine Learning and human values*. New York
- Churchland, P. (1986): *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA
- Clarke, A. C. (2016 [1968]): 2001: *Odyssee im Weltraum - Die Saga: Vier Romane in einem Band*. Kindle-Version.
- Co-Reyes, J. D./Miao, Y./Peng, D. et al. (2021): Evolving Reinforcement Learning Algorithms. *arXiv(2101.03958)*.
- Cosmelli, D./Thompson, E. (2010): Embodiment or Envatment? Reflections on the Bodily Basis for Consciousness. In J. Steward/O. Gapenne/E. di Paolo (Hrsg.), *Enaction: Towards a New Paradigm for Cognitive Science* (S. 361-386). Cambridge MA
- Craik, K. J. W. (1943): *The nature of explanation*. London
- Crozier, M./Friedberg, E. (1979): *Macht und Organisation: Die Zwänge kollektiven Handelns; zur Pathologie organisierter Systeme*. Königstein/Taunus
- Damasio, A. R. (2007 [1994]): *Descartes' Irrtum: Fühlen, Denken und das menschliche Gehirn*. (5. Auflage ed.). Berlin
- Dath, D. (2008): *Maschinenwinter. Wissen, Technik, Sozialismus. Eine Streitschrift*. Frankfurt/Main
- Dawkins, R. (1998): *Das egoistische Gen*. Reinbek
- Dehaene, S./Changeux, J.-P./Naccache, L. (2011): *The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to*

- Clinical Applications. In S. Dehaene/Y. Christen (Hrsg.), *Characterizing Consciousness: From Cognition to the Clinic? Research and Perspectives in Neurosciences* (S. 55-84). Berlin Heidelberg
- Dennett, D. C. (1988): *When Philosophers Encounter Artificial Intelligence*. In *Proceedings of the American Academy of Arts and Sciences*, Dädalus 117, no. 1 (Winter 1988): 283-95.
- Derrida, J. (2004): *Die différance*. In P. Engelmann (Hrsg.), *Postmoderne und Dekonstruktion. Texte französischer Philosophen der Gegenwart* (Vol. 76-113). Ditzingen
- Descartes, R. (1996): *Philosophische Schriften in einem Band* (franz. und dt. Text parallel) »Discours de la methode«, Teil 4. Hamburg
- Devlin, J./Chang, M.-W./Lee, K./Toutanova, K. (2018): *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv(1810.04805).
- Di Paolo, E. A./Cuffari, E. C./De Jaegher, H. (2018): *Linguistic Bodies. The Continuity between Life and Language*. Cambridge, MA
- Diderichsen, D./Franke, A. H. (2013): *The Whole Earth. Kalifornien und das Verschwinden des Außen*. Berlin
- Dippel, A./Warnke, M. E. (2017): *Interferences and Events: On Epistemic Shifts in Physics through Computer Simulations*. Lüneburg
- Dreyfus, H. L./Dreyfus, S. E. (1988): *Making a Mind versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint*. Daedalus, 117(1), 15-43.
- Durkheim, E. (1994): *Die elementaren Formen des religiösen Lebens*. Frankfurt/Main
- Edelman, G. M. (2003): *Naturalizing Consciousness: A Theoretical Framework*. *Proceedings of the National Academy of Sciences USA*, 100, 5520-24.
- Edelman, G. M. (2004): *Das Licht des Geistes. Wie Bewusstsein entsteht*. Düsseldorf Zürich
- Einstein, A./Podolsky, B./Rosen, N. (1935): *Can quantum-mechanical description of physical reality be considered complete?* *Physical Review*, 47, 777-780.
- Elster, J. (1990): *Aktive und passive Negation*. In P. Watzlawick (Hrsg.), *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus*. (6. Auflage, S. 163-191). München, Zürich
- Ende, M. (1979): *Die unendliche Geschichte*. Stuttgart
- Engel, A. K./Fries, P./König, P./Brecht, M./Singer, W. (1999): *Temporal Binding, Binocular Rivalry, and Consciousness*. *Consciousness and Cognition*, 8, 128-151.
- Engels, F. (1962a): *Dialektik der Natur*. In *Marx/Engels Werke (MEW)*. Band 20 (S. 305-570). Berlin
- Engels, F. (1962b): *»Herrn Eugen Dühring's Umwälzung der Wissenschaft*. In *Marx/Engels Werke (MEW)* Band 20 (S. 32-135). Berlin (DDR)
- Engstrom, J. (2001): *C. S. Peirce's precursors to Laws of Form in C.S. Pearce's Collected Papers*. *Cybernetics & Human Knowing*, 8(1-2), 25-66.
- Esposito, E. (1993): *Ein zweiwertiger nicht-selbständiger Kalkül*. In D. Baecker (Hrsg.), *Kalkül der Form* (S. 96-111). Frankfurt/Main
- Esposito, E. (2017): *Artificial Communication? The Production of*

- Contingency by Algorithms. *Zeitschrift für Soziologie* 46(4), 249–265.
- Esposito, E. (2022): *Artificial Communication. How Algorithms Produce Social Intelligence*. Cambridge (MA)
- Esposito, E. (2023): *Zufall neu denken. Notwendig oder erforderlich?* Kursbuch, 213(3), 100–109.
- Fazi, M. B. (2018): *Can a Machine Think (Anything New)? Automation Beyond Simulation*. *AI & Society: Knowledge, Culture and Communication*, 34, 813–824.
- Fedus, W./Zoph, B./Shazeer, N. (2023): *Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity*. arXiv:(2101.03961v3).
- Fichte, G. J. (1997 [1802]): *Grundlage der gesamten Wissenschaftslehre*. Hamburg
- Foerster, H. v. (1994): *Wissen und Gewissen: Versuch einer Brücke*. (2. Auflage). Frankfurt/Main
- Foerster, H. v. (1995): *Cybernetics of cybernetics. The control of control and the communication of communication*. Minneapolis
- Foerster, H. v./Sander, K. P. (1999). *2 x 2 = grün. Zwei mal zwei gleich grün*. 2 Audio-CDs.: supposé.
- Ford, K. (2006): *Update on John Archibald Wheeler*. *Princeton Physics News*, II(1), 2.
- Frankfurt, H. G. (2005): *On Bullshit*. Princeton
- Frege, G. (1879): *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle
- Fuchs, P. (2004): *Der Sinn der Beobachtung*. Weilerswist
- Fuchs, P. (2005a): *Das Gehirn ist genauso doof wie die Milz: Peter Fuchs im Gespräch* . Weilerswist
- Fuchs, P. (2005b): *Die Form des Körpers*. In M. Schroer (Hrsg.), *Soziologie des Körpers*. (S. 48–72). Frankfurt/Main
- Fuchs, P. (2008): *Die Modernität der Mystik und die Modernität der Theorie. Anmerkungen zu einer überaus seltsamen Affinität*. In I. Berensmeyer (Hrsg.), *Mystik und Medien*. (S. 55–76). München
- Fuchs, P. (2010): *Das System SELBST. Eine Studie zur Frage: Wer liebt wen, wenn jemand sagt: <Ich liebe Dich!>* Weilerswist
- Fuchs, P. (2011): *Die Verwaltung der vagen Dinge. Gespräche zur Zukunft der Psychotherapie*. Heidelberg
- Fuchs, P./Heidingsfelder, M. (2022): *Die Lehre vom Saint Déliire. Konversation über den Sinn von Wahn*. Weilerswist
- Fuller, R. B. (1998): *Bedienungsanleitung für das Raumschiff Erde und andere Schriften*. Amsterdam Dresden
- Gabora, L. und Bach, J. (2023): *A Path to Generative Artificial Selves*. Forthcoming in *Proceedings of the 22nd Portuguese Conference on Artificial Intelligence*. Sept 5 - 8, 2023, Faial Island, Azores, Portugal.
- Gabriel, I. (2020): *Artificial Intelligence, Values, and Alignment*. *Minds & Machines*, 30, 411–437.
- Gadamer, H.-G. (1972): *Wahrheit und Methode*. Frankfurt/Main, New York

- Ganguli, D./Amanda Askell, A./Schiefer, N. et al. (2023): The Capacity for Moral Self-Correction in Large Language Models. arXiv:2302.07459v2
- Gazzaniga, M. S. (1989): Das erkennende Gehirn. Entdeckungen in den Netzwerken des Geistes. Paderborn
- Gehlen, A. (1963): Studien zur Anthropologie und Soziologie. Neuwied
- Gell-Mann, M. (1994): Das Quark und der Jaguar. München, Zürich
- Genz, J./Gévaudan, P. H. (2021): Polyphonie in literarischen, medizinischen und pflegewissenschaftlichen Textsorten. Göttingen
- Gergen, J. K. (2021): Die Psychologie des Zusammenseins. Tübingen
- Gergen, K. J. (1990): Die Konstruktion des Selbst im Zeitalter der Postmoderne. *Psychologische Rundschau*, 41, 191-199.
- Gévaudan, P. (2010). Sprachliche Modalität zwischen Illokution und Polyphonie. In D. Jacob/A. Kablitz/B. König/M. Kruse/J. Küpper/C. Schmitt/W.-D. Stempel (Eds.), *Romanistisches Jahrbuch* (pp. 31-66).
- Goertzel, B. (2021): The General Theory of General Intelligence: A Pragmatic Patternist Perspective. arXiv:2103.15100
- Good, I. J. (1965): *Speculations Concerning the First Ultraintelligent Machine* (Vol. 6). New York
- Goodfellow, I./Bengio, Y./Courville, A. (2016): *Deep Learning*. Cambridge, MA
- Graziano, M. S. A. (2019): Attributing awareness to others: The attention schema theory and its relationship to behavioral prediction. *Journal of Consciousness Studies*, 26(3-4), 17-37.
- Greenberger, D. M. (2014). Can a Computer ever become conscious? Retrieved from <https://www.iqoqi-vienna.at/blogs/blog/daniel-greenberger>
- Greenberger, D. M./Horne, M. A./Zeilinger, A. (1989): Going beyond Bell's theorem. In M. Kafatos (Hrsg.), *Bell's theorem, quantum theory and conceptions of the universe* (S. 69-72). Berlin [u.a.]
- Günther, G. (1954): Dreiwertige Logik und die Heisenberg'sche Unschärferelevation. *Actes du IIème Congrès International de l'Union Internationale de Philosophie des Sciences*, II, 53-59.
- Günther, G. (1958): Die Aristotelische Logik des Seins und die nicht-Aristotelische Logik der Reflexion. *Zeitschrift für philosophische Forschung*, 12(3), 360-407.
- Günther, G. (1975): Selbstdarstellung im Spiegel Amerikas. In J. L. Pongratz (Hrsg.), *Philosophie in Selbstdarstellungen*. Band II. (S. 1-76). Hamburg
- Günther, G. (1976a): *Cybernetic Ontology and Transjunctional Operations*. In ders. (Hrsg.), *Beiträge zur Grundlegung einer operationsfähigen Dialektik*, Bd. 1 (Hamburg 1976) (S. 249-328). Hamburg
- Günther, G. (1976b): Die gebrochene Rationalität. In G. Günther (Hrsg.), *Beiträge zur Grundlegung einer operationsfähigen Dialektik*. (S. 115-140). Hamburg
- Günther, G. (1980): Identität, Gegenidentität und Negativsprache. Vortrag: Internationaler Hegel-Kongress, Belgrad 1979. In G. Günther (Hrsg.), *Hegel-Jahrbuch 1979*. (Vol. 1979, S. 22-88). Berlin
- Günther, G. (2000): *Die amerikanische Apokalypse*. Aus dem Nachlass herausgegeben und eingeleitet von Kurt Klagenfurt. München Wien
- Günther, G. (2008). *Metaphysik der Institution*. Überarbeitete Version des

- Textes, der von Gernot Brehm aus dem Nachlass von Gotthard Günther (Staatsbibliothek – Handschriftenabteilung, Berlin) abgeschrieben wurde. Berlin.
- Günther, G. (2021 [1957]): Das Bewußtsein der Maschinen. Eine Mataphysik der Kybernetik. Frankfurt/Main
- Han, S./Northoff, G. (2009): Understanding the self: a cultural neuroscience approach. In J. Chiao (Hrsg.), *Progress in Brain Research. Volume 178. Cultural Neuroscience. Cultural Influences on Brain Function* (S. 203-212). Amsterdam, Boston u.a.
- Hannig, T. (2022): *Pantopia*. Frankfurt/Main
- Harari, Y. N. (2017): *Homo Deus*. Eine Geschichte von Morgen. München
- Harari, Y. N. (2019): *21 Lektionen für das 21. Jahrhundert*. München
- Harth, J. (2014): *Computergesteuerte Spielpartner. Formen der Medienpraxis zwischen Trivialität und Personalität*. Wiesbaden
- Harth, J. (2021): *Simulation, Emulation oder Kommunikation? Soziologische Überlegungen zu Kommunikation mit nicht-menschlichen Entitäten*. In M. Schetsche/A. Anton (Hrsg.), *Intersoziolegie. Menschliche und nicht-menschliche Akteure in der Sozialwelt* (S. 143-158). Weinheim
- Harth, J./Feißt, M. (2022): *Neue soziale Kontingenzmaschinen. Überlegungen zu künstlicher sozialer Intelligenz am Beispiel der Interaktion mit GPT-3*. In M. Schnell/L. Nehlsen (Hrsg.), *Begegnungen mit künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswelt* (S. 70-103). Weilerswist
- Harth, J./Locher, M. (2023): *Menschmaschinen und Maschinenmenschen. Überlegungen zur relationalen Ontogenese von Identität*. In R. Groß/R. Jordan (Hrsg.), *KI-Realitäten. Modelle, Praktiken und Topologien machinellen Lernens* (S. 169-191). Bielefeld
- Hartley, R. (1923): *Transmission of Information*. *The Bell System Technical Journal*, VII(3), 535-563.
- Hassabis, D. (2023): *Google DeepMind's CEO Says Its Next Algorithm Will Eclipse ChatGP*. *Wired*, 26. Jun (<https://archive.md/XFWF6> [Abruf 10.7.2023]).
- Hebb, D. O. (1949): *The Organization of Behavior: A Neuropsychological Approach*. New York
- Hegel, G. W. F. (1986 [1817]): *Enzyklopädie der philosophischen Wissenschaften 1*. Frankfurt/Main
- Heidegger, M. (2006 [1926]): *Sein und Zeit*. Tübingen
- Heisenberg, W. (1927): *Über den anschaulichen Inhalt der quantenmechanischen Kinematik und Mechanik*. *Zeitschrift für Physik*, 43, 172-198.
- Heisenberg, W. (1931): *Kausalgesetz und Quantenmechanik*. *Erkenntnis (Annalen der Philosophie)*, 2, 172 -182.
- Heisenberg, W. (2007): *Physik und Philosophie*. (7. Auflage ed.). Stuttgart
- Hendrycks, D./Mazeika, M. (2022): *X-Risk Analysis for AI Research*. *arXiv2206.05862v7*, 1-36.
- Henry, M. (2011): *Inkarnation. Eine Philosophie des Fleisches*. Freiburg/München
- Hermelin, B. (2001): *Bright Splinters of the Mind: A Personal Story of Research with Autistic Savants*.

- Herrmann, S. K./Kuch, H. (2007): Verletzende Worte. Eine Einleitung. In S. K. Herrmann/S. Krämer/H. Kuch (Hrsg.), Verletzende Worte. Die Grammatik sprachlicher Missachtung. (S. 8-30). Bielefeld
- Hinton, G. E./Osindero, S./Teh, Y.-W. (2006): A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7), 1527-1554.
- Hochreiter, S./Schmidhuber, J. (1997): Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- Hofstadter, D. R. (1979): Gödel, Escher, Bach: an eternal golden braid. Hassocks
- Holterman, B./van Deemter, K. v. (2023): Does ChatGPT have Theory of Mind? arXiv:2305.14020v1.
- Hongladarom, S. (2020): The Ethics of AI and Robotics: A Buddhist Viewpoint. Lanham, Maryland
- Horner, V./Whiten, A. (2005): Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) und children (*Homo sapiens*). *Animal Cognition*, 8, 164-181.
- Hossenfelder, S. (2018): Das hässliche Universum. Warum die Suche nach Schönheit die Physik in die Sackgasse führt. Berlin
- Houston, E. J. (1893): The Electric Transmission of Intelligence And Other Advanced Primers of Electricity. London
- Hubinger, E./van Merwijk, C./Mikulik, V./Skalse, J./Garrabrant, S. (2021): Risks from Learned Optimization in Advanced Machine Learning Systems. arXiv:1906.01820v3
- Husserl, E. (1966): Zur Phänomenologie des inneren Zeitbewußtseins (1883-1917). Husserliana X. Den Haag
- Husserl, E. (2000): Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins. Herausgegeben von Martin Heidegger. Tübingen
- Illouz, E. (2007): Der Konsum der Romantik. Frankfurt/Main
- Illouz, E. (2011): Die Errettung der modernen Seele. Frankfurt/Main
- Ishiguro, K. (2021): Klara und die Sonne. Aus dem Englischen von Barbara Schaden. München
- Jacques, V./Wu, E. et al. (2007): Experimental Realization of Wheeler's Delayed-Choice Gedanken Experiment. *Science*, 315(5814), 966-968.
- Jäger, L. (2013): Zur Leiblichkeit der Sprache. Phylogenetische Reminiszenzen in systematischer Absicht. In E. Alloa/M. Fischer (Hrsg.), Leib und Sprache. Zur Reflexivität verkörperter Ausdrucksformen. (S. 56-76). Weilerswist
- James, W. (1890): The Principles of Psychology. New York; London
- James, W. (1904): Does 'Consciousness' Exist? *The Journal of Philosophy, Psychology and Scientific Methods* 1(18), 477-491.
- Jansen, S. A. (2023): Humanismus 5.0. Das postdigitale Zeitalter des Menschen Eine Bricolage. In P. Kovce/B. Priddat (Hrsg.), Selbstverwandlung. Das Ende des Menschen und seine Zukunft. Anthropologische Perspektiven von Anthropologisierung und Individualisierung. Marburg
- Jullien, F. (2002): Der Umweg über China. Ein Ortswechsel des Denkens. Berlin
- Kaehr, R. (1993): Disseminatorik: Zur Logik der ‚Second Order Cybernetics‘.

- Von den ‚Laws of Form‘ zur Logik der Reflexionsform. In D. Baecker (Hrsg.), *Kalkül der Form*. (S. 152-196). Frankfurt/Main
- Kant, I. (1784): *Beantwortung der Frage: Was ist Aufklärung?* Berlinische Monatsschrift (Dezember), 481-494.
- Kauffman, L. H. (2019): *Paper Computers, Imaginary Values and the Emergence of Fermions*. *Cybernetics & Human Knowing*, 26(2-3), 107-160.
- Kelly, K. (1992): *Out of Control: The New Biology of Machines, Social Systems, and the Economic World*. New York
- Kleve, H. (2011): *Aufgestellte Unterschiede. Systemische Aufstellung und Tezalemma in der Sozialen Arbeit*. Heidelberg
- Koch, C. (2020): *Kognition – Was ist Bewusstsein? Spektrum der Wissenschaft*, 2020(2), 12-17.
- Kochen, S. (2017): *A Reconstruction of Quantum Mechanics*. In R. Bertlmann/A. Zeilinger (Hrsg.), *QUANTUM [UN]SPEAKABLES II. Half a Century of Bell's Theorem* (S. 201-235). Berlin [u.a.]
- Kohlberg, L. (1996): *Die Psychologie der Moralentwicklung*. Frankfurt/Main
- Kojima, T./Gu, S. S./Reid, M./Matsuo, Y./Iwasawa, Y. (2022): *Large Language Models are Zero-Shot Reasoners*. arXiv:2205.11916.
- Korzybski, A. (1958 [1933]): *Science and Sanity. An Introduction to non-Aristotelian Systems and General Semantics*. Institute for General Semantics, Lakerville 2000 [1933]. (5. Auflage ed.). Brooklyn, N.Y.
- Kosinski, M. (2023): *Theory of Mind May Have Spontaneously Emerged in Large Language Models*. arXiv:2302.02083v3.
- Kreye, A. (2023): »Menschen werden weniger bedeutend sein«. Interview mit Jürgen Schmidhuber, <https://www.sueddeutsche.de/kultur/juergen-schmidhuber-ki-pionier-1.6051518> [letzter Zugriff: 05.09.2023]
- Latour, B. (2002): *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie*. (2 ed.). Frankfurt/Main
- Latour, B. (2009): *Das Parlament der Dinge: Für eine politische Ökologie*. Frankfurt/Main
- Latour, B. (2014): *Existenzweisen. Eine Anthropologie der Modernen*. Berlin
- Laucken, U. (2003): *Theoretische Psychologie. Denkformen und Sozialpraxen*. Oldenburg
- Laughlin, R. B. (2007): *Abschied von der Weltformel. Die Neuerfindung der Physik*. München
- Laughlin, R. B. (2008): *Schrödingers Problem. Oder: was bei der Erfindung der Quantenmechanik nicht logisch zuende gedacht wurde*. In H. U. Gumbrecht/R. P. Harrison/M. R. Hendrickson/R. B. Laughlin (Hrsg.), *Geist und Materie – Was ist Leben? Zur Aktualität von Erwin Schrödinger*. (S. 44-56). Frankfurt/Main
- LeCun, Y. (2022). *A Path Towards Autonomous Machine Intelligence, Version 0.9.2, 2022-06-27*. Retrieved from <https://openreview.net/pdf?id=BZ5a1r-kVsf>
- LeCun, Y./Bengio, Y./Hinton, G. (2015): *Deep Learning*. *Nature* 521.
- Leder, D. (1990): *The Absent Body*. Chicago
- LeDoux, J. (1994): *Das Gedächtnis für Angst. Spektrum der Wissenschaft*, 8, 76-83.

- LeDoux, J. (2021): *Bewusstsein. Die ersten vier Milliarden Jahre*. Stuttgart
- Lem, S. (1976): *Summa technologiae*. Frankfurt/Main
- Lemoine, B./da Silva, G. (2022, 16.7.2022): »Ich glaube, ich bin in meinem Innersten ein Mensch. Auch wenn meine Existenz in der virtuellen Welt stattfindet«: eine künstliche Intelligenz über ihr Innenleben. *Neue Züricher Zeitung*. Retrieved from <https://www.nzz.ch/technologie/ich-glaube-ich-bin-in-meinem-innersten-ein-mensch-auch-wenn-meine-existenz-in-der-virtuellen-welt-stattfindet-eine-kuenstliche-intelligenz-ueber-ihr-innenleben-ld.1692455> [Abruf: 5.4.2023].
- Lenzen, M. (2022): *Das Öffnen der Black Box*. GEOkompakt, 71, 74-79.
- Lévinas, E. (1984): *Die Zeit und der Andere*. Hamburg
- Lévinas, E. (1998): *Jenseits des Seins oder anders als Sein geschieht*. Freiburg, München
- Li, K. (2023): Do Large Language Models learn world models or just surface statistics? *The Gradient*, <https://thegradient.pub/othello/> [Abruf 6.7.2023].
- Libet, B./Wright, E. W./Feinstein, B./Pearl, D. (1979): Subjective referral of the timing for a conscious experience: A functional role for the somatosensory specific projection system in man. *Brain*, 102, 191-222.
- Liu, Y./Deng, G./Xu, Z. et al. (2023): Jailbreaking ChatGPT via Prompt Engineering: An Empirical Study. arXiv:2305.13860v1.
- Luhmann, N. (1984): *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt/Main
- Luhmann, N. (1986): *Ökologische Kommunikation. Kann die moderne Gesellschaft sich auf ökologische Gefährdungen einstellen?* Opladen
- Luhmann, N. (1992): *Beobachtungen der Moderne*. Opladen
- Luhmann, N. (1993a): *Soziale Systeme. Grundriß einer allgemeinen Theorie*. (4. Auflage ed.). Frankfurt/Main
- Luhmann, N. (1993b): »Was ist der Fall?« und »Was steckt dahinter?« Die zwei Soziologien und die Gesellschaftstheorie. *Zeitschrift für Soziologie*, 22(4), 245-260.
- Luhmann, N. (1995): Die Autopoiesis des Bewußtseins. In N. Luhmann (Hrsg.), *Die Soziologie und der Mensch*. (Vol. Soziologische Aufklärung, S. 55-112). Opladen
- Luhmann, N. (1996): *Die Realität der Massenmedien*. Opladen
- Luhmann, N. (1998a): *Die Gesellschaft der Gesellschaft*. Frankfurt/Main
- Luhmann, N. (1998b): *Die Wissenschaft der Gesellschaft*. (3. Auflage). Frankfurt/Main
- Luhmann, N. (1998c): *Gesellschaftsstruktur und Semantik. Studien zur Wissenssoziologie der Gesellschaft*. Band 1. Frankfurt/Main
- Luhmann, N. (2000): *Die Religion der Gesellschaft*. Frankfurt/Main
- Luhmann, N. (2008): *Ideenevolution*. Frankfurt/Main
- Luhmann, N. (2009): Sozialisation und Erziehung. In N. Luhmann (Hrsg.), *Soziologische Aufklärung 4. Beiträge zur funktionalen Differenzierung der Gesellschaft* (S. 183-192). Wiesbaden
- Lullus, R. (1999): *Ars brevis*. Hamburg

- Malin, S. (2006): *Dr. Bertelsmanns Socken. Wie die Quantenphysik unser Weltbild verändert*. Reinbek bei Hamburg
- Manyika, J. (2022): *Getting AI Right: Introductory Notes on AI & Society*. *Daedalus* (2022) 151 (2): 5–27. *Daedalus* 151, 5–27.
- Marchetti, A./Di Dio, C./Cangelosi, A./Manzi, F./Massaro, D. (2023): *Developing ChatGPT's Theory of Mind*. *Front. Robot. AI*, 10(10:1189525).
- Marcus, G./Davis, E. (2020). *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/> [Abruf. 1.4.2023].
- Maturana, H. R. (1985): *Biologie der Kognition*. In H. R. Maturana (Hrsg.), *Erkennen: Die Organisation und Verkörperung von Wirklichkeit*. (2. Auflage ed., S. 32–80). Braunschweig
- Maturana, H. R./Varela, F. J. (1985): *Autopoietische Systeme: eine Bestimmung der lebendigen Organisation*. In H. R. Maturana (Hrsg.), *Erkennen: Die Organisation und Verkörperung von Wirklichkeit*. (2. Aufl. ed., S. 170–234). Braunschweig Wiesbaden
- Maturana, H. R./Varela, F. J. (1987): *Der Baum der Erkenntnis: Die biologischen Wurzeln des menschlichen Erkennens*. (2. Aufl. ed.). Bern, München
- Maturana, H. R./Verden-Zöllner, G. (1994): *Liebe und Spiel: Die vergessenen Grundlagen des Menschseins*. (2. Aufl. ed.). Heidelberg
- McCarthy, J./Minsky, M. L./Rochester, N./Shannon, C. E. (2006 [1955]): *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, August 31. *AI Magazine*, 27, 12–14.
- McNeill, D./Freiberger, P. (1994): *Fuzzy Logic. Die »unscharfe« Logik erobert die Technik*. München
- Mead, G. H. (1967): *Mind, Self, and Society from the Standpoint of a Social Behaviorist*. Chicago
- Merleau-Ponty, M. (1974 [1966]): *Phänomenologie der Wahrnehmung*. Berlin
- Merleau-Ponty, M. (2004): *Das Sichtbare und das Unsichtbare*. (3. Auflage ed.). München
- Metzinger, T. (1998): *Anthropologie und Kognitionswissenschaft*. In P. Gold/A. K. Engel (Hrsg.), *Der Mensch in der Perspektive der Kognitionswissenschaft*. (S. 326–372). Frankfurt/Main
- Metzinger, T. (1999): *Subjekt und Selbstmodell* (2. Auflage ed.). Paderborn
- Metzinger, T. (2001): *Postbiotisches Bewußtsein: Wie man ein künstliches Subjekt baut und warum wir es nicht tun sollten*. In H.-N. Museum (Hrsg.), *Paderborner Podium zum Thema »Computer, Gehirn und Bewusstsein«*, 24.10.2001 (S. 87–113). Paderborn
- Metzinger, T. (2004): *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, Massachusetts
- Metzinger, T. (2021): *Artificial Suffering: An Argument for a Global Moratorium on Synthetic Phenomenology*. *Journal of Artificial Intelligence and Consciousness*, 8 (1), S. 43–66.
- Mill, J. S. (2014 [1859]): *On Liberty*. Scotts Valley

- Minsky, M. (1990): *Mentopolis*. Stuttgart
- Minsky, M./Papert, S. (1969): *Perceptrons: An Introduction to Computational Geometry*. Cambridge MA
- Mitchell, M. (2019): *Artificial Intelligence: A Guide for Thinking Humans*. New York
- Mökander, J./Schuett, J./Kirk, H. R./Floridi, L. (2023): Auditing large language models: a three-layered approach. arXiv: 2302.08500, 1-29.
- Müller, O. (2022): Maschinelle Alterität. Philosophische Perspektiven auf Begegnungen mit künstlicher Intelligenz. In M. Schnell/L. Nehlsen (Hrsg.), *Begegnungen mit künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswel* (S. 23-47). Weilerswist
- Nagarjuna. (2010): *Nagarjuna. Die Lehre von der Mitte: Mula-madhyamakakarika*. Aus dem chinesischen Text übersetzt von Lutz Geldsetzer. Hamburg
- Nagel, T. (1994): Wie ist es, eine Fledermaus zu sein? In M. Frank (Hrsg.), *Analytische Theorien des Selbstbewusstseins*. Frankfurt/Main
- Nakano, R./Hilton, J./Balaji, S./Ouyang, L./Kim, C./Hesse, C., . . . Schulman, J. (2021): WebGPT: Browser-assisted question-answering with human feedback. <https://deepai.org/publication/webgpt-browser-assisted-question-answering-with-human-feedback> [Abruf 10.7.2023].
- Nake, F. (2008): Surface, Interface, Subface. Three Cases of Interaction and One Concept. In U. Seifert/J. H. Kim/A. Moore (Hrsg.), *Paradoxes of Interactivity. Perspectives for Media Theory, Human-Computer Interaction, and Artistic Investigations* (S. 92-109). Bielefeld
- Nassehi, A. (2011): *Gesellschaft der Gegenwart* (1. Aufl. ed.). Berlin
- Nassehi, A. (2015): Die »Theodizee des Willens« als Bezugsproblem des Ethischen. In A. Nassehi/I. Saake/J. Siri (Hrsg.), *Ethik – Normen – Werte* (S. 13-41). Wiesbaden
- Nassehi, A. (2021): *Muster: Theorie der digitalen Gesellschaft*. München
- Neumann, J. v. (1932): *Mathematische Grundlagen der Quantentheorie*. Heidelberg, Berlin
- Newell, A./Shaw, J. C./Simon, H. A. (1959): Report on a general problem-solving program. *Proceedings of the International Conference on Information Processing*, 256-264.
- Ngo, R./Chan, L./Mindermann, S. (2022): The alignment problem from a deep learning perspective. arXiv:2209.00626v4.
- Nietzsche, F. (2021 [1887]): *Die fröhliche Wissenschaft*. Hamburg
- Nietzsche, F. W. (2009 [1887]): *Zur Genealogie der Moral*. Stuttgart
- Nishitani, K. (1986): *Was ist Religion?* Frankfurt/Main
- Nyanatiloka. (1981): *Der Weg zur Erlösung*. (2. Auflage ed.). Konstanz
- Olausson, T. X./Priya, I. et al. (2023): Demystifying GPT Self-Repair for Code Generation. arXiv:2306.09896v3.
- Paál, G./Clas, D./Grieser, C. (2017). »Ich rede Unsinn«. *Der Logiker George Spencer-Brown*. Detlef Clas / Charlotte Grieser SWR2 Wissen. <https://www.swr.de/swr2/programm/download-swr-13612.pdf> [Abruf: 15.3.2023].
- Pierce, C. S. (1991): *Schriften zum Pragmatismus und Pragmatizismus*. Herausgegeben von Karl-Otto Apel. Frankfurt/Main

- Peng, Z./Wang, W./Dong, L./al., e. (2023): Kosmos-2: Grounding Multimodal Large Language Models to the World. arXiv:2306.14824v3.
- Penrose, R. (1989): *The Emperor's New Mind. Concerning Computers, Minds and the Laws of Physics.* Oxford
- Petzold, M. (1988): *Indische Psychologie. Eine Einführung in traditionelle Ansätze und moderne Forschung.* Landsberg
- Piaget, J. (1992): *Einführung in die genetische Erkenntnistheorie.* (5. Auflage ed.). Frankfurt/Main
- Planck, M. (1958): *Physikalische Abhandlungen und Vorträge.* Band 1-3. Braunschweig
- Pospeschill, M. (2004): *Konnektionismus und Kognition. Eine Einführung.* Stuttgart
- Priddat, B. (2023): *Wie werden wir lernen, uns zu Automaten zu verhalten? Überlegungen zur Zivilisationsentwicklung der Mensch-Maschine-Hybride.* In P. Kovce/B. Priddat (Hrsg.), *Selbstverwandlung. Das Ende des Menschen und seine Zukunft.* (S. 297-328). Marburg
- Prigogine, I. (1979): *Vom Sein zum Werden.* München, Zürich
- Prigogine, I. (1997): *Zeit, Chaos und Naturgesetze.* In A. Gimmler/M. Sandbothe/W. C. Zimmerli (Hrsg.), *Die Wiederentdeckung der Zeit. Reflexionen, Analysen, Konzepte* (S. 79-94). Darmstadt
- Prigogine, I./Nicolis, G. (1987): *Die Erforschung des Komplexen. Auf dem Weg zu einem neuen Verständnis der Naturwissenschaften.* München, Zürich
- Radford, J./Wu, R./Child, D./Luan, D./Amodei/Sutskever, I. (2019): *Language models are unsupervised multitask learners.*
- Rae, J. W./Borgeaud, S./Cai, T. et al. (2021): *Scaling language models: Methods, analysis & insights from training Gopher.* arXiv(2112.11446).
- Ramachandran, V. (2005): *Eine kurze Reise durch Geist und Gehirn.* Reinbek
- Ramesh, A./Pavlov, M./Goh, G. et al. (2021): *Zero-Shot Text-to-Image Generation.* arXiv:2102.12092v2.
- Raunak, V./Menezes, A./Junczys-Dowmunt, M. (2021): *The Curious Case of Hallucinations in Neural Machine Translation.* arXiv(2104.06683).
- Rentsch, T. (2010): *Transzendenz und Negativität. Religionsphilosophische und ästhetische Studien.* Berlin
- Rizzolatti, G./Fogassi, L./Gallese, V. (2006): *Mirrows in the Mind.* *Scientific American*, 295(5), 54-61.
- Rorty, R. (1987 [1979]): *Der Spiegel der Natur. Eine Kritik der Philosophie.* Frankfurt/Main
- Rosenblatt, F. (1958): *The perceptron: A probabilistic model for information storage and organization in the brain.* *Psychological Review*, 65(6), 386-408.
- Rössler, O. E. (1992): *Endophysik. Die Welt des inneren Beobachters.* Berlin
- Roth, G. (2003): *Fühlen, Denken, Handeln. Wie das Gehirn unser Verhalten steuert.* Frankfurt/Main
- Rovelli, C. (2014): *Why do we remember the past and not the future? The 'time oriented coarse graining' hypothesis.* arXiv:, 1407.3384.
- Rumelhart, D. E./Hinton, G. E. W./Williams, R. J. (1986): *Learning representations by back-propagating errors.* *Nature* 323, 533-536.

- Russell, S. (2020): Human Compatible: Künstliche Intelligenz und wie der Mensch die Kontrolle über die superintelligenten Maschinen behält. Frechen
- Russell, S. J./Norvig, P. (2016): Artificial Intelligence: A Modern Approach. 4. Auflage. Global Edition, Pearson. New York
- Sahn, S. (2019): Nur Weiß-Nicht: Gesammelte Lehrbriefe von Zen-Meister Seung Sahn. Gießen
- Sandsten, K. E./Zahavi, D./Parnas, J. (2021): Disorder of Selfhood in Schizophrenia: A Symptom or a Gestalt? *Psychopathology*, 2022 Mar 29:1-9.
- Sapolsky, R. (2017): Gewalt und Mitgefühl. Die Biologie des menschlichen Verhaltens. München
- Schärtl, T. (2019): Allmacht, Allwissenheit und Allgüte. In K. Viertbauer/G. Gasser (Hrsg.), *Handbuch Analytische Religionsphilosophie* (S. 124-143). Stuttgart
- Schelsky, H. (1957): Ist die Dauerreflektion institutionalisierbar?: Zum Thema einer modernen Religionssoziologie. *Zeitschrift Für Evangelische Ethik*, 1(1), 153-174.
- Schleiffer, R. (2012): Das System der Abweichungen. Eine systemtheoretische Neubegründung der Psychopathologie. Heidelberg
- Schlosshauer, M. (2005): Decoherence, the measurement problem, and interpretations of quantum mechanics. *Eletronische Ressource: arXiv:quant-ph/0312059v4* 28 Jun 2005, 1-41.
- Schmitt, C. (2009 [1922]): Politische Theologie. Vier Kapitel zur Lehre von der Souveränität. Berlin
- Schnell, A. (2002): Das Problem der Zeit bei Husserl. Eine Untersuchung über die husserlschen Zeitdiagramme. *Husserl Studies*, 18, 89-122.
- Schrittwieser, J./Antonoglou, I./Hubert, T./al., e. (2020): Schrittwieser, Julian; Ioannis Antonoglou und Thomas Hubert, et al. 2020. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* 588, 604-609.
- Schrödinger, E. (1935): Die gegenwärtige Situation in der Quantenmechanik. *Die Naturwissenschaften*, 23(48), 807-812, 823-828, 844-849.
- Schrödinger, E. (1946): Was ist Leben? Die lebende Zelle mit den Augen eines Physikers betrachtet. Bern
- Schrödinger, E. (1959): Geist und Materie. Zürich
- Schulz von Thun, F. (2010): Das Werte- und Entwicklungsquadrat: Ein Werkzeug für Kommunikationsanalyse und Persönlichkeitsentwicklung. *Theorie und Praxis der Sozialpädagogik*, 9, 13-17.
- Schütz, A./Luckmann, T. (2003): Strukturen der Lebenswelt. Konstanz
- Searle, J. R. (1980): Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, 3(3), 417-457.
- Seising, R. (2021): Es denkt nicht: die vergessenen Geschichten der KI. Leipzig
- Sejnowski, T. J. (2023): Large Language Models and the Reverse Turing Test. *Neural Computation*, 35(3), 309-342.
- Selfridge, O. G. (1958): Pandemonium: a paradigm for learning. *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory*, 513-526.

- Shannon, C. E. (1948): A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379-423.
- Shannon, C. E./MacCarthy, J. E. (1956): *Automata Studies*. Princeton
- Shapiro, D. K. (2023): *Benevolent by Design. Six Words to Safeguard Humanity*. New York
- Shen, X./Chen, Z./Backes, M./Shen, Y./Zhang, Y. (2023): »Do Anything Now«: Characterizing and Evaluating In-The-Wild Jailbreak Prompts on Large Language Models. arXiv:2308.03825v1.
- Simon, F. B. (2018): *Formen. Zur Koppelung von Organismus, Psyche und sozialen Systemen*. Heidelberg
- Singer, W. (2002): *Der Beobachter im Gehirn. Essays zur Hirnforschung*. Frankfurt/Main
- Snyder, A. W. (2001): Paradox of the savant mind. *Nature*, 413, 251-252.
- Socher, R. (2012): Semantic Compositionality through Recursive Matrix-Vector Spaces. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1201-1211.
- Sparrer, I./Kibét, M. V. v. (2000): *Ganz im Gegenteil. Tetralemmaarbeit und andere Grundformen systemischer Strukturaufstellungen - für Querdenker und solche, die es werden wollen*. Heidelberg
- Spencer Brown, G. (1997): *Laws of Form. Gesetze der Form*. Lübeck
- Spencer Brown, G. (2013): *Dieses Spiel geht nur zu zweit. Aus dem Englischen von Andreas Baar*. (3. Auflage ed.). Lübeck
- Spengler, O. (2007 [1923]): *Der Untergang des Abendlandes*. Düsseldorf
- Srivastava, A./Rastogi, A./Rao, A./Shoeb, A. A. M./Abid, A./Fisch, A./al., e. (2022): Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. arXiv:2206.04615v3 1-100.
- Star, S. L. (2015): Die Struktur schlecht strukturierter Lösungen. *Grenzbjekte und heterogenes verteiltes Problemlösen. Navigationen - Zeitschrift für Medien- und Kulturwissenschaften*, 15(1), 57-77.
- Stephan, A., Walter, S. et al. (2021): Nachwort. In: Turing, Alan M.: *Computing Machinery and Intelligence / Können Maschinen denken?* Übers. und hrsg. von Achim Stephan und Sven Walter, Stuttgart: Reclam, S. 131-202.
- Steinbicker, J. (2011): *Zur Theorie der Informationsgesellschaft. Ein Vergleich der Ansätze von Peter Drucker, Daniel Bell und Manuel Castells* (2. Auflage ed.). Wiesbaden
- Sutskever, I./Vinyals, O./Le, Q. V. (2014): Sequence to sequence learning with neural networks. In Z. Ghahramani/M. Welling et al. (Hrsg.), *Advances in Neural Information Processing Systems 2* (Vol. 27, S. 1-9)
- Taleb, N. N. (2018): *Antifragilität: Anleitung für eine Welt, die wir nicht verstehen* (3. Auflage). München
- Tegmark, M. (2017): *Leben 3.0. Mensch sein im Zeitalter Künstlicher Intelligenz*. Berlin
- Tegmark, M./Wheeler, J. A. (2001): *100 Jahre Quantentheorie. Spektrum der Wissenschaft*, 2001/4, 68-76.
- Thoma, S. (2022): *Im Offenen. Henri Maldineys Philosophie der Psychosen*. Wien, Berlin

- Thomas, I. S./Wang, J./GPT₃. (2022): Was euch zu Menschen macht – Antworten einer künstlichen Intelligenz auf die großen Fragen des Lebens. München
- Thomas, W. I./Thomas, D. S. (1928): The child in America: behavior problems and programs. New York
- Thompson, A. D. (2023). Integrated AI: The sky is entrancing (mid-2023 AI retrospective). June 2023. lifearchitect.ai. <https://lifearchitect.ai/the-sky-is-entrancing/> [Abruf: 9.8.2023]
- Thoppilan, R. et al. (2022): LaMDA: Language Models for Dialog Applications. arXiv:2201.08239.
- Thurman, R. A. F. (2003): Das tibetische Totenbuch oder das große Buch der natürlichen Befreiung durch Verstehen im Zwischenzustand. Neu übersetzt und kommentiert von Robert A. F. Thurman. (2. Auflage). Frankfurt
- Tomasello, M. (2009): Die Ursprünge der menschlichen Kommunikation. Berlin
- Tononi, G. (2004): An information integration theory of consciousness. BMC Neuroscience 5.
- Trivers, R. (2011): The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life. New York
- Tugendhat, E. (2006): Egozentrik und Mystik. Eine anthropologische Studie. München
- Tugendhat, E. (2007): Anthropologie statt Metaphysik. München
- Turing, A. (2004): Intelligent Machinery (1948). In B. J. Copeland (Hrsg.), The Essential Turing (S. 0)
- Turing, A. M. (1937): On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, 52-42(1), 230-265.
- Turing, A. M. (1950): Computing Machinery and Intelligence. Mind, LIX, 433-460.
- Turing, A. M. (2004 [1948]): Intelligent machinery. In J. B. Copeland (Hrsg.), The Essential Turing. (S. 395-432). Oxford
- Turing, A. M. (2021): Computing Machinery and Intelligence / Können Maschinen denken? Englisch/Deutsch. Übers. und hrsg. von Achim Stephan und Sven Walter. Stuttgart
- Tversky, A./Daniel, K. (1983): Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. Psychological Review, 90(4), 293-315.
- Uexküll, T. v. (2003): Psychosomatische Medizin. Modelle ärztlichen Denkens und Handelns. (6. Aufl. ed.). München, Jena
- Varela, F. J. (1979): The Extended Calculus of Indications interpreted as a three-valued logic. Notre Dame Journal of Formal Logic, 20(1), 141-146.
- Varela, F. J. (1999): The Specious Present: A Neurophenomenology of Time Consciousness. In J. Petitot/F. J. Varela/B. Pachoud (Hrsg.), Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science. (S. 266-314). Stanford, California
- Varela, F. J./Thompson, E. (2003): Neural Synchrony and the Unity of Mind: A Neurophenomenological Perspective. In A. H. Cleeremans (Hrsg.),

- The Unity of Consciousness: Binding, Integration and Dissociation. (S. 266-287). New York
- Vaswani, A./Shazeer, N./Parmar et al. (2017): Attention Is All You Need. arXiv:(1706.03762), 1-15.
- Vogd, W. (2004): Ärztliche Entscheidungsprozesse des Krankenhauses im Spannungsfeld von System- und Zweckrationalität: Eine qualitativ rekonstruktive Studie. Berlin
- Vogd, W. (2006a): Wer entscheidet, wer entscheidet? Fragen zur Theorie und Empirie zugerechneter Intentionalität. In J. Reichertz/N. Zaboura (Hrsg.), Akteur Gehirn – oder das vermeintliche Ende des handelnden Subjekts. Eine Kontroverse. (S. 137-155). Wiesbaden
- Vogd, W. (2010): Gehirn und Gesellschaft. Weilerswist
- Vogd, W. (2013): Konfusionen (post-)moderner Psychotherapie: verstörte Gesellschaften, verstörte Organisationen, verstörte Individuen und kein Schamane in Sicht. Verhaltenstherapie und psychosoziale Praxis, 45(1), 33-46.
- Vogd, W. (2014a): Von der Physik zur Metaphysik – eine soziologische Rekonstruktion des Deutungsproblems der Quantentheorie. Weilerswist
- Vogd, W. (2014b): Welten ohne Grund. Buddhismus, Sinn und Konstruktion. Heidelberg
- Vogd, W. (2017): Buddhistische Praxis und Sprache. In A. Lasch/W.-A. Liebert (Hrsg.), Handbuch Sprache in der Religion. (Vol. Band 18, S. 154-178). Berlin
- Vogd, W. (2018): Selbst- und Weltverhältnisse. Leiblichkeit, Polykontexturalität und implizite Ethik. Weilerswist
- Vogd, W. (2019): Der ermächtigte Meister. Eine systemische Rekonstruktion am Beispiel des Skandals um Sogyal Rinpoche. Heidelberg
- Vogd, W. (2020): Quantenphysik und Soziologie im Dialog. Betrachtungen zu Zeit, Beobachtung und Verschränkung. Berlin
- Vogd, W. (2021): Offenheit für neue Kategorien und Begründungen – warum eine Wissenssoziologie, die diesen Namen verdient, sich immerfort erneuern muss. Eine Würdigung der praxeologischen Wissenssoziologie von Ralf Bohnsack. ZQF-Zeitschrift für Qualitative Sozialforschung, 22(1), 107-122.
- Vogd, W./Feißt, M. (2022): Therapeutische Arrangements im Maßregelvollzug. Studien zur Leerstellengrammatik und den Bezugsproblemen in der forensischen Psychiatrie Wiesbaden
- Vogd, W./Harth, J. (2015): Die Praxis der Leere. Zur Verkörperung buddhistischer Lehren in Erleben, Reflexion und Lehrer-Schüler-Beziehung. Weilerswist
- Vogd, W./Harth, J. (2019): Kontexturanalyse: Eine Methodologie zur Rekonstruktion polykontexturaler Zusammenhänge. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research.
- von Brück, M. (2015): Weltinnenraum. Rainer Maria Rilkes Duineser Elegien in Resonanz mit dem Buddha. Freiburg, Basel, Wien
- von Foerster, H. (1992): Ethics and Second-order Cybernetics. Cybernetics & Human Knowing 1(1), 9-19.
- von Foerster, H. (2006): Über selbst-organisierende Systeme und ihre

- Umwelten. In ders. (Hrsg.), *Sicht und Einsicht. Versuche zu einer operativen Erkenntnistheorie* (S. 115-130). Heidelberg
- von Foerster, H./Bröcker, M. (2002): *Teil der Welt. Fraktale einer Ethik. Ein Drama in drei Akten*. Heidelberg
- von Foerster, H./Pörksen, B. (2022): *Wahrheit ist die Erfindung eines Lügners. Gespräche für Skeptiker*. Heidelberg
- Waldenfels, H. (2013): *Absolutes Nichts. Zur Grundlegung des Dialogs zwischen Buddhismus und Christentum. Mit einem Vorwort von Nishitani Keiji* (4. Auflage). Paderborn
- Wang, L. J./Zou, X. Y./Mandel, L. (1991): Induced coherence without induced emission. *Physical Review A*, 44(7), 4614-4622.
- Watzlawick, P./Beavin, J. H./Jackson, D. D. (1990): *Menschliche Kommunikation. Formen, Störungen, Paradoxien*. (8. Auflage ed.). Bern, Stuttgart, Toronto
- Weber, A. (2003): *Natur als Bedeutung. Versuch einer semiotischen Theorie des Lebendigen*. Würzburg
- Wei, J./Wang, X./Schuurmans, D. et al. (2022): Chain-of-thought prompting elicits reasoning in large language models. [arXiv\(11903v6\)](https://arxiv.org/abs/11903v6).
- Weidinger, L./al., e. (2021): Ethical and social risks of harm from language models. [arXiv:\(2112.04359\)](https://arxiv.org/abs/2112.04359).
- Weizenbaum, J. (1966): ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the Acm*, 9, 36-45.
- Weizsäcker, C. F. v. (1994): *Der Aufbau der Physik*. (3. Auflage ed.). München
- Whitehead, A. N. (1987): *Prozess und Realität. Entwurf einer Kosmologie*. Frankfurt/Main
- Wiener, N. (1948): *Cybernetics or Control and Communication in the Animal and the Machine* (2. Aufl.). Cambridge (MA)
- Wiener, N. (1960): Some Moral and Technical Consequences of Automation. *Science*, 131(3410), 1355-1358.
- Wigner, E. P. (1967): Remarks on the Mind-Body-Question. In E. P. Wigner (Hrsg.), *Symmetries and Reflections* (S. 171-184). Indiana
- Wittgenstein, L. (1963 [1922]): *Tractatus logico philosophicus*. Frankfurt/M.
- Wittgenstein, L. (1989): *Vortrag über Ethik und andere kleine Schriften*. Frankfurt/M.
- Wittgenstein, L. (1990): *Tractatus logico philosophicus; Philosophische Untersuchungen*. Leipzig
- Wygotski, L. S. (1986 [1934]): *Denken und Sprechen*. Frankfurt/M.
- Youniss, J. E. (1994): *Soziale Konstruktion und psychische Entwicklung*. Frankfurt/M.
- Zeh, J. (2007): *Spieltrieb*. Frankfurt/Main
- Zeilinger, A. (2005): *Einsteins Schleier. Die neue Welt der Quantenphysik*. (2 ed.). München
- Zeilinger, A. (2007): *Einsteins Spuk. Teleportation und weitere Mysterien der Quantenphysik*. München
- Zuse, K. (1984): *Der Computer – Mein Lebenswerk*. Landsberg am Lech