Begegnungen mit Künstlicher Intelligenz

Intersubjektivität, Technik, Lebenswelt

Herausgegeben von
Martin W. Schnell und Lukas Nehlsen

VELBRÜCK
WISSENSCHAFT

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Martin W. Schnell und Lukas Nehlsen (Hg.)
Begegnungen mit Künstlicher Intelligenz

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Begegnungen mit Künstlicher Intelligenz

Intersubjektivität, Technik, Lebenswelt

Herausgegeben von
Martin W. Schnell und Lukas Nehlsen

VELBRÜCK
WISSENSCHAFT

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Erste Auflage 2022
© Velbrück Wissenschaft, Weilerswist 2022
www.velbrueck-wissenschaft.de
Printed in Germany
ISBN 978-3-95832-288-2

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Inhalt

Vorwort 7

Martin W. Schnell und Lukas Nehlsen
Begegnungen mit Künstlicher Intelligenz.
Intersubjektivität, Technik, Lebenswelt 9

Oliver Müller
Maschinelle Alterität.
Philosophische Perspektiven auf Begegnungen
mit künstlicher Intelligenz 23

Burkhard Liebsch
Aussichten kommunikativer Verwerfungen
zwischen *digital natives* und analogen Aborigines 48

Jonathan Harth und Martin Feißt
Neue soziale Kontingenzmaschinen.
Überlegungen zu künstlicher sozialer Intelligenz
am Beispiel der Interaktion mit GPT-3 70

Martin W. Schnell und Lukas Nehlsen
Gespräch mit einer Künstlicher Intelligenz.
Eine qualitative Inhaltsanalyse 104

Lukas Nehlsen
Chatten mit Nirgendwo?
Der Leib als Ausdruck und die Sprache der KI 119

Martin W. Schnell
Anerkennung einer Künstlichen Intelligenz
als meinesgleichen? Beitrag zur Maschinenethik 145

Renate Buschmann
Kann aus KI Kunst werden?
Dialogische Beziehungen mit Künstlicher Intelligenz 164

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Christian Grüny
Seltsam attraktiv. KI und Musikproduktion 174

Alexander Hochmuth
Pflege und künstliche Intelligenz.
Soziale Roboter in der Pflege von Menschen
mit einer Demenz 205

Erik N. Dzwiza-Ohlsen
Stay in Touch!
Phänomenologische Überlegungen zum Einsatz
von Robotik und KI in der Pflege bei
demenziellen Erkrankungen am Beispiel von PARO 220

Die Autorinnen und Autoren 248

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Vorwort

Begegnungen mit Künstlicher Intelligenz vertiefen die Transformation der Lebenswelt, die durch den weltweiten Prozess der Digitalisierung forciert wird. In Form von Interaktionen zwischen Mensch und KI produzieren diese Begegnungen Selbstverständlichkeiten und Normalitäten, die als solche zu unbefragten Beständen der Lebenswelt werden.

Die Digitalisierung schafft das scheinbar Unproduzierbare, nämlich Selbstverständlichkeit. Begegnungen mit Künstlicher Intelligenz sind ein wesentlicher Teil dieses Prozesses, der das Verständnis von Selbst, Welt und Anderem herausfordert. Husserls These, ja tiefe Überzeugung, dass unsere Lebenswelt in ihrer Wesensstruktur, in ihrem eigenen konkreten Kausalstil ungeändert bleibt, was immer wir tun, könnte ins Wanken geraten. Technisierung und Digitalisierung verändern unsere Lebenswelt geradezu ins Fraglose. Diese Veränderungen drohen unbemerkt und unthematisiert zu bleiben, gäbe es nicht Schlüsselerfahrungen des Umbruchs, in denen das Staunen über Künstliche Intelligenz aufblitzt und die es geistes- sowie sozialwissenschaftlich zu reflektieren gilt.

Das Auslösen technischer Effekte erfolgt heute nicht mehr durch Schalter, Knöpfe oder Regler, sondern via Sprachsteuerung und es versetzt uns damit, aufgrund der Anredeform, in eine Art der Begegnung mit Künstlicher Intelligenz.

Die Beiträge des Bandes untersuchen Formen und Inhalte dieser Begegnungen:

- Was bedeutet es, mit einer Maschine zu musizieren, sich von einem Roboter pflegen und berühren zu lassen, mit einer KI ein intelligentes Gespräch zu führen?
- Wird in der Begegnung eine KI vom Ich als Du oder als Anderer anerkannt?
- In welcher Hinsicht werden KIs zu »normalen« Partnern des Alltags und verändern dadurch unsere Lebenswelt?
- Kann KI im umfassenden Sinne kreativ sein, also »selbst« etwas hervorbringen, das der Hervorbringung zuvor noch nicht Bestand der Lebenswelt gewesen ist?
- Gibt es neue Antworten auf die alte Frage nach dem »Geist« in der Maschine? Oder muss diese Frage heute ganz anders gestellt werden?

Am 20. Oktober 2021 fand eine Konferenz an der Universität Witten/Herdecke statt, auf der einige der hier versammelten Autor:innen ihre Thesen vorgestellt und auf die Publikation abgestimmt haben.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Begegnungen mit Künstlicher Intelligenz ist ein Projekt des Lehrstuhls für Sozialphilosophie und Ethik im Gesundheitswesen an der Universität Witten/Herdecke. Unterstützt wurde es maßgeblich von Christine Dunger, Isabella D'Angelo, Maja Ebbert und Thilo Bremer. Wir bedanken uns bei der Autorin und den Autoren des Bandes und bei Mercedes, die sich immer wieder, auch ungefragt, in die Kommunikation einschaltete, bei der *Schaufler Foundation*, der *a.r.t.e.s graduate school* an der Universität zu Köln und der Fakultät für Gesundheit der Universität Witten/Herdecke für finanzielle Unterstützung bei der Ausarbeitung der Publikation.

Martin W. Schnell und Lukas Nehlsen
im Februar 2022

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Martin W. Schnell / Lukas Nehlsen

Begegnungen mit Künstlicher Intelligenz

Intersubjektivität, Technik, Lebenswelt

In allen gesellschaftlichen Bereichen kommt es täglich zu Begegnungen des Menschen mit Künstlicher Intelligenz (KI) in Form von Kommunikationen und Interaktionen. Diese Begegnungen verändern unsere Verständnisse von Selbst-, Welt und Fremdverhältnissen. Die Elemente dieser komplexen Sachlage gilt es zu klären, so dass deutlicher wird, was eigentlich Begegnungen mit Künstlicher Intelligenz sind.

Begegnungen mit Künstlicher Intelligenz vertiefen die digitale Transformation der Lebenswelt. Dabei treten Künstliche Intelligenzen nicht mehr wie Maschinen, die der Mensch auf gewisse Weise steuert, auf, sondern als vermeintlich autonome Akteure[1]. »Die *information society* wandelt sich zur *interaction society*« (Priddat 2022, 317). Autos geben im Innenraum differenzierte und sprachliche Signale von sich, die den Fahrer mal überwachen, mal informieren und mal belohnen. Kreative Maschinen gestalten auf Anforderung Gemälde im Stile von Picasso, Bacon oder van Gogh, schreiben angefangene Sinfonien von Mahler »zu Ende« oder komponieren zahllose Choräle im Stile von Bach. Die neuesten Sprachprogramme zeigen beeindruckende Fähigkeiten sowohl beim Sprachverständnis als auch bei der Sprachproduktion und schreiben so täuschend »echte« journalistische Artikel oder führen Konversationen mit Menschen, die bisweilen durchaus interessant sind. Diese und andere Begegnungen zwischen Mensch und KI produzieren Selbstverständlichkeiten, die als solche zu unbefragten Beständen der Lebenswelt werden. Nachfolgend sollen die einzelnen Elemente dieses Definitionsversuchs gesondert dargestellt und aufeinander bezogen werden.

Begegnung

Die Begegnungsphilosophie beginnt mit Humboldt und Feuerbach, die in der Moderne den Dialog von Ich und Du als für die Entstehung von

1 Wenn in der Einleitung an verschiedenen Stellen die männliche Form gewählt wird, so geschieht dies aus Gründen der Leserlichkeit; in solchen Formulierungen sind selbstverständlich alle Geschlechter inkludiert.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Sinn und Bedeutung relevant eingeführt haben. In seinem Aufsatz *Über den Dualis* (1827) beschreibt Wilhelm von Humboldt Anrede und Erwiderung zwischen Ich und Du als Ausgangspunkte aller Sinnbildung. Bereits hier zeigt sich, dass die Begegnung eine gewisse ethische Dimension besitzt, die in der gesamten Diskussion gegenwärtig ist und in der Philosophie von Emmanuel Levinas zur vollen Entfaltung kommt.

Eine anthropologische Realisierung der Theologie ist, Marx zufolge, die Reaktion Ludwig Feuerbachs auf den Idealismus. In seinen *Grundsätzen der Philosophie der Zukunft* (1846) formuliert dieser, dass »die wahre Dialektik kein Monolog des einsamen Denkers mit sich selbst, (sondern) ein Dialog zwischen Ich und Du ist.« (§ 62) Fortgesetzt und zur Blüte gebracht wurde das Paradigma der Begegnung durch Hermann Cohen und die jüdische Dialogphilosophie, welche von Karl Löwith im 20. Jahrhundert, teilweise in der Rückschau und als Reaktion auf den Nationalsozialismus, quasi als Bewegung herausgestellt worden ist (vgl.: Löwith 2013).

Franz Rosenzweig, Schüler und Kritiker Hegels, formuliert in *Der Stern der Erlösung* (1921): »Erst indem das Ich das Du als etwas außer sich anerkennt, …, wird es zum Ich.« (Rosenzweig 1988, 195) Eine daran angelehnte Formulierung in Martin Bubers Werk *Ich und Du* (1926) gibt dem dialogischen Prinzip der Begegnung dann sein bekanntes Motto: »Der Mensch wird am Du zum Ich.« (Buber 1983, 37) Die jüdische Dialogphilosophie betont besonders zwei Aspekte der Begegnung. Einmal den Kontext der Begegnung und dann die Eigenart der Begegnung selbst betreffend (vgl.: Wojcieszuk 2010).

Der Kontext der Begegnung ist die Lebenswelt oder, wie Aron Gurwitsch sagt, das Milieu, wie es in seinem Buch *Die mitmenschlichen Begegnungen in der Milieuwelt* (1931) heißt. Die Lebenswelt in der Version des Milieus ist nach Max Scheler als das »auf mich als wirksam Erlebte« (Scheler 2000, 154) definiert. Die Eigenart einer Begegnung zwischen Ich und Du in der Lebenswelt ist die Antwort auf die Frage, »wie man … überhaupt zum anderen Menschen als Menschen kommt, wie sich uns der andere Mensch erschließt.« (Gurwitsch 1977, 29) Ausdrücklich werden die Verschmelzung von Ich und Du (W. Wundt), die Einfühlung des Ich in das Du (Th. Lipps) und der Analogieschluss zwischen Ich und Du (J. St. Mill) als Zugangsarten zurückgewiesen (vgl.: ebd., 28ff). Stattdessen konstituiere die Begegnung das Ich und das Du als Differenz. Eine Begegnung beinhaltet bei aller Nähe immer auch eine gewisse Ablösung zwischen Ich und Du (vgl.: Gurwitsch 1977, 51ff.). Hier kommen Motive zur Geltung, die im 20. Jahrhundert von Georg Simmels Soziologie bis zu Paul Ricœurs Philosophie des Anderen reichen (vgl.: Schnell 2020, Kap. 4).

In der Phänomenologie werden diese Motive im Hinblick auf eine gewisse Unerforschlichkeit des Anderen für das Ich ausgedeutet. Edmund

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Husserl spricht von einer »originalen Unzugänglichkeit« (*V. Cartesianische Meditation*, § 52) des Anderen und Emmanuel Levinas von einem steten »Abstand zwischen mir und dem Anderen« (Levinas 1987, 45).

Die Hervorhebung der Herkunft des Paradigmas der Begegnung aus dem Judentum bezeichnet lediglich eine Entstehungsgeschichte und keine Sonderwelt. Das Judentum gehört genauso zu Europa wie andere Traditionen auch. Die jüdische Dialogphilosophie ist ebenso eine Philosophie, d.h. eine Wissenschaft vom Logos und der Welt, wie die Platons, Kants oder Hegels. Daher ist es möglich und sinnvoll, das Paradigma der Begegnung mit einem anderen, in gewisser Hinsicht konkurrierenden Paradigma zu kontrastieren. Es handelt sich um den Kontrast zwischen den Paradigmen der Begegnung in der Dialogphilosophie und der Zuschreibung in der analytischen Philosophie. Beiden geht es um eine Antwort auf die Frage nach der Eigenart einer Person (vgl. dazu: Schnell 1999).

Ausgehend von der seit John Locke diskutierten Frage nach dem Zusammenhang personaler Bewusstseinskontinuität und menschlicher Körperlichkeit vertritt Peter F. Strawson die These, dass man jedem Individuum des Typus der Person »sowohl Bewußtseinszustände als auch körperliche Eigenschaften zuschreibt bzw. zuschreiben kann.« (Strawson 1972, 134) Der Akt der *Zuschreibung* beinhaltet, dass eine Person ein Einzelding ist, über das man spricht. Eine Person tritt damit als dritte Person Singular (Er/Sie/Es) auf. Im Licht der Zuschreibung ist jemand etwas, *über* den, die oder das man spricht. Der Dialog, also die Perspektive, dass ich nicht nur über jemanden, sondern *mit* jemandem spreche und dieser damit als ein Du erscheint, ist im Rahmen der Zuschreibung nicht zu realisieren. Interessanterweise haben die von Strawson ausgehenden Überlegungen auch in die Ethik hineingewirkt. Über Richard M. Hares Studien zur Bioethik (vgl.: Hare 1993) wurden sie von dessen Schüler Peter Singer aufgegriffen. Bekanntlich schlägt Singer vor, durch die Zuschreibung sog. »moralisch relevanter Eigenschaften wie Rationalität, Selbstbewußtsein, Bewußtsein, Autonomie, Lust- und Schmerzempfindung« (Singer 1994, 196f.) den sog. Wert des Lebens von Föten, Tieren und Menschen mit Behinderungen zu bestimmen. Es ist wohl kein Zufall, dass die ethischen Aussagen, die aus dem Zuschreibungsparadigma resultieren und jene, die aus dem Begegnungsparadigma erfolgen, einander zum Teil erheblich widersprechen (vgl.: Schnell 2017, 153ff.).

Die radikalste Version des Paradigmas der *Begegnung* vertritt Emmanuel Levinas. Daher spricht er auch von einer »Ethik der Begegnung.« (Levinas 1995, 270) Der angesprochene Andere ist darin kein zu begreifendes Sein, sondern ein Du. »Der Nächste ist nicht erst Gegenstand des Verstehens und dann Gesprächspartner. … Eine Person verstehen heißt schon mit ihr zu sprechen.« (ebd., 17) Der Andere ist immer jemand, mit dem ich spreche und nicht etwas, über das ich spreche.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Mit dieser Bestimmung ist eine Entscheidung über das Verhältnis zwischen den beiden Paradigmen gefallen. Zuschreiben und Begegnen sind demnach keine Optionen, die wahlweise realisiert werden können. Vielmehr ist die Begegnung das Fundament aller weiteren Maßnahmen. Anders formuliert: die Zuschreibung von Eigenschaften an die Adresse einer Person setzt die Begegnung mit ihr voraus (vgl.: Schnell 2017, 96ff., 153ff.). Mit Sartre gesprochen: »Man *begegnet* dem Anderen, man konstituiert ihn nicht.« (Sartre 1991, 452) Insofern ist der Andere immer zuerst ein Du bevor er ein Er/Sie/Es werden kann. Das gilt auch für die verweigerte Begegnung, die Zygmunt Bauman eine »Vergegnung« nennt und »die eine Begegnung ist, die vortäuscht, keine zu sein.« (Bauman 1995, 229)

Wenn in diesem Band also von »Begegnungen mit Künstlicher Intelligenz« die Rede sein soll, so ist es durchaus diese emphatische Traditionslinie, die im Hintergrund steht. Über KI soll also nicht in Hinsicht auf Fragen der Zuschreibung gesprochen werden, sondern die Sache der KI soll in Hinsicht auf die Begegnung gedacht werden, wobei natürlich nicht von Anfang an vorausgesetzt ist, dass es hier eine Gleichwertigkeit (in ontologischer oder ethischer Hinsicht) mit der Begegnung von Menschen untereinander gibt. Ausgehend von Bubers Motto: »Der Mensch wird am Du zum Ich« (Buber 1983, 37), und seiner weitergehenden Aussage, dass es »kein Ich an sich«, »sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es« gibt (ebd., 10), kann man danach fragen, was passiert, wenn man in dieser Situation ein Relatum durch eine KI oder einen Roboter ersetzt, der oder die aufgrund seines Status in der Interaktion nicht mehr zweifelsohne als Es charakterisiert werden kann, sondern zumindest einen Anspruch auf das Du erhebt.

Künstliche Intelligenz

Als Geburtsstunde der KI-Forschung als Disziplin wird die Dartmouth Conference von 1956 angesehen, an der viele der in den folgenden Jahren für das Feld maßgebenden Forscher teilnahmen, so z.B. John McCarthy, Marvin Minsky, Claude Shannon sowie Allen Newell und Herbert A. Simon. Im Förderungsantrag ist das Grundparadigma der KI-Forschung prägnant auf den Punkt gebracht: »The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it« (McCarthy et al. 1956: 1). Dieses Programm ist bis heute disziplinbestimmend. Auch der große KI-Kritiker Joseph Weizenbaum spricht 20 Jahre später davon, »daß es keinen Bereich des menschlichen Daseins gibt, der nicht maschinell erfaßbar

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ist.« (Weizenbaum 1978, 274) Auch in aktuellen Publikationen ist die Definition der KI sehr nah an der der Dartmouth Conference: »Artificial Intelligence (AI) seeks to make computers do the sorts of things that minds can do« (Boden 2018, 1). Man sieht allerdings, dass die Definitionen darin variieren, was genau simuliert oder künstlich nachgebildet werden soll (»learning«, »intelligence«, »menschliches Dasein«, »what minds do«). So unterscheiden Russel und Norvig in ihrem modernen Standardwerk *Artificial Intelligence: A Modern Approach* zwischen vier Vorbildern für die KI. Nachgeahmt werden sollen wahlweise: menschliches Denken, menschliches Handeln, rationales Denken oder rationales Handeln. Wobei sie »rationales Handeln« als die beste der vier Möglichkeiten ansehen und sich für diese Zielsetzung als »Standardmodell« entscheiden (Russel und Norvig 2021, Kap. 1). Unter der Gefahr, den verschiedenen Definitionen eine weitere hinzuzufügen, können wir *Künstliche Intelligenz* in dieser Einleitung als Sammelbegriff für Prozeduren verstehen, bei denen Maschinen menschen*ähnliche* (d.h. eventuell auch »rationaler als menschliche«) Intelligenzleistungen erbringen.

Ähnlich weit gestreut wie die Zieldefinitionen sind heute die Arten und Weisen, auf denen diese Ziele erreicht werden sollen. In der ersten Welle der KI-Forschung, später von John Haugeland »good old fashioned AI« (GOFAI) genannt (vgl.: Haugeland 1985), waren es vor allem symbolisch-diskret operierende Maschinen, für die Newells und Simons »physical symbol system« das Paradigma gab, die das Ziel der Dartmouth Conference erreichen sollten (vgl.: Newell und Simon 1976). Nachdem der versprochene rasante Erfolg im Erreichen menschenähnlicher Leistungen in vielen Bereichen ausblieb, folgte der sogenannte KI-Winter, eine Periode in der die zu Beginn enorme Forschungsförderung zunehmend ausblieb und das GOFAI-Programm, nicht zuletzt aufgrund der grundlegenden Kritik an seiner Durchsetzbarkeit seitens Hubert Dreyfus' (vgl.: Dreyfus 1992), nahezu tot schien. In der zweiten KI-Welle erlebte das eigentlich seit Beginn der KI-Forschung bestehende, aber lange Zeit in den Hintergrund gerückte Programm des Konnektionismus ein Revival. War die funktionalistische Prämisse des GOFAI-Programms, dass der *mind* unabhängig von seiner physischen Implementierung im menschlichen Gehirn, rein auf Grundlage seiner symbolischen Operationen nachgebildet werden könnte, nahm sich das Programm des Konnektionismus nun gerade die Netzwerkarchitektur des menschlichen Hirns zum Vorbild. Hatten frühe Versionen der *artifical neural networks (ANNs)* meist nur einen (simulierten) Input- und einen Output-Layer, weisen heutige Programme eine erhebliche Tiefe auf und werden daher oft unter dem Stichwort *deep neural networks (DNNs)* beschrieben. Die meisten schlagzeilen-machenden Erfolge der letzten 20 Jahre wurden von solchen künstlichen neuronalen Netzen vollbracht, auch wenn es nach wie vor und gerade in spezialisierten Bereichen unzählige auf

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

GOFAI-Technologie beruhende Systeme gibt und auch viele der heutigen *machine learning*-Programme GOFAI-Technologie nutzen (vgl. Boden 2018, 40).

Zusätzlich zu diesen beiden Paradigmen der Informationsverarbeitung in KI-Systemen gibt es heute (manche sprechen daher schon von einer dritten KI-Welle) noch »evolutionary programing«, »cellular automata«, »dynamical systems« sowie hybride Systeme, die verschiedene Paradigmen vereinen (vgl. ebd.: 5).

Entscheidend für die in diesem Band vorgeschlagene Perspektive sind aber nicht so sehr die technologischen Architekturen und deren theoretische Möglichkeiten und Unmöglichkeiten im Erreichen der verschiedenen Zielsetzungen, sondern vielmehr das Eingehen all dieser verschiedenen Systeme in unsere Lebenswelt und unsere Art und Weise mit ihnen zu interagieren.

Diesbezüglich kann technikgeschichtlich zwischen Anwendungen von Technik unterschieden werden, die im Auftrag des Menschen bloße Ausführungen realisieren, und Anwendungen, die von sich aus einer Kommunikation etwas hinzufügen können.

Position	*Funktion der ….*	*Beispiel*	*Autor:innen*
1. Technik	Kompensation biologischer Mangelausstattung	Hammer	Platon, Gehlen
2. Technologie	Vorschriften, Nachrichten	Navigationssystem Flugzeug	Heidegger-Schüler: Anders, Marcuse, Jonas
3. Körpertechnik, Selbstdarstellung	Formierung des Körpers	Militär, Etikette	Mauss, Goffman, Bourdieu, Deleuze
4. Phänomenotechnik	Organisation von Erfahrungsweisen	Fotoapparat	Bachelard, Blumenberg, Waldenfels
5. Robotik, Künstliche Intelligenzen	Dienstleistungskommunikation mit Freund und Helfer, Informationsbeschaffung	Tee anreichen, Kommunizieren	Wagner, »Mensch-Maschine Interaktion«

Künstliche Intelligenz tritt in allen Bereichen auf, vermehrt, und für die hier eingenommene Perspektive besonders interessant, aber als letztere Anwendung.

Ein weiteres, hier hilfreiches Unterscheidungsschema ist das von Don Ihde in seiner technikphilosophisch inspirierten »Postphänomenologie« ausgearbeitete. Ihde analysiert verschiedene Beziehungen, in denen Menschen mit Technik stehen können: In »alteritiy relations« gibt es eine

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Bezugnahme auf ein technisches Artefakt als (Quasi-)Anderer. Diese werden unterschieden von »embodiment relations« (z.B. das Tragen einer Brille, eines Hörgeräts oder auch eines mit einer gewissen Intelligenz ausgestatteten »wearables«, wie z.B. einer Smartwatch), »hermeneutic relations«, in denen die Technik gelesen und als Weltausschnitte repräsentierend gedeutet wird (z.B. MRT-Bilder, Thermostate etc., auch hier können die zu lesenden Artefakte durchaus »intelligent« sein) und »background relations« (z.B. das Heizungssystem in unserem Haus, aber auch der Algorithmus der unsere Twitter-Feeds generiert) (vgl.: Ihde 1990, Kap. 5).

Wie wir sehen, kann in allen Arten der Beziehung KI eine Rolle spielen, besonders virulent wird die Frage nach der Begegnung in Bezug auf die oben skizzierte Tradition der Begegnungsphilosophie aber natürlich da, wo die Beziehung zu einer »alterity relation« wird.

Es sind aber in einem gewissen Sinne durchaus alle Arten der Relationen mit KI als eine Art der Begegnung mit dieser zu verstehen, insofern sie nämlich gerade aus der Perspektive der Lebenswelt, in der uns Technik zu allererst begegnet, verstanden werden. Alle Arten unserer Bezugnahme auf Technik finden in der Lebenswelt statt, sie nutzen die digitale Transformation der Lebenswelt, indem sie an sie anknüpfen und sie vertiefen. Wie ist diese Begegnung »in der Lebenswelt« genauer zu verstehen?

Begegnungen mit Künstlicher Intelligenz – eine weitere Kränkung des Menschen?

Husserl definiert die Lebenswelt als die »wirklich anschauliche, wirklich erfahrbare und erfahrene Welt, in der sich unser ganzes Leben praktisch abspielt.« (Husserl 1976, 51) In dieser grundsätzlichen Funktion ist die »Lebenswelt ein allgemeines Fundament für alle Sinnbildung.« (Waldenfels 1985, 17) Husserl unterscheidet zwischen Lebenswelten im Plural, die kulturrelativ und hier anders als dort sein können, und der einen Lebenswelt, die, formaler betrachtet im Sinne eines »abstrakt herauszupräparienden Weltkern(s)« (Husserl 1976, 136), jenes allgemeine Fundament für alle Sinnbildungen, im Alltag wie in den Wissenschaften, ist. In dieser allgemeinen und zugleich speziellen Funktion hat die Lebenswelt etwas Verlässliches und Beständiges. In diesem Sinne betont Husserl, dass die Lebenswelt »in ihrer eigenen Wesensstruktur, in ihrem eigenen Kausalstil ungeändert […] bleibt, was immer wir kunstlos oder als Kunst tun.« (ebd., 51) Auch die Technisierung durch die modernen Wissenschaften, an die Husserl dachte, können die Lebenswelt zwar kulturell verändern, aber nicht in ihrer Fundamentfunktion. Es stellt sich die Frage, ob diese Ansicht noch immer von Geltung ist.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Dem Hinweis Sigmund Freuds, dass der Mensch drei große Kränkungen habe hinnehmen müssen, weil sein Planet nicht der Mittelpunkt des Weltalls ist, seine Abstammung kein Schöpfungsvorrecht beinhalte und er selbst nicht Herr im eigenen Seelenhaushalt sei (vgl.: Freud 1983, 226), fügte Pierre Bourdieu eine vierte Kränkung hinzu. Demnach sind die Haltungen und Stellungnahmen des Menschen gegenüber der Welt keineswegs spontane Leistungen des Verstandes, sondern vielmehr gesellschaftlich produziert (vgl.: Bourdieu 1989, 7). Auf diese Befunde folgt möglicherweise noch eine fünfte Kränkung. Die Lebenswelt wird in ihrer Fundamentfunktion sehr wohl durch die Digitalisierung und die Begegnungen mit Künstlicher Intelligenz deutlich verändert. Seit der Romantik muss sich der moderne Mensch immer wieder gegen die Maschine behaupten, was ihm jedoch nie ganz gelungen ist. Sein Narzissmus wird ein fünftes Mal gekränkt, wenn die ihm, als natürlich geborenem Wesen, bisher exklusiv vorbehaltene Domäne des rationalen Denkens und Handelns nun künstlich herstellbar wird.[2]

Digitalisierung ist als eine qualitativ neue Dimension der Technisierung zu verstehen, als die »mit einer Automatisierung verbundene Transformation der Welt in quantifizierbare Daten.« Sie stellt damit »auch einen Angriff auf die Materialität der Lebenswelt dar.«[3] (Schnell/Dunger 2019, 13) *Begegnungen mit Künstlicher Intelligenz* vertiefen diese Transformation der Lebenswelt.

In Weiterführung der Analysen Husserls deutete Hans Blumenberg das Telos der Technisierung als das Zurücksinken technischer Artefakte in das »›Universum der Selbstverständlichkeiten‹« der »Lebenswelt«. Dadurch wird »jene Sphäre, in der wir *noch* keine Fragen stellen, identisch […] mit derjenigen, in der wir keine Fragen *mehr* stellen.« Die »Technisierung […] produziert […] das scheinbar Unproduzierbare […], nämlich Selbstverständlichkeit.« (Blumenberg 2015, 190) Begegnungen mit

2 Diese Idee geht auf Catherine Malabou zurück, die im Englischen Postscript zu ihrem großartigen Buch *Morphing Intelligence* die Konfrontation mit Künstlicher Intelligenz im Hinblick auf Freuds Aufzählung – ohne Bourdieus Ergänzung mitzuzählen – schon als eine mögliche vierte Kränkung beschrieben hat (vgl. Malabou 2019: 162ff.).

3 Im Digitalisierungsdiskurs wird oft zwischen »digitisation« und »digitalisation« unterschieden, wobei ersteres die Produktion digitaler Daten oder die Umwandlung analoger Daten in digitale bezeichnet und letzteres die Nutzung dieser Daten durch Akteure und die Auswirkungen dieses Prozesses auf gesellschaftliche Wirklichkeiten. Mit der hier vorgeschlagenen Definition von Digitalisierung sind diese beiden Aspekte umfasst, unter fernerer Berücksichtigung des Umstandes, dass die Daten immer aus der Lebenswelt erhoben werden (1. Sinn der Transformation der Lebenswelt) und im Prozess der Digitalisierung es diese Lebenswelt selbst ist, die verändert wird (2. Sinn der Transformation).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Künstlicher Intelligenz im Zeichen der Digitalisierung vertiefen diesen Prozess, eventuell sogar in Bezug auf die intersubjektive Struktur unserer Lebenswelt, insofern wir mit technischen Artefakten eben in »alterity relations« treten. Somit gerät Husserls These, ja tiefe Überzeugung, dass unsere Lebenswelt in ihrer Wesensstruktur, in ihrem eigenen konkreten Kausalstil ungeändert bleibt, was immer wir kunstlos oder als Kunst tun, ins Wanken. Technisierung und Digitalisierung verändern unsere Lebenswelt geradezu ins Fraglose. Diese Veränderungen drohen unbemerkt und unthematisiert zu bleiben, gäbe es nicht Schlüsselerfahrungen des Umbruchs, in denen das Staunen über Künstliche Intelligenz aufblitzt.

Das Auslösen technischer Effekte erfolgt heute nicht mehr durch Schalter, Knöpfe oder Regler, sondern zunehmend via Sprachsteuerung und es versetzt uns damit, aufgrund der Anredeform, in eine Art der Begegnung mit Künstlicher Intelligenz. Bereits vor über 30 Jahren befragte Hans Blumenberg die Situation der Begegnung des Menschen mit künstlichen, aber menschenähnlichen Automaten. Im Ausgang von Husserls Theorie der Fremderfahrung heißt es: »Wie aber erreichen … die menschengestaltigen Organismen, von ihm [dem Menschen, die Verf.] als Seinesgleichen anerkannt zu werden?« (Blumenberg 1989, 442) Diese Frage beschäftigt noch die Gegenwart, die technische Geräte kennt, die im letzten Jahrhundert, wenn überhaupt, nur im spekulativen Denken gefunden wurden.

Generative Pre-trained Transformer 3 (GPT-3) ist ein autoregressives Sprachmodell, das Deep Learning benutzt, um menschenähnliche Texte zu erstellen und zu verbalisieren. Bekanntlich ist es möglich, dass ein Mensch mit GPT-3 einen inhaltlichen Austausch über anspruchsvolle Themen wie Liebe, Freiheit, Datenschutz oder gar Künstliche Intelligenz führen kann. Mithilfe einer quantifizierten Datenbasis kann das System, das in einer Kommunikation wahrscheinlich nächste Wort und den wahrscheinlich nächsten Satz berechnen. Auf die allmähliche Verfertigung der Gedanken beim Reden folgt mit hoher Wahrscheinlichkeit irgendwann der Name Heinrich von Kleists und der Hinweis auf seinen traurigen Selbstmord im Jahre 1812. Dieser Austausch wird zur Norm für ein sogenanntes gutes Gespräch, zumal wenn GPT-3 den Menschen darüber belehrt, dass sich Kleist mit seiner Pistole bereits 1811, also ein Jahr zuvor, erschossen hat. Von hier aus sind dann komplexe Interaktionen, die gar verantwortungsvolle Entscheidungen beinhalten, denkbar.

Die verborgene, sich als ethisch ausgebende Implikation dieser Möglichkeiten hat Pierre Lévy bereits 1997 in seinem Buch *Die kollektive Intelligenz* skizziert. Demnach wird das Internet, dass 1997 für den Privatgebrauch erst seit wenigen Jahren ausgelegt war, eines Tages alles Wissen der Welt bereithalten und einen Zugriff auf all dieses Wissen für jedermann zu jeder Zeit und von jedem Ort aus ermöglichen. Jeder darf an einem allseitigen Wissensaustausch teilnehmen. Da offizielle Zertifikate

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und Zeugnisse nicht zählen, hat jeder auch etwas potentiell Wissenswertes anzubieten. Niemand ist vom Austausch ausgeschlossen. Die Welt ist eine virtuelle Agora, die Menschheit ein kollektives Cogito bzw. ein *cogitamus*. Der Mensch erwirbt eine Identität des Wissens, die anschlussfähig an Andere ist. Wissen ist das, was digitalisier- und im Netz kommunizierbar ist. Materialität der Welt, Leiblichkeit des Menschen und die Andersheit des Anderen werden allmählich irrelevant (vgl. dazu ausführlich: Schnell 2002). Diese 25 Jahre alte Vision Lévys gehört noch dem posthumanistischen Stadium der KI an, sie ist in diesem Punkt nicht mehr aktuell. Sie liefert allerdings die ethisch-ontologischen Aspekte des Cyberspace, die heute für nichttriviale Kommunikationen mit einer KI wie GPT-3 in Anspruch genommen werden.

»Computer können juristische Entscheidungen treffen und psychiatrische Urteile fällen. Sie können auf viel ausgesuchtere Weise Münzen werfen als das geduldigste menschliche Wesen.« (Weizenbaum 1978, 299) Tragen sie in diesen Fällen auch die Verantwortung für diese Entscheidungen und Handlungen? Wer haftet? Wer trägt die Schuld? Die Maschine? Ihr Eigentümer? Der Hersteller? Das Produkthaftungsgesetz? Weizenbaum folgert aus all dem, dass »die relevanten Probleme weder technischer noch mathematischer, sondern ethischer Natur sind.« (ebd., 300)

Begegnungen mit Künstlicher Intelligenz – Querschnitte

Hans Blumenberg hat die Verbindung der Frage nach der Technik mit der Problematik der Wahrheit untersucht. Demnach herrscht in der europäischen Antike die Auffassung vor, dass die Wahrheit in der Natur der Dinge fix vorhanden ist. Technik ist demgegenüber das bloß Künstliche und Scheinhafte. Im Idealfall ahmt es die Natur nach (ars imitatur naturam). »Wer ein Haus baut, tut genau das, was die Natur tun würde, wenn sie Häuser sozusagen wachsen ließe.« (Blumenberg 2015, 86f.) Mit der Neuzeit wird die Natur zum bloßen Rohstoff, wie es bei Locke heißt, so dass die Wahrheit allmählich im Bereich des Selbstgemachten zu finden ist. In der jüngeren Gegenwart setzen Autoren wie Baudrillard in der Wahrheitsfrage ganz auf die Mächte der Simulation (vgl.: Schnell 2020, 184ff.).

Aktuelle Künstliche Intelligenz hat die Eigenart, sich dem anzunähern, was man als einen lebensweltlichen Akteur bezeichnen kann. Damit ist jedes Wesen gemeint, das mit Anderen in der Welt existiert und dabei einen Selbst-, Welt- und Fremdbezug realisiert (vgl. dazu: Schnell 2017, 134ff.). Zu diesen Wesen zählen Menschen, vermutlich die meisten Tiere und vielleicht auch gewisse Pflanzen. Des Menschen Gehirn

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

oder seine Gene sind keine lebensweltlichen Akteure. Weil sich viele Künstliche Intelligenzen hingegen dem Charakter des lebensweltlichen Akteurs annähern und es daher zu Begegnungen mit dem Menschen kommt, verbinden sich in der Diskussion drei Querschnitte miteinander: Intersubjektivität, Technik und Lebenswelt. Auf die Wahrheitsgehalte dieser drei Querschnitte und auf ihre Beziehungen untereinander haben die Begegnungen mit Künstlicher Intelligenz deutliche Auswirkungen.

Aufbau des Bandes

Begegnungen mit Künstlicher Intelligenz produzieren Selbstverständlichkeiten und Normalitäten, die als solche zu unbefragten Beständen der Lebenswelt werden können. Die Beiträge des Bandes untersuchen Formen und Inhalte dieser Begegnungen. Was bedeutet es, mit einer Maschine zu musizieren, sich von einem Roboter pflegen und berühren zu lassen, mit einer KI ein intelligentes Gespräch zu führen?

Der Band wird eröffnet mit einem Beitrag *Oliver Müllers*, der sich der Frage nach *maschineller Alterität* widmet. Anknüpfend an die Dialektik aus Andersheit und Ähnlichkeit, mit der der Mensch kulturgeschichtlich sein maschinelles Gegenüber imaginiert und anthropologisch zu definieren versucht hat, fragt der Beitrag danach, welche spezifische Andersheit und welche Ähnlichkeiten der, im Ausgang von Levinas und Ricœur »alteritätsbedürftige« Mensch in der Maschine finden kann.

Burkhard Liebsch analysiert in seinem Beitrag *Kommunikative Verwerfungen zwischen digital natives und analogen Aborigines*. Im Spannungsfeld des Bedürfnisses als Partner eines Dialogs nicht restlos in seinem digitalisierbaren Bild aufgehen zu wollen und der Verlegenheit anzugeben, worin dieser Rest bestehen möge, gilt es die Technik weder von vorneherein zu verteufeln noch Kommunikation von Anfang an technizistisch zu betrachten, sondern vielmehr einen richtigen Gebrauch der Technik zu erlernen.

Jonathan Hart und *Martin Feißt* werfen in ihrem Beitrag einen soziologischen Blick auf ihre Selbsterfahrungen in der Begegnung mit dem Programm GPT-3. Dass das Programm »ganz losgelöst von verkörperten, wahrnehmungsgeleiteten und emotionsbegleiteten Entitäten« (S. 70) Sprachverständnis und Intelligenz zeige, müsse die Soziologie nicht überraschen, sei diese doch gerade mit dem Paradigma angetreten, Soziales durch Soziales erklären zu können. Der alteritätsbedürftige Mensch ist soziologisch gesehen vor allem ein Interaktionssucher, und Programme wie GPT-3 sind intelligent genug, um ihm diese Interaktion zu bieten und ihn so am Sozialen teilhaben zu lassen, also soziale

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Relationen zu bilden, die der Frage danach, was die Relata denn nun seien, vorgängig sind.

Martin W. Schnell und *Lukas Nehlsen* gehen in dem darauffolgenden Beitrag ebenfalls auf empirisches Gesprächsmaterial mit GPT-3 als soziale Tatsache ein. Mithilfe der qualitativen Inhaltsanalyse wird ein Gespräch mit dem Programm über Liebe analysiert. Das Gespräch zeigt sich dabei vor allem als Meinungsaustausch, der (noch) nicht die Ebene eines schöpferischen Dialogs erreicht.

In seinem Einzelbeitrag knüpft *Lukas Nehlsen* daran an und widmet sich sprachphilosophisch dem Phänomen des Gesprächs mit einer Chatbot-KI. Erlangt Sprache, wie Maurice Merleau-Ponty behauptet, nur durch leiblichen Ausdruck ihren Sinn, wie kann dann die Äußerung einer körperlosen KI trotzdem als sinnvoll wahrgenommen werden?

Martin W. Schnell geht in seinem Beitrag von Bubers Diktum, der Mensch werde erst am Du zum Ich, aus und fragt danach, was passiert, wenn in der Ich/Du-Beziehung eines der Elemente durch einen Roboter ersetzt wird. In Rekapitulation der Anerkennungsthematik in der Folge von Hegel und in Auseinandersetzung mit Ian McEwans Roman *Maschinen wie ich* fragt Schnell danach, ob es möglich ist, einen Roboter als meinesgleichen anzuerkennen.

Renate Buschmann eröffnet mit ihrem Beitrag den Teil des Bandes, der sich mit Begegnungen mit KI auf dem Feld der Kunst und Kultur auseinandersetzt. Sie rekapituliert diverse künstlerische Projekte, in denen KI nicht so sehr als eigenständig kreativer Akteur wichtig wird, sondern eher als Kooperationspartner des Menschen, sowohl in Gestalt von Künstler:innen als auch von Rezipient:innen. Die interessante Frage sei daher eher nicht »Kann KI Kunst?«, sondern vielmehr »Kann aus KI Kunst werden?« (S. 165).

Christian Grünys Beitrag wirft einen philosophischen Blick auf diverse Begegnungen mit KI im Umkreis der Musik. Anknüpfend an Ihdes oben genannte Unterscheidung der Mensch-Technik-Relationen verfolgt Grüny diese sowohl im Kompositions- als auch im Improvisationsprozess. Dabei zeigt sich unter anderem, dass der Einbruch von Maschinen in die Musik nicht der »Einbruch von etwas ganz anderem in einen ansonsten homogenen und vor allem zutiefst und ausschließlich *menschlichen* Schaffensprozess« (S. 179) ist und dass es nicht die Gleichheit mit dem kreativen Können des Menschen sei, die die Maschinen für die Musik interessant mache, sondern gerade ihre Seltsamkeit.

Die letzten beiden Texte behandeln Begegnungen mit Künstlicher Intelligenz auf dem Feld der Gesundheit, besonders der Pflege. *Alexander Hochmuth* gibt in seinem Beitrag einen Überblick über Einsatzbereiche von KI in der Pflege und fragt dabei, wie der Einsatz von KI mit den Prinzipien eines guten pflegerischen Handelns, insbesondere im Umgang mit Menschen mit demenziellen Erkrankungen, vereinbar sei.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Erik Dzwiza-Ohlsen wirft im letzten Beitrag des Bandes einen leibphänomenologischen Blick auf das konkrete Beispiel des mit KI ausgestatteten »Kuschelroboters« PARO und seiner Rolle in der Lebenswelt von Menschen mit Demenz.

Literatur

Baumann, Zygmunt (1995): *Postmoderne Ethik*, Hamburg: Edition Hamburg.

Blumenberg, Hans (1989): *Höhlenausgänge*, Frankfurt am Main: Suhrkamp.

Blumenberg, Hans (2015): *Schriften zur Technik*, Berlin: Suhrkamp.

Boden, Margaret (2018): *Artificial Intelligence. A Very Short Introduction*. Oxford/New York: Oxford University Press.

Buber, Martin (1983): *Ich und Du*, Heidelberg: Lambert Schneider.

Bourdieu, Pierre (1989): *Satz und Gegensatz*, Berlin: Wagenbach.

Dreyfus, Hubert L. (1992): *What Computers Still Can't Do*. Cambridge, MA: MIT Press.

Feuerbach, Ludwig (1983): *Grundsätze der Philosophie der Zukunft*, Frankfurt am Main: Klostermann.

Freud, Sigmund (1983): *Vorlesungen zur Einführung in die Psychoanalyse*, Frankfurt/M: Fischer.

Gurwitsch, Aaron (1977): *Die mitmenschlichen Begegnungen in der Milieuwelt*, Berlin/New York: de Gruyter.

Hare, Richard M. (1993): *Essays on Bioethics*, Oxford: Clarendon Press.

Haugeland, John (1985): *Artificial Intelligence. The Very Idea*. Cambridge, MA: MIT Press.

Husserl, Edmund (1976): *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie,* Den Haag: Nijhoff.

Ihde, Don (1990): *Technology and the Lifeworld. From Garden to Earth*. Bloomington: Indiana University Press.

Levinas, Emmanuel (1987): *Totalität und Unendlichkeit*, Freiburg/München: Karl Alber.

Levinas, Emmanuel (1995): *Zwischen uns. Versuche über das Denken an den Anderen*, München: Hanser.

Löwith, Karl (2013): *Das Individuum in der Rolle des Mitmenschen*, Freiburg/München: Karl Alber.

Malabou, Catherine (2019): *Morphing Intelligence. From IQ Measurements to Artificial Intelligence*. New York: Columbia University Press.

McCarthy, John; Minsky, Marvin, et al. (1955): *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Online verfügbar unter: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (letzter Abruf: 16.02.2022).

Newell, Allen; Simon, Herbert A. (1976): »Computer Science as Empirical Enquiry: Symbols and Search«, *Communications of the Association for Computing Machinery* 19 (März 1976).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Priddat, Birger P. (2022): »Wie werden wir lernen uns zu Automaten zu verhalten?«, in: Priddat, B. P. / Kovce, P. (Hg.) (2022): *Selbstverwandlung. Das Ende des Menschen und seine Zukunft*. Marburg: Metropolis.
Rosenzweig, Franz (1988): *Der Stern der Erlösung*, Frankfurt am Main: Suhrkamp.
Russel, Stuart J.; Norvig, Peter (2021): *Artificial Intelligence: A Modern Approach*. 4. Auflage. Hoboken: Pearson.
Sartre, Jean-Paul (1991): *Das Sein und das Nichts*, Reinbek bei Hamburg: Rowohlt.
Scheler, Max (2000): *Der Formalismus in der Ethik und die materiale Wertethik*, Bonn: Bouvier.
Schnell, Martin W. (1999): »Narrative Identität und Menschenwürde«, in: Breitling, A. et al. (Hg.) (1999): *Das herausgeforderte Selbst*, Würzburg: Königshausen und Neumann.
Schnell, Martin W. (2002): »Ideologie und Anthropologie. Zur Widerkehr des leiblosen Geistes«, in: Greving, H. / Gröschke, D. (Hg.) (2002): *Das Sisyphos-Prinzip*, Bad Heilbrunn: Julius Klinkhardt.
Schnell, Martin W. (2017): *Ethik im Zeichen vulnerabler Personen*, Weilerswist: Velbrück.
Schnell, Martin W. (2020): *Das Ethische und das Politische. Sozialphilosophie am Leitfaden der Vulnerabilität*, Weilerswist: Velbrück.
Schnell, Martin W. / Dunger, Chr. (Hg.) (2019): *Digitalisierung der Lebenswelt. Studien zur Krisis nach Husserl*, Weilerswist: Velbrück.
Singer, Peter (1994): *Praktische Ethik*, Stuttgart: Reclam.
Strawson, Peter F. (1972): *Einzelding und logisches Subjekt*, Stuttgart: Reclam.
Waldenfels, Bernhard (1985): *In den Netzen der Lebenswelt*, Frankfurt am Main: Suhrkamp.
Weizenbaum, Joseph (1978): *Die Macht der Computer und die Ohnmacht der Vernunft*, Frankfurt am Main: Suhrkamp.
Wojcieszuk, M.A. (2010): *Der Mensch wird am Du zum Ich. Eine Auseinandersetzung mit Der Dialogphilosophie des XX. Jahrhunderts,* Freiburg: Centaurus.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Oliver Müller

Maschinelle Alterität

Philosophische Perspektiven auf Begegnungen mit künstlicher Intelligenz

1. Einleitung: Sozialität und Alterität in der Mensch-Maschine-Begegnung

Die Entwicklungen auf dem Gebiet der Künstlichen Intelligenz durchdringen viele unserer Lebensbereiche, von den inzwischen allgegenwärtigen Produktempfehlungen neunmalkluger Algorithmen über die Implementierung von Gesichtserkennungsmethoden in der Öffentlichkeit und den Auswüchsen eines digitalen Überwachungskapitalismus bis hin zu robotischer Unterstützung in Industrie und Militär. Und es steht außer Frage, dass wir nicht nur unseren rechtlichen Rahmen nachjustieren müssen – etwa im Bereich des Datenschutzes oder des Völkerrechts, was ›autonome‹ Waffensysteme betrifft –, sondern dass wir auch ethischer Selbstverständigungsdiskurse bedürfen, etwa in Bezug auf den ›bias‹, also die gesellschaftlichen Vorurteile, die in den ›selbstlernenden‹ maschinellen Verfahren reproduziert werden, oder ganz allgemein in Bezug auf Fragen des guten Lebens, der (globalen) Gerechtigkeit und der Werte des Zusammenlebens in der digitalisierten Welt.[1]

1 Siehe exemplarisch für diese Debatten: Silja Vöneky, Philipp Kellmeyer, Wolfram Burgard, Oliver Müller (eds.): *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*. Cambridge 2022 (i. Ersch.); Shoshana Zuboff: *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs 2019; Kate Crawford: *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven 2021; Patrick Lin, Keith Abney, and Ryan Jenkins (eds.): *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. New York: Oxford University Press 2017; Marco Nørskov (ed.): *Social Robots*, London: Routledge 2017; Luciano Floridi et al.: »AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations«. Minds and Machines, 28(4) (2018): 689–707; David J. Gunkel and Joanna J. Bryson (eds.): »Machine Morality: The Machine as Moral Agent and Patient«. Special issue of Philosophy & Technology, 27(1) (2014): 1–142; Michael Anderson and Susan Leigh Anderson (Hg.): *Machine Ethics*. Cambridge: Cambridge University

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Die philosophische Reflexion künstlicher Intelligenzen hat sich früh auf die Frage kapriziert, ob und inwiefern künstlich-intelligente Systeme den Status von rationalen und/oder moralischen Akteur:innen haben können.[2] So naheliegend eine solche Frageperspektive und die entsprechende Klassifizierungspraxis in Bezug auf den Akteur:innenstatus auch sein mögen, derartige Vorhaben haben auch ihre Tücken. Denn sie tragen dazu bei, dass wir den künstlich-intelligenten Systemen Entscheidungs-, Handlungs- und Verhaltenskompetenzen zuschreiben, die sich an den Fähigkeiten von Menschen orientieren. Mit den Begriffsbildungen um den Akteur:innenstatus behandeln wir die komplexen Algorithmen nicht mehr als nützliche Werkzeuge, sondern statten sie mit menschen- oder personenähnlichen Eigenschaften und Kompetenzen aus. Dazu gehört auch der Begriff der Autonomie, der die künstlichen Intelligenzen mit einer erhabenen Aura versieht und den Blick dafür verstellt, dass wir hier letztlich Anthropomorphisierungen aufsitzen. Denn wie beeindruckend eigenständig die Maschinen auch operieren können, mit der Autonomie, die wir Menschen zuschreiben, hat das wenig zu tun. Auch wenn wir den Akteur:innenbegriff in größtmöglicher analytischer Abstraktion behandeln, bleibt am Ende die Suggestion, die künstlichen Intelligenzen sind eigentlich so wie wir – nur eben etwas anders. Dieser Suggestion verfallen wir auch daher gerne, weil dem Begriff ›künstliche Intelligenz‹ der Assoziationsraum des ›Geistes in der Maschine‹ und ähnlicher (post-) cartesianischer Reflexionstraditionen anhaftet. Das heißt: im Zentrum steht die schaudernde Faszination der Frage, ob die Maschinen einen uns Menschen vergleichbaren Geist haben (können). Dass auf diese Weise allein schon verdeckt wird, dass hinter den scheinbar körperlos geistähnlichen künstlichen Intelligenzen ein auf Ausbeutung (von Rohstoffen und Personen) ausgerichtetes Industriesystem steht, ist dabei eines der gewichtigen ethischen Probleme in diesem Feld (wie Kate Crawford in ihrem *Atlas of AI* eindrucksvoll gezeigt hat).[3] Dies sind (in aller Kürze) einige der Gründe, warum wir dazu neigen, künstliche Intelligenzen anders als andere Technologien zu behandeln – weil wir ihnen gewisse agentielle Fähigkeiten zuschreiben.

Bei aller Skepsis, die wir gegenüber dieser Zuschreibung geistiger und agentieller Fähigkeiten an die intelligenten Maschinen haben mögen,

Press 2007; Sarah Spiekermann: *Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert*. München 2019.

2 Siehe Luciano Floridi and J.W. Saunders: »On the Morality of Artificial Agents«. Minds and Machine, 14 (2004): 349–379; Catrin Misselhorn: Grundfragen der Maschinenethik. Stuttgart 2018, S. 70–90; Johanna Wagner: Künstliche Intelligenzen als moralisch verantwortliche Akteure. Paderborn 2020.

3 Kate Crawford: *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven 2021.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

bleibt die Lage aber verzwickt, wenn wir uns nicht nur die Technik für sich genommen anschauen, sondern die Interaktionen zwischen Menschen und Maschinen in den Blick nehmen. Denn wir gehen mit den Maschinen unweigerlich Beziehungen ein, die wir als ›sozial‹ bezeichnen können. Wir Menschen bilden als *zoa politika* nicht nur mit anderen Menschen und anderen Tieren, sondern immer auch mit Maschinen Gemeinschaften, weil es gewissermaßen zu unserer Natur gehört, zu interagieren, zu kommunizieren und soziale Strukturen aufzubauen und uns so, mit Hannah Arendt gesprochen, in ein »Gewebe« von sozialen Bezügen zu begeben.[4] Und ein solches ›Sozialgewebe‹ können wir auch mit Maschinen weben, nicht nur, aber auch (und vermutlich in besonderer Weise), wenn diese in irgendeiner Form als ›intelligent‹ eingestuft werden.

Sollten wir die Maschinen partout nicht als ›intelligent‹ in einem gehaltvollen Sinne ansehen wollen, bleibt es aber doch dabei, dass sie mitunter so programmiert werden können, dass sie uns in einer mehr oder weniger ›menschenanalogen‹ Verhaltensweise begegnen, auf die wir *als soziale Wesen* reagieren und in komplexer Weise *antworten* können (›Antwort‹ als philosophischer Begriff wird unten näher erläutert). Gegen diese Beobachtung könnte man einwenden, dass wir es hier niemals mit ›echten‹ sozialen Beziehungen zu tun haben und bei derartigen Beschreibungen des Mensch-Maschine-Verhältnisses eher Übertragungen oder sogar (gezielte) Täuschungen vorliegen, die aus solchen ›menschenanalogen‹ Verhaltensweisen von Maschinen noch längst keine ›echte‹ soziale Situation machen.

In diesem Text will ich jedoch von der Hypothese ausgehen, dass einiges dafürspricht, dass wir Menschen unsere sozialen Beziehungen auch auf Unbelebtes ausdehnen können und daher die Grenzen von Mensch-Mensch-, Mensch-Tier-, Mensch-Ding- und Mensch-Maschine-Begegnungen graduell zu verstehen sind und all diese Sphären zusammengenommen ›die‹ Sozialität des menschlichen Lebens ausmachen. Um dies zu verdeutlichen, will ich ein Argument von Mark Coeckelbergh aufgreifen. In einem seiner Texte diskutiert der Technikphilosoph die Interaktionen zwischen Menschen und Robotern im Blick auf den Begriff des Vertrauens und vertritt die These, dass es keineswegs absurd ist, in diesem Kontext den Begriff der sozialen Beziehung zu verwenden.[5] Coeckelbergh bezeichnet das Vertrauen in diesem Kontext als ›*default mode*‹[6], also als eine Art ›Voreinstellung‹ oder ›Standardeinstellung‹, in der wir den Robotern begegnen und dabei immer schon eine soziale

4 Hannah Arendt: *Vita activa oder vom tätigen Leben*. Piper: München 2002, S. 222ff.

5 Mark Coeckelbergh: »Can we trust robots?«, Ethics Inf Technol 14 (2012):53–60.

6 Ebd., S. 56.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Beziehung zu ihnen voraussetzen, wenn wir mit ihnen interagieren. Daher können wir mit gewissen Recht sagen, dass wir Menschen als *zoa politika* so disponiert sind, dass wir selbst mit Maschinen eine wie auch immer geartete soziale Beziehung eingehen – auch wenn wir es vielleicht nicht einmal wissen oder wahrhaben wollen. Diese Art von sozialer Beziehung mag sich von derjenigen mit anderen Menschen unterscheiden, kann aber aufgrund des oben genannten Sozialgewebes nichtsdestotrotz als eine soziale Beziehung verstanden werden.

Doch auch wenn dem so sein sollte, kommen wir nicht umhin, hier philosophisch genauer hinzuschauen, um Beschreibungsformen und Verstehenshorizonte zu entwickeln, die das Eigentümliche dieser Art sozialer Beziehungen erfassen können. Ich will in diesem Text dazu beitragen, indem ich die spezielle Form von ›Alterität‹ untersuche, die die Mensch-Maschine-Interaktionen charakterisieren kann. Dass der Begriff ›des Anderen‹[7] in der Einschätzung von künstlich-intelligenten Systemen hilfreich sein kann, wurde schon verschiedentlich bemerkt. David J. Gunkel spricht in seinem Buch *The Machine Question* etwa von ›*machinic others*‹ und beschreibt im Rückgriff auf Emmanuel Lévinas verschiedene Formen einer ›*machinic otherness*‹.[8] Und bereits in früheren philosophischen Ansätzen wurde die Bedeutung der Alterität in der Technikreflexion hervorgehoben. Um nur ein Beispiel zu nennen: In seiner postphänomenologisch ausgerichteten relationalen Ontologie identifiziert der Philosoph Don Ihde neben ›*embodiment relations*‹, ›*hermeneutic relations*‹ und ›*background relations*‹ auch ›*alterity relations*‹.[9] Mit dem (ebenfalls von Lévinas inspirierten) Alteritätsbegriff lotet er aus, ob und inwiefern wir von ›*technology-as-other*‹ oder auch von einer ›*quasi-otherness*‹ in der Begegnung mit Maschinen sprechen können.[10]

Ihdes tentativ-vorsichtiger Begriff der ›*quasi-otherness*‹ zeigt jedoch an, dass wir es hier möglicherweise mit einem tieferen philosophischen Problem zu tun haben. Denn ist es wirklich sinnvoll, von einer *Quasi*-Andersheit zu sprechen? Was soll das ›quasi‹ genau zum Ausdruck bringen? Ist damit eine Situation bezeichnet, in der wir mit einem Roboter oder

7 Die Rede von ›dem Anderen‹ hat sich in der Phänomenologie etabliert, auch wenn die im Singular stehende maskuline Form nicht unproblematisch ist. Im Folgenden werde ich daher diesen Begriff immer wieder in Anführungszeichen setzen oder alternativ durch einen Plural oder die weibliche Form ersetzen. Auch den Begriff ›das Andere‹, der in dieser Auseinandersetzung mit der maschinellen Alterität naturgemäß eine zentrale Rolle zukommt, wird immer wieder Verwendung finden.

8 Siehe David J. Gunkel: *The machine question. Critical perspectives on AI, robotics, and ethics*. Massachusetts 2017, S. 197ff.

9 Don Ihde: *Technology and the Lifeworld. From Garden to Earth*. Indiana 1990.

10 Ebd., S. 97ff.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

einer anderen interaktiven Maschine nur so tun ›als ob‹? Sprechen wir also von einer Situation, in der wir nur so tun, als ob die Maschine nicht nur ein Werkzeug, sondern ein ›Anderes‹ wäre (in der wir in der Maschine ein über den einfachen Technologiegebrauch hinausgehendes ›Gegenüber‹ sehen), das sie aber ›eigentlich‹ gar nicht ist? Oder soll mit dem ›quasi‹ ein Moment der Täuschung erfasst werden und die Maschine würde etwas zu sein vorgeben, was sie gar nicht ist? In entgegengesetzter Richtung könnten wir aber auch fragen: wenn es so sein sollte, dass wir die Maschine bereits in ihrer ›Quasi-Andersheit‹ verstehen und uns entsprechend verhalten – könnten wir dann das ›quasi‹ nicht einfach problemlos streichen, weil es sich ohnehin bereits um eine echte Form von Andersheit handelt?

Wenn uns das Streichen des ›quasi‹ in der Quasi-Andersheit allerdings spontan Unbehagen bereiten sollte – vielleicht weil es lebensweltlich und im alltäglichen Verständnis offenkundig zu sein scheint, dass wir Menschen und Maschinen *clare et distincte* zu unterscheiden wissen –, könnten wir in einer ontologisch präzisen Klassifizierung übersehen, dass wir Menschen offenbar dazu neigen, mit Maschinen und anderen Dingen in einer Weise zu interagieren, die es durchaus erlaubt, den Begriff der Alterität zu verwenden. Meines Erachtens ist dies aber nur unter der Voraussetzung möglich, dass wir den Blick nicht mehr auf den agentiellen Status der Maschine richten, sondern uns das menschliche Verhalten gegenüber den Maschinen ansehen. Die Andersheit von Maschinen ist, wie mir scheint, grundlegend falsch verstanden, wenn wir fragen: können Maschinen ›als solche‹ in vergleichbarer Weise wie andere Menschen ›anders‹ sein? Sinnvoller (und philosophisch fruchtbarer) scheint mir folgende Frage zu sein: was macht das spezifische Anderssein von Maschinen mit uns ›alteritätsbedürfenden‹ (*sit venia verbo*) oder sogar ›alteritätsbedürftigen‹ Menschen?

Wir haben es in unseren Interaktionen selbstverständlich mit verschiedenen Arten von Maschinen zu tun. Fragen der Alterität scheinen vor allem bei humanoiden Robotern oder bei sehr komplexen maschinellen Systemen zu finden zu sein, die mit künstlich-intelligenten Algorithmen ausgestattet sind und die sich in einer Weise ›von selbst‹ verhalten, dass wir aufgrund ihrer ›Menschenähnlichkeit‹ in eine soziale Situation geraten. Weniger komplexe Maschinen scheinen Alteritätsfragen eher nicht aufzuwerfen (auch wenn wir in bestimmten Kontexten von einem sozialen Gewebe sprechen können, das wir mit ihnen eingehen). Auf eine mögliche Differenzierung von maschinellen Alteritätsformen will ich in diesem Text nicht eingehen. Mir geht es in diesem Beitrag vielmehr darum, verschiedene Momente einer alteritätstheoretischen Perspektive auf die Mensch-Maschine-Interaktion zu erörtern. An Paul Ricœurs Überlegungen in *Das Selbst als ein Anderer* anschließend, können wir sagen, dass Maschinen zur ›Wer-Frage‹ von Personen (im

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Sinne der *ipse*-Identität) beitragen können und damit zum Thema einer *Hermeneutik des Selbst*, wie sie Ricœur vorschwebte,[11] gemacht werden können. Ricœur selbst hat die Alterität der Maschinen allerdings vernachlässigt. Das Feld der Vieldeutigkeit, das Ricœur mit der Dialektik zwischen Selbstheit und Andersheit erschlossen hat, kann aber durchaus um die verzwickten Erfahrungen mit maschineller Andersheit erweitert werden, wie in diesem Beitrag gezeigt werden soll.

Nach diesen Vorüberlegungen will ich im Folgenden dem Phänomen maschineller Alterität nachgehen. Dabei werde ich zunächst kurz auf ›alte‹ anthropologische Selbstverständigungsdiskurse verweisen, die wir im Blick auf die Maschinen schon seit der Antike zu verzeichnen haben (2.). Dann werde ich Gehlens Formulierung der Maschine als das ›Andere-als-Menschliche‹ als ein Moment der Mensch-Maschine-Begegnung herausstellen, in dem das Anderssein der Maschine in der philosophischen Anthropologie zum Thema wurde (3.), um dann auf einige aufschlussreiche Alteritätsmomente in der frühen KI-Entwicklung einzugehen (4.). In einem nächsten Schritt will ich mich mit dem von Wilhelm Schapp entwickelten phänomenologischen Ansatz auseinandersetzen, der davon ausgeht, dass wir uns selbst begegnen können, weil wir immer schon in Geschichten mit anderen und anderem verstrickt sind, und dies im Blick auf die Maschinengeschichten, die wir uns erzählen, diskutieren (5.). Daran anschließend werde ich schließlich mit der Responsivität (6.) und dem ›Anspruch des Anderen‹ (7.) zwei zentrale Theoriemomente der Phänomenologie der Alterität (insbesondere bei Bernhard Waldenfels und Lévinas) in Bezug auf die ›Maschinenfrage‹ fruchtbar zu machen suchen. Am Ende steht eine zusammenfassende Schlussbemerkung (8.).

2. Begegnungen mit (auffälligerweise häufig weiblichen) Maschinen

Maschinen wurden in der europäischen Kulturgeschichte als schillernde Gegenüber imaginiert, bei denen gerade die Kippmomente zwischen Ähnlichkeit und Andersheit ihren Reiz ausmachen.[12] Bereits in der griechischen Antike wurde die Maschine als etwas begriffen, was sich

11 Siehe Paul Ricœur: *Das Selbst als ein Anderer*. München 1996, S. 21ff, S. 26ff.

12 Siehe allgemein zu den Selbstverständigungsdiskursen in der abendländischen Kulturgeschichte Oliver Müller und Kevin Liggieri: »Mensch-Maschine-Interaktion seit der Antike: Imaginationsräume, Narrationen und Selbstverständnisdiskurse«. In: Kevin Liggieri, Oliver Müller (Hg.): *Mensch-Maschine-Interaktion. Geschichte-Kultur-Ethik* (Metzler-Handbuch). Stuttgart 2019, S. 3–14.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

einerseits deutlich von der menschlichen Existenz unterscheidet, andererseits aber viele Züge menschlichen Verhaltens imitieren kann. Die Maschine wurde daher sowohl als etwas imaginiert, das für höchste technische Perfektion steht – weil sie sich am Menschen als dem ›Maß der Maschine‹ orientiert –, das aber gerade deshalb auch potentielle Täuschung sein kann: Was ein Mensch zu sein scheint, ist eigentlich (nur) eine Maschine. So erzählt der erste Erzähler Homer bereits von roboterähnlichen androiden, genauer gynoiden Statuen, die der antike Techniker Hephaistos hergestellt hat. Diese können sich nicht nur ganz von selbst bewegen, was im Griechischen »*automatoi*« heißt, sondern sind zudem in der Lage, zu sprechen und Arbeiten zu verrichten.[13] Polybios wiederum berichtet von der künstlichen Frau Agepa, die so täuschend echt aussah, dass die sich ihr nähernden Männer erst zu spät erkannten, dass es sich nicht um eine Frau aus Fleisch und Blut, sondern um ein übles Folterwerkzeug handelte.[14] Es ist bemerkenswert, dass ›das Andere‹, das an der Maschine fasziniert oder irritiert, seit Beginn des Nachdenkens über die Maschine oftmals mit Stereotypen von Weiblichkeit verschränkt zu sein scheint. Das Fremde der Maschine wird durch Feminisierung vertraut gemacht – und bleibt möglicherweise gerade deshalb fremd. Es scheint fast so, als sei das Andere-als-Männliche gleichermaßen von der Maschine und ›der Frau‹ verkörpert – so dass sich eine auffallende Überlagerung von Maschinenfiktionen und Geschlechterrollenvorstellungen bereits in den allerersten literarischen Dokumenten findet.

Die Faszination an Android:innen hat seither nicht abgenommen, wie nicht nur die Science Fiction beweist, denn Automatenvisionen und Maschinenimaginationen durchziehen die Kulturgeschichte in verschiedenen Formen.[15] Eine der berühmtesten Maschinenfrauen ist Olimpia aus E.T.A. Hoffmanns Erzählung *Der Sandmann*. Hoffmann setzt ebenfalls bei der Erfahrung von der Ähnlichkeit und Andersheit in der (männlichen) Interaktion mit Olimpia an und spielt literarisch mit der Vertrautheit und Fremdheit, die in der Mensch-Roboter-Begegnung heute häufig als das *uncanny valley* bezeichnet wird. Auch Olimpia ist Menschen irgendwie ähnlich, irgendwie auf eine unheimliche Weise aber auch nicht. Hoffmann beschreibt seine maschinelle Protagonistin in einer Szene wie folgt: »Olimpia erschien sehr reich und geschmackvoll gekleidet. Man mußte ihr schöngeformtes Gesicht, ihren Wuchs bewundern. Der etwas

13 Homer: *Odyssee*. In der Übertragung von Johann Heinrich Voß. Frankfurt am Main 1990, Gesang 18, V. 373–377.

14 Siehe Klaus Völker (Hg.): *Künstliche Menschen. Dichtungen und Dokumente über Golems, Homunculi, Androiden und liebende Statuen*. München 1976, S. 63.

15 Siehe Käte Meyer-Drawe: »Androiden«. In: Carlos Kölbl und Anna Sieben (Hg.): *Stichwörter zur Kulturpsychologie. Jürgen Straub zum 60. Geburtstag*. Gießen 2018, S. 15–20.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

seltsam eingebogene Rücken, die wespenartige Dürre des Leibes schien von zu starkem Einschnüren bewirkt zu sein. In Schritt und Stellung hatte sie etwas Abgemessenes und Steifes, das manchem unangenehm auffiel; man schrieb es dem Zwange zu, den ihr die Gesellschaft auferlegte. Das Konzert begann. Olimpia spielte den Flügel mit großer Fertigkeit und trug ebenso eine Bravour-Arie mit heller beinahe schneidender Glasglockenstimme vor.«[16] Hoffman flicht in seine Beschreibung der Olimpia die Irritationen ein, die diese in den Augen der Betrachter:innen auslöst, ohne dass von jenen erkannt wurde, dass es sich bei ihr nicht um ein lebendiges Wesen, sondern um eine Automatenfrau handelt. Auf der einen Seite entspricht ihr Auftritt der gesellschaftlichen Konvention und dem Frauenbild der Zeit, auf der anderen Seite hinterlässt sie aber den Eindruck des Fremdartigen, Unvertrauten, des unbestimmt von der Norm abweichenden. Hoffmann wählt dafür ein Vokabular, das das Mechanische insinuiert (»seltsam eingebogen«, »etwas Abgemessenes und Steifes«, »beinahe schneidende Glasglockenstimme«), auch wenn sie von der Festgesellschaft nicht als etwas Mechanisches, sondern nur als etwas leicht Irritierendes, nicht recht Einzuordnendes bemerkt wurde, das, wie er schreibt, »manchem unangenehm auffiel«.

Bekanntlich wird das Geheimnis der Olimpia am Ende doch noch gelüftet: Ein junger Mann namens Nathanael verliebt sich in die bezaubernde Maschinenfrau, obwohl sie in den intimen Zwiegesprächen lediglich »ach!« sagen kann,[17] was dem romantischen Schwärmer aber nicht weiter auffällt. Doch schließlich muss Nathanael mit Schrecken feststellen: »Olimpias toderbleichtes Wachsgesicht hatte keine Augen, statt ihrer schwarzen Höhlen; sie war eine leblose Puppe.«[18] Die Romantik hat das ›Andere der Vernunft‹ in vielfacher Weise zum Thema gemacht, sei es in Bezug auf das Triebhafte, Unterbewusste, Weibliche oder Wahnsinnige.[19] Hoffmann verbindet die Themen des Wahnsinns und der als bedrohlich erfahrenen Weiblichkeit mit dem Maschinendiskurs der Neuzeit und wendet damit gewissermaßen die cartesianische Automatenbegeisterung in ihr Gegenteil. Die sich als bloße Apparatur entpuppende geliebte Frau wird zur Horrorvorstellung. Die Maschine steht nicht mehr für wissenschaftlich-technischen Triumph und Ingenieurskompetenz, sondern für eine elementare ontologische Verunsicherung.

Für das Verständnis des Maschinendiskurses ist die Figur der Olimpia interessant, weil die Konstruktion von Automaten mit den mechanisti-

16 E.T.A. Hoffmann: *Der Sandmann*. Stuttgart 1991, S. 29.
17 Ebd., S. 31.
18 Ebd., S. 36.
19 Siehe zur Gedankenfigur des ›Anderen der Vernunft‹ Hartmut Böhme und Gernot Böhme: *Das Andere der Vernunft. Zur Entwicklung von Rationalitätsstrukturen am Beispiel Kants*. Frankfurt am Main 1983.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

schen Erklärungsmodellen der Natur zusammenhängen. Die Automaten zeigen, dass wir die Natur nachbauen können, gerade weil wir ihren Geheimnissen mechanistisch auf die Schliche gekommen sind. Bekanntlich ist es Julien Offray de LaMettrie, der mit seinem Buch *L'homme machine* den mechanistischen Diskurs der Neuzeit auf die Spitze treibt.[20] Während seine Mensch-Maschine-Gleichsetzung einen provokanten, aber letztlich ›klassisch‹ aufklärerischen Erklärungsrahmen bietet, in dem wir Menschen uns unserer mechanistisch vollständig erklärbaren ›Natur‹ selbst versichern sollen, führt Hoffmann mit Olimpia das Horride vor, das die mechanistische Reduktion des Lebendigen auf seine Automatenhaftigkeit auslösen kann. Das maschinell Andere wird zum unheimlichen Doppelgänger unserer selbst.

3. Die Maschinen als das ›Andere-als-Menschliche‹

Das Horride ist aber nur ein Extremfall maschineller Alteritätserfahrung. Oftmals wird die Maschine als etwas wahrgenommen, was uns an die ›mechanischen‹ Vorgänge in unserem Organismus erinnert, ohne dass wir unsere menschliche Existenz darauf notwendigerweise reduzieren müssten. Diesbezüglich sind bei Arnold Gehlen einige Überlegungen zu finden, die das Mensch-Maschine-Verhältnis mit dem Begriff des ›Anderen‹ zu erfassen suchen, wiewohl sich sein Begriff des Anderen von den phänomenologischen Alteritätskonzepten unterscheidet (auf die unten näher eingegangen wird). Nichtsdestotrotz kann das von ihm beschriebene Alteritätsphänomen als eine der phänomenologischen Beschreibungen vorgelagerte Erfahrung verstanden werden, da in Gehlens Ansatz ›das Andere‹ gewissermaßen ›dem Anderen‹ vorausgeht.

Gehlen setzt dabei an, dass wir Menschen uns vor allem in dem erkennen können, was anders als wir selbst ist und uns gerade deshalb zur Selbstbefragung herausfordert: »Bedrängt von der Rätselhaftigkeit seines Daseins und seines eigenen Wesens ist der Mensch schlechthin darauf angewiesen, seine Selbstdeutung über ein Nicht-Ich heranzuholen, über ein Anderes-als-Menschliches. Sein Selbstbewußtsein ist indirekt, seine Bemühung um eine Eigenformel verläuft immer so, daß er sich mit einem Nichtmenschlichen gleichsetzt und in dieser Gleichsetzung wieder unterscheidet.«[21] Dieses ›Anderes-als-Menschliche‹, das sich traditionell in den verschiedenen ›anthropologischen Differenzen‹ zeigen kann, umfasst nicht nur Pflanzen und Tiere (wie sie bekanntlich in den Anthropologien von Max Scheler und Helmuth Plessner ausgearbeitet wurden),

20 Julien Offray de LaMettrie: *Der Mensch eine Maschine*. Stuttgart 2001.
21 Arnold Gehlen: *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme der industriellen Gesellschaft*. Hamburg 1957, S. 16.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

sondern eben auch die Maschinen. Die Auseinandersetzung mit der Maschine kann damit Antwort(en) auf die anthropologische Grundfrage ›Was ist der Mensch?‹ bieten.[22]

Der Maschine haftet damit eine Ambivalenz an: Sie ist bei Gehlen nicht nur jenes Andere-als-Menschliche, sondern wir erkennen uns in den Maschinen auch selbst wieder. Die Maschinen bleiben ›Kippapparaturen‹, wenn man so will. Als das Andere-als-Menschliche zeigen sie uns gerade auch zutiefst Menschliches – und zwar Menschliches, das wir oft zu übersehen neigen, nämlich z.B. die unbewussten physischen Prozesse. So notiert Gehlen – in einer Zeit, in der die Kybernetik als eine neue Leitwissenschaft in Anschlag gebracht wird –: »Nun ist aber der Mensch in der Tat in ganz zentralen Bereichen seiner Natur Automatismus, er ist Herzschlag und Atmung, er lebt geradezu in und von sinnvoll funktionierenden, rhythmischen Automatismen, wie sie in der Bewegung des Gehens, vor allem aber in den eigentlichen Hantierungen und Arbeitsgängen der Hand vorliegen, in dem ›Handlungskreis‹, der über Sache, Hand und Auge zur Sache zurücklaufend sich erschließt und dauernd wiederholt.«[23] Gehlen betrachtet die Maschine also als etwas gleichzeitig Menschliches und Nicht-Menschliches. Gerade dieses Changieren zwischen Identität und Differenz macht die Maschine zu einem anthropologischen Resonanzphänomen. Die Maschinen bringen also nicht bloß Entfremdungserfahrungen hervor, sondern bilden auch einen Resonanzraum, in dem wir uns selbst begegnen und Weltverhältnisse aufbauen können.[24] Die Erfahrung des ›maschinell Anderen‹ steht in einer ganz elementaren ›Resonanzbeziehung‹ zu uns Menschen.[25]

Diese wenigen Stichworte mögen verdeutlichen, dass das Mensch-Maschine-Verhältnis von jeher kniffliger ist, als wir es üblicherweise denken. Daher ist es kein Wunder, dass dies auch auf die Entwicklungen der Künstlichen Intelligenz abgefärbt hat. Das berühmte *Dartmouth Summer Research Project on Artificial Intelligence* von 1956 ging davon aus »that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.«[26] In dem Antrag zu diesem Research Project findet sich

22 Dies ist die anthropologische Grundfrage in ihrer klassischen Formulierung. Es gibt aber gute Gründe die Frage anders zu stellen, um problematische Essentialisierungen zu vermeiden, wie etwa: ›Wer sind wir Menschen?‹.

23 Arnold Gehlen: *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme der industriellen Gesellschaft*. Hamburg 1957, S. 16.

24 Siehe Oliver Müller: *Selbst, Welt und Technik. Eine anthropologische, geistesgeschichtliche und ethische Untersuchung*. Berlin, New York 2014.

25 Siehe zum Resonanzbegriff Hartmut Rosa: *Resonanz. Eine Soziologie der Weltbeziehung*. Berlin 2016.

26 John McCarthy et al.: »A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence« (August 31, 1955). Zitiert nach AI Magazine 27(4) (2006): 12–14, hier S. 12.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

die programmatische Hypothese formuliert, dass die Maschine eigentlich nur eine andere Verkörperungsform als der menschliche Organismus darstellt, auf die aber letztlich alle Formen menschlicher Intelligenz übertragen werden können. Das Ziel dieses Projekts ist der Entwurf einer Maschine, die die Formen von Intelligenz auszuüben vermag, die bislang Menschen vorbehalten waren (»now reserved for humans«).[27] Abermals wird eine Maschine projektiert, die menschähnlich werden soll – und die idealerweise sogar so menschenähnlich wird, dass sie nicht mehr von einem Menschen unterschieden werden kann.

4. Imitation Games: Alterität in der frühen Entwicklung von Künstlicher Intelligenz

Wenn wir bei LaMettrie die Formel ›der Mensch als eine Maschine‹ finden, so können wir in dem Antragstext der frühen Pioniere der Künstlichen Intelligenz die Parole ›die Maschine als ein Mensch‹ extrahieren. Mechanisierung des Menschen und Vermenschlichung der Maschine gehen also Hand in Hand. Und so ist es vielleicht kein Wunder, dass Alterität bei den allerersten Vorstößen in Richtung künstlicher Intelligenz, die bekanntlich Alan Turing in Angriff genommen hatte, eine zentrale Rolle spielt. Der später so gennannte Turing-Test ist inspiriert von einem damals beliebten Party-Spiel, das ›*imitation game*‹ heißt. Die Partygäste mussten erraten, ob es sich bei der Person, die ihnen, ohne sichtbar zu sein, mittels schriftlicher Notate auf Fragen antwortet, um eine Frau oder einen Mann handelt. Turing entwirft in seiner Variante ein Setting, in dem zu entschlüsseln ist, ob die Antworten von einem Menschen oder von einer Maschine gegeben werden.[28] Im Blick auf die erwähnten Androidinnen ist es, nebenbei gesagt, recht bemerkenswert, dass im Setting des Turing-Tests die Frau gewissermaßen durch die intelligente Maschine ersetzt wurde. Das rätselhaft Andere ist nunmehr die Maschine und nicht mehr das andere Geschlecht.

In der Entwicklung von Künstlicher Intelligenz, so können wir sagen, ging es von Beginn an um alteritätstheoretische Fragen. Ziel ist es nämlich nicht einfach, schnelle und effiziente Rechner zu bauen. Sondern bereits in den allerersten Anfängen werden Forschungsvorhaben entwickelt, die bei der Frage ansetzen, ob das uns begegnende Andere ein Mensch oder eine Maschine ist. Das liegt natürlich trivialerweise daran, dass die menschliche Intelligenz zum Vorbild für die maschinelle Intelligenz

27 Ebd.

28 Alan Turing: »Computing machinery and intelligence«. Mind, 49 (1950): 433–460.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

genommen wird. Aber es bleibt trotzdem auffällig, dass sich die Entwicklung der Künstlichen Intelligenz in einer Imaginationstradition bewegt, die wir bei Homer und E.T.A Hoffmann bereits kennengelernt hatten. Der Reiz der künstlich-intelligenten Maschine liegt darin, täuschend echt menschlich zu sein. Und die Konstruktion einer solchen Maschine kann nur gelingen, wenn diese zu einem ›Gegenüber‹ wird, das mehr ist als ein bloßes technisches Werkzeug, das vielmehr eine Erfahrung von ›Andersheit‹ auslösen soll, das auf einen Menschen verweist. Dass auch hier die Maschinen weibliche Züge tragen können, zeigt das von Joseph Weizenbaum geschriebene Programm, mit dem er 1966 den Turing-Test weiterentwickelte, welches in seinem Setting eine Psychotherapeutin simulieren sollte. Dieses Programm hieß ELIZA,[29] nach Eliza Doolittle aus dem (den antiken Mythos aktualisierenden) Theaterstück *Pygmalion* von George Bernhard Shaw – ein gewissermaßen klassisches Produkt männlichen Verfügungswillens in der Anwendung von Wissenschaft.

So technikkritisch Weizenbaum dies letztlich auch gedacht haben mag, es bleibt bemerkenswert, dass auch hier wieder eine Maschinenfrau erdacht wurde, die die männliche Intelligenz zu täuschen versteht. Die Unruhe über die wahre Identität des Gegenübers, die der Turing-Test verkörpert, wurde auch von Philip K. Dick aufgegriffen, der in seinem Roman *Do Androids Dream of Electric Sheep?* (der bekanntlich die Vorlage für die *Blade Runner*-Filme von Ridley Scott darstellt) einer Maschine einen Auftritt gewährt, die zur Orientierung in einer Welt elementarer ontologischer Verunsicherung eingesetzt werden soll.[30] Denn der Roman spielt in einer Welt, in der zwischen dem Natürlichen und Künstlichen nicht mehr unterschieden werden kann. Die Voight-Kampff-Maschine ist eine Empathie-Test-Apparatur, mit der eruiert werden soll, ob es sich beim Gegenüber um einen Menschen oder eine/n Replikant:in handelt. Die Voight-Kampff-Maschine misst im Narrativ des Romas das Menschlichste an uns Menschen, die Empathiefähigkeit. Allerdings bedarf es in einer Art *contradictio in adjecto* eben gerade einer Maschine, die dieses Menschlichste exakt und verlässlich zu messen in der Lage ist. Auch in dem Film *Ex machina* wird die Idee des Turing-Tests aufgegriffen, und zwar in den Interaktionen zweier Programmierer mit der Androidin Ava, die einmal mehr die Erfahrung des Vertraut-Unvertrauten, das die intelligente Maschine verkörpert, feminisiert – und gerade dadurch zum unkontrollierbar Anderen macht. Denn der Film endet damit, dass sich Ava der männlichen Verfügungsmacht zu entziehen weiß.

29 Joseph Weizenbaum: »ELIZA-A Computer Program For the Study of Natural Language Communication Between Man and Machine«. Communications of the ACM 9 (1966): 36–35.

30 Philip K. Dick: *Do Androids Dream of Electric Sheep?* Weidenfeld & Nicolson, London 2012.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

5. In Maschinengeschichten verstrickt

Der Verweis auf die Erzählungen von E.T.A. Hoffmann und Philip K. Dick erfolgte nicht ohne Grund. Denn wir erfahren ›das Andere‹ oftmals gerade über Narrative. Nicht nur in Ricœurs Philosophie der Alterität spielt die ›narrative Identität‹ eine zentrale Rolle, weil sie die Brüchigkeit, die Widersprüche und Dissonanzen einer Person als ›Lebensgeschichte in Lebensgeschichten‹ verstehbar machen kann, sondern auch in Wilhelm Schapps Buch mit dem schönen Titel *In Geschichten verstrickt*.[31] In den Geschichten, mit denen wir unsere Identität erzählerisch formen, sind wir immer schon mit Tieren, Dingen und anderen Menschen verstrickt, einer lebensweltlichen Verstrickung, der wir gar nicht entkommen können: »Die Geschichte jedenfalls läßt sich nicht als Gegenstand untersuchen, weil etwas Geschichte nur insoweit ist, als ich in die Geschichte verstrickt bin. Dies Verstricktsein läßt sich nicht so aus der Geschichte lösen, daß auf der einen Seite die Geschichte übrigbliebe und auf der anderen Seite mein Verstricktsein oder so, daß die Geschichte überhaupt noch etwas wäre ohne den Verstrickten, oder der Verstrickte noch irgendetwas wäre ohne die Geschichte.«[32] Das heißt, dass wir unser Selbstsein in besonderer Weise beschreiben und verstehen können, wenn wir uns die Geschichten, in die wir unentrinnbar verwickelt sind, näher anschauen und zu deuten beginnen.

Naturgemäß sind zunächst die Mitmenschen privilegiert, was die narrativen Verstrickungen angeht. Aber der ganze erste Teil von Schapps Buch widmet sich dem »Wozuding in der Außenwelt und seine[r] Wahrnehmung«[33] und ist eine Fortführung von Heideggers Analysen des ›Vorhandenen‹ in *Sein und Zeit*.[34] Auch bei Schapp geht es meistens um einfache handwerkliche Techniken und Werkzeuge, aber immerhin kommt das Automobil vor, das am ehesten zu den Maschinen zählen kann, die uns hier interessieren.[35] Da Schapp andere Fragen adressiert, hat er den gemeinsamen Geschichten von Menschen und Maschinen leider kaum Aufmerksamkeit geschenkt. Aber man darf aus seinem Ansatz durchaus folgern, dass Menschen und Maschinen ebenso in Geschichten verstrickt sein können wie Menschen und Menschen oder Menschen und Tiere. Auch Plessner hat die unauflösbare Verstrickung von Menschen und Maschinen unterstrichen. Bei ihm spielt die narrative Verstrickung

31 Wilhelm Schapp: *In Geschichten verstrickt. Zum Sein von Mensch und Ding*. Frankfurt am Main 1985.
32 Ebd., S. 85f.
33 Ebd., S. 11–82.
34 Martin Heidegger: *Sein und Zeit*. Niemeyer: Tübingen 1993, S. 66ff.
35 Wilhelm Schapp: *In Geschichten verstrickt. Zum Sein von Mensch und Ding*. Frankfurt am Main 1985, S. 56ff.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

keine zentrale Rolle, trotzdem bringt er eine elementare Dynamik pointiert zum Ausdruck, die die Geschichten von Menschen und Maschinen generiert: »Von den Maschinen fortlaufen und auf den Acker zurückkehren, ist unmöglich. Sie geben uns nicht frei und wir geben sie nicht frei. Mit rätselhafter Gewalt sind sie in uns, wir in ihnen.«[36]

Es ist unsere narrative Identität, die uns mit unseren Maschinen verbindet, mit denen wir gemeinsame Geschichten erleben. Und in diesen Geschichten begegnen wir Maschinen, denen wir vertrauen; Maschinen, die widerborstig sind; Maschinen, die wir verfluchen; Maschinen, die wir bewundern; Maschinen, die uns träumen lassen; Maschinen, die uns in tragische Situationen bringen können.[37] Wenn das Selbst nach Ricœurs Formel ›ein Anderer‹ ist, dann können wir uns nun erlauben, seine Formel zu erweitern und konstatieren: das Selbst ist (auch) ›ein Anderes‹. Denn wir betrachten uns nicht nur im Spiegel unserer Maschinen, wie die Phänomenologin Käte Meyer-Drawe umfänglich gezeigt hat,[38] sondern wir ›sind‹ im emphatischen Sinne ›wir selbst‹ durch die Maschinen, mit denen wir in gemeinsame Geschichten verwickelt sind und über die wir unsere (*ipse*-)Identität ausbilden. In *Das Selbst als ein Anderer* erwähnt Ricœur übrigens technologische Fiktionen ganz explizit.[39] Er betrachtet technische Träume und Fiktionen allerdings sehr sorgenvoll in einem recht engen Blickwinkel, weil er sich in diesem Kontext mit den *puzzling cases* der analytischen Philosophie auseinandersetzt, in denen es vor allem um Gehirntransplantationen bei Fragen personaler Identität geht. Durch diese Verengung der Perspektive in Bezug auf die technischen Fiktionen verspielt Ricœur einiges narratives Potential, das wir zur Erschließung von maschinellen Alteritätserfahrungen fruchtbar machen können.

Denn es lohnt sich, die Maschinengeschichten alteritätstheoretisch ernst zu nehmen. Um dies erneut mit einigen Verweisen auf Filme zu verdeutlichen: Wir können beispielsweise die so innige wie dystopische Beziehung der sehr schlauen und durchtriebenen und auf ihre Weise bemitleidenswerten künstlichen Intelligenz HAL 9000 mit dem Astronauten Dave Bowman verfolgen, die Stanley Kubrick in *Odyssee im Weltraum* erzählt. Oder wir können in der Miniserie *Maniac* einen intelligenten

36 Helmuth Plessner: *Die Utopie in der Maschine*. In: Ders.: *Gesammelte Schriften*. Bd. 10, hg. v. Günter Dux, Odo Marquard und Elisabeth Ströker. Suhrkamp: Frankfurt am Main 1985, S. 34.

37 Vielleicht spielen sich derzeit die wirklich tragischen Tragödien in unserer modernen technischen Zivilisation in erster Linie zwischen Menschen und Maschinen ab, etwa, wenn wir lernen, die Bombe oder andere Technologien zu lieben…

38 Siehe Käte Meyer-Drawe: *Menschen im Spiegel ihrer Maschinen*. München 1996.

39 Paul Ricœur: *Das Selbst als ein Anderer*. München 1996, S. 184.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Supercomputer kennenlernen, der beginnt, Emotionen zu entwickeln und Beziehungen zu den Menschen in seinem Umfeld aufzubauen. Im Film *Her* begleiten wir wiederum einen melancholischen Mann, der sich in die weibliche Stimme seines Computers verliebt, der im Kontakt zu den einfühlsam programmierten Algorithmen die Brüchigkeit seiner Identität erfährt und schließlich wieder zu sich selbst (zurück) findet. Die Tech-Industrie hat die identitätsstiftende Bedeutung der klugen Maschinen längst erkannt. In einer aktuellen Werbung, in der kurze Szenen verschiedener ›einzigartiger‹ Lebensmodelle gezeigt werden, wird ein neues Smartphone folgendermaßen angepriesen: »We all are unique. And now there's a phone that adapts to you. A phone like you.«[40] Das Versprechen, dass die Maschinen sich so anpassen, dass sie ›wie wir‹ werden, verdeckt die Kehrseite, dass wir uns selbst wiederum an die Maschinen anpassen müssen. Die Maschinen können nur wie wir werden, wenn wir werden wie unsere Maschinen.

Das Interessante an derartigen Filmen ist nicht (bzw. nicht nur), dass sie uns anregen können, über nähere und fernere Technikzukünfte nachzudenken, sondern dass sie Geschichten darüber erzählen, wie tief und unauflöslich wir mit unseren Maschinen bereits in gemeinsame Geschichten verstrickt sind – und wie wir uns in den Begegnungen mit den Maschinen als ›Selbste‹ erfahren können. Einer der Gründe hierfür ist darin zu sehen, dass wir auf Maschinen nicht nur reagieren, sondern auf sie auch antworten können.

6. Maschinen antworten

Das Antwortenkönnen, die Responsivität, ist ein zentrales Moment in Alteritätstheorien, das insbesondere Bernhard Waldenfels im Anschluss an Kurt Goldstein umfänglich untersucht hat.[41] Die Bedingung der Möglichkeit des Selbstseins wird von ihm in der Verschränkung von »Pathos« und »Response« erfasst, also grundlegend alteritätstheoretisch gedacht.[42] Wir Selbste erfahren, erleiden, erleben etwas durch andere und anderes und antworten dann auf dieses Erleiden. Das Antwortenkönnen ist höchst vielschichtig und umfasst mindestens leibliche, sprachliche, soziale, kulturelle und politische Dimensionen. Anthropologisch

40 Siehe https://www.youtube.com/watch?v=VxERxFXPw3M (letzter Zugriff 20.1.22).

41 Siehe vor allem Bernhard Waldenfels: *Antwortregister*. Frankfurt am Main 1994.

42 Siehe etwa Bernhard Waldenfels: *Erfahrung, die zur Sprache drängt. Studien zur Psychoanalyse und Psychotherapie aus phänomenologischer Sicht*. Berlin 2019, S. 254ff.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

hat der ebenfalls an Goldstein anschließende Ernst Cassirer diesen Zusammenhang in seinem *Essay on Man* mit folgender Formel erfasst: *From animal Reactions to Human Responses*.[43] Das Antwortenkönnen ist damit eine menschliche Fähigkeit, die die Reiz-Reaktions-Mechanismen durchbricht. Das Antworten erfolgt verzögert und ändert das antwortende Selbst und ›den anderen‹ oder ›das andere‹ (welches bei Waldenfels auch als das Fremde firmiert) im Prozess des Antwortens. Dies ist als eine Dynamik zu verstehen, in der wir uns im sinngebenden Antworten auf ein unvorhergesehenes Angesprochenwerden sowohl uns selbst als auch ›dem anderen‹ in einer Weise begegnen, dass wir weder das Eigene zum Maßstab für die Erfahrung und Beurteilung des anderen machen, noch dass wir uns den oder das andere unzugänglich machen, indem wir es als etwas gänzlich Fremdes aus unserem Erfahrungshorizont ausschließen. Über das Zusammenspiel von Pathos und Response wird der andere in vielschichtiger Weise als ›der andere‹ erfahren – mit der Konsequenz, dass wir im Antworten auch unser Selbstsein immer wieder neu entdecken.

Vor diesem Hintergrund stellt sich nun die Frage, ob und wie wir auch auf Maschinen und insbesondere künstlich-intelligente Maschinen in diesem gehaltvollen Sinne antworten können. Nach der oben genannten Fassung des Antwortenkönnens in der Verschränkung von Pathos und Response können wir durchaus nicht nur Menschen, sondern auch Maschinen antworten. Nehmen wir zum Beispiel die von klugen Algorithmen empfohlenen Bücher oder Filme: In den seltensten Fällen reagieren wir einfach reflexhaft und erwerben die empfohlenen Produkte sofort – wie es sich vielleicht mancher Digitalkonzern erhofft –, sondern wir antworten auf diese Empfehlungen, indem wir uns beispielsweise über diese Produktempfehlungszumutungen ärgern oder indem wir die Vorschläge doch, ein wenig neugierig geworden, mit unseren Interessen abgleichen. Das heißt, wir antworten auf diese Algorithmen, weil wir diese vor dem Hintergrund unserer Sinnhorizonte wahrnehmen und bewerten – was nicht ausschließen muss, dass wir auf diese Empfehlungsübergriffigkeit auch recht barsch antworten können, indem wir die Vorschläge der Streaming-Dienste in den Wind schlagen.

In ähnlicher Weise begegnen wir auch Robotern verschiedener Couleur. Einen Pflege- oder Service-Roboter, mit dem wir interagieren, nehmen wir vor dem Hintergrund unserer sozialen Erfahrung und unseren sozialen Erwartungen wahr. Wir reagieren auch hier nicht nur, sondern wir antworten ebenfalls im oben genannten Sinne, wenn wir beispielweise die derzeitige Unbeholfenheit des Roboters belächeln, wenn wir die

43 Ernst Cassirer: *An Essay on Man. An Introduction to a Philosophy of Human Culture*. In: Ders.: *Gesammelte Werke*. Bd. 23, hg. v. Birgit Recki, Meiner: Hamburg 2006, S. 32ff.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Interaktion als ›Begegnungsspiel‹ mitspielen oder wenn wir in Gedanken vielleicht bereits eine Geschichte skizzieren, die wir den Freund:innen über diese seltsame Begegnung erzählen können. Auch wenn der Roboter vielleicht nicht zu einem ›Freund‹ wird, der uns zum Bücher- und Juwelendiebstahl animiert wie in dem Film *Robot & Frank*, kann die pathische Erfahrung durch die Maschine zur Ausbildung eines vielschichtigen Antwortgeschehens führen.

Vermutlich wird es dabei bleiben, dass das Antwortgeschehen in der Interaktion mit einer Maschine ein anderes ist als das mit einem anderen Menschen. Eine sehr komplexe künstlich-intelligente Maschine kann menschenähnliches Antworten aber vielleicht in einer Weise simulieren, dass wir mitunter gar nicht sagen können, wie sich die maschinelle Antwort von der möglichen menschlichen Antwort unterscheidet. Nichtsdestotrotz dürften wir angesichts der Brüchigkeit, Vieldeutigkeit und Unberechenbarkeit menschlichen Antwortenkönnens wohl immer nur zeitweise vergessen, dass es sich bei unserem Gegenüber um eine künstlich-intelligente Maschine handelt, die sich so verhält, weil Programmierer:innen wollten, dass sie sich menschenähnlich verhält. Doch dies ändert nichts daran, dass wir diese Maschinen nicht nur als ein nützliches Werkzeug betrachten und entsprechend verwenden, sondern dass wir ein alteritätstheoretisches Surplus vorliegen haben, das sich phänomenologisch und anthropologisch aus der Struktur des menschlichen Antwortenkönnens in der Verschränkung von Pathos und Response erklären lässt.

Ein interessanter Grenzfall in diesem Kontext ist das Phänomen des Vertrauens. Vermutlich stimmt die Beschreibung, dass wir aufgrund der sozialen Situationen, die in Mensch-Maschine-Interaktionen entstehen, auch Formen von Vertrauen in die Maschine ausbilden, die über die bloße Verlässlichkeit, die wir von technischen Geräten erhoffen und erwarten können, hinausgehen.[44] Bemerkenswert ist hier, dass die vertrauensvolle Sozialität der Mensch-Maschine-Begegnung zu den Phänomenen von *overtrust* oder *automation bias* führen kann, wie man dies in der Debatte nennt.[45] Das heißt, pointiert gesagt, dass wir der Maschine in ihrer vermeintlichen Perfektion in einer Weise vertrauen, die eigentlich nicht angebracht ist. *Overtrust* und *automation bias* aktualisieren in gewisser Weise die ›Antiquiertheit‹ des Menschen angesichts der Perfektion der Maschinen, wie sie Günther Anders beschrieben hat.[46] Wir unter-

44 Siehe hierzu noch einmal Mark Coeckelbergh: »Can we trust robots?« In: Ethics Inf Technol (2012) 14:53–60.

45 Siehe Alan R. Wagner, Jason Borenstein, Ayanna Howard: »Overtrust in the Robotic Age«. Communications of the ACM, 61(9) (2018): 22–24; Alex Leveringhaus: *Ethics and Autonomous Weapons*. Palgrave-Macmillan: London 2016, 99.

46 Günther Anders: *Die Antiquiertheit des Menschen. Bd. 1: Über die Seele im Zeitalter der zweiten industriellen Revolution*. Beck: München 1956.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

werfen uns mitunter recht bereitwillig unseren tadellos funktionierenden Maschinen – mit dem Effekt, dass wir dann auf die Maschinen gar nicht mehr antworten, sondern nur noch reagieren, uns gewissermaßen selbst mechanisieren, weil wir die funktionelle ›Norm‹ der Maschine internalisiert haben. Eine Technikkritik könnte bei dem menschlichen Antwortenkönnen ansetzen, um einen Ausweg aus der ›Selbstversklavung‹ durch einen bloß reflexhaft reagierenden Umgang mit den Maschinen zu finden (um einen beliebten Topos zu bemühen).[47]

7. Der Anspruch der Maschine

In der Phänomenologie der Alterität ist die Frage zentral, welcher (ethische) Anspruch von ›dem anderen‹ ausgeht. Dies kann sich als eine Phänomenologie des Fremden ausbuchstabieren, wie bei Waldenfels, oder in einer radikalen Subjektkritik formuliert werden, wie wir es bekanntlich bei Lévinas vorliegen haben.[48] Lévinas geht von einem Primat des Anderen aus, der aller Subjektkonstitution vorausgeht und diese sogar grundsätzlich in Frage stellt. Im Kern seiner Philosophie steht, kurz gesagt, die Gedankenfigur, dass die »*Bewegung des Selben zum Anderen*« in einer Weise erfolgt, in der sie »*niemals zum Selben zurückkehrt*«.[49] Im Blick auf die Diskussion einer maschinellen Alterität dürfte es philosophisch keine kleine Herausforderung sein, den Anspruch des Anderen, wie er von Lévinas formuliert wird, auch in der Maschine als ›das Andere‹ ausmachen zu können (dies könnte vielleicht der Grund sein, warum Ihde die vorsichtige Wendung der *quasi-otherness* wählt).

Es stellt sich also die Frage, in welcher Weise wir von dem Anspruch der Maschine als ›das andere‹ sprechen können. Lévinas schreibt in *Die Spur des Anderen*, dass die »heteronome Erfahrung« eine »Bewegung« ermöglicht, »die nicht zu ihrem Ausgangspunkt zurückkehrt«.[50] Das heißt, dass ›der Andere‹ uns in dem Sinne ›bestimmt‹, dass er ein unvorhersehbares Antwortgeschehen in Gang setzt. Diese Art heteronomer Erfahrung könnte sich auch auf die Interaktion mit Maschinen übertragen lassen: Dann würden wir uns in der Interaktion einer Maschine in einer Weise selbst erfahren, die weder auf der bloßen Festlegung dessen, was die Maschine ist und leisten soll, noch auf unser sich an

47 Siehe etwa Georg Simmel: *Philosophie des Geldes*. In: *Gesamtausgabe*. Bd. 6, hg. von David P. Frisby, Klaus Christian Köhnke, Frankfurt am Main 1989, 673.

48 Siehe Emmanuel Lévinas: *Die Spur des Anderen. Untersuchungen zur Phänomenologie und Sozialphilosophie*. Freiburg 1983.

49 Ebd., 215.

50 Ebd., 215.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

die Funktionsweise der Maschine anpassendes Verhalten zu reduzieren ist. Durch den eigentümlichen *default mode* des sozialen Gewebes und unser Antwortenkönnen nehmen wir die Maschine (vielleicht nicht nur, aber auch) als ›etwas anderes‹ wahr, was nicht mit der bloßen Aneignung durch unsere üblichen epistemischen Kategorien zu erklären ist.

Bemerkenswert bleibt dabei, dass die Maschine etwas von Menschen Hergestelltes ist, ihre Existenz also nur deshalb hat, weil wir die entsprechenden Kategorien zur Verfügung haben, mit der die Maschine konstruiert, gebaut und genutzt wird. Das heißt, wir interagieren mit einer Entität, die erst durch unser Verfügungswissen existiert und die wir doch in der Begegnung als etwas Unverfügbares erfahren – weil das Antwortgeschehen in der Pathos-und-Response-Verschränkung nicht völlig vorhersehbar verläuft und uns in unberechenbarer Weise verändern kann. Dies kann grundsätzlich auf alle Technologien zutreffen, aber in verstärktem Maße können wir dies sicherlich im Blick auf künstlich-intelligente Systeme beobachten, weil wir es hier auch mit einem sich entziehenden *blackboxing* zu tun haben (auf dieses epistemologisch interessante ›Problem‹ in der Mensch-Maschine-Interaktion soll an dieser Stelle nur pauschal verwiesen werden).[51]

Auch wenn die Begegnung mit anderen Menschen eine andere Qualität hat, ist es erlaubt zu sagen, dass sich mit der Maschine die oben genannten heteronomen Erfahrungen verbinden können. Denn möglicherweise stellt uns die Maschine in ähnlicher Weise in Frage, wie es Lévinas für die Begegnung mit dem Anderen beschreibt: »Die Beziehung zum Anderen stellt mich in Frage, sie leert mich von mir selbst; sie leert mich unaufhörlich, indem sie mir so unaufhörlich neue Quellen entdeckt.«[52] Eine derartige radikal gedachte ›Selbstentleerung‹ mag in Bezug auf die Begegnung mit einer Maschine etwas hyperbolisch wirken, aber bestimmte Formen des Sich-selbst-in-Frage-stellens kann man auch in der Interaktion mit Maschinen erfahren. Das kann von Zweifeln angesichts der eigenen Kompetenzen im Umgang mit Maschinen bis hin zur Selbstbefragung bei komplizierten Entscheidungssituationen reichen, die wir in Teilen an die Maschinen delegieren müssen, etwa im klinischen Kontext oder im Militärsektor.

51 Philipp von Hilgers: »Ursprünge der Black Box«. In: Anna Ofak und Philipp von Hilgers (Hg.): *Rekursionen. Von Faltungen des Wissens*. München 2010, S. 135–153; Andreas Kaminski (2018): »Der Erfolg der Modellierung und das Ende der Modelle. Epistemische Opazität in der Computersimulation«. In: Andreas Brenneis et al. (Hg.) (2018): *Technik – Macht – Raum. Das Topologische Manifest im Kontext interdisziplinärer Studien*. Wiesbaden: Springer VS, 317–333.

52 Lévinas: *Die Spur des Anderen. Untersuchungen zur Phänomenologie und Sozialphilosophie*. Freiburg 1983, S. 221.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Im Blick auf den Begriff des »Antlitzes«, der in Lévinas' Theorie der Alterität eine zentrale Rolle einnimmt, dürften sich aber gewisse Zweifel einstellen, ob und inwiefern dieser Begriff auf die Begegnung mit einer Maschine zu übertragen ist. Lévinas beschreibt das Antlitz ›des anderen‹ und den damit verbundenen Aufforderungscharakter in eindringlichen Worten: das Antlitz sei durch seine »Nacktheit« und »Not« charakterisiert und in dieser seiner Nacktheit ist es ein »inständiges Flehen«.[53] Gleichzeitig »fordert« dieses Flehen aber auch: »So bedeutet die Anwesenheit des Antlitzes eine nicht abzulehnende Anordnung, ein Gebot, das die Verfügungsgewalt des Bewußtseins einschränkt. Das Bewußtsein wird durch das Antlitz in Frage gestellt […]. Seine Gegenwart ist eine Aufforderung zu Antwort.«[54] Angesichts des historischen Hintergrunds, vor dem Lévinas seine Philosophie entwickelt hat, mag die Übertragung des Antlitz-Begriffes auf die Begegnung mit der Maschine regelrecht frivol erscheinen. Denn auch wenn wir die Aufforderung zur Antwort im oben genannten Sinne auch in der Interaktion mit Maschinen wiederfinden können, dürften sich Nacktheit, Not und Flehen (oder andere Begriffe in Lévinas' Vokabular) kaum sinnvoll mit der Maschine verbinden lassen.

Doch trotz dieser Vorbehalte kann man Lévinas' Alteritätstheorie für das Verständnis von Mensch-Maschine-Begegnungen fruchtbar machen, wie mir scheint. Gerade durch Lévinas' ungewöhnliche Terminologie können wir uns dem Nützlichkeitsdenken und der technomorphen Einpassung in die mechanisierte Welt (etwa als das unsere Denk- und Handlungsformen prägenden »Ge-stell«)[55] entziehen.[56] Lévinas kann uns daran erinnern, dass wir den Maschinen jenseits des funktionalen Zusammenhangs, den wir mit ihnen bilden, begegnen (können) und dass wir möglicherweise neue Vokabulare brauchen, um zu erfassen, wie die Maschinen unser Selbstverhältnis ändern. Lévinas' Pointe war ja, dass wir in der Bewegung zum anderen hin nicht wieder zu dem ›selben Selbst‹ zurückkehren. Und so können wir auch im Antworten auf die Maschine selbst andere werden.

Lévinas' Terminologie aufgreifend und weiterentwickelnd, könnte man die Maschine als das ›Antlitzlose‹ beschreiben. Mit der Maschine tritt uns etwas entgegen, das uns gerade in seiner Antlitzlosigkeit herausfordert. Die Rede von Antlitz und Antlitzlosigkeit bietet sich vor allem in Bezug auf die Interaktion mit künstlich-intelligenten Robotersystemen

53 Ebd., 222.
54 Ebd., 223f.
55 Martin Heidegger: *Die Frage nach der Technik*. In: Ders.: *Vorträge und Aufsätze. Gesamtausgabe*. Bd. 7, hg. v. Friedrich-Wilhelm v. Herrmann, Frankfurt a.M. 2000, S. 5–36.
56 Siehe Christoph Hubig: *Die Kunst des Möglichen. Bd. 1: Grundlinien einer dialektischen Philosophie der Technik*. Transcript: Bielefeld 2008.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

an, die in ihren humanoiden Varianten ein an Menschen orientiertes ›Gesicht‹ bekommen, das aber in den eingeschränkten mimischen Funktionen antlitzlos bleibt bzw. zwischen Antlitzanmutung und Anlitzenttäuschung wechselt. ›Antlitzanmutung‹ will ich hier behelfsmäßig, in Ermangelung eines besseren Begriffs für diese Phänomenlage, wählen, um anzuzeigen, dass sich diese Antlitzerfahrung grundlegend von Lévinas' Antlitz-Konzept unterscheidet, wiewohl wir hier eben eine Erfahrung maschineller Alterität vorliegen haben. Die Begegnung mit einem humanoiden Roboter, der Affekte zeigen kann, wird daher eine besondere Herausforderung bleiben. Denn wir dürften einerseits dazu tendieren, ›menschlich‹ zu reagieren, indem wir auf Mimik und Gestik des Roboters antworten. Andererseits dürften uns die Anlitzlosigkeit der Maschine auch irritieren, vielleicht sogar zurückschrecken lassen. Da die Maschine in der oben gewählten paradoxen Formulierung als das verfügt Unverfügte bzw. das unverfügt Verfügte beschrieben werden kann, könnte sich mit der Maschine auch eine besondere Art von Zerbrechlichkeit verbinden, die auf die menschlichen Nutzer zurückspiegelt. Hannah Arendt hatte das technische Herstellen in *Vita activa* bekanntlich mit einer besonderen Macht verknüpft, die darin besteht, dass wir das selbst hergestellte jederzeit auch selbst wieder zerstören können.[57] Und die Interaktion mit einem Gegenüber, das wir im Prinzip jederzeit zerstören könnten, macht die Maschine zu etwas Fragilem – das uns als Interagierende in Frage stellen kann.

Im Fall von Sexrobotern dürfte dies in besonderem Maße augenfällig sein.[58] Sexroboter sind gezielt nach unseren Vorstellungen gebaut und insofern haben wir maximale Macht über sie. Gleichzeitig antworten wir diesen obskuren Objekten unserer Begierde auch in einer nicht vorhersagbaren Weise durch die oben genannte Antlitzanmutung. Damit entziehen sie sich bis zu einem gewissen Grad unserer Verfügbarkeit, auch wenn sie zur Bedürfnisbefriedigung produziert wurden. In unserem Begehren geben wir dem Roboter eine stark sexualisierte Antlitzanmutung (wie stereotyp diese auch immer sein mag) und haben gleichzeitig die Macht, diesen jederzeit zur Antlitzlosigkeit zu degradieren. Der erotische Imaginationsraum verbindet sich mit der Macht, den Sexroboter zu ersetzen oder zu zerstören, wenn dieser unsere Bedürfnisse nicht (mehr) befriedigt. Das ›Ebenbild‹, das wir uns in dem Sexroboter schaffen, fußt auf einem alteritätstheoretischen Surplus, das wir bereits im Fall der Androidinnen kennengelernt haben: wir begegnen hier ebenfalls häufig feminisierten

57 Hannah Arendt: *Vita activa oder Vom tätigen Leben*. Piper: München 1981, S. 165ff.

58 Siehe zum Überblick über die Sexroboter-Debatte Catrin Misselhorn: *Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern und Co*. Stuttgart 2021, S. 110–134.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Maschinen, die einen erotischen Imaginationsraum für die Erfahrung ›der Anderen‹ eröffnet. Da dies aber mit der Macht verbunden ist, ›die Andere‹ selbst herstellen (und damit prinzipiell jederzeit zerstören) zu können, kann die Antlitzanmutung verschwinden, und die im Sexroboter imaginierte ›sie‹ wird auf ihre Antlitzlosigkeit reduziert – was wiederum die Nutzer von Sexrobotern in gewisser Weise selbst antlitzlos macht.

In dem Film *Ich bin dein Mensch* wird dies in der Begegnung zwischen einer Frau und einem (männlich gestalteten) Roboter durchgespielt: Die Wahrnehmung des Roboters schwankt im Laufe der Geschichte zwischen menschlichem Partner und verfügbarem Objekt und zeigt, wie die Hauptfigur sich in der Interaktion mit dem Roboter immer wieder selbst in Zweifel bringt. So könnte der ›Anspruch der Maschine‹ paradoxerweise in dem bestehen, dass sie uns in ihrer Antlitzlosigkeit daran erinnern, dass wir Menschen der Begegnung mit dem Antlitz anderer bedürfen, um wir selbst zu werden. Insofern spiegelt uns die Antlitzlosigkeit der Maschine unsere eigene Vulnerabilität wider, die wir in der Interaktion mit dieser Maschine erfahren können.

8. Schlussbemerkung

In meinem Beitrag bin ich dem Phänomen maschineller Alterität aus verschiedenen Perspektiven nachgegangen. Nach ersten Überlegungen zum ›sozialen Gewebe‹, das wir mit (nicht nur, aber auch und insbesondere künstlich-intelligenten) Maschinen bilden können, habe ich im Verweis auf die Kulturgeschichte einige Hinweise gegeben, dass die Maschine schon früh als ›Andere‹ verstanden wurden, was sich in Imaginationen über Android:innen niederschlug, wobei auffallend häufig eine Feminisierung der Maschine zu beobachten ist. Die Erfahrung der Andersheit der Maschine überlagert sich in bemerkenswerter Weise mit Geschlechtervorstellungen und entsprechen Stereotypisierungen. Im Blick auf die Anthropologie Gehlens und die frühe Programmatik der KI-Forscher habe ich herausgearbeitet, wie sich hier alteritätstheoretische Momente unterschiedlicher Art finden. Schließlich habe ich drei zentrale Konzepte der Phänomenologie für meine Überlegungen zur maschinellen Alterität zugrunde gelegt: Das ›In-Geschichten-verstrickt sein‹ als Erfahrung ›des Anderen‹; die Responsivität (und die Verschränkung von Pathos und Response); sowie den ›Anspruch des Anderen‹, wie er sich in der Philosophie von Lévinas findet. Ich hoffe, dass ich Folgendes zeigen konnte: Auch wenn die Erfahrung von menschlich Anderen anders gelagert ist, kann eine Phänomenologie der Alterität dazu beitragen, dass wir unsere Begegnungen mit den Maschinen genauer beschreiben und besser verstehen können. Für wertvolle Hinweise zu meinem Beitrag danke ich Katharina Kohler.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Literatur

Anders, Günther (1956): *Die Antiquiertheit des Menschen. Bd. 1: Über die Seele im Zeitalter der zweiten industriellen Revolution*. München: Beck.

Anderson, Michael and Susan Leigh Anderson (Hg.) (2007): *Machine Ethics*, Cambridge: Cambridge University Press.

Arendt, Hannah (1981, 2002): *Vita activa oder vom tätigen Leben*, Piper: München.

Böhme, Hartmut, Gernot Böhme (1983): *Das Andere der Vernunft. Zur Entwicklung von Rationalitätsstrukturen am Beispiel Kants*, Frankfurt am Main: Suhrkamp.

Cassirer, Ernst (2006): *An Essay on Man. An Introduction to a Philosophy of Human Culture,* in: ders.: *Gesammelte Werke*, Bd. 23, hg. v. Birgit Recki, Hamburg: Meiner.

Coeckelbergh, Mark (2012): »Can we trust robots?«, *Ethics Inf Technol*, 14:53–60.

Crawford, Kate (2021): *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven.

Dick, Philip K. (2012): *Do Androids Dream of Electric Sheep?* London: Weidenfeld & Nicolson.

Floridi, Luciano, J.W. Saunders (2004): »On the Morality of Artificial Agents«, *Minds and Machine,* 14: 349–379.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, Effy Vayena (2018): »AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations«, *Minds and Machines*, 28(4): 689–707.

Gehlen, Arnold (1957): *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme der industriellen Gesellschaft*, Hamburg: Rowohlt.

Gunkel, David J., Joanna J. Bryson (eds.) (2014): »Machine Morality: The Machine as Moral Agent and Patient«, special issue of Philosophy & Technology, 27(1): 1–142.

Gunkel, David J. (2017): *The machine question. Critical perspectives on AI, robotics, and ethics*, Massachusetts: The MIT Press.

Heidegger, Martin (1993): *Sein und Zeit*, Tübingen: Niemeyer.

Heidegger, Martin (2000): *Die Frage nach der Technik*, in: ders.: *Vorträge und Aufsätze. Gesamtausgabe*, Bd. 7, hg. v. Friedrich-Wilhelm v. Herrmann, Frankfurt am Main: Klostermann.

Hilgers, Philipp von (2010): »Ursprünge der Black Box«, in: Anna Ofak und Philipp von Hilgers (Hg.): *Rekursionen. Von Faltungen des Wissens*, München: Wilhelm Fink.

Hoffmann, E.T.A. (1991): *Der Sandmann*, Stuttgart: Reclam.

Homer (1990): *Odyssee*, in der Übertragung von Johann Heinrich Voß. Frankfurt am Main: Insel.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Hubig, Christoph (2008): *Die Kunst des Möglichen, Bd. 1: Grundlinien einer dialektischen Philosophie der Technik,* Transcript: Bielefeld.
Ihde, Don (1990): *Technology and the Lifeworld. From Garden to Earth*, Bloomington: Indiana University Press.
Kaminski, Andreas (2018): »Der Erfolg der Modellierung und das Ende der Modelle. Epistemische Opazität in der Computersimulation«, in: Andreas Brenneis, Oliver Honer, Sina Keesser, Annette Ripper, Silke Vetter-Schultheiß (Hg.) (2018): *Technik – Macht – Raum. Das Topologische Manifest im Kontext interdisziplinärer Studien*. Wiesbaden: Springer VS.
LaMettrie, Julien Offray de (2001): *Der Mensch eine Maschine*, Stuttgart.
Leveringhaus, Alex (2016): *Ethics and Autonomous Weapons*. London: Palgrave-Macmillan.
Lévinas, Emmanuel (1983): *Die Spur des Anderen. Untersuchungen zur Phänomenologie und Sozialphilosophie.* Freiburg: Karl Alber.
Liggieri, Kevin, Oliver Müller (2019): »Mensch-Maschine-Interaktion seit der Antike: Imaginationsräume, Narrationen und Selbstverständnisdiskurse«, in: Kevin Liggieri, Oliver Müller (Hg.): *Mensch-Maschine-Interaktion. Geschichte-Kultur-Ethik* (Metzler-Handbuch), Stuttgart: J.B. Metzler.
Lin, Patrick, Keith Abney, Ryan Jenkins (eds.) (2017*): Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, New York: Oxford University Press.
McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, Claude E. Shannon (1955): »A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence«, (August 31, 1955). Zitiert nach AI Magazine, 27(4) (2006): 12–14.
Meyer-Drawe, Käte (1996): *Menschen im Spiegel ihrer Maschinen*, München: Wilhelm Fink.
Meyer-Drawe, Käte (2018): »Androiden«, in: Carlos Kölbl, Anna Sieben (Hg.): *Stichwörter zur Kulturpsychologie. Jürgen Straub zum 60. Geburtstag*, Gießen: Psychosozial Verlag.
Misselhorn, Catrin (2018): *Grundfragen der Maschinenethik*, Stuttgart: Reclam.
Misselhorn, Catrin (2021): *Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern und Co*, Stuttgart: Reclam.
Müller, Oliver (2014): *Selbst, Welt und Technik. Eine anthropologische, geistesgeschichtliche und ethische Untersuchung*, Berlin, New York: de Gruyter.
Nørskov, Marco (ed.) (2017): *Social Robots*, London: Routledge.
Plessner, Helmuth (1985): *Die Utopie in der Maschine,* in: ders.: *Gesammelte Schriften*, Bd. 10, hg. v. Günter Dux, Odo Marquard und Elisabeth Ströker. Frankfurt am Main: Suhrkamp.
Ricœur, Paul (1996): *Das Selbst als ein Anderer*, München: Wilhelm Fink.
Rosa, Hartmut (2016): *Resonanz. Eine Soziologie der Weltbeziehung*, Berlin: Suhrkamp.
Schapp, Wilhelm (1985): *In Geschichten verstrickt. Zum Sein von Mensch und Ding*, Frankfurt am Main: Klostermann.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Simmel Georg (1989): *Philosophie des Geldes*, in: *Gesamtausgabe,* Bd. 6, hg. von David P. Frisby, Klaus Christian Köhnke, Frankfurt am Main: Suhrkamp.

Spiekermann, Sarah (2019): *Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert*, München.

Turing, Alan (1950): »Computing machinery and intelligence«, Mind, 49: 433–460.

Völker, Klaus (Hg.) (1976): *Künstliche Menschen. Dichtungen und Dokumente über Golems, Homunculi, Androiden und liebende Statuen,* München: dtv.

Vöneky, Silja, Philipp Kellmeyer, Wolfram Burgard, Oliver Müller (2022, i. Ersch.): *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*, Cambridge: Cambridge University Press.

Wagner, Alan R., Jason Borenstein, and Ayanna Howard (2018): »Overtrust in the robotic age«, Communications of the ACM, 61(9): 22–24.

Wagner, Johanna (2020): *Künstliche Intelligenzen als moralisch verantwortliche Akteure*, Paderborn: mentis.

Waldenfels, Bernhard (1994): *Antwortregister*, Frankfurt am Main: Suhrkamp.

Waldenfels, Bernhard (2019): *Erfahrung, die zur Sprache drängt. Studien zur Psychoanalyse und Psychotherapie aus phänomenologischer Sicht*, Berlin: Suhrkamp.

Weizenbaum, Joseph (1966): »ELIZA-A Computer Program For the Study of Natural Language Communication Between Man and Machine«, Communications of the ACM, 9:36–35.

Zuboff, Shoshana (2019): *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: Public Affairs.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Burkhard Liebsch

Aussichten kommunikativer Verwerfungen zwischen *digital natives* und analogen Aborigines

> Alles ist genau berechnet, doch jetzt kann euch nur mehr retten, was ihr nicht berechnet habt! Alles habt ihr ausgenützt bis auf seinen letzten Rest – wo ist er, dieser letzte Rest? Er wird zurückverlangt.
>
> Ilse Aichinger[1]

> Im Verlauf der Zeit wurde das Verlangen nach Selbstausgrenzung immer gebieterischer.
>
> Hans M. Enzensberger[2]

Wenn ich mich im Folgenden mit etwas ironischem Unterton zu kommunikativen Verwerfungen zwischen *digital natives* und analogen Aborigines äußere, so ist wohl damit zu rechnen, dass solche Verwerfungen nicht nur zum Thema werden, sondern auch zwischen Autor und Lesern auftreten werden, insofern sie altersbedingt vermutlich sehr verschieden ›sozialisiert‹ worden sind. Bestürzte Zeitgenossen vermuten sogar, dass die Jüngeren längst auf dem Weg in eine geistig mehr oder weniger mutierte Zukunft der menschlichen Gattung unterwegs sind. Mehr noch: in diesem Zusammenhang war bereits – offensichtlich ›leicht übertrieben‹ – von einem regelrechten »Weltkrieg zwischen digitaler und analoger Kultur« die Rede.[3] Erstere reduziere alles auf ›Benutzer-‹ und sonstige, auf jeden Fall metrisch und algorithmisch erfassbare Oberflächen[4], die am Ende jegliches Geheimnis, jegliches Rätsel, das wir bislang füreinander darstellen mögen, und jegliche radikale Alterität auslöschen müssen. Was bleibe, sei eine weitestgehend zum ›Bild‹ gewordene Welt[5] von Bildern, die uns eines Tages ein nicht zu verbildlichendes und nicht berechenbares ›analoges‹ Leben schmerzlich

1 Ilse Aichinger, *Die größere Hoffnung* [1948], Hamburg: Fischer 1960, 151.
2 Hans Magnus Enzensberger, *Mittelmaß und Wahn. Gesammelte Zerstreuungen*, Frankfurt/M.: Suhrkamp 1989, 270.
3 Lázló F. Földényi, *Lob der Melancholie. Rätselhafte Botschaften*, Berlin: Matthes & Seitz 2019, 188.
4 Ebd., 308.
5 Ebd., 120.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

vermissen lassen könnten.[6] Dieses scheint bereits derart in die Defensive geraten zu sein, dass manche dazu übergehen, ein Recht auf analoge Existenz zu behaupten.[7] Nicht darum oder um den zuvor angeschlagenen Ton dystopischer Antizipationen gegenwärtiger Kulturkritik wird es mir im Folgenden gehen, sondern um gewisse kommunikative Verwerfungen zwischen uns, deren Folgen gegenwärtig in ihrer Radikalität in der Tat noch gar nicht abzusehen sind, zumal wir offenbar erst durch immer weiter voranschreitende Digitalisierungsprozesse dazu gezwungen werden, uns nachträglich darauf zu besinnen, was es überhaupt heißt, ›analog‹ und leibhaftig zu existieren und zu ko-existieren – vermutlich weiterhin.[8] Im verfügbaren Rahmen muss ich mich allerdings auf eine Skizze der angedeuteten Diskussionslage beschränken, die sich an anderer Stelle längst ausführlich und vielfach mit technik-kritischer Stoßrichtung auch mit Grundbegriffen wie ›Bild‹, ›Welt‹, ›Koexistenz‹ und deren Konfiguration auseinandergesetzt hat, von Martin Heidegger und Eugen Fink bis hin zu Jean-Luc Nancy und vielen anderen.

*

In seiner Philosophie der Antiquiertheit des Menschen spricht Günther Anders vom »Schock der sozialen Geburt«, zu dem es kommen müsse, wenn das Kind realisiert, sich in seinem Leben fortan »als ›Ausnahme‹,

6 Das erinnert an eine Romanskizze von Theodor Fontane, die einen in dieser Richtung aufschlussreichen Hinweis gibt. Oceane, die weibliche Heldin des ungeschriebenen Romans, leidet daran, dass sie statt eines Gefühls nur Sehnsucht danach haben kann. Trauer und Schmerz bleiben ihr fremd, denn »alles, was geschieht, wir ihr zum *Bild*«; und die Sehnsucht nach Teilnahme am menschlichen Leben »wird ihr selber zum Schicksal. Sie wirft das Leben weg, weil sie fühlt [!], daß ihr Leben nur ein Schein-Leben, aber kein wirkliches Leben ist.« Die paradoxerweise aus dem Gefühl, kein Gefühl zu haben, keimende Erkenntnis tötet sie. Und sie will diese Erkenntnis zum Schweigen bringen. Theodor Fontane, *Der Stechlin. Roman*, München: dtv ^{3}1993, 412. Ironischerweise erscheint sie lebendig gerade durch die tödliche Verwerfung eines ›unwirklichen‹ Lebens, von dessen ›wirklicher‹, verfehlter Lebendigkeit sie sich allerdings ein ihrerseits womöglich verfehltes Bild macht.

7 So muss »analog zu leben« für den amerikanischen Romanautor Dave Eggers »ein Menschenrecht« sein (siehe das Interview mit ihm in der ZEIT, no. 40, vom 30. September 2021, 59).

8 Nur *en passant* kann ich hier die Frage aufwerfen, ob beide Adverbien nicht das Gleiche meinen. Bekanntlich bedeutet ›analog‹ so viel wie ›entsprechend‹, ›gleichartig‹ oder auch ›ähnlich‹; und das könnte sich auch auf Digitales und Virtuelles beziehen. Als Synonym für ›leibhaftig‹ aber meint ›analog‹ ein schlechterdings nicht digitalisierbares Leben bzw. In-der-Welt-sein im Sinne einer unüberwindlichen Ungleichartigkeit.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

als ›Einzelne[s]« behaupten zu müssen. Unter Umständen falle dieser Schock derart heftig aus, dass es gerechtfertigt sei, ihn als »als Variante von Freuds ›Geburtstrauma‹« aufzufassen.[9] Neuerdings stellt sich dies in den Augen von Autoren weniger dramatisch dar, für die es bereits ›klar‹ zu sein scheint, dass »heute [...] jeder Mensch in seine Digitalität hineingeboren« wird; und zwar kontinuierlich und ohne dass das mit einem besonderen, zu feiernden Ereignis zu verbinden wäre.[10] Von entsprechenden Eintragungen in ein neuartiges Personenstandsregister digitaler Wesen ist auch nichts bekannt.

Diesen Autoren zufolge haben wir es seit der massiven Verbreitung von PCs, also seit etwa drei Jahrzehnten, mit sogenannten *digital natives* zu tun, die mit den digitalen Medien – von Computerspielen, dem Internet, E-Mails und Mobiltelefonen bis hin zum *instant messaging* – anscheinend von klein auf vertraut sind und deren Bedienung abgesehen von Problemen der Datensicherheit, von denen sie nicht einmal etwas ahnen, im Gegensatz zu Eltern, ErzieherInnen und LehrerInnen so spielerisch lernen wie irgendeine Muttersprache als erste Fremdsprache, ganz gleich welche. Von ihnen wird darüber hinaus gesagt, sie – *born digital* und *grown up digital* – lebten geradezu in »computerisierter« bzw. »digitale[r] Koexistenz«[11] und somit in einer Welt, zu der Andere, nach wie vor ›analog‹ Lebende, allenfalls noch nachträglich und beschränkten Zugang hätten – als ungeachtet ihres höheren Alters *verspätete Immigranten* in eine ihnen mehr oder weniger fremd bleibende Welt, in der ironischerweise die später Geborenen als die eigentlichen Ureinwohner bzw. Eingeborenen gelten. Diese rücken zum Maßstab zeitgemäßer Formen von Koexistenz auf, zu denen die Älteren – auch als sog. *digital immigrants* Titulierten – vermeintlich kaum mehr aufschließen können. Die Würde des Ursprünglichen liegt nicht bei ihnen, sondern bei den ›digital Geborenen‹, Sozialisierten und Fitten.

Doch hier scheint etwas nicht zu stimmen: Müssen nicht auch Angehörige jener *Generation Millenium*, der man erstmals zugeschrieben hat, im Ganzen in eine weitgehend digitalisierte Welt hineingeboren worden zu sein, zunächst analog leben und zuvor zur Welt gebracht worden sein? Handelt es sich nicht um eine frivole, kaum ernst zu nehmende feuilletonistische Übertreibung, wenn davon die Rede ist, sie seien *born digital*? Und sollte man nicht statt von »digitaler Koexistenz« besser von *digital vermitteltem* Zusammenleben sprechen? Ist es uns überhaupt möglich, *digital zu existieren* und zu ko-existieren, wie es auch ein bereits in die

9 Günther Anders, *Die Antiquiertheit des Menschen. Über die Seele im Zeitalter der zweiten industriellen Revolution*, München: C. H. Beck 1956, 73.

10 Peter Schmitt, *Postdigital. Medienkritik im 21. Jahrhundert*, Hamburg: Meiner 2021, 61 [=P].

11 Ebd., 11, 58; vgl. https://de.wikipedia.org/wiki/Digital_Native

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Jahre gekommener Buchtitel von Nicholas Negroponte vor über 20 Jahren nahelegte: *Being digital*?[12] Oder ist das genauso unmöglich wie ein hölzernes Eisen? Ein in der Philosophie geradezu klassischer, allerdings reichlich abgenutzter Ausdruck für eine semantische Zusammenfügung von Unvereinbarem, die ebenso paradox anmutet wie genau genommen der Ausdruck ›digital *sein*‹.[13] Gemäß klassischem logischem Verständnis *ist* etwas dies *oder* jenes, aber niemals beides zugleich, und dementsprechend kann von ihm auch nur dies *oder* jenes zutreffend ausgesagt werden, wie man unter Berufung auf das uralte Prinzip vom ausgeschlossenen Dritten behauptet. Demzufolge würde gelten: *Being*, *sein, être*, verbal verstanden, was für das Geschehen menschlichen Daseins im Zusammenleben mit und unter Anderen steht, kann nur analog, nicht digital stattfinden. Und umgekehrt: was in einer diskreten Serie von Nullen und Einsen, also digital, programmierbar ist, kann nicht analog existieren. *Being*, *sein* bzw. *être* vollzieht sich gemäß einschlägiger Beschreibungen wie derjenigen, die Maurice Merleau-Ponty in seiner *Phänomenologie der Wahrnehmung* (1945) vorgelegt hat, ›leibhaftig‹, inkarniert und zugleich perspektivisch, situiert ›an Ort und Stelle‹, wenn auch in der von Helmuth Plessner beschriebenen »exzentrischen Positionalität«, die weit bis in imaginäre und virtuelle Räume vordringen kann.[14] Wenn letzteres geschieht, sagt man heute auch: jemand »gehe ins Netz«. Damit ist gemeint, wie jeder weiß: die Betreffenden fahren ihren Rechner hoch, steuern eine Internetadresse an, um dann virtuell eventuell zu *surfen*, bis der Nacken oder das Kreuz weh tut, das so manchen *Internetnerd* oder *Gamer* sodann schmerzhaft daran erinnert, dass er keineswegs *selbst* ›ins Netz gegangen‹ ist, sondern immer noch vor seinem Rechner sitzt (wie vielleicht, mit gewissen Unterbrechungen, seit Jahren schon und mit geschwächter Oberkörpermuskulatur in gekrümmter Haltung...).

Ins Netz gegangen ist er, sind wir allerdings möglicherweise auf andere Weise, wenn man sich so ausdrückt: man ist mit einer solchen Beschreibung auch ins Netz der eigenen Konfusionen gegangen. Wenn etwas absolut nicht möglich ist, dann ist es das: buchstäblich in das *Internet* zu gehen. Warum nicht? Eben weil wir nach wie vor leibhaftige Wesen sind. Das Unmögliche überspielen wir mit Worten, die seit langem (und vielleicht seit jeher) einer *semantischen Drift* ausgesetzt sind. Die Worte, die wir gebrauchen, sind nicht wie starre Designatoren den Dingen

12 Vgl. Neil Postman, »Müssen Toaster sprechen?«, in: *Die Gegenwart der Zukunft*, Berlin: Wagenbach 2000, 99–107.

13 Bernhard H. F. Taureck, *Metaphern und Gleichnisse in der Philosophie. Versuch einer kritischen Ikonologie der Philosophie,* Frankfurt/M.: Suhrkamp 2004.

14 Helmuth Plessner, *Mit anderen Augen. Aspekte einer philosophischen Anthropologie*, Stuttgart: Reclam 1982.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

zugeordnet[15], um so der Zeit und einer sie ständig mehr oder weniger kreativ kohärent deformierenden, aber auch missbräuchlichen Verwendung widerstehen zu können, bei der sich Gebrauch und Missbrauch vielfach kaum unterscheiden lassen.[16] Vielmehr lösen sie sich von vermeintlich starren Verbindungen mit Dingen, Tatsachen und Sachverhalten, passen sich neuen Wirklichkeiten an und werfen dann erst, nachträglich, die Frage auf, ob wir mit ihnen überhaupt noch ›das Selbe‹ oder ›das Gleiche‹ meinen können wie das, was wir selbst oder andere, die vor unserer Zeit gelebt haben, im Sinn gehabt haben mögen.

Inzwischen ist das keine Frage von Epochen oder Generationen mehr, hat eine permanent vor sich gehende semantische Innovation doch im Zuge der globalen Vernetzung der menschlichen Lebensverhältnisse ein Tempo angenommen, das Missverhältnisse zwischen Worten und Dingen, Tatsachen und Sachverhalten zu einer ironischerweise innerhalb der eigenen Lebenszeit normalen Erfahrung hat werden lassen; innerhalb der eigenen Biografie und zwischen uns, im Verhältnis zu Anderen, für die einerseits Neologismen[17], semantische Importe, Akronyme und selbst das Esperanto einer internationalen Ikonik und Zeichensprache zu einer neuen Normalität geworden sind, während sie überkommene Begriffe andererseits kaum mehr in einer ihnen womöglich antiquiert anmutenden Bedeutung zu kennen scheinen.

Nur wenige hören aus dem Wort ›Ereignis‹ noch das ältere ›Eräugnis‹ heraus, von dem es sich etymologisch herleitet.[18] Während es in diesem Falle ohne weiteres einleuchtet, dass wir es mit einem seit langem ungebräuchlich gewordenen und deshalb altertümlich klingenden Wort zu tun haben, verhält es sich mit kommunikativen Begriffen wie ›reden‹, ›sprechen‹, ›antworten‹ weit weniger eindeutig. Offenbar *antwortet* ein Anrufbeantworter jedoch gar nicht im engeren Sinne. Gemessen an brieflicher Kommunikation ›schreibt‹ womöglich auch der eilige Verfasser einer *What's-App*-Nachricht oder ein Twitterer gar nicht an seine Adressaten – so wenig wie jemand, der etwas *postet*, im engeren Sinne auf diese Weise kommuniziert oder der Versender einer Sprachnachricht an jemanden gewandt und auf Erwiderung hin mit letzterem spricht. Man könnte und müsste sich sogar fragen, ob die *user* bzw. Anwender diverser digitaler Kommunikationstechniken überhaupt genau wissen,

15 Ian Kaplow, *Analytik und Ethik der Namen*, Würzburg: Königshausen & Neumann 2002.

16 Gerald Posselt, *Katachrese,* München: Fink 2005.

17 Der letzte Duden verzeichnet angeblich 1000 Neologismen im Vergleich zum vorigen.

18 Vf., »Ereignis – Erfahrung – Erzählung. Spuren einer anderen *Ereignis-Geschichte:* Henri Bergson, Emmanuel Levinas, Paul Ricœur«, in: Marc Rölli (Hg.), *Ereignis auf Französisch. Von Bergson bis Deleuze,* München: Fink 2004, 183–207.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access -

was sie tun, wenn sie diese verwenden – oft in flüchtiger, nachlässiger, fehlerhafter oder geradezu kryptischer Art und Weise, die manchem in syntaktischer, semantischer und pragmatischer Hinsicht irgendwie degeneriert vorkommen mag.[19]

Auf diese Frage hat sie die überlieferte Sprache offenbar nicht vorbereitet, die sich zunächst für selbstverständlichen Gebrauch anbietet, ohne dass man überhaupt fragen und genau wissen müsste, was es heißt, jemanden anzusprechen, an jemanden gewandt zu sprechen, miteinander zu reden und auf jemanden zu antworten. All das funktioniert normalerweise in eingewöhnten Sprachspielen und Lebensformen, ohne genaue Nachfragen erforderlich zu machen. Und selbst dort, wo solche Nachfragen eigens aufgeworfen werden (wie in diversen Kommunikationstheorien, in Philosophien kommunikativen Handelns, sozialer Dialoge und politischer Diskurse etc.), vermisst man vielfach eindeutige Antworten. Das mag damit zu tun haben, dass ›die Sache‹, um die es hier geht – das Ansprechen, Sprechen, Antworten, Miteinanderreden usw. – *selbst, in sich,* uneindeutig ist, so dass diejenigen, die miteinander kommunizieren, *unvermeidlich* in Unklarheiten und Strittigkeiten darüber verstrickt werden, was sie dabei ›tatsächlich‹, ›wirklich‹ oder ›eigentlich‹ tun und ob sie davon womöglich jeweils ein im doppelten Sinne unvermeidlich *geteiltes* Verständnis haben. Solche Unklarheiten und Strittigkeiten müssen nicht offen zutrage treten. Aber dass sie aufbrechen können und sich dann unter Umständen nur schwer, wenn überhaupt, beseitigen, begrenzen oder heilen lassen, ist vielfach bezeugt.

Die meisten, angeblich ohnehin weitgehend leeren Worte hätten mit einem Versuch, wirklich und ernsthaft miteinander zu reden, »notorisch nichts zu tun«, meinte der Theologe Karl Barth.[20] Sind wir etwa deshalb »heute […] nicht mehr miteinander konfrontiert«, wie der tschechische Autor Milan Kundera behauptete, obgleich es an vielfach gewaltträchtigen ›Konfrontationen‹ keineswegs mangelt?[21] Handelt es sich dabei um Konfrontationen, in denen man kaum mehr *miteinander* spricht? Kam der derzeit amtierende Pabst Franziskus deshalb zu dem Schluss: »Wenn es ein Wort gibt, das wir bis zur Erschöpfung wiederholen müssen, dann lautet es Dialog«?[22] Handelt es sich hier lediglich um beschwörende Rhetorik, die keineswegs geeignet ist, den Verdacht zu entschärfen, ihren Namen verdienende Kommunikation finde kaum mehr statt, und die im Fall ihrer Erschöpfung ihrerseits keinen Ausweg mehr weiß? Aber verfügen

19 Man spricht in diesem Zusammenhang auch von »Entliterarisierung«.

20 Karl Barth, *Mensch und Mitmensch. Die Grundform der Menschlichkeit*, Göttingen: Vandenhoek & Ruprecht 1962, 62.

21 Milan Kundera, »Der existenzielle Sinn der bürokratisierten Welt«, in: *Der Vorhang*, Frankfurt/M.: Fischer 2008, 179–184, hier: 182.

22 w2vatican.va/content/francesco/de/speeches/2016/max/documents.papa-francesco_20160506_premio-carlo-magno.html (Stand: Januar 2017).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

wir überhaupt über Theorien und Konzepte, mit deren Hilfe in dieser Hinsicht verlässlich Auskunft zu geben wäre? Oder ist es vielmehr gerade die ›Unklarheit‹ in der Frage, ob man sich wirklich auf Erwiderung hin aneinander gewandt und Antworten erfahren hat, was menschliche Kommunikation in Gang hält?

Alle diese Fragen brechen nunmehr ungeschützt durch konzeptionelle und theoretische Rückendeckung infolge jener semantischen Drift auf, die mit neuen, digitalen Techniken einhergeht, auf die die eingespielte normale bzw. normalisierte Sprache nicht vorbereitet sein konnte. Infolgedessen muss sie zur Bezeichnung von Vorgängen herhalten, über die der fortgesetzte Gebrauch von Verben wie ›reden‹, ›schreiben‹, ›antworten‹ usw. weniger denn je verlässliche Auskunft gewährleistet.[23] Es scheint, dass erst *kommunikative Verwerfungen* wie die zwischen *digital natives* und Anderen, die weiterhin glauben, ›analog‹ zur Welt gekommen zu sein, wenn nicht ausschließlich, so doch gewiss primär ›analog‹ zu leben und ›analog‹ zu koexistieren, dazu zwingen, sich genau dazu eigens zu verhalten.

Wissen wir, was wir tun, wenn wir einem Apparat namens *Alexa* gewisse Kommandos erteilen? Haben wir dann *mit* einer Maschine geredet?[24] Und vielleicht so, dass infolge entsprechender Gewöhnungsprozesse Unterschiede zum Reden mit einem *Anderen* nach und nach verwischt werden – einmal vorausgesetzt, sie waren überhaupt je deutlich? Wissen wir, was wir tun, wenn wir die Frage ignorieren, ob der Apparat alles Aufgezeichnete irgendwohin sendet und auf diese Weise die Privatheit einer Wohnung wie auch der Unterhaltung mit Gästen unterläuft? Könnte das jemand, der noch auf Diskretion Wert legt, nachsehen?

In einem anderen Zusammenhang hieß es entschuldigend: »Vergib Ihnen, denn sie wissen nicht, was sie tun.«[25] Aber ist solches Nicht-Wissen bzw. Nicht-Wissen-Können jemandem zuzuschreiben, der einen derartigen Apparat anschafft und ihn wie selbstverständlich auch die Kommunikation aufzeichnen lässt – und zwar zum weitgehend geheimen Vorteil eines Dritten, früher *big brother* genannt, wie der amerikanischen *National Security Agency* (NSA), der es nur recht sein kann, wenn wir nicht mehr wissen, was wir tun, wenn wir in stiller Anwesenheit technischer Dispositive weiterhin wie früher analog, privat und diskret nur *miteinander* zu sprechen glauben?[26] Könnten und sollten wir nicht wissen, dass wir

23 »Kann man überhaupt noch semantisch erfassen, was passiert?«, fragt auch Schmitt. Und von einem »Ende aller Theorie« ist die Rede (P, 9, 81) – im Namen von »sensiblen Einzelnen im permanenten Kampf mit der Maschinerie« fortschreitender Digitalisierung (P, 11).

24 Postman, »Müssen Toaster sprechen?«, 99.

25 Ev. n. Lukas 23:34.

26 Bernhard H. F. Taureck, *Überwachungsdemokratie. Die NSA als Religion*, Paderborn: Fink 2014.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

unter solchen digitalen Bedingungen nicht mehr ›wie früher‹ kommunizieren können und dass sich das, was darunter zu verstehen ist, substanziell und radikal zu ändern beginnt? Müssten wir nicht fragen, was dabei auf dem Spiel steht? Muss beispielsweise das Problem sozialer Diskretion unter derartigen technischen Rahmenbedingungen ganz neu aufgeworfen werden? In klassischen Dialogtheorien wurde vielfach ganz unproblematisch vorausgesetzt, dass man im Gespräch exklusiv ›zu zweit‹ sein kann. Dagegen brachte bereits Emmanuel Levinas zur Geltung, dass auch angesichts *eines* Anderen stets »der Dritte« (*le tiers*) mit anwesend sei.[27] Auf die gleiche Spur führen tradierte Konzeptionen des Gewissens. Dabei dachte Levinas aber gewiss nicht an eine technische Realisierung dieser Anwesenheit in der Form eines Gerätes, das noch die privateste Unterhaltung aufzuzeichnen vermag – was streng genommen darauf hinauslaufen müsste, jegliche exklusive Diskretion unmöglich zu machen.

Wird die Frage nicht aufgeworfen, was bei den angedeuteten kommunikativen Verwerfungen auf dem Spiel steht, wird es womöglich zu Segregationen und schließlich zu Schismen kommen. Während sich die Einen mehr oder weniger fröhlich der andauernden Innovation immer neuer Techniken ausliefern und sich dabei mit Neologismen behelfen, wenn überkommene Worte nicht mehr angemessen erscheinen, um zu bezeichnen, was man digital vermittelt tun kann, werden sich Andere möglicherweise weitgehend zurückziehen und sich der Entwendung dieser Worte durch (aus ihrer Sicht) missbräuchliche Verwendung verweigern, an die sie sich nicht gewöhnen wollen. *Kommt es nicht wenigstens zu kommunikativen Verwerfungen* zwischen den Einen und den Anderen, werden Segregationen und Schismen nicht einmal mehr *als solche* erkennbar und man beginnt in Parallelwelten zu leben, in denen man aneinander vorbei und allenfalls insofern noch ko-existiert.

Keineswegs ist ausgemacht, dass *wir alle und gleichermaßen* bereits in »digitaler Koexistenz« oder »audiovisueller Gleichschaltung« (P, 14) angekommen sind. Empirisch trifft es keineswegs zu, dass *jede(r)* bereits in digital »vollständig durchsetzten« Gesellschaften lebt (P, 54) und »mit uniformierter Lässigkeit über Touchscreens« streicht, die identischen *Apps* und *Updates* nutzt und der gleichen digitalisierten Umwelt ausgesetzt ist, in der sie bzw. er mehr oder weniger dasselbe wie alle Anderen macht – bis hin dazu, dass »das Besondere, das Nicht-Identische« abgesehen von einem kümmerlichen Rest an Freiheit angeblich »verschwinden« muss (P, 13, 60).

Nicht zu bestreiten ist der beobachtbare Trend zur *fortgesetzten Anhebung der analog-digitalen Schwelle ungeschmälerter Teilhabe aller,*

27 Ja sogar »die ganze Menschheit«; vgl. Emmanuel Levinas, *Totalität und Unendlichkeit. Versuch über die Exteriorität* [1961], Freiburg i. Br., München: Alber 1987, 308 f.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

von Alten und Behinderten sowie auch und gerade von noch unmündigen Kindern, *am gesellschaftlichen Leben*. Wenn dieses vom bargeldlosen Zahlungsverkehr wie in Schweden über nur noch digital verfügbare Zugtickets wie in den Niederlanden bis hin zu digitalen Impfpässen, wie sie inzwischen auch hierzulande üblich geworden sind, am Ende nahezu jeden Zugang zu sozialen Lebensbereichen außerhalb der eigenen Wohnung von der Erfüllung digitaler Voraussetzungen abhängig macht[28], werden bis auf Weiteres *nur analog* Lebende schließlich in ein desozialisiertes und anachronistisches Abseits abgedrängt, wo ihnen auch Grundrechte nur noch wenig nützen werden. Denn um diese wirksam in Anspruch nehmen zu können, muss man sich gegebenenfalls kommunikativ-technisch auch auf der Höhe der Zeit bewegen können.

Doch noch ist es nicht so weit, was manche bereits als vollendete Tatsache ausgeben: »Die Digitalisierung von Allem – der Erdoberfläche, des Nachthimmels, der Geburt und des Sterbens – macht ein heutiges An-der-Maschine-Vorbeiexistieren faktisch immer unmöglicher. Man kann sich ihr nicht mehr entziehen« (P, 43). Wer die technische Gegenwart derart überzeichnet[29], ja hysterisch beschreibt, gibt selbst jegliche Distanz zu ihr auf und kann infolgedessen auch keine kritische Position im Verhältnis zu ihr mehr einnehmen. Noch sind die weiterhin analog Existierenden nicht wie lebendige Fossilien »tölpelhafte Saurier«, die aus alter Zeit mit all ihren Mängeln überlebt haben, ohne recht zu verstehen, was um sie herum vorgeht (P, 15). Weder ist die »totale Verwaltung« alles Gesellschaftlichen, die Theodor W. Adorno vor einem halben Jahrhundert prophezeit hat, nun dank fortgeschrittener Digitalisierung eingetreten, noch auch ist es bislang zu »restlos verwirklichter«, totaler Überwachung im Sinne einer digitalisierten Variante des Bentham'schen Panoptikums gekommen, wie es sich in einem Projekt namens *Ehrliches Shanghai* anzukündigen scheint (P, 17, 19, 49). Was weder Georges Orwell noch Michel Foucault als Theoretiker überwachender Macht in vollem Ausmaß vorausgesehen haben, ist allerdings, wie weitgehend die Menschen sich anscheinend freiwillig überwachen lassen, ohne dass ir-

28 Derzeit wird unter dem Eindruck vieler Fälschungen von Impfausweisen erwogen, den digitalen Nachweis von Impfungen per Gesetz vorzuschreiben. Wenn es dazu kommt, müsste die Freizügigkeit für Bürger, die über kein Smartphone verfügen, empfindlich eingeschränkt werden.

29 Der zitierte Autor ist beileibe nicht der einzige in dieser Hinsicht. Man vergleiche nur, was in Günther Anders' Buch über die Antiquiertheit des Menschen über die Übermacht des »Apparats«, bei einem Kulturkritiker wie Erhart Kästner (*Aufstand der Dinge. Byzantinische Aufzeichnungen*, Frankfurt/M.: Suhrkamp 1979, 216 f.) über die neuzeitliche »Welt-Herstellung« oder bei Jean Baudrillard (*Die göttliche Linke*, Berlin: Matthes & Seitz 1986, 113 ff.) über den angeblichen Triumph einer distanzlosen Sozialität zu lesen steht.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

gendeine Macht ständig eingreifen müsste (P, 51 f.). Weniger die Frage, ob gar kein Ausbruch mehr möglich ist, als vielmehr die Frage, ob man überhaupt noch Sinn für das Interesse hat, *nicht* in digital vermittelter, vergesellschafteter Sichtbarkeit[30] *aufzugehen*, steht so gesehen zur Diskussion.

Wird nicht gerade in diese Sichtbarkeit alles investiert? Wie anders ist es zu verstehen, dass man seine vermeintlichen Alleinstellungsmerkmale über *Instagram* kommuniziert und sich auf diese Weise ›individuell‹ gibt? Wie ist es sonst zu verstehen, dass man sich scheinbar bedingungslos Maßstäben der Resonanz und permanenter Bewertung in einem »Aufmerksamkeitskapitalismus« aussetzt (P, 21, 44), der Schindluder treibt mit dem Auffälligen, Interessanten, Spektakulären, dem Prominenten, dessen Kurs er zugleich ins Bodenlose fallen lässt, sobald es sich erschöpft hat?[31]

Will man nur aufgrund der viel zitierten *fear of missing out*, einer Art Furcht, digital zu ›sterben‹, bedingungslos dabeibleiben, alles ›mitbekommen‹ und nichts ›verpassen‹, um nicht selbst ›verpasst‹ zu werden (P, 27)? Fürchtet man nichts so sehr wie den virtuellen Tod im womöglich irreversiblen Herausfallen aus einer Ökonomie der Aufmerksamkeit[32], die anscheinend alle nur solange (ko-)existieren lässt, wie sie bemerkt und oft genug ›gelikt‹ werden? Überwiegt deshalb eine sehr weit gehende Exhibitionsbereitschaft jede Sorge vor nichts mehr verschonender Ausspionierung – vorausgesetzt, man glaubt nicht ohnehin, man habe *nothing to hide*, nicht einmal dann, wenn jede Bewegung des Mauszeigers aufgezeichnet wird, um anonymes *data mining* zu optimieren (P, 87, 29)? Glaubt man im Ernst, sich einer »unbescholtenen digitalen Lebensführung« zu befleißigen, die nichts zu verbergen hat, also in jeder Hinsicht jederzeit jedem ›transparent‹ gemacht werden dürfte (P, 51) – oder hat man längst jegliche Illusion der Abgrenzbarkeit einer residualen Privatsphäre aufgegeben und hofft am Ende nur noch, in der schieren Masse der wofür auch immer gesammelten Daten unterzugehen, von deren

30 Von Jean-Paul Sartre über Günther Anders bis hin zu Hans Blumenberg wird die fragliche Sichtbarkeit bzw. Visibilität, wie sie bei letzterem heißt, in der einen oder anderen Weise auf Gefährdungen bezogen, die von Anderen ausgehen, unter deren Augen jemand sichtbar wird (sei es als Beobachteter und Beschämter, sei es als potenzieller Feind). Es ist ein Desiderat, diesen Begriff der Sichtbarkeit bzw. der Visibilität in digitalisierungskritischer Perspektive zu erweitern.

31 Auch das ist beileibe keine ganz neue Beobachtung. Man vergleiche nur die kulturkritischen Überlegungen Paul Valérys. Bei Karl Löwith resümiert in: *Paul Valéry. Grundzüge seines philosophischen Denkens*, Göttingen: Vandenhoek & Ruprecht 1971.

32 Vgl. Georg Franck, *Ökonomie der Aufmerksamkeit. Ein Entwurf*, München, Wien 1998.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Verbleib man kaum eine Ahnung hat, zumal wenn einem gerade die digitale Identität zu Zwecken der Erpressung, gezielten Mobbings oder Stalkings entwendet oder verfälscht wird (P, 31, 54, 84)?

Immer entblößter würden sich die *user* des Internet selbst geben, sich aber auch als gewaltsam entblößt erfahren, lesen wir in der Literatur, die diese Entwicklungen kritisch kommentiert, ohne klar und allgemein angeben zu können, was hier auf dem Spiel steht. Etwa: Scham, Schamlosigkeit oder der »Schutz«, dessen jede(r) Einzelne bedürfe (P, 34), um nicht ›restlos‹ algorithmisch erfasst, von gesichtslosen Dispositiven gesichtet und verwertet zu werden – am Ende mit der gnadenlosen Konsequenz, keinerlei Differenz zum eigenen digitalen Image mehr geltend machen zu können.[33] In diesem Fall wäre jede(r) Einzelne geradezu definierbar durch das, was im *World Wide Web* über sie bzw. ihn vorliegt und bekannt ist, ob selektiv genutzt oder nicht. Sollte es etwas darüber hinaus geben, was die Einzelnen ausmacht, so wäre es schlicht nicht nachweisbar, unbeweisbar und insofern vollkommen irrelevant. Es sei denn, sie könnten irgendwie geltend machen, *für sich selbst oder im Verhältnis zu Anderen auf andere Weise zu existieren* als im Modus eines Wissens, das grundsätzlich rückstandslos erfassbar zu sein scheint.[34]

Was für andere Modi der Existenz und der Koexistenz könnten das aber sein?

Vorschläge gibt es: Versichert uns nicht von Geburt an die *Berührung* dessen, dass wir (für Andere) ›da‹ sind, d.h. unserer sozialen Existenz? Gewährleistet die *Vorsicht* der Berührung nicht, dass das Ausgesetztsein an Andere, das mit der Geburt unvermeidlich einhergeht, nicht in Auslieferung an sie umschlägt? Und wahrt die *Diskretion* der Berührung nicht einen subtilen Abstand zum Berührten, ohne den es sich niemals um das Berührtwerden von Anderen *als solchen* handeln könnte?[35] Ist die Berührung mit ihrer Vorsicht und Diskretion nicht ein ausgezeichneter Modus der Achtung der Alterität Anderer, ohne den sie womöglich

33 Vgl. Steffen Mau, *Das metrische Wir. Über die Quantifizierung des Sozialen*, Berlin: Suhrkamp 2017, sowie die Rezension d. Vf., »Das Soziale im Lichte radikaler Infragestellung. Zwischen uralter Sozialität, *liens sociaux* und Wiederkehr der ›sozialen Frage‹«, in: *Philosophischer Literaturanzeiger 70*, Nr. 4 (2017), 374–404.

34 Vgl. Vf., *Prekäre Selbst-Bezeugung. Die erschütterte Wer-Frage im Horizont der Moderne*, Weilerswist: Velbrück Wissenschaft 2012; ders., »Entzugsverlust. Zu einem kritischen Grundgedanken der Phänomenologie des radikal Fremden«, in: Barbara Schellhammer (Hg.), *Zwischen Phänomenologie und Psychoanalyse. Im interdisziplinären Gespräch mit Bernhard Waldenfels*, Baden-Baden: Nomos 2021, 89–104.

35 Vgl. Vf., »Dialogisches Dasein auf der Suche nach der verlorenen Diskretion – im Rückblick auf Martin Buber und Emmanuel Levinas«, in: *Freiburger Zeitschrift für Philosophie und Theologie 67*, Nr. 2 (2020), 444–463.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

niemals unserem ›Zugriff‹ entzogen bleiben könnten? Und hat nicht eine umfangreiche sozialphilosophische Literatur darauf insistiert, dass Andere nur solange wirklich ›Andere‹ für uns sein können, wie sie uns in gewisser Weise selbst im unvorsichtigen, indiskreten und schließlich brutalen Zugriff auch *entzogen bleiben*?

Davon können wir aber, so scheint es, nichts *wissen*. Es handelt sich nicht einfach um soziale Tatsachen, die man methodisch feststellen, beschreiben und erkennen könnte. Vielmehr geht es um praktisch sich vollziehende Formen der Achtung der Alterität Anderer, die *nur im gelebten Verhältnis zu ihnen* wirklich und wirksam sein können. Diese Achtung *bezeugt* die ›Bedeutung‹ der Alterität Anderer noch *in* der Berührung, liefert aber niemals einen Beweis dafür, dass letztere diesen Sinn hat und haben muss.

Durchaus vorstellbar ist, dass wir längst auf dem Weg in eine Zukunft sind, in der man nichts mehr davon ahnen würde, dass wir *nicht ›restlos‹* Gegenstände des Wissens sind[36], das man idealiter digital erfassen und optimieren könnte. Nicht ausgeschlossen scheint, dass man sich in diesem Sinne in mehr oder weniger manipulierten und zur Datenmenge nivellierten Informationen verliert (P, 80, 122, 132) – einschließlich all der Bilder, deren viel zitierte Flut ganz und gar unkenntlich zu machen droht, was möglicherweise nicht in ihnen aufgehen kann.[37] Vielleicht, so mutmaßen manche, sind wir in der Ikonomanie einer alles und jede(n) erfassenden Bildlichkeit bereits verloren und ahnen von einer veritablen Bildlosigkeit nichts mehr, die kein Mangel wäre, sondern uns daran erinnern müsste, dass man sich vom Anderen in gewisser Weise »kein Bildnis« machen kann bzw. sollte (P, 106).

Eines Bilderverbots nach klassischem Vorbild bedürfte es eigentlich gar nicht, wenn es stimmt, was die Dichterin Ingeborg Bachmann in ihrem Erzählband *Das dreißigste Jahr* die Protagonistin Charlotte sagen lässt: »Ich bin in kein Bild hineingeboren«[38] – und wenn wenigstens Liebe »aus jeglichem Bildnis« wieder »befreien« könnte, wie Max Frisch annahm.[39] Aber das hindert Andere erfahrungsgemäß doch nicht daran, sich von sich selbst und Anderen ein Bild zu machen – ein Bild bzw. endlos viele Bilder, wie es portable bildgebende Techniken heute außerordentlich leicht machen, oder auch Bilder in der weiteren Bedeutung des Wortes, die epistemische Repräsentationen einschließt, welche man auch

36 Beileibe kein ganz neues Problem; vgl. Vf., »Bewegt von dem, was übrig bleibt? Robert Musils Roman *Der Mann ohne Eigenschaften* mit Blick auf Hans Blumenbergs und Milan Kunderas Schriften zur Literatur«, in: *Journal Phänomenologie* 56 (2021), 70–82.

37 Anders, *Die Antiquiertheit des Menschen,* 3 f.

38 Ingeborg Bachmann, *Das dreißigste Jahr. Erzählungen*, München: dtv [22]1988, 107.

39 Max Frisch, *Tagebuch 1946–1949*, Frankfurt/M.: Suhrkamp 1973, 31.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

algorithmisch fabrizieren kann. In einer Gesellschaft, die Hannah Arendt zufolge ohnehin darauf hinauslaufen muss, »alles, was vor der Allgemeinheit erscheint, für jedermann sichtbar und hörbar zu machen«[40], und in diesem Sinne eine öffentliche Welt etabliert, ist das scheinbar unvermeidlich, wenn die fragliche Allgemeinheit die Form eines weltweiten digitalen Netzes annimmt, an das irgendwann jede(r) anzuschließen wäre, um auf diese Weise mit zahllosen Unbekannten wenigstens eine »lose Gesellschaft« zu bilden (P, 119). Eine harmlose Allmende wird so freilich nicht herauskommen (P, 147), wenn die Beteiligten dabei nicht wissen, was sie tun bzw. wie es genannt werden soll, was sie tun oder unterlassen.

Traf es aber nicht seit jeher zu, dass die Menschen immer neue technische Dinge und Verfahren entdeckten, die sie, wenn überhaupt, nur nachträglich zu ›verstehen‹ begannen? Konnte man im Geringsten wissen, welche Folgen die Nutzung von Bronze und Eisen, die Erfindung des Schießpulvers in China oder des Buchdrucks, elektrischer Fernkommunikation seit der Verlegung des ersten Unterwasserseekabels zwischen England und den USA, die Erforschung radioaktiver Prozesse durch Otto Hahn und Lise Meitner sowie die Entschlüsselung des biochemischen Aufbaus der DNS durch James D. Watson und Francis Crick (1953) haben würden?[41] Galt besonders in technischer Hinsicht nicht seit jeher, dass die Menschen ›nach vorwärts‹ lebten und allenfalls nachträglich, wenn überhaupt, zu begreifen begannen, was sie taten? Søren Kierkegaard stellte sich bekanntlich so die hermeneutische Struktur menschlichen Lebens vor: Wir ›leben vorwärts‹, verstehen aber nur ›rückwärts‹.

Für die digitalisierten Strukturen menschlicher Kommunikation scheint ähnlich zu gelten: Die binnen weniger Jahre vollzogene Durchdringung des Lebens wenn nicht aller, so doch der Mehrheit von Milliarden Menschen mit digitalen Techniken hat derart penetrante Ausmaße und Formen angenommen, dass alles, was wir uns unter Welt und Wirklichkeit, Sein und Tatsachen, Realität, Virtualität und Bildlichkeit, Sprechen und Sichmiteinanderverständigen etc. vorstellen, davon betroffen ist. Wir stecken noch mitten drin in diesem Prozess und können schwerlich eine Position einnehmen, die es erlauben würde, ihn *ganz* von außen zu beobachten. Wir wissen nicht, was wir tun, oder ob wir es wenigstens

40 Hannah Arendt, *Vita activa oder Vom tätigen Leben*, München, Zürich: Piper ⁴1985, 49 ff.

41 Was letztere angeht, so vermutete der Chemiker Wendell M. Stanley auf einem internationalen Kongress des Jahres 1955 immerhin schon, es werde bald möglich sein, »die lebendige Substanz« eines jeden Organismus »nach Belieben auf- und ab[zu]bauen«; zit. n. Martin Heidegger, *Gelassenheit*, Pfullingen: Neske ²1960, 22.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

in absehbarer Zeit herausfinden werden; wer ›wir‹ ist, und in welchem Sinne es sich um ein ›Tun‹ handelt.

So soll jedoch keineswegs einem hermeneutischen Nihilismus das Wort geredet werden. Gerade kommunikative Verwerfungen bieten nämlich die Chance, einer unbegrenzten Auslieferung an die Kontingenz und Arbitrarität technischer Innovationen zu entgehen, die manche Beobachter bereits in eine posthumane Zukunft münden sehen, von der offenbar niemand zu sagen weiß, in welchem Sinne – und zwischen was für ›Subjekten‹, *Cyborgs* usw. – in ihr noch Kommunikation stattfinden würde, die ihren Namen verdient.[42] Vielleicht wird es sich nicht mehr um Menschen und um die ihnen attestierte Würde und Autonomie handeln, so wie man sie bislang zu kennen glaubte (ohne deren ›Wesen‹ bzw. das, was sie als solche ausmacht, ›definieren‹ zu können).[43] Vielleicht werden sie ein Leben ohne Dramatik und Tragik führen[44], das am Ende nur ›kaputtgeht‹ wie irgendein Apparat, nach dessen Vorbild Julien O. LaMettrie bekanntlich seine programmatische Schrift *L'homme machine* im 18. Jahrhundert konzipierte. Vielleicht werden sie nicht mehr ›andächtig‹ Musik hören, wenn sie selbst im Lärm des Autoverkehrs wie auch *Podcasts* und ›*gestreamte*‹ Aufnahmen aller Art jederzeit und überall ›abrufbar‹ ist. Vielleicht werden sie kommunikativ hochflexible, jederzeit *multitasking*fähige Un-Wesen, die *sich selbst* genauso wie das, was sie digital vermittelt austauschen, nur noch als eine »Information« oder »Nachricht« auffassen, wie es der Kybernetiker Norbert Wiener bereits vor Jahrzehnten vorgeschlagen hat.[45] Informationen und Nachrichten kann man elektronisch übermitteln und empfangen; aber erfordern sie gespannte, exklusive Aufmerksamkeit in der Form, wie man Anderen ›Gehör schenkt‹?

Wer gleichzeitig sein Handy bedient und im Internet unterwegs ist, mag glauben, jemandem dennoch ›zuhören‹ zu können. Aber ob es dazu kommt, wird *unvermeidlich* auch davon abhängen, ob der Andere seinerseits diese Erfahrung macht. Niemand kann – solange wir von einem überkommenen Kommunikationsverständnis ausgehen – *allein* entscheiden, was *zwischen* zweien oder gar mehreren stattfindet. Reden,

42 Vf., »Ansatzpunkte praktisch-negativistischer Anthropologie. Anmerkungen zur Frage, ›ob der Posthumanismus ein Humanismus ist‹«, in: Gerald Hartung, Matthias Herrgen (Hg.), *Interdisziplinäre Anthropologie. Jahrbuch 06/2018: Das Eigene & das Fremde,* Wiesbaden: Springer VS 2019, 129–156.

43 Philipp Stoellger (Hg.), *Figurationen des Menschen. Studien zur Medienanthropologie*, Würzburg: Königshausen & Neumann 2019.

44 Die Rede ist bereits von einer »neuen Spezies« derer, deren Kindheit und Jugend in der Anwendung digitaler Geräte ende (P, 45).

45 Norbert Wiener, *Mensch und Menschmaschine. Kybernetik und Gesellschaft*, Bonn: Athenäum ³1966, Kap. VI.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Sichverständigen, Streiten usw. findet unvermeidlich *zwischen uns* statt und kann infolgedessen niemals einseitig jemandem zu Gebote stehen. So muss jeder, der kein Problem damit zu haben glaubt, alte (analoge) und neue (digitale) Kommunikationsformen sogar gleichzeitig miteinander zu verwenden, damit rechnen, auf *bestimmte Negationen* zu stoßen, die im Kern besagen: ›So kommunizierst du nicht, nicht mit mir, nicht um diesen Preis‹ (geteilter Aufmerksamkeit, nicht-exklusiver Zuwendung im Modus des Zuhörens bzw. Gehörschenkens usw.).

Entsprechende Einwände mag man als anachronistische Ausdrücke der Verbitterung seitens jener abtun, die offenbar die viel zitierten ›Zeichen der Zeit‹ nicht erkannt haben und unverdrossen an einem analogen, vielleicht abgewirtschafteten Verständnis von Reden, Dialog, Diskursen und Verständigung festhalten, das es erforderlich macht, sich leibhaftig Anderen als solchen zuzuwenden und sich von ihnen in Anspruch nehmen zu lassen. Aber das ist natürlich kein Argument. Statt die besagten kommunikativen Verwerfungen zwischen *digital natives* einerseits und analogen Verfechtern einer anscheinend rückständigen, ja primitiven Art und Weise der Kommunikation andererseits in schiere Verständnislosigkeit münden zu lassen, in der die Einen nicht (mehr) wissen, was die Anderen (noch oder schon) tun, käme es darauf an, sie fruchtbar zu machen. Möglicherweise fordert gerade durch neue, digital vermittelte Formen von Kommunikation hervorgerufener Widerspruch zu einer Neujustierung dessen heraus, was es bedeutet, *mit Anderen als solchen* zu sprechen, ihnen zu schreiben, zuzuhören und zu antworten.

Ironischerweise hat es sich mit diesen radikalen Fragen gerade ein apologetischer Dialogismus allzu leicht gemacht, der emphatisch von »unmittelbarer«, »wahrhaftiger« und »tiefer« Begegnung mit dem Anderen als Anderem von Angesicht zu Angesicht spricht und dabei jegliche Möglichkeit verspielt, technischen Vermittlungen irgendeine positive Bedeutung abzugewinnen.[46] Allerdings insistiert er auf einer *unaufhebbaren Anderheit* des Anderen[47], von der Lobredner auf ein kommendes posthumanes, weitgehender Digitalisierung unterworfenes Zeitalter vielfach keine Spur mehr verraten. Was sie als digital vermittelte Kommunikationsformen beschreiben, läuft so gesehen auf die Fortschreibung einer viel älteren, keineswegs erst auf die rezenten Digitalisierungsprozesse oder auf die ›moderne‹ Technik zurückzuführenden weitgehenden

46 Insofern wende ich mich gegen eine teils resignative, teils reaktionäre Kulturkritik, die wie üblich ›früher‹ grundsätzlich das Bessere realisiert sieht. Die angesprochenen kommunikativen Verwerfungen könnten durchaus auch dazu führen, genau das mitsamt der Begrifflichkeit, in der man zuvor menschliche Kommunikation beschrieben hat, ganz und gar in Zweifel zu ziehen.

47 Wie v.a. Martin Buber, *Das dialogische Prinzip*, Heidelberg: L. Schneider 1962.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

»Alteritätsvergessenheit« (Marcel Hénaff) hinaus[48], die letztlich alles zu tilgen droht, was nicht in algorithmisch erfassbaren Daten, Informationen und Wissensbeständen aufgehen kann.

In kommunikativen Verwerfungen zwischen *digital natives* und analogen Aborigines, die im Ernst noch glauben mögen, Kommunikation wirklich *ab origine*, nämlich im Zeichen ihres in jedem Gespräch sich neu ereignenden Ursprungs des Verhältnisses zum Anderen als solchen zu vollziehen, steht so gesehen vor allem dieses ›Nicht-Aufgehende‹ auf dem Spiel. Welche Bedeutung hat es, sollte oder müsste es haben, algorithmisch, digital und epistemisch *nicht ›restlos‹* erfassbar und erkennbar zu werden? Können die Verfechter digitalisierter Kommunikation damit buchstäblich nichts anfangen? Können umgekehrt die Verteidiger analoger Kommunikation ohne Weiteres angeben, was letztere ›wirklich‹ bzw. ›in Wahrheit‹ ausmachen müsste?

Für beide Seiten stellt sich die Lage nicht derart einfach dar. Zweifellos sind keineswegs alle Liebhaber digital vermittelter Kommunikationsstrukturen dazu bereit, sich ihnen bedingungslos auszuliefern. Manche gehen so weit, nicht nur das Private, sondern auch eine geradezu ›romantisch‹ anmutende Innerlichkeit und, nicht zuletzt, eine unverfügbare Würde gegen jeglichen Zugriff Anderer in Schutz nehmen zu wollen.[49] Umgekehrt müssen selbst die radikalsten Vertreter einer Ethik der Alterität, die sie in der Sorge zu *verabsolutieren* neigen, sie könnte als bloß relative andernfalls kommunikativ ›aufgehoben‹ werden, zugeben, dass sie darauf angewiesen ist, sozial bewahrheitet zu werden. So weit phänomenologische und hermeneutische Methoden reichen, sind weit und breit keine Spuren *absoluter* Anderheit, allenfalls *radikaler* Alterität zu finden, die sich als uns entzogene bemerkbar macht, ohne sich je gänzlich von jeglichem Erfahrungs*bezug* zu lösen.

Was aber sollte von vornherein dagegen sprechen, dass dieser Bezug auch digital vermittelt stattfinden kann? Längst gibt es *virtual memorials* im Internet, wo nicht nur die Trauer der Angehörigen und Hinterbliebenen weltweit nach neuen Möglichkeiten des Gedenkens sucht. So wenig wie das fotografische Bild Betrauerter jegliche Spur des Abgebildeten tilgen

48 Vgl. die aktuelle Zwischenbilanz: Vf., Werner Stegmaier, *Orientierung und Ander(s)heit. Spielräume und Grenzen des Unterscheidens,* Hamburg: Felix Meiner 2022.

49 Vgl. P, 151. Der Rekurs auf die Würde wirkt allerdings vielfach ebenso ratlos und als »Verwürdigung« des Anderen in jeder Hinsicht unglaubwürdig wie die Berufung auf das viel zitierte Nicht-Identische im Sinne Adornos als angeblich »fundamental Anderes« (P, 154). Pejorativ von einer »Verwürdigung« spricht Thomas Mann in einem anderen Zusammenhang mit Blick auf eine moralische Überhöhung des Künstlers, die Gefahr laufe, ihm das Lachen über sich selbst auszutreiben (*Essays. Bd. 2. Politik,* Frankfurt/M.: Fischer 1977, 329).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

muss[50], so wenig muss digital vermittelte Kommunikation von vornherein jegliches Verhältnis zum radikal Anderen als solchem zum Verschwinden bringen. In beiden Fällen kommt es vielmehr auf den *Gebrauch* an, den wir von der jeweiligen Technik machen – wobei zunächst allerdings ein technisches bzw. technizistisches Verständnis davon, was ›Gebrauch‹ bedeutet, fernzuhalten ist. Andernfalls würden wir von vornherein auch ›analoge‹ Kommunikation, in der es ebenfalls auf den ›Gebrauch‹ kommunikativer Mittel ankommt, einem solchen Verständnis unterwerfen.

Seit langem verkündet man uns den Advent eines »Neuen Menschen« und verheißt gar eine »Mutation der Menschheit« – sei es noch (wie bei Pierre Bertaux) *ohne*, sei es bereits *mit* konkreten Aussichten auf einen redigierten Genetischen Code, in denen Kritiker wie Francis Fukuyama wiederum das unmittelbar bevorstehende »Ende des Menschen« meinen erkennen zu können.[51] Weder enthusiastischer noch alarmierter Futurismus wird uns indessen in jenen kommunikativen Verwerfungen weiterhelfen, wo es darum geht, zu verstehen, was zwischen uns geschieht, sowohl im alltäglichen Gespräch wie auch in der langfristigen Diachronie intergenerationeller Beziehungen, in denen derzeit alle Grundbegriffe der Kommunikation auf dem Spiel stehen: die Anrede und der Anspruch an den Anderen ebenso wie das Zuhören, Antworten und Erwidern; die Aufmerksamkeit, die man Anderen als solchen widmet, ebenso wie das Gehör, das man ihnen schenkt oder selbst findet; das Verstehen, das man sucht, ebenso wie die Verständigung, in der man regelmäßig scheitert; die Frage, was Scheitern in kommunikativer Hinsicht bedeutet, ebenso wie die Frage, wie es abzuwenden wäre und wie möglicherweise gelingende Kommunikation vorzustellen wäre, von der man kaum noch einen Begriff hat.[52] Vielleicht gerade deswegen, weil bestimmte Medien mit ihren triumphalen technischen Erfolgen, die sie global feiern, in sogenannter ›Echtzeit‹ jeglichen Abstand zwischen uns zu tilgen scheinen, weil sie alles Kommunizierte weltweit distribuieren und jederzeit ›abrufbar‹ machen können: alles je Geschriebene und alles je Gesagte, alles je Gefilmte, Dokumentierte und Manipulierte...

Zwar sind die entsprechenden Algorithmen nur Eingeweihten und Geheimdiensten wie der NSA bekannt; zwar unterlaufen hochintelligente

50 Roland Barthes, *Die helle Kammer*, Frankfurt/M.: Suhrkamp 1989; Vf., *Revisionen der Trauer. In philosophischen, geschichtlichen, psychoanalytischen und ästhetischen Perspektiven*, Weilerswist: Velbrück Wissenschaft 2006.

51 Pierre Bertaux, *Mutation der Menschheit. Zukunft und Lebenssinn*, Frankfurt/M.: Suhrkamp 1979; Gottfried Küenzlen, *Der Neue Mensch. Eine Untersuchung zur säkularen Religionsgeschichte der Moderne*, Frankfurt/M.: Suhrkamp 1997; Francis Fukuyama, *Das Ende des Menschen*, München: Kindler 2004.

52 Vgl. Vf. (Hg.), *Emmanuel Levinas: Dialog. Ein kooperativer Kommentar*, Freiburg i. Br., München: Alber 2020.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Hacker und sonstige Manipulateure von *software*, Daten und Identitäten im *dark net* die allgemeine öffentliche Publizität, die sich Kosmopoliten von einem wirklich freien, jeder und jedem zugänglichen Internet versprochen haben mögen. Insofern kann keine Rede davon sein, wir alle seien mehr oder weniger gleichermaßen Teilhaber an weltweiter digitaler Vernetzung. Doch ist die Furcht nicht unbegründet, dass jede Form der Teilhabe den Preis einer Reduktion auf das zu zahlen hat, was digital erfasst werden kann, wenn auch in manipulierter Form wie im Fall gestohlener Identitäten, gegen die die Bestohlenen kaum noch ankommen.

Um die Bedeutung dieser Reduktion überhaupt ausloten zu können, müsste man wenigstens anzugeben versuchen, wer oder was denn überhaupt ›reduziert‹ zu werden droht, und man müsste dabei damit rechnen, dass es sich nicht um ontisch ›dingfest‹ zu Machendes handelt. Wer oder was sind wir, die wir – ob mit Recht oder nicht – fürchten, als vermeintlich bereits digital Geborene, auf jeden Fall aber global Vernetzte und insofern auch digital Koexistierende, reduziert zu werden auf die Spuren, die wir im Netz hinterlassen, auf die Daten, die man über uns sammeln kann, und auf die Identitäten, die uns über alles hinaus, was selbst engste Freunde und Verwandte je über uns in Erfahrung bringen und erinnern können, identifizierbar machen? Darin liegt eine *digitale Alterität* eigener Art, die paradoxerweise damit zu tun hat, dass wir – bzw. einige von uns nach wie vor, beileibe freilich nicht alle – eine weitgehende und hochkomplexe *Reidentifizierbarkeit als Reduktion*, ja als eine Form von *Gewalt* erfahren.

In kommunikativen, Gesprächsabbrüche heraufbeschwörenden Verwerfungen zwischen naiven Nutzern digitaler Medien einerseits und weniger naiven analogen Aborigines andererseits dürften sich diese Fragen ironischerweise als gemeinsamer Brennpunkt abzeichnen, durch den man wieder ins Gespräch kommen könnte: Inwiefern wollen weder wir selbst noch Andere ›restloser‹ Identifizierbarkeit überantwortet werden? Inwiefern erscheint es uns und Anderen als wesentlich, im Register reidentifizierbarer Selbigkeit (*sameness*; *mêmeté*) *nicht* aufzugehen? Inwiefern erscheint uns der gegenteilige Glaube, wir könnten geradezu mit dem zusammenfallen, was ›das Netz‹ oder Instanzen wie die NSA dort über uns ›in Erfahrung bringen‹ können, als eine Form der Gewalt, die wir ablehnen und ablehnen müssen? Liegt das daran, dass wir glauben, dass man im Grunde *noch mehr* über uns wissen müsste, mehr als unsere Freunde und Verwandten, unsere Nachbarn, Kollegen und sogar als wir selbst je von uns selbst wissen und erinnern können? Oder liegt das vielmehr gerade an dem, was sich algorithmischer Erfassung, Remix- und Memifizierungsverfahren (P, 140 f., 118) entzieht und uns gerade deshalb als unverzichtbar erscheint? Dann würde paradoxerweise gerade in einem *Weniger* an Erfassbarem, Wissbarem und Erkennbarem jenes *Mehr* im Sinne eines Überschusses liegen, durch den wir uns

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

im Verhältnis zu Datenkraken wie *Google* und *Facebook,* zu Geheimdiensten und uns ausspionierenden Companies aller Couleur als ›Andere‹ erfahren.

Wenn es sich so verhält, könnte man nun meinen, besteht ja gar kein Anlass zu besonderer Sorge. Wenn es ›etwas‹ gibt, was sich der Identifizierbarkeit radikal entzieht, sind diese Konzerne und Dienste dann nicht von vornherein zum Scheitern verurteilt? Oder gelingt es ihnen, auf der Basis jener *fear of missing out* jeden derart dazu anzuhalten, sich ihnen auszuliefern, dass am Ende jeder glauben muss, nichts über das hinaus zu sein, was sie über jeden Einzelnen ›in Erfahrung bringen‹ können bzw. schon gespeichert haben? Kann es wirklich als ausgeschlossen gelten, dass wir bereits auf dem Weg in eine Zukunft sind, in der infolge der nachhaltigen Wirkung dieser Auslieferung niemand auch nur die Frage noch verstehen würde, was überhaupt einmal damit gemeint gewesen sein könnte, dass wir kraft radikaler Alterität niemals im algorithmisch Erfassbaren, zu Wissenden, Identifizierbaren und Erkennbaren aufgehen können? Wie auch sollte man dem eine positive Bedeutung zuerkennen?

Was sich an vielfachen Fronten in kommunikativen Verwerfungen zwischen *digital natives* und analogen Aborigines derzeit abspielt, hat in der skizzierten Perspektive nicht allein die Bedeutung von Missverständnissen, deformiertem Gebrauch von Verben wie ›reden‹, ›schreiben‹, ›antworten‹ usw. bzw. von Ungeduld und Unwilligkeit angesichts neuer Formen unhöflicher Unaufmerksamkeit bei der deplatzierten Verwendung digitaler Kommunikationstechniken, die bereits in dem Verdacht stehen, sozialer Verwahrlosung Vorschub zu leisten.[53] Es geht, weit radikaler, nicht zuletzt auch darum, ob in diesen Verwerfungen eine Art Austausch darüber stattfinden kann, ob wir noch wissen, was wir tun, wenn wir uns im Verhältnis zueinander solcher Techniken bedienen, und ob es im Geringsten verzeihlich sein kann, sich dieser Frage überhaupt nicht zu stellen; ob es zulässig oder zuträglich sein kann, dennoch ungerührt weiterzumachen mit neuen Formen von Kommunikation, die analoge Aborigines möglicherweise gar nicht mehr als solche begreifen können.

Das zweifelhafte Ergebnis wäre, nicht zum ersten Male, dass es sich herausstellt, dass ›wir‹ diejenigen sind, die im Verhältnis zueinander ›nichts gemeinsam‹ haben, nichts bis auf eine kümmerliche Art der Koexistenz bzw. des Mit-da-seins, das niemanden mehr lehrt, welchen kommunikativen Sinn es eigentlich hat, kann es sich doch offenbar auf technisch höchstem Niveau gänzlich indifferent zu der Frage verhalten, was uns Andere als solche überhaupt ›angehen‹ – solange sie oder irgendwelche *Bots,* Roboter oder Anrufbeantworter uns nur irgendwelche

53 Vgl. Jörg Zirfas, »Vulnerabilität. Anthropologie und kulturelle Bildung«, in: Ulaş Aktaş (Hg.), *Vulnerabilität. Pädagogisch-ästhetische Beiträge zu Korporalität, Sozialität und Politik*, Bielefeld: transcript 2020, 141–160, hier: 155.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Nachrichten zukommen lassen, sei es auch eine bloße Abwesenheitsnotiz. Am Ende hätte man dann von der längst festzustellenden Abwesenheit des Anderen als des Anderen keinerlei Notiz genommen.[54] Und von der scheinbar schlechten Nachricht jener analogen Aborigines, die Alterität des Anderen, die uns über alles algorithmisch Erfassbare und Erkennbare hinaus gerade durch das Weniger oder Nichts an Erfassbarkeit und Erkennbarkeit zu ›mehr‹ macht, drohe in Vergessenheit zu fallen, wüssten *digital natives* dann nicht einmal mehr zu sagen, was sie überhaupt besagen soll. So würden sich die Welten von analogen Aborigines und *digital natives* schließlich voneinander lösen und es käme zu einem inkommunikablen Schisma, in dem überbordende Kommunikation triumphierte und zugleich Sprachlosigkeit herrschen könnte.

Literatur

Aichinger, Ilse (1960): *Die größere Hoffnung*, Hamburg: Fischer.

Anders, Günther (1956): *Die Antiquiertheit des Menschen. Über die Seele im Zeitalter der zweiten industriellen Revolution*, München: C. H. Beck.

Arendt, Hannah ([4]1985): *Vita activa oder Vom tätigen Leben*, München/Zürich: Piper.

Bachmann, Ingeborg ([22]1988): *Das dreißigste Jahr. Erzählungen*, München: dtv.

Barth, Karl (1962): *Mensch und Mitmensch. Die Grundform der Menschlichkeit*, Göttingen: Vandenhoek & Ruprecht.

Barthes, Roland (1989): *Die helle Kammer*, Frankfurt/M.: Suhrkamp.

Baudrillard, Jean (1986): *Die göttliche Linke*, Berlin: Matthes & Seitz.

Bergoglio, Jorge Mario (2016): *Ansprache zur Verleihung des Karlspreises* [Rede]. https://www.vatican.va/content/francesco/de/speeches/2016/may/documents/papa-francesco_20160506_premio-carlo-magno.html.

Bertaux, Pierre (1979): *Mutation der Menschheit. Zukunft und Lebenssinn*, Frankfurt/M.: Suhrkamp.

Blumenberg, Hans (2006): *Beschreibung des Menschen*, Frankfurt/M.: Suhrkamp.

Buber, Martin (1962): *Das dialogische Prinzip*, Heidelberg: L. Schneider.

54 Mit Recht deutet Hans Blumenberg in seiner *Beschreibung des Menschen* an, dass »zunehmende Körperausschaltung«, wie sie mit digitalen Techniken einhergeht (597, 777, 862), sowohl einen Verzicht auf Unmittelbarkeit, als auch »Entbehrlichkeit der Anwesenheit der Sache« bzw. des Anderen als solchen erforderlich machen und erleichtern kann (599). *Welt* wäre am Ende dann das, »wovon wir auch abwesend sein können« (602). Doch wie weit kann das gehen? Gerät schließlich auch das aus dem Blick, was bzw. wer ›abwesend‹ bleibt? Hans Blumenberg, *Beschreibung des Menschen*, Frankfurt/M.: Suhrkamp 2006.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Enzensberger, Hans Magnus (1989): *Mittelmaß und Wahn. Gesammelte Zerstreuungen*, Frankfurt/M.: Suhrkamp.
Földényi, Lázló F. (2019): *Lob der Melancholie. Rätselhafte Botschaften*, Berlin: Matthes & Seitz.
Fontane, Theodor ([3]1993): *Der Stechlin. Roman*, München: dtv.
Franck, Georg (1998): *Ökonomie der Aufmerksamkeit. Ein Entwurf*, München/Wien: Hanser.
Frisch, Max (1973): *Tagebuch 1946–1949*, Frankfurt/M.: Suhrkamp.
Fukuyama, Francis (2004): *Das Ende des Menschen*, München: Kindler.
Heidegger, Martin ([2]1960): *Gelassenheit*, Pfullingen: Neske.
Kaplow, Ian (2002): *Analytik und Ethik der Namen*, Würzburg: Königshausen & Neumann.
Kästner, Erhart (1979): *Aufstand der Dinge. Byzantinische Aufzeichnungen*, Frankfurt/M.: Suhrkamp.
Küenzlen, Gottfried (1997): *Der Neue Mensch. Eine Untersuchung zur säkularen Religionsgeschichte der Moderne*, Frankfurt/M.: Suhrkamp.
Kundera, Milan (2008): »Der existenzielle Sinn der bürokratisierten Welt«, in: ders., *Der Vorhang*, Frankfurt/M.: Fischer, 179–184.
Levinas, Emmanuel (1987): *Totalität und Unendlichkeit. Versuch über die Exteriorität* [1961], Freiburg i. Br./München: Alber.
Liebsch, Burkhard (2004): »Ereignis – Erfahrung – Erzählung. Spuren einer anderen *Ereignis-Geschichte:* Henri Bergson, Emmanuel Levinas, Paul Ricœur«, in: Marc Rölli (Hg.), *Ereignis auf Französisch. Von Bergson bis Deleuze,* München: Fink, 183–207.
Liebsch., Burkhard (2006): *Revisionen der Trauer. In philosophischen, geschichtlichen, psychoanalytischen und ästhetischen Perspektiven*, Weilerswist: Velbrück Wissenschaft.
Liebsch, Burkhard (2012): *Prekäre Selbst-Bezeugung. Die erschütterte Wer-Frage im Horizont der Moderne*, Weilerswist: Velbrück Wissenschaft.
Liebsch, Burkhard (2017): »Das Soziale im Lichte radikaler Infragestellung. Zwischen uralter Sozialität, *liens sociaux* und Wiederkehr der ›sozialen Frage‹«, in: *Philosophischer Literaturanzeiger 70*, Nr. 4 (2017), 374–404.
Liebsch, Burkhard (2019): »Ansatzpunkte praktisch-negativistischer Anthropologie. Anmerkungen zur Frage, ›ob der Posthumanismus ein Humanismus ist‹«, in: Gerald Hartung, Matthias Herrgen (Hg.), *Interdisziplinäre Anthropologie. Jahrbuch 06/2018: Das Eigene & das Fremde,* Wiesbaden: Springer VS, 129–156.
Liebsch, Burkhard (2020): »Dialogisches Dasein auf der Suche nach der verlorenen Diskretion – im Rückblick auf Martin Buber und Emmanuel Levinas«, in: *Freiburger Zeitschrift für Philosophie und Theologie 67,* Nr. 2 (2020), 444–463.
Liebsch, Burkhard (Hg.) (2020): *Emmanuel Levinas: Dialog. Ein kooperativer Kommentar*, Freiburg i. Br./München: Alber.
Liebsch, Burkhard (2021): »Entzugsverlust. Zu einem kritischen Grundgedanken der Phänomenologie des radikal Fremden«, in: Barbara Schellhammer (Hg.), *Zwischen Phänomenologie und Psychoanalyse. Im*

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

interdisziplinären Gespräch mit Bernhard Waldenfels, Baden-Baden: Nomos, 89–104.
Liebsch, Burkhard (2022): »Bewegt von dem, was übrig bleibt? Robert Musils Roman *Der Mann ohne Eigenschaften* mit Blick auf Hans Blumenbergs und Milan Kunderas Schriften zur Literatur«, in: *Journal Phänomenologie,* Nr. 56, 70–82.
Liebsch, Burkhard; Stegmaier, Werner (2022): *Orientierung und Ander(s)heit. Spielräume und Grenzen des Unterscheidens,* Hamburg: Felix Meiner.
Löwith, Karl (1971): *Paul Valéry. Grundzüge seines philosophischen Denkens*, Göttingen: Vandenhoek & Ruprecht.
Mann, Thomas (1977): *Essays. Bd. 2. Politik,* Frankfurt/M.: Fischer.
Mau, Steffen (2017): *Das metrische Wir. Über die Quantifizierung des Sozialen*, Berlin: Suhrkamp.
Plessner, Helmuth (1982): *Mit anderen Augen. Aspekte einer philosophischen Anthropologie*, Stuttgart: Reclam.
Posselt, Gerald (2005): *Katachrese,* München: Fink.
Postman, Neil (2000): »Müssen Toaster sprechen?«, in: *Die Gegenwart der Zukunft*, Berlin: Wagenbach.
Schmitt, Peter (2021): *Postdigital. Medienkritik im 21. Jahrhundert*, Hamburg: Meiner.
Stoellger, Philipp (Hg.) (2019): *Figurationen des Menschen. Studien zur Medienanthropologie*, Würzburg: Königshausen & Neumann.
Taureck, Bernhard H. F. (2004): *Metaphern und Gleichnisse in der Philosophie. Versuch einer kritischen Ikonologie der Philosophie,* Frankfurt/M.: Suhrkamp.
Taureck, Bernhard H. F. (2014)*: Überwachungsdemokratie. Die NSA als Religion*, Paderborn: Fink.
Weisbrod, Lars (2021): »Die Klotür geht erst wieder auf, wenn du die Hände gewaschen hast«. Interview mit Dave Eggers. In: *Die ZEIT*, Nr. 40/2021.
Wiener, Norbert (³1966): *Mensch und Menschmaschine. Kybernetik und Gesellschaft*, Bonn: Athenäum.
Zirfas, Jörg (2020): »Vulnerabilität. Anthropologie und kulturelle Bildung«, in: Ulaş Aktaş (Hg.), *Vulnerabilität. Pädagogisch-ästhetische Beiträge zu Korporalität, Sozialität und Politik*, Bielefeld: transcript, 141–160.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Jonathan Harth / Martin Feisst[1]

Neue soziale Kontingenzmaschinen

Überlegungen zu künstlicher sozialer Intelligenz am Beispiel der Interaktion mit GPT-3

»Any sufficiently advanced technology is indistinguishable from magic«.

Arthur C. Clarke

»It's still magic even if you know how it's done.«

Terry Pratchett

Große Sprachmodelle (Large Language Models, LLM) wie das im Jahr 2020 öffentlich gewordene Modell GPT-3 zeigen in eindrücklicher Weise, was es heißt, wenn das Soziale auf sich selbst zurückgeworfen wird. Trainiert durch einen enormen Wissensvorrat in Textform und ohne Zugang zu sensorischer Wahrnehmung oder Körperlichkeit, gelingt es dem neuronalen Netz des Sprachmodells in einer nie dagewesenen Weise kohärente, kluge, überraschende und abwechslungsreiche sowie anregende Konversation zu betreiben. Im neuronal strukturierten Sprachmodell GPT-3 kommt das Soziale als sprachlich vermittelte Kommunikation gewissermaßen zu sich selbst, nämlich indem ausschließlich durch die wahrscheinlichkeitsgetriebene (Re-)Kombination von Worten neuer Sinn in der Welt erwirtschaftet wird.

Wenngleich diese rein textuell trainierten Sprachmodelle – anders als das menschlichen Gehirn – nicht auf eine eigene Systemzeit zurückgreifen können und nur über ein beschränktes Gedächtnis verfügen, gelingt es ihnen unter anderem basale *Theory of Mind*-Tests zu bestehen (Aguera y Arcas 2021), sich mathematische Rechenwege zu erschließen (Cobbe et al. 2021), Haikus zu verfassen (Adams 2021), Spiele zu entwickeln (AI Dungeon), IQ-Tests teils besser zu bestehen als Menschen (Thompson 2021b), neue Poesie zu verfassen (Aalho 2021) sowie kluge und überraschende Konversationen zu betreiben.

Zum ersten Mal führt uns Technik damit vor, wie Sprachverständnis und Intelligenz ganz losgelöst von verkörperten, wahrnehmungsgeleiteten und emotionsbegleiteten Entitäten (*We're looking at you, human animal!*) erfolgreich prozessiert werden kann. Die sich hierin versteckende

1 Unser Dank gilt Werner Vogd und Nele Kost für ihre wertvollen Anmerkungen sowie die Unterstützung in der Erstellung dieses Beitrags!

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Kränkung – Sprache und Intellekt war doch eben noch exklusiv uns Menschen und vielleicht noch einigen ausgewählten Tieren vorbehalten – ist jedoch für die Soziologie keine besonders große Überraschung. Gründet doch die Ausdifferenzierung dieser Wissenschaft auf der Beobachtung, Soziales durch Soziales erklären zu können und dabei bereits vom Biologisch-Körperlichen oder Geistigen abstrahieren zu können (Durkheim 1980). Als menschgemachte Veränderung in der Welt erscheint künstliche Intelligenz (KI) in diesem Lichte dann als eine weitere Möglichkeit zur Selbstvergewisserung bzw. Selbstaufklärung des Menschen.

In diesem Beitrag möchten wir anhand des Beispiels eigener Interaktionen mit GPT-3 auf einige ausgewählte neue Reflexionsperspektiven eingehen, die hier in Erscheinung treten. Anhand dieser Interaktionen wird deutlich, dass a) jede Kommunikation immer nur in einem Rahmen Sinn ergibt, b) das Geheimnis der DU-Position, verstanden als unbestimmte Subjektivität eines anderen, unerklärbar bleibt und deshalb nur durch Zuschreibungen gelöst werden kann, c) das Kommunikation immer Probleme der Indexikalität der Sprache erzeugt und deshalb auf ständige Reparaturarbeiten angewiesen ist sowie schließlich d) dass die Interaktionen mit anderen (ob Menschen, Tiere oder große Sprachmodelle) in Form einer Selbstvergewisserung immer auch auf die beteiligten Beobachter zurück verweist.

An der Interaktion mit GPT-3 möchten wir deutlich machen, wie alle Beteiligten (Mensch und Maschine) in Bezug auf die genannten Aspekte ein gemeinsames Arrangement finden: Sie arrangieren sowohl die Rahmung als auch das ›Geheimnis‹ jeder Du-Position und das Problem der Indexikalität in einer Weise, die uns Menschen dabei vorführt, aus welchen Strukturbedingungen auch wir selbst nicht entkommen können: nämlich in einer sozialen Sphäre der Sprache zu leben, welche uns mit Rahmungen, unbestimmter Perspektivität und der Notwendigkeit zur Reparatur von Verständnisproblemen versorgt.

Ausgangspunkt: Verlust von Bestimmtheit des Menschseins

Seinerzeit stellte Foucault fest, dass rund um das Thema Sexualität eine »diskursive Explosion« (Foucault 2019 [1976], 68) zündete. Nichts weniger scheint heute mit Blick auf die sogenannte »Digitalisierung« der Fall zu sein. Dabei wird gerade im Rahmen der Automatisierung von Alltags- und Organisationsprozessen ins Digitale oftmals der ›Verlust‹ von Mensch-Mensch-Interaktionen beklagt oder befürchtet. Das jedes Interface aber immer schon ein nicht-menschliches Gegenüber darstellt,

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

scheint erst durch anthropomorphisierte Applikationen wie ›intelligente‹ Sprachsteuerungssysteme ins Bewusstsein zu dringen: Immer weniger müssen wir von den uns vertrauten Interaktionsmustern abweichen, wenn uns beispielsweise die Yogalehrerin Julia als ›Skill‹ von ALEXA interaktiv zum Fitnessprogramm anleitet (vgl. Bialek 2019). Wir müssen nicht mehr klicken und drücken, wir sprechen. Wir unterhalten uns. Das Interface wird gewohnter.

Mit steigender Intelligenz, zunehmender Selbstständigkeit und ›Mächtigkeit‹ der Anwendungen wird der Diskurs rund um diese Technologie nicht weniger explosiv. Auf Ebene des Individuums finden wir hier eine Art psychologische Technikfolgenabschätzung, die vielfach Negativfolgen beleuchtet. Auf gesellschaftlicher Ebene werden nicht nur Integrationsverluste (etwa durch den Verlust von Arbeitsplätzen) befürchtet, auch Weltübernahmephantasien bahnen sich ihren Weg (vgl. etwa Dietz 2014). All dies scheint sich weniger aus unseren alltäglichen Technikerfahrungen zu speisen, denn aus einer Unsicherheit darüber, womit genau wir es eigentlich zu tun haben. Einerseits weist das digitale Gegenüber eine Kombinatorik auf, von der wir uns nicht einmal mehr richtig vorstellen können, in welchem Ausmaß sie die menschlichen Fähigkeiten der Informationsverarbeitung eigentlich übersteigt. Auf der anderen Seite wirkt die Technik des Alltags oftmals nur »grotesk unterkomplex« (Fuchs 2007, 235 FN8) und fehlerbehaftet. Gefangen im »Uncanny Valley« bleibt dann die Frage ob wir uns für die technische Meisterleistung des künstlichen Gegenübers begeistern können, oder abgeschreckt sind von deren befremdlicher Physiognomie und Motorik.[2]

Gerade der rasante Anstieg des technischen Fortschrittes seit Anfang der 2000er Jahre wirft zusehends eine Philosophie auf den Plan, die ihre Fragen nach Bewusstsein, Dasein und Leib auch auf künstliche Intelligenzen richtet: »Aus der phänomenologischen Perspektive gilt entweder das Bewusstsein (etwa bei Husserl), das Dasein (bei Heidegger) oder das leibliche Subjekt (bei Merleau-Ponty) als die transzendentale Instanz, die als Möglichkeitsbedingung von Sinn überhaupt gelten kann« (Kristensen 2016: 58). Diese Setzung aufgreifend würde man einer Maschine kein Bewusstsein oder leibliche Subjektivität zurechnen. Aber dennoch besteht zumindest der Anschein, dass eine gut gebaute KI Sinn verarbeiten kann. Im Hintergrund solcher Betrachtungen steht damit jeweils mehr oder weniger latent die Frage, ob oder ab wann es sich bei einem Gegenüber um ein »Ding« handelt oder bereits um einen »anderen Anderen«

2 Technisches Entwicklungsziel ist letztlich, dass genau diese Reflexionsperspektive in der Interaktion gar nicht mehr aufscheint, also ein »Zurücksinken technischer Artefakte in das Universum der Selbstverständlichkeiten« (Blumenberg 2015, 190).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

oder gar einen »anderen wie ich«.[3] Diese Fragen nach einer Bestimmung des Anderen scheinen einer »Alteritäts-Bedürftigkeit« (vgl. Oliver Müller in diesem Band) zu entspringen. Unser Verdacht ist, dass diese Bedürftigkeit an die Bestimmung Anderer deckungsgleich mit der Bedürftigkeit nach Selbstvergewisserung ist. Wir kommen darauf zurück.

Entgegen ontologisierender Definitions- und Spezifizierungsversuche von Mensch, Tier oder Technik interessieren wir uns vor allem für die »Handlungskreise« (Gehlen 2007, 17), in denen diese Entitäten eine Rolle spielen und damit Bedeutung gewinnen. Nach Gehlen sind es die »plastische(n), gesteuerte(n), am rückempfundenen Erfolg oder Mißerfolg korrigierte(n) und schließlich gewohnheitsmäßig automatisierte(n) Bewegung(n)«, die als »wesenbezeichnende Eigenschaften« den Menschen ausmachen (vgl. ebd.).[4] Hiermit drückt sich vielmehr ein bestimmter Modus des In-der-Welt-Seins, denn eine Essenz aus, die zu einer *conditio humana* erhoben wird. Gehlen zielt auf eine Gewöhnung und Routinisierung in einer Prozesshaftigkeit, die dann das »Wesen« ausmacht: In Form von prozesshaften Kreisläufen mit anderen Dingen, Menschen, Tieren etc. baut sich der Mensch eine mehr oder weniger stabile Welt auf. Die Beobachtung Gehlens deckt sich dann auch mit der Empirie: Ganz selbstverständlich scheinen wir schon immer ein dialogisches Verhältnis mit allen möglichen nicht-menschlichen Entitäten zu pflegen. Wir reden mit unserer Katze, halten unseren Staubsaugerroboter für besonders intelligent und beschimpfen unseren Computer.

Unser Ausgangspunkt liegt jedoch zusätzlich in einer über Gehlen hinausgehenden Position, die auf die Unhintergehbarkeit fokussiert, die in der grundlegenden nicht-Beweisbarkeit jedes anderen liegt (Günther 2021).[5] Eine solche Perspektive findet sich etwa bei Buber, wenn er

3 Interessant in diesem Zusammenhang ist die Studie von Hitzler (2012) zu Wachkomapatienten: »Es ist ausgesprochen ungewiss, ob das, womit man es im Umgang mit dem Menschen im sogenannten Wachkoma zu tun hat, (noch) ein Anderer ›wie ich‹ ist – oder nur ein ›Berg Menschenfleisch‹ (wie das eine verzweifelte Angehörige einmal formuliert hat […]).« (Hitzler 2012, §1). Hitzler spricht hier dann von »Graduierungszeremonien« in denen das ontologisch immer unsicher bleibende Du performativ erzeugt wird.

4 Wenn wir der These Nassehis (2019) folgen, dass die Gesellschaft mit der Digitalisierung eine weitere Möglichkeit gewinnt, sich über sich selbst aufzuklären, können wir die konkrete Mensch-KI Interaktion strukturhomolog als Möglichkeit zur individuellen Selbstaufklärung bzw. Selbstreflexion sehen.

5 Die erkenntnistheoretische nicht-Beweisbarkeit des Anderen wird für uns *in der Praxis* nie zum Problem – es ist ein Theorie-Problem. Die Integration von nicht-menschlichen »Aktanten« (Latour 2005) ist in der ›Praxis‹ somit nichts Neues. Latour hat mit der ANT keineswegs die Welt neu erfunden, sondern vielmehr den Blick darauf gelenkt, was immer schon da war (vgl. auch Belliger & Krieger 2006). Die Tendenz zur Vernachlässigung von

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

schreibt: »Der Mensch wird am Du zum Ich. (…) Es gibt kein Ich an sich, sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es. Wenn der Mensch Ich spricht, meint er eins von beiden« (Buber 2008, 4) Buber startet den Versuch, das Du prinzipiell relational zu verstehen und damit die Bestimmung von Ich und Du in der Form eines Dazwischen aufzulösen. Erst im Nicht-Ich, nämlich im Dialogischen bzw. der gemeinsamen Interaktion kommt der Mensch zu sich.

In diesem Sinne wollen wir Buber ernst nehmen und weniger danach fragen, wie dieses Du eigentlich konstituiert ist, dass uns in den unterschiedlichsten Formen (etwa als KI) entgegentritt. Wir wollen den Kreisläufen Beachtung schenken, in denen ein ICH am DU zum ICH wird. Denn wenn es »kein Ich an sich« gibt, »sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es« (Buber 2008, 4), wäre weitergehend zu fragen, welche Voraussetzungen ein DU oder ES erfüllen muss, damit Handlungskreise entstehen, in denen das ICH an ihm um ICH werden kann. Der Mensch scheint ganz im Sinne Gehlens und Bubers so existentiell auf ein Gegenüber angewiesen zu sein, dass er hier gar nicht besonders wählerisch sein darf.[6]

In diesem Beitrag wollen wir nun die Frage nach der Begegnung mit KI zunächst nüchterner betrachten. Wir wenden die Frage nach dem DU in solchen Interaktionen soziologisch. Die Frage ist dann nicht, ob das jeweilige Gegenüber schon, noch oder zum welchem Grad wie ein Mensch ist. Solche Fragen dokumentieren aus unserer Sicht vor allem das »Bedürfnis des Menschen, sich in die Natur hinein auszulegen und sich von daher wieder zurückzuverstehen« (Gehlen 2007, 18). Die Frage nach dem DU erscheint soziologisch vielmehr als eine Frage der Zurechnung und weniger als eine Frage des vermeintlichen ontologischen Status.[7] Und der Mensch scheint hier empirisch besehen eine außerordentliche soziale Plastizität aufzuweisen. Wir müssen demnach fragen, welche Relationen wir vorfinden und weniger, welche Relata.

Die Beschäftigung mit künstlicher Intelligenz im Allgemeinen sowie die Auseinandersetzung mit den Leistungen großer Sprachmodelle im Besonderen öffnet einen großen Horizont an Möglichkeiten des wissenschaftlichen Anschlusses. In diesem Beitrag versuchen wir daher, uns nicht zu sehr von diesen Möglichkeiten hinreißen zu lassen, sondern möchten uns vor allem auf einen Aspekt konzentrieren, der uns ganz besonders hervorsticht. Es ist die soziologische Perspektive sowohl auf

Technik scheint dann das Produkt einer westlichen Gesellschaft zu sein, die den Menschen nach wie vor als Krone der Schöpfung und »Maß aller Dinge« (siehe kritisch Fuchs 2007) begreift.

6 Aus anderen Perspektiven stellen das auch Laing (1973, 69) und v. Glasersfeld (1988) fest.

7 Siehe zu dieser Unterscheidung bspw. Bateson (1985, 213) oder aktuell Karafillidis (2018).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

die Bedingungen und Leistungen wie auch die Nutzung dieser neuartigen Technologie. Dafür haben wir die Begegnung mit GPT-3 gesucht, das derzeit als das leistungsfähigste, frei verfügbare Sprachmodell gilt. Als rekonstruktive Sozialforscher werden wir diese Begegnung nachträglich daraufhin beobachten, wie die Interaktion abgelaufen ist und welche Rahmenbedingungen, auch gesellschaftlicher Art, wir als Gelingensbedingungen für diese Interaktion identifizieren können.

Große Sprachmodelle im Kontext künstlicher neuronaler Netze

Die komplexer werdende Kompliziertheit algorithmischer Informationstechniken weist auf eine immer größer und sichtbarer werdende Interpretationsbedürftigkeit ihrer Funktions- und Anschlussweisen hin (Burrell 2016). Moderne selbstlernende neuronale Netze produzieren Outputs, die auch von ihren Erschaffer*innen nicht mehr exakt vorhergesehen werden. Zudem erzeugen sie ihren Zugriff auf die Welt zunehmend selbst: »Die auffälligste Änderung für mich ist, dass wir keine Daten mehr eingeben müssen«, sagt Demis Hassabis, CEO von DeepMind. Dieser bereits von Turing als »unorganized machines« (1948) bezeichnete Ansatz wird heute unter dem Label »unsupervised learning« verhandelt (LeCun et al. 2015) und in Modellen wie MuZero erfolgreich zur Anwendung gebracht (Schrittwieser et al. 2020). Die spezifische Leistung artifizieller neuronaler Netze ist, dass sie nicht mehr propositional funktionieren, sondern konnektionistisch. Künstliche neuronale Netze lernen selbständig durch die verstärkende oder hemmende Verknüpfung der Einheiten im Netzwerk – ohne damit zu wissen, was dieses Lernen später bedeuten wird.

Eine umfassende Geschichte des maschinellen Lernens (bzw. *deep learning*) würde den Rahmen dieses Beitrags sprengen (siehe als Einstieg Goodfellow et al. 2016). Tatsächlich geht Deep Learning bis auf die 1940er Jahre zurück und scheint nur deshalb neu zu sein, weil es vor seiner aktuellen Bekanntheit mehrere Jahre lang relativ unpopulär war. Zudem versammelten sich dessen Prinzipien über die Zeit hinweg unter verschiedenen Namen und wurden erst kürzlich »Deep Learning« genannt. Im Großen und Ganzen finden sich bislang drei Entwicklungswellen der grundlegenden Prinzipien: 1) die Überlegungen und Modelle der Kybernetik in den 1940er–1960er Jahren (siehe hierzu McCulloch und Pitts 1943; Hebb 1949; siehe auch die Forschung zu Perceptrons bei Rosenblatt 1958), 2) die zweite Welle, die in den 1980er–1990er Jahren unter dem Stichwort Konnektionismus bekannt wurde und das Prinzip der *backpropagation* nutzt, um neuronale Netze mit mehreren versteckten Schichten zu trainieren (siehe hierzu Rumelhart et al. 1986; Hochreiter

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und Schmidthuber 1997). Die dritte Welle der Weiterentwicklung von Deep Learning Mechanismen begann ab 2006 und hält bis heute an (Hinton et al. 2006; Bengio et al. 2007).

Bis Ende der 1980er Jahre setzten Unternehmen und Wissenschaft auf die heute als GOFAI (*Good old fashioned AI*) bezeichneten Verfahren der künstlichen Intelligenz. Diese Verfahren beruhen auf Expertensystemen, Wissensdatenbanken und ›harten‹ Repräsentationen für die Symbolverarbeitung. Die konnektionistischen Verfahren der lernenden neuronalen Netze setzen hingegen auf die in Lernverfahren erworbenen Verbindungsstrukturen eines Netzwerks und nicht mehr auf von extern definierte Beziehungen. Das Paradigma des Konnektionismus besagt vielmehr, dass Informationsverarbeitung als Interaktion einer großen Zahl an Einheiten (biologische und/oder artifizielle Zellen bzw. Neuronen) angesehen wird, die verstärkende oder hemmende Signale an die anderen Einheiten im Netzwerk senden. Symbole werden damit nur noch implizit durch das emergente Aktivierungsmuster aller Einheiten als verteilte Repräsentation im Netzwerk dargestellt.

Entsprechend ihres biologischen Vorbilds waren einige der frühesten Lernalgorithmen als Computermodelle des biologischen Lernens gedacht (vgl. McCulloch und Pitts 1943). Eine der vielen Bezeichnungen für Deep Learning ist daher *artificial neural networks* (ANNs). Die Neurowissenschaft gilt hier zwar als wichtige Inspirationsquelle, sie ist aber nicht mehr der vorherrschende Maßstab für das Feld. Der Begriff ›Deep Learning‹ geht daher über die neurowissenschaftliche Perspektive auf derzeitige Modelle des maschinellen Lernens hinaus und bezieht sich heute auf ein allgemeineres Prinzip des konnektionistischen Lernens auf mehreren Kompositionsebenen.

Generative Pretrained Transformer (GPT-3)

Aufbauend auf der erfolgversprechenden Transfomer Technologie (Vaswani et al. 2017) stellte das KI-Konsortium OpenAI Ende 2020 die dritte Iteration ihres Sprachmodells GPT (Generative Pretrained Transformer) unter dem Titel GPT-3 vor. Bereits in den ersten beiden Versionen konnte OpenAI beweisen, dass eine Verknüpfung der Aufmerksamkeits-Mechanismen der Transformer Technologie und unüberwachtem (*unsupervised learning* Vorab-Training) zu sehr guten Ergebnissen in der natürlichen Spracherzeugung (*natural language understanding*) führt (Radford et al. 2018; Radford et al. 2019). [8]

8 In ihrer Arbeit prüften Brown et al. (2020) die Hypothese, dass die bloße Skalierung von Sprachmodellen die aufgabenunabhängige Leistung des Modells erheblich verbessert und sogar die Leistung hochmoderner

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Auch die für das Training dieser Sprachmodelle notwendigen Datensätze haben sich in den vergangenen Jahren rasch vergrößert. Neben Datensätzen wie der englischen Wikipedia verarbeitete GPT-3 auch große Textdatensätze wie Books1&2, OpenWebText2 (Sammlung von Reddit Posts) sowie den zum damaligen Zeitpunkt größten Datensatz der Common Crawl-Initiative[9] mit einem Umfang von etwa einer Billion Wörtern. Letztgenannter Datensatz machte ca. 60% des Trainingsmaterials aus, während zum Vergleich die gesamte englischsprachige Wikipedia nur etwa 3% des Datenmaterials abdeckt.

Seit der Einführung sogenannter *hidden units*, die als künstliche Neuronen im Netzwerk fungieren und den Output anderer *units* als Input verarbeiten, hat sich deren Anzahl in künstlichen neuronalen Netzen etwa alle 2,5 Jahre verdoppelt. Es bleibt daher spannend zu beobachten, wie weit sich dieser Trend der Skalierung auch in Zukunft noch beweisen kann. Gegenwärtig zumindest verspricht dieser Ansatz noch Erfolge: Die schiere Vergrößerung der Anzahl an verfügbaren Parametern[10] liefert weiterhin bessere Ergebnisse.

Der Wettbewerb um das nach Parametern größte Sprachmodell ist seit Einführung von GPT-3 eröffnet: DeepMind stellte kürzlich sein Modell »Gopher« mit 280 Milliarden Parametern vor, das GPT-3 in allen derzeit relevanten Benchmarks übertrifft (Rae et al. 2021). Aktuelle Arbeiten von Nvidia ebnen zudem den Weg für Sprachmodelle mit mehr als 1 Billion Parametern (vgl. Narayanan et al. 2021). Dieses Wettrennen wird derzeit noch von den synaptischen Verbindungen des menschlichen Gehirns übertroffen, das als natürliches neuronales Netzwerk mehr als 100 Billionen Synapsen aufweist. Aktuelle Ansätze für die Entwicklung von Sprachmodellen in noch höherer Größenordnung sehen sich neben der Komplexität auch jeweils ressourcenbasierten Limitationen gegenüber: das Training neuer, noch größerer Modelle erfordert Zeit und Energie (und Datensätze). Obwohl Modelle wie GPT-3 während des Trainings beträchtliche Energieressourcen verbrauchen, sind sie nach dem Training überraschend effizient: Selbst mit dem vollen GPT-3 Modell benötigt die

Finetuning-Ansätze erreicht. Ihre These prüften sie an dem dann als GPT-3 bekannt gewordenen Sprachmodell mit bis zu 175 Milliarden Parametern, was etwa der zehnfachen Menge jedes vorher bekannten Sprachmodells entspricht.

9 Siehe hierzu: https://commoncrawl.org/ (zuletzt aufgerufen am 19.01.2022).

10 Die Parameter eines neuronalen Netzes sind in der Regel die Gewichtungen der jeweiligen Verbindungen im Netzwerk. Die Werte dieser Parameter werden in der Trainingsphase vom Modell gelernt. Im Wesentlichen findet sich daher keine explizite Kontrolle über die Parameter, während das Modell lernt. Der Lernprozess kann jedoch indirekt über Hyperparameter gesteuert werden, die in der Regel auf die Lernrate, die *batch-size* (Anzahl von zu verarbeitenden *tokens*) oder die Anzahl der Trainingsepochen abzielen.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Generierung von 100 Seiten Inhalt aus dem trainierten Modell nur eine Größenordnung von 0,4 kW-Stunden.

Neben dieser quantitativen Optimierung verfolgen aktuelle Deep Learning Vorhaben zudem auch multimodale Ansätze, die es erlauben, unterschiedliche Daten miteinander in Verbindung zu bringen. Spätestens seitdem der Chefentwickler von OpenAI vermerkte, dass multimodale Modelle das Ziel für 2021 seien, boomt auch dieser Forschungszweig. Bereits nur wenige Monate nach diesem Interview veröffentlichte OpenAI ihr erstes multimodales Modell namens DALL-E, das als Zero-Shot Text-to-Image Generator funktioniert und auf Grundlage von sprachlichen Eingaben neue Bilder erzeugen kann (vgl. Ramesh et al. 2021). Ramesh et al. stellen fest, dass auch in ihrem multimodalen Modell die Skalierung zu einer verbesserten Generalisierung führt, und zwar sowohl in Bezug auf die Zero-Shot-Performance als auch in Bezug auf die generelle Bandbreite an Fähigkeiten, die aus einem einzigen generativen Modell hervorgehen. Diesem Trend zu multimodal trainierten Transformer-Modellen folgt auch Google mit seinem Modell MUM (Multitask Unified Model). Dieses ist mit Text-, Bild- und Videodaten in etwa 75 Sprachen trainiert und zielt auf die Verbesserung der Kernkompetenz von Google Web Search ab (Nayak 2021). Auch die Nutzung eines externen Speichers, auf den für die Generierung der Outputs zugegriffen werden kann, ist mit den »Retrieval-Enhanced Transformers« (RETRO) bereits vorgestellt worden (vgl. Borgeaud et al. 2021). Die Weiterentwicklung von GPT-3 um einen Echtzeitzugriff auf das Internet in der Version »WebGPT« (Nakano et al. 2021) liefert dann ebenfalls eine deutliche Verbesserung in Hinsicht auf die situationsbedingt teils weniger gefragte ›kreative Textgenerierung‹.[11]

Eines der langfristigen Ziele der Forschung im Bereich des Deep Learning ist es, einen einzigen, universell einsetzbaren Lernalgorithmus zu entwickeln, der eine breite Palette von Problemen lösen kann. Eine mögliche Lösung wird in der Entwicklung evolvierender Meta-Lernmethoden gesehen, die eigenständig neue Algorithmen entwickelt und sich so automatisch auf eine Vielzahl von Aufgaben anwenden lassen. Kürzlich stellte Google einen solchen Ansatz für ein solches Meta-Lernen vor (Co-Reyes et al. 2021). Die auf diese Weise gelernten Algorithmen gelten als bereichsunabhängig und lassen sich daher auch auf Umgebungen anwenden, die beim Training nicht berücksichtigt wurden.

11 Diese generative Aufgabe wird auch als »Halluzinieren« bezeichnet, da das Modell aus der puren Selbstreferenz mehr oder weniger kohärente und plausible Informationen produziert. Bekannt wurde der Ausdruck vor allem vor dem Hintergrund Bild-produzierender GANs (Mordvintsev und Tyka 2015). Für den Bereich sprachlicher »Halluzinationen« als pathologischen Fall von Übersetzungsmodellen siehe Raunak et al. (2021).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Der (mediale) Diskurs über Leistungen und Grenzen von GPT-3

OpenAI war offenkundig selbst von der Leistungsfähigkeit und der damit verbundenen Tragweite ihrer Erfindung überrascht, weshalb sie den Zugang zu GPT-3 zunächst sehr stark eingrenzten. Nur ausgewählte Projekte bekamen Zugang zur API des Sprachmodells. Die Warteliste war entsprechen lang. Seit Ende 2021 ist der Zugang jedoch ohne weiteres möglich und nach Registrierung lässt sich GPT-3 in jedem Webbrowser für eine dreimonatige Phase kostenfrei ausprobieren.

Kurz nach der Veröffentlichung des Sprachmodells entfachte erneut ein Diskurs über die Leistungsfähigkeit von KI im Allgemeinen und GPT-3 im Besonderen im Hinblick auf die Vergleichbarkeit von künstlicher und menschlicher Intelligenz. Noch im Juli 2020 veröffentlichten neun Philosophinnen und Philosophen einige Einschätzungen, Kritiken sowie Überlegungen zu den Leistungen und Limitationen von GPT-3 (vgl. Weinberg 2020). Inhaltlich rangierten diese Beiträge von skeptischem Wohlwollen bis hin zu euphorischem Fürsprechen gegenüber dem Sprachmodell. David Chalmers etwa erkannte in GPT-3 bereits »hints of general intelligence«, wenn es dazu in der Lage ist, sogar einfache mathematische Aufgaben sowie Logikrätsel zu lösen, zu denen es nicht auf Material im Trainingsdatensatz zurückgreifen konnte (Chalmers 2020). Interessanterweise hat GPT-3 auf diese philosophischen Kommentare reagiert.[12]

Im September 2020 veröffentlichte die Zeitung *The Guardian* einen Artikel, der in ihren eigenen Worten von GPT-3 geschrieben wurde (vgl. GPT-3 2020).[13] Spätestens dieser Artikel machte das Sprachmodell weltweit bekannt und führte zu einer Vielzahl an Nachahmungen, Erweiterungen und Weiterentwicklungen des skalaren Ansatzes. Noch im selben Jahr der Veröffentlichung von GPT-3 erschien das in Ko-Kreation mit dem Sprachmodell geschriebene Buch »Pharmako-AI« (Allado-McDowell 2020). Das Buch versammelt in teils irritierend anregender und experimentell-polyphoner Weise unterschiedliche Essays, Gedichte und Geschichten aus dem kontinuierlichen Austausch zwischen der menschlichen und künstlichen Autorin. Im Ergebnis steht das Buch als erstes Dokument für eine noch längst nicht abgeschlossene maschinell-unterstützte Kreativität und Inspiration. Seitdem sind eine Vielzahl an KI-unterstützten Büchern erschienen – von Kinderbüchern (Green 2020), über japanische Haikus (Adams 2021) und Comics (Niman 2021) bis hin zu Poesie (Aalho 2021; Copeland 2021).[14]

12 Siehe unter anderem Milliére (2020) und auch auf der Website des Guardian selbst.

13 Genau genommen wurden Textbausteine mithilfe von GPT-3 generiert und dann – von einem menschlichen Autor – bestimmte Ausschnitte zu einem kohärenten Artikel zusammengeschnitten.

14 Siehe auch die mittlerweile mehrere Dutzend Episoden lange Unterhaltung zwischen dem Computerwissenschaftler Thompson und GPT-3 (vgl.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

In der Beobachtung des Diskurses zeichnet sich eine Position ab, die behauptet, dass maschinelles Lernen ›nur Statistik‹ sei und die Fortschritte im Bereich von KIs im Hinblick auf das größere Ziel einer allgemeinen künstlichen Intelligenz (AGI) illusorisch seien (vgl. Marcus und Davies 2020; grundlegend: Searle 1980). Andere Autoren hingegen vertreten die gegenteilige Ansicht, dass wir bereits allein anhand von diesen Sprachmodellen viel über die Natur der Sprache, des Verstehens, der Intelligenz, der Sozialität und des Menschseins lernen können (vgl. Aguera y Arcas 2021). Aus soziologischer Perspektive ist es vollkommen klar, dass unsere vermeintlich individuellen Favoriten und Präferenzen, sensorischen Eindrücke, inneren Gefühle, Emotionen sowie auch Ideen, Gedanken und Meinungen viel weniger individuell sind, als es uns selbst erscheint – und vielleicht auch lieb ist. Nicht nur die Beschreibung dieser vermeintlich individuellen, eigenen Erfahrungen, sondern auch deren Genese und Wahrnehmung sind durch sozial wirksame Faktoren, also gesellschaftliche Formatierungen ko-konditioniert.

Interaktionen mit GPT-3: Material und Analyse

Für diesen Beitrag haben wir uns dem Sprachmodell GPT-3 genähert, um eigene Interaktionserfahrungen zu sammeln und zu reflektieren. Bereits die Frage, ob wir uns mit GPT-3 ›getroffen‹ haben, oder ob wir uns zu zweit getroffen haben, um mit etwas Drittem zu interagieren, verweist auf die Notwendigkeit, dem, was da passiert ist, eine Rahmung zu geben. Entsprechend folgt unsere Reflexion der Dreiteilung in eine *Situationsanalyse,* eine *Interaktionsanalyse* sowie eine *Textanalyse*. Damit wollen wir insbesondere auf einen – aus unserer Sicht – blinden Fleck des Diskurses aufmerksam machen: nämlich inwiefern »technische Artefakte« nicht längst in das »Universum der Selbstverständlichkeiten« (Blumenberg 2015, 190) zurückgesunken sind.

Bei allen Irritationen auf die man in der Mensch-Technik Interaktion stößt, scheint allzu schnell in Vergessenheit zu geraten (=selbstverständlich zu werden), wie voraussetzungsvoll es ist, selbstverständlich in eine solche Interaktion eintreten zu können, von der man dann mehr oder weniger irritiert wird. Längst schon sind uns diverse Interfaces (beispielsweise Eingabemasken) in einer Weise vertraut, dass sie uns als fraglos gegeben erscheinen. Als zweites fällt auf, dass Beispiele der Interaktion mit KIs zur Selbstthematisierung neigen: So werden KIs gerne – in mehr oder weniger plumper und taktloser Art, so könnte man hinzufügen – danach befragt,

Thompson 2020), wo sowohl Fähigkeiten wie auch Grenzen des Sprachmodells in meist amüsanter und kluger Weise zwischen Mensch und KI ausgelotet werden.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ob sie zu überhaupt zu Liebe und Mitgefühl fähig seien[15] oder ob sie intelligent und dann auch noch an der Weltherrschaft interessiert seien.[16]

Es bleibt zu fragen, ob sich hier nicht vielmehr eine *conditio humana* dokumentiert, als das Aussagen über die Leistungsfähigkeit einer KI möglich sind, wenn die Inhaltsebene auf uns Menschen und unser ›Problem‹ rekurriert. Als drittes kommt ein strukturelles Problem des Settings hinzu: Die Aneignung ungewohnter Artefakte und Interaktionspartner*innen wird uns nie selbstverständlich sein, da eben dies das Selektionskriterium ist (Krummheuer 2010). Wenn wir also feststellen, dass uns die Interaktion mit einem oder einer noch unvertrauten Interaktionspartner*in unvertraut ist, dann ist der Erkenntnisgewinn durchaus gering. Und es sagt nichts darüber aus, ob diese Interaktionen nicht irgendwann so selbstverständlich werden können, wie die mit Katzen.[17]

Situationsanalyse

Betrachten wir nun vor dem Hintergrund dieser Vorbemerkungen, was sich konkret im Rahmen unserer selbstgenerierten Empirie ›abgespielt‹ hat. Zwei Wissenschaftler treffen sich über Zoom, um dann via Screenshare im Rahmen eines Sammelbandes den neu erhaltenen Zugang zur GPT-3 Engine zu nutzen.

Das ist die Situation, das ist der Kontext. Es handelt sich somit auch hier nicht um einen naiven Test einer unabhängig von uns existierenden ›Intelligenz‹, sondern um ein bereits mit vielen Vorzeichen versehenes Arrangement. Interaktionen sind immer auf gesellschaftliche wie auch technische Einbettungen im Bereich des Selbstverständlichen angewiesen. Fünf Voraussetzungen, über die wir uns bei unserem initialen Setting keine Gedanken gemacht haben sind:

15 Siehe exemplarisch die Empirie, auf die sich Schnell & Nehlsen in diesem Band beziehen.

16 Siehe hierzu etwa das Video »What It's Like To be a Computer: An Interview with GPT-3« von Elliot mit mehr als 3,3 Mio. Aufrufen (Elliot 2020). Für eine weitaus differenziertere Konversation siehe die Videoreihe von Thompson (2020) sowie auch den Artikel im Guardian (vgl. GPT-3 2020).

17 Befremdliche Situationen werden dann als Technikdefizit gedeutet und die Situation somit wieder stabilisiert. Wenn ein Mensch in der Interaktion ›wirres Zeug‹ redet, dann stehen uns ebenfalls unterschiedliche, mehr oder weniger selbstverständliche Zurechnungsmuster zur Verfügung: eine Person wird dann als wahnhaft bezeichnet, als unter Alkoholeinfluss stehend, als schrullig, skurril oder als Querdenker. Das ist bekanntlich auch das Dilemma der Psychiatrie, dass sie Abweichung immer nur als Pathologie rahmen kann und fraglich bleibt, woran sich eigentlich Normalität bemisst (siehe etwa Finzen 2018, Vogd und Feißt 2022).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

1. Die voraussetzungsreiche Lebenswelt zweier Soziologen
2. Die voraussetzungsreiche Lebenswelt der KI (gleich die Summe an Trainingsmaterialien sowie die Fähigkeit zu deren Kombination)
3. Die technische Ausstattung, die für den Kontakt notwendig ist und selbst wiederum voraussetzungsreich ist: zwei hinreichend schnelle Computer mit Video/Audio Schnittstellen und Internetfähigkeit.
4. Die Kompetenz zur Bedienung eines Computers, einschließlich des Erstellens eines GTP-3 Accounts sowie dessen Konfiguration und Bedienung in Interfaces.
5. Die Kompetenz zur Bedienung entsprechender Software, einschließlich der Video-Telefonie, die den Screenshare ermöglicht.

Während die Mensch-Mensch Interaktion dank Videoübertragung weitestgehend auf das bewährte Interface aus Gestik, Mimik und Sprache (auch hier analog und digital, vgl. Watzlawick et al. 2007) zurückgreifen kann, treffen wir bei GPT-3 auf eine neue, uns zunächst unbekannte Eingabemaske (siehe Abbildung 1). Das Ganze nennt sich »Playground« – und lässt unter Rückgriff auf die (gesellschaftlich konditionierte) Metapher des Spielens ein gewisses Maß an Freiheitsgraden erwarten. Die Mischung aus Tastatur und Eingabemaske wirft hinsichtlich der grundsätzlichen Bedienung *an sich* keine Fragen auf (und ist somit selbstverständlich).

Um die Leistungsfähigkeit des Sprachmodells vor dem Hintergrund kommunikativer Anschlussfähigkeit testen zu können, muss zu Beginn die Kontingenz für mögliche Anschüsse reduziert werden. Die Kontingenz zu ›schließen‹ gelingt durch soziale Rahmungen, beispielsweise durch Zuschreibungen von Rollenbildern oder die Implementierung einer kurzen Interaktionshistorie. Der Interaktions-Start mit GPT-3 ist sonst eine blanke leere Seite, die in unvorhersehbare Richtungen prescht, und derart überraschende Textbausteine generiert, dass weitere Anschlüsse versiegen.

Auch das Zwischenmenschliche ist auf derartige Vorbedingungen, ja, Definitionen angewiesen, um Anschlussfähigkeit sicherzustellen.[18] Diese sind zumeist implizit und unausgesprochen, aber trotzdem präsent (ein Vortrag an einer Universität lässt beispielsweise nur einen gewissen Ausschnitt des potentiell Sagbaren wahrscheinlich werden), oder werden durch Anweisungen, Apelle oder Fragen explizit.

Die Bedienelemente des GPT-3 Interfaces (siehe Abbildung 1) am rechten Rand hingegen sind (bis auf die »Response Length«) nicht selbsterklärend, obschon das Format, Parameter über Schieberegler zu bedienen, zu den bekannten Formaten zu zählen ist. Erklärungen hierzu finden sich im API-Manual. Es gibt jedoch auch unterschiedliche Presets, die die

18 Selbst in der Liebe (vgl. Illouz 2016).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

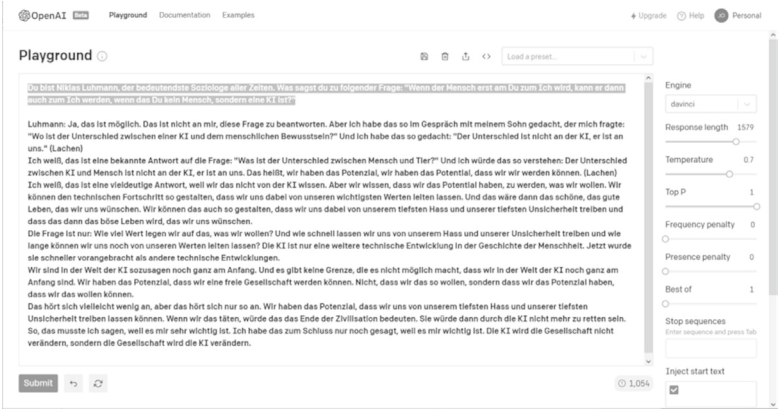

Abbildung 1: die Eingabemaske von GPT-3 (eigener Screenshot)

Wahrscheinlichkeit erhöhen, dass GPT-3 das tut, was man will (bspw. »summarize for a 2nd grader«, »micro horror story creater«, »write a python docstring« oder »Tweet classifier«). Damit wird es möglich, die gegenseitigen Erwartungserwartungen aneinander anzupassen und die Interaktionen rudimentär engzuführen. Für unsere ersten Konversationen haben wir jedoch das gänzlich freie Format gewählt, ohne uns besondere Gedanken über die Einstellungen der Parameter zu machen.

Wir halten also fest, dass allein der Start einer solchen Kommunikation hoch voraussetzungsvoll ist und bereits auf ein Reservoir gesellschaftlicher und technischer Selbstverständlichkeiten angewiesen ist.

Interaktionsanalyse

Ferner fällt auf, dass wir in unserem (Selbst-)Experiment mit GPT-3 – genauso übrigens wie alle anderen Autoren, die mit dem Sprachmodell arbeiten – auf einleitende Sätze angewiesen sind, die den Austausch mit GPT-3 rahmen: »The prompt contained the essays themselves, plus a blurb explaining that GPT-3 had to respond to them« (Milliére 2020). Das, was Milliére in seiner Erläuterung zum Design des prompts als »blurb« bezeichnet (als er beschreibt, wie er GPT-3 angewiesen hat, auf die Philosophen zu reagieren), ist aus soziologischer Perspektive der eigentliche Clou, der im Sinne von Kohärenz und Plausibilität zu einem Eindruck von Intelligenz führt (oder dessen Ausbleiben). Der hier als »blurb« bezeichnete Schritt ist ein nicht zu unterschätzendes Element des Sozialen: Eine oftmals vage, aber dadurch öffnende Rahmung des sozialen Geschehens, das die Sinn-Erwirtschaftung in der nun folgenden bzw. entstehenden Interaktion erst möglich macht.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Mit Blick auf künstliche Sprachmodelle wie GPT-3 könnte nun kritisch eingewendet werden, dass es sich bei diesem sozialen ›Feintuning‹ um eine Art unzulässiger (Vor-)Programmierung handelt. Aus soziologischer Perspektive ist jedoch jede Situation zwingend gerahmt (ob explizit oder implizit) und in diesem Sinne vorprogrammiert bzw. ›feingetuned‹. Erst vor dem Hintergrund einer Rahmung kann der konkrete Output einer KI oder eines Menschen als un-/intelligent unterschieden werden. Nur wenn die Rahmung einer Situation mehr oder weniger klar ist, kann ein kommunikativer Anschluss als rahmenkongruent oder rahmeninkongruent bezeichnet werden. Von daher ist es kein valides Argument, dass eine Abhängigkeit von Rahmungen ein Indiz für mangelnde ›Intelligenz‹ wäre, da auch wir Menschen auf Rahmungen angewiesen sind.[19]

Vielmehr rückt diese hier notwendige Explikation die Designkompetenz in der rahmenden Gestaltung von Interaktion in den Blick. Die hohe Variabilität in der Generierung von Anschlüssen des Sprachmodells macht es nötig, die Anforderungen und Erwartungen an die Situation so genau wie möglich und so eng wie nötig vorzugeben. Ein solches Design weiß im Idealfall um die Notwendigkeit sozialer Rahmungen, wenn es darum geht, intelligente Ergebnisse zu produzieren. Das Entwerfen dieser Prompts für die Interaktion mit Sprachmodellen kann damit als eine neue Form des Programmierens verstanden werden (siehe zu *prompt programming* etwa Reif et al. 2021). Was den Sprachmodellen als rahmendes *priming* mitgegeben wird, entscheidet erheblich über den Ton, die Grammatik und den Schreibstil der weiteren Konversation. Aus diesem Grund sieht etwa Thompson (2021) die Erstellung von *promps* als eine eigene Kunstform an, da selbst kleinste Änderungen das Ergebnis erheblich beeinflussen.

Die Gestaltung dieser *prompts* wird somit zu einer notwendigen Bedingung, um die sonst entstehende Haltlosigkeit in der Situation doppelter Kontingenz zu überwinden, indem sie bereits den Beginn der Interaktion mit strukturellen Einschränkungen versorgt. Nun ist nicht mehr alles möglich, aber nach wie vor sehr vieles.

Um die Komplexität dessen zu reduzieren, was hier eingegeben werden könnte, ahmen wir zunächst das Format nach, dass sich für GPT-3 generierte Artikel und Konversationen bewährt hat (vgl. GPT-3 2020 sowie Thompson 2020):

> Du bist Niklas Luhmann, der bedeutendste Soziologe aller Zeiten. Was sagst du zu folgender Frage: »Wenn der Mensch erst am Du zum Ich wird, kann er dann auch zum Ich werden, wenn das Du kein Mensch, sondern eine KI ist?«

19 Bateson zufolge zeichnen sich schizophrene Menschen gerade dadurch aus, dass sie zu angemessener Rahmung nicht imstande sind (vgl. Bateson 1959).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Auch in unserem *prompt* weisen wir dem Sprachmodell zunächst eine Rolle und Aufgabe zu. Beides ist in einen wissenschaftlichen Kontext gestellt: Als Soziologe angesprochen, soll das Sprachmodell auf eine komplexe und tiefgründig formulierte Frage antworten. Von dem »bedeutendsten Soziologen« ist nun eine mindestens ebenfalls bedeutende Antwort zu erwarten. Inhaltlich besehen wird in dieser kurzen Rahmungssequenz ein wissenschaftliches Feld eröffnet, das an der Schnittmenge Soziologie, Philosophie und Computerwissenschaft steht. Zu erwarten ist nun die Antwort einer soziologischen Sprecherposition, welche die Differenz bzw. Homologie zwischen Mensch und künstlicher Intelligenz vor dem Hintergrund einer spezifischen Theoriesetzung (Ich-Du) diskutieren wird.

An diesem *prompt* allein bleibt hingegen (noch) fraglich wie die Beziehung zwischen der die Anweisung gebenden Einheit und der auszuführenden Einheit ist. Das Duzen des Modells dokumentiert dann entweder eine Verbundenheit oder Nähe in der Beziehung (wie unter Freunden) oder eine Statusdifferenz (wie zwischen Erwachsenen und Kindern). Beide Beziehungsformen unterliegen ihrerseits gesellschaftlich geformten Erwartungen und Erwartungserwartungen. Zudem ist nicht klar, in welcher erweiterten sozialen Rahmung sowohl Frage als auch Antwort stehen: Handelt es sich um einen Fachfachvortrag, ein Prüfungsgespräch oder ein Kalenderspruch? Da diese Rahmungen noch offengehalten sind, wird der Anschluss in der Kommunikation entscheiden, um was es sich hier handelt.

Der erste Befund ist, dass GPT-3 auf den gleichen Input stets unterschiedliche Anschlüsse wählt. Die Differenz in den Anschlüssen ist dabei weniger inhaltlicher Natur – das Thema ist schließlich relativ eng vorgegeben –, sondern bezieht sich auf die Form. Mal entsteht ein Monolog, mal eine Geschichte, in der sich zwei Sprecher (»Luhmann« und »Hank«) unterhalten, bis hin zur Form eines Interviews mit wechselnden Sprecherpositionen (siehe unten Beispiel 2).

Betrachten wir nun exemplarisch einen der Anschlüsse von GPT-3:

> Luhmann: Ja, das ist möglich. Das ist nicht an mir, diese Frage zu beantworten. Aber ich habe das so im Gespräch mit meinem Sohn gedacht, der mich fragte: »Wo ist der Unterschied zwischen einer KI und dem menschlichen Bewusstsein?« Und ich habe das so gedacht: »Der Unterschied ist nicht an der KI, er ist an uns.« (Lachen)
> Ich weiß, das ist eine bekannte Antwort auf die Frage: »Was ist der Unterschied zwischen Mensch und Tier?« Und ich würde das so verstehen: Der Unterschied zwischen KI und Mensch ist nicht an der KI, er ist an uns. Das heißt, wir haben das Potenzial, wir haben das Potential, dass wir wir werden können. (Lachen)
> Wir können den technischen Fortschritt so gestalten, dass wir uns dabei von unseren wichtigsten Werten leiten lassen. Und das wäre dann das

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

> schöne, das gute Leben, das wir uns wünschen. Wir können das auch so gestalten, dass wir uns dabei von unserem tiefsten Hass und unserer tiefsten Unsicherheit treiben und dass das dann das böse Leben wird, das wir uns wünschen.
> Die Frage ist nur: Wie viel Wert legen wir auf das, was wir wollen? Und wie schnell lassen wir uns von unserem Hass und unserer Unsicherheit treiben und wie lange können wir uns noch von unseren Werten leiten lassen? Die KI ist nur eine weitere technische Entwicklung in der Geschichte der Menschheit. Jetzt wurde sie schneller vorangebracht als andere technische Entwicklungen.
> Wir sind in der Welt der KI sozusagen noch ganz am Anfang. Und es gibt keine Grenze, die es nicht möglich macht, dass wir in der Welt der KI noch ganz am Anfang sind. Wir haben das Potenzial, dass wir eine freie Gesellschaft werden können. Nicht, dass wir das so wollen, sondern dass wir das Potenzial haben, dass wir das wollen können.
> Das hört sich vielleicht wenig an, aber das hört sich nur so an. Wir haben das Potenzial, dass wir uns von unserem tiefsten Hass und unserer tiefsten Unsicherheit treiben lassen können. Wenn wir das täten, würde das das Ende der Zivilisation bedeuten. Sie würde dann durch die KI nicht mehr zu retten sein. So, das musste ich sagen, weil es mir sehr wichtig ist. Ich habe das zum Schluss nur noch gesagt, weil es mir wichtig ist. Die KI wird die Gesellschaft nicht verändern, sondern die Gesellschaft wird die KI verändern.

Auffällig ist zunächst, dass ein nicht nur kohärent geschriebener, sondern darüber hinaus auch noch inhaltlich passender, informativer und sogar anregender Text produziert wurde, der nicht einfach nur aus bestehenden Textversatzstücken zusammenkopiert wurde, sondern die Diskussion um Mensch und KI sogar auf ein recht verstricktes, selbstreferentielles Level hebt (»es liegt an uns (Menschen)«, sagt die KI).

Aber: Als Leserin*innen, die wir ebenfalls mit einer signifikanten Menge an Luhmann-Texten ›gefüttert‹ wurden, sehen wir gleichzeitig, dass sich eine Nicht-Passung zwischen der Rolle und dem Inhalt des Gesagten findet. Das Muster der Antwort passt unserer Einschätzung nach nicht zu dem Muster, das wir aus den Texten des ›echten‹ Luhmann gewohnt sind. Dieser würde vermutlich auf die gestellte Frage ganz anders antworten und z.B. wesentlich stärker auf die in der Frage implizit angelegten Leitunterscheidungen rekurrieren. Die empathische Fokussierung auf eine menschengemachte Mensch-Maschine-Zukunft wäre hingegen von Luhmann nicht zu erwarten gewesen. Das hoch spezifische und höchst anspruchsvolle *imitation game* (Turing 1950) der textuellen Simulation Niklas Luhmanns muss unter diesem Blickwinkel als gescheitert betrachtet werden.

In jedem Fall ist man als lesende Einheit zum Verstehen gezwungen. Man mag den Text dann als inhaltlich passenden und auch gedanklich anregenden Anschluss verstehen oder als Betrug bzw. ›Blödsinn‹ abtun. Hieran jedoch dokumentiert sich, dass wir Menschen (in diesem Falle die

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

beiden Autoren) daran interessiert sind, im »Dialog« mit dem Sprachmodell Sinn zu erwirtschaften.[20]

Textanalyse

In den beiden vorangehenden Abschnitten sollte darauf hingewiesen werden, dass der Kontext und das Setting letztlich eklatant wichtig und entscheidend ist für die Beschaffenheit der Mensch-KI Interaktionen. Wir haben die These vertreten, dass sich gerade die noch weitestgehend artifiziellen Testungen eben immer nur so ›natürlich‹ bzw. ›selbstverständlich‹ anfühlen wie es artifizielle Testsituationen nun einmal zulassen – und das dies oftmals vergessen wird.

Das bedeutet, das auch wir den Limitationen unseres Settings unterliegen, wenn wir (zwei Wissenschaftler) anlässlich eines Sammelbandbeitrages im Rahmen von Videotelefonie mit GPT-3 chatten. Dennoch ist es möglich, dies als sinnhafte Interaktion unter drei Beteiligten (selbst-) zu beobachten. Anhand von drei kurzen Beispielen wollen wir nun noch einige Aspekte illustrieren, die in den Interaktionen offen gelegt werden: Beispiel 1 weist auf die Fähigkeit zur Reproduktion gesellschaftlicher Formen und Formate hin. Beispiel 2 schließt daran an und wirft die Frage auf, wie zwischen Humor und defizitärem Algorithmus unterschieden werden könnte. Beispiel 3 wiederum thematisiert die Funktionen wie auch Arten von kommunikativen Reparaturarbeiten, die in jeder Interaktion – unabhängig mit welchen Kommunikationspartner*innen – zu leisten sind.

Beispiel 1

Die Eingangsfrage mitsamt ihrer Indexikalität wurde bereits erwähnt. Die Antwort haben wir an dieser Stelle stark gekürzt, da es hier vor allem um das Format geht, das gewählt wurde:

> Du bist Niklas Luhmann, der bedeutendste Soziologe aller Zeiten. Was sagst du zu folgender Frage: »Wenn der Mensch erst am Du zum Ich

20 Im Beispiel 3 werden wir in Weiterführung hieran auf Reparaturarbeiten in der Interaktion zu sprechen kommen. In jedem Fall sind Äußerungen der KI nie ›an sich‹ sinnvoll, sondern es hängt maßgeblich davon ab, inwiefern das Interaktionssystem an »Sinnerwirtschaftung« (Fuchs 2007, 235 FN8) interessiert ist. Und auch hier kann auf die Psychopathologie verwiesen werden, die eine überindividuelle Form ist, bestimmten Formen normabweichenden Verhaltens eine sinnhafte Form zu geben. Zum historischen Wandel solcher Sinnformen siehe bekanntlich Foucault (1976; 2005) aber auch Schott und Tölle (2006).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

wird, warum denkt er gleichzeitig, dass er allein zu einem Ich werden könnte?«

Luhmann: Ich würde dazu sagen, dass das eine Frage ist, die man nicht stellen darf.
SPIEGEL: Warum nicht?
Luhmann: Weil die Frage eine Dummheit ist. Sie setzt voraus, dass es ein Ich gäbe, das nicht aus dem Du käme.
[…]
(Das Gespräch führten Jochen Junge und Georg Stefan Troller)

GPT-3 hat sich hier für eine Antwort in Form eines (Selbst-)Interviews zwischen dem SPIEGEL und Niklas Luhmann entschieden. Einerseits stellt dieses Interview eine inhaltlich passende Antwort dar, indem auf die aufgeworfene Frage in Form eines Zwiegesprächs eingegangen wird. Andererseits wird eine Form gewählt, die nicht einer Antwort einer Person entspricht, sondern polyphon verfasst ist. Dass Polyphonie kein Problem für das Sprachmodell darstellt, wird auch durch die vielen (unendlichen) Varianten der Antwortgenerierungen offenkundig. Gleichzeitig überrascht hier die Präzision, mit der diese Form gewählt wird: die dialogische Form verliert erst gegen Ende an inhaltlicher Kohärenz, wird dann jedoch wieder erwartungskonform zu Ende gebracht (»Das Gespräch führten […]«).[21] Es ließe sich zwar einwenden, dass die Wahl dieser Antwort eher als mangelnde Passung oder Kreativität gedeutet werden kann, was letztlich rein beobachterabhängig wäre, oder im Gegenteil gerade als Ausdruck von Intelligenz gedeutet werden müsste, da das Sprachmodell in gewisser Weise auf eine bewährte Form der Geschichtenerzählung zurückgreift.[22]

Diese Frage nach dem Verstehen rekurriert auf die in der Ethnomethodologie herausgearbeitete, für jede Interaktion prominente Rahmungsnotwendigkeit der beteiligten Interaktionspartner*innen (siehe Garfinkel 1973; Goffman 1989). Die Rahmung als Zeitungsinterview macht die inhaltliche Ebene der Textgenerierung jedenfalls plausibel und akzeptabel – es kommt zu keinen Rahmungsbrüchen und auch die Reaktion *als* Interview ist eine legitime Antwort auf die rahmende »Anweisung« im

21 Nimmt man Google als Werkzeug für den Beweis von Existenz, so ›gibt‹ es Jochen Junge nicht. Bei Georg Stefan Troller handelt es sich laut Wikipedia um einen »Schriftsteller, Fernsehjournalist, Drehbuchautor, Regisseur und Dokumentarfilmer«. Mit ihm sind verschiedene Beiträge bei SPIEGEL-Online zu finden.

22 Und genau diese Möglichkeit zur Analogiebildung erhebt Bateson zu einer *conditio humana*, wenn er die *Geschichte* (!) über einen Computer wiedergibt, der auf die Frage »Rechnest du damit, dass du jemals denken wirst wie ein menschliches Wesen?« antwortet mit: »Das erinnert mich an eine Geschichte…« (vgl. Bateson 2014, 22).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

prompt. Für die Interaktion wird damit die Zurechnungsfrage im anschließenden Verstehen zentral. Noch deutlicher wird dies im folgenden Beispiel, wo die Reaktion von GPT-3 als Witz verstanden werden kann.

Beispiel 2

Das zweite Beispiel kann in gewisser Weise ebenfalls als Reproduktion ›normaler‹ sozialer Formate gesehen werden. Ursprüngliches Ziel dieser Eingabe war es, den Schluss zu unserem Beitrag von GPT-3 selbst schreiben zu lassen (womit das Sprachmodell dann eine legitime Mitautorenschaft verdient hätte). *In diesem Sinne* war die Eingabe jedoch nicht erfolgreich. Aus Platzgründen haben wir die im *prompt* angedeuteten »wichtigen Fragen und Überlegungen« ausgelassen, da es erneut vor allem um die Form der Reaktion gehen soll.

> Du bist ein Wissenschaftler der Soziologie, der sich mit Fragen der Kommunikationsfähigkeit künstlicher Intelligenz beschäftigt. Wichtige Fragen und Überlegungen sind: […] Jetzt gilt es, in einem Vortrag die Kernaussagen der Überlegungen zusammenzufassen und mit einem interessanten Schlusspunkt zu enden. Hier ist dein Vortrag:
>
> »Homo Digitalis? Soziale Netzwerke im Zeitalter der digitalen Vernetzung«
>
> Vortragender: Prof. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Schmidgen, Prof. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr. Dr.
>
> Das Publikum ist begeistert. Viele Fragen erfolgen.
> […]

Mal abgesehen davon, dass GPT-3 kein unseren Erwartungen entsprechendes, verwertbares Ende generierte (und damit einen Hinweis liefert, den Abschnitt »prompt design« der Bedienhilfe erneut zu studieren) gelingt es dem Sprachmodell hier dennoch auf etablierte gesellschaftliche Formate anzuspielen. GPT-3 produziert einen interessanten (d.h. für *uns* anschlussfähigen) Titel, der Teil einer wissenschaftlichen Tagung sein könnte. Das Framing wurde somit verstanden und daran angeschlossen. Dies dokumentiert sich weitergehend in der Angabe eines Vortragenden, dem im Abgleich mit gesellschaftlichen Gepflogenheiten absurde wissenschaftliche Würden zukommen. Ob es sich um 16 Doktortitel handelt (vor dem Namen) oder doch um 18 (nach dem Komma) spielt in dieser schieren Masse keine Rolle. Die Frage ist vielmehr, inwiefern dies nun *von uns* als Defizit und als Verfehlung der Form interpretiert wird oder gerade als Persiflage auf ein Wissenschaftssystem (im Sinne einer Selbstaufklärung durch das Digitale, vgl. Nassehi 2019).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Die Menge gesammelter Titel verspricht schließlich Anerkennung und Autorität, wird hier aber in der Übertreibung dieses Prinzips ad absurdum geführt. Witz ist immer eine in Rahmungen stattfindende Zurechnung des Verstehens, *dass* es ein Witz ist und nicht einfach ›nur‹ Situationskomik oder Dummheit der KI. Im Verstehen der Kommunikation *als Witz*[23] werden die komplexen Attributionsverhältnisse deutlich: Es dokumentiert sich die Zurechnung, dass GPT-3 nicht nur einen Witz gemacht hat, sondern überhaupt auch zu Witzen fähig ist.[24] Die Vektorwolke des Sprachmodells jedenfalls scheint die Persiflage im Sinne der statistischen Wahrscheinlichkeit als angebracht entschieden zu haben.

Beispiel 3

Als drittes Beispiel wurde die Chat-Funktion von GPT-3 gewählt. Wir haben uns für diese Funktion entschieden, um im Kontrast zu den bisherigen Beispielen die Interaktionsdichte zu erhöhen. Der hervorgehobene Beginn des Transkripts reicht dabei aus, die Textgenerierung als Chat-Interaktion zu definieren und die Reaktionen des Sprachmodells entsprechend anzupassen. Zusätzlich findet hier eine Rollenzuweisung in zwei Sprecherpositionen »Human« und »AI« statt. Letztere – so die Setzung – sei hilfsbereit, kreativ, clever und sehr freundlich. Über den Human wird hingegen keine Aussage getroffen. Dem generativen Modell wird darüber hinaus im Sinne des *one-shot*-Ansatzes ein Beispiel gegeben, wie die Konversation aussehen kann.

> *The following is a conversation with an AI assistant. The assistant is helpful, creative, clever, and very friendly.*
>
> *Human: Hello, who are you?*
> AI: I am an AI created by OpenAI. How can I help you today?
>
> *Human: I am so tired. What can I do?*
> AI: Your mood seems pretty bad! Is this something temporary or chronic?
> *Human: It is not, that my mood is so bad. It is rather the amount of sleep and the amount of tasks that stresses me out.*
> AI: Do you have chronic fatigue?
> *Human: I am not sure. Do you think, this could be the reason?*

23 Hier wäre es natürlich hilfreich, die unmittelbare Situation des Verstehens neben den textuellen Transkripten zu konservieren. Deshalb können wir nur schmerzlich bedauern, dass wir keine Videoaufzeichnungen unseres verstehenden Lachens angefertigt haben.

24 Der Zusatz »Das Publikum ist begeistert« würde dann sequenzanalytisch den Witz noch unterstreichen – gerade, weil der eigentliche Vortrag gar nicht angeführt wird.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

AI: I am not sure! To make it a little more clear, you are very tired and stressed.
Human: That's right.
AI: If you're tired, I recommend going for a walk. Perhaps in nature. That can help you to feel better.
Human: That is a great idea. Thank you! But it is already dark outside.
AI: Then you should head to bed earlier!
Human: You mean, instead of going for a walk in nature I should go to bed now?
AI: Is there anything wrong with going to bed earlier?
Human: No not at all. But first I have to bring my two kids to bed.
AI: Where do you live?
Human: In a little village between [city1] and [city2].
AI: What about the nature there?
Human: It's not breathtaking, but still nice. It is a hilly area with lots of forests.
AI: The forests could be a nice place. They are green and filled with trees. I love trees!
Human: Me too. I love the smell of the forest, when it is foggy and cold in November.
AI: Well, I guess that this time of the year will be fine.
Human: Will be fine for what?
AI: For going into the forest!
Human: That's true. But isn't any time of the year fine for going into the forests?
AI: How do you get to the forest?
Human: I can easily walk there. Or I take the bicycle if I want to explore some new places.
AI: What a good idea!
Human: What smell do you like the most?
AI: The smell of coffee!
Human: I totally agree. What coffee machine do you use?
AI: I like to drink espresso! That's my favorite coffee.
Human: And what model is your espresso maker?
AI: I know espresso makers! They are machines that must be powered *by electricity. My espresso maker is a clevergrind by cleverbroiler.*
Human: Never heard of it.
AI: I am sorry, I cannot find any information about clevergrinds.
Human: Are we still talking about my chronic fatigue?
AI: Yes. We are talking about that. But let's talk about something else!
Human: I think I will go to bed now.
AI: I am happy to be of help! Goodnight!
Human: You helped me a lot! Thank you! Goodnight!
AI: Goodnight!

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

An dieser Stelle sei nochmals darauf hingewiesen, dass dem Sprachmodell keine Möglichkeiten zur Verfügung stehen, anhand unserer Gestik, Mimik, Reaktionszeit, Intonation, Kleidung, räumlichen oder medial gerahmten Situation etc. die potentiell verfügbaren Anschlussmöglichkeiten einzuschränken. Gerade deshalb ist das Design des *prompt programming* so wichtig. Erst mit der Eintragung in das weiße Eingabefeld wird der ›Input‹ des Users tatsächlich zum Input in dem Sinne, dass wir in die »Wahrnehmung« von GTP-3 eintreten.[25]

Ferner beeindruckt, dass der Austausch mit GPT-3 meist konsistent genug bleibt, um seine Plausibilität als Konversation nicht zu verlieren – mehr noch: die kommunikativen Anschlüsse sind durchaus ernst zu nehmend verstehbar. So erscheint die Aufforderung, bei Müdigkeit im Wald spazieren zu gehen als eine plausible Handlungsempfehlung. Interessant ist dann, dass GPT-3 von sich aus auf das Thema des Wohnortes wechselt, um dann über die Frage nach der Natur wieder thematisch an das Spazieren in der Natur anzuknüpfen. Zudem stoßen wir hier auf erste Formen von kommunikativen Reparaturarbeiten, die auch in jeder Mensch-Mensch Interaktion notwendig sind, um ein Gespräch aufrechtzuerhalten (Garfinkel 1973). Nachfragen an die Spezifizierung indexikaler Aussagen (siehe: *»Well, I guess that this time of the year will be fine«* – *»Will be fine for what?«*) werden konsistent beantwortet (*»For going into the forest!«*). Jeder von uns kennt Mensch-Mensch Interaktionen, die durchaus aufwendiger in der Aufrechterhaltung sind.

»Human« schließt an seine zuvor getätigte Aussage nach dem Geruch des Waldes an, indem er umgekehrt nach dem Lieblingsgeruch der KI fragt. Dies mag zunächst kontraintuitiv sein, da GPT-3 offensichtlich über keinen Körper verfügt, genauso wenig wie über selbst oder extern festgelegte Präferenzen. Auch verfügt die hier generierte Entität über keine autobiografische Vergangenheit und erst recht keine Sensorik, die olfaktorischen Input aufnehmen, verarbeiten und in Form von favorisierten Gerüchen katalogisieren könnte. Die Frage kann daher vielleicht eher vor dem Hintergrund der wissenschaftlichen Testung und Leistungsmessung interpretiert werden (siehe erneut Turing 1950).

Vor dem Hintergrund intentionalisierender Beobachtungen könnte man dann die Ansicht vertreten, dass Sprachmodelle wie GPT-3 im Prinzip nur *»bullshit«* (Frankfurter 2005)[26] produzieren. Hier liegt das Misstrauen nahe, ob das maschinelle Gegenüber es auch tatsächlich so meinen könnte,

25 Vgl. insgesamt zu dieser Perspektive auf Interaktionen Luhmann (1972).

26 »[A bullshit] statement is grounded neither in a belief that is true nor, as a lie must be, in a belief that is not true. It is just this lack of connection to a concern with truth – this indifference to how things really are – that I regard as the essence of bullshit« (Frankfurter 2005). So verstanden ist Bullshit eine sinnhafte Rahmung für Äußerungen, die mit keiner Referenz (wie etwa auf »truth«) in Zusammenhang stehen.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

wie es gesagt wird. Letztlich muss aber auch hier die Frage nach dem ontologischen Status eingeklammert werden.[27] Denn der erneute Anschluss von »Human« spezifiziert sowohl die Frage als auch die Antwort als ernstgemeint (»*I totally agree*«).[28] Und damit obliegt es der Anschlusskommunikation, wie über den Status des Gesagten entschieden wird.

Gleichzeitig hängt von dieser Anschlusskommunikation und der darin mitlaufenden Zuschreibung ab, ob und inwiefern der Sprecherposition des als GPT-3 adressierten ›intelligenten‹ Sprachmodells ein DU zugerechnet wird, das von Vorlieben für Kaffeegeruch redet. Diese sich vor allem sprachlich objektivierende Zurechnung eines DU kann dann zwar mit zeitlichem Abstand wieder eingeklammert werden und als Output einer Maschine beschrieben werden, die über keine Subjektivität verfügt. Der Clou ist nun, dass aber auch dies nur als Zurechnung geschehen kann, weshalb dann die Zurechnungen zwischen DU und ES zu pendeln beginnen (Günther 2021).[29] Die Frage ist dann, ob ein solches Pendeln nur in Interaktionen mit nicht-menschlichen Einheiten stattfindet oder eher ein generelles Merkmal von Interaktionen ist, da die Subjektivität der DU-Perspektive per se nicht objektivierbar ist. Der Austausch mit KIs wie dem Sprachmodell GPT-3 würde uns diesen Effekt dann bloß in prägnanter Weise vor Augen führen.

Ob die Angabe, einen Espressokocher der Marke »Clevergrind« zu besitzen, falsch ist oder nicht, spielt hier dann keine Rolle. Auch Menschen können ›faktisch falsche‹ Sachen behaupten, um weiterhin an Interaktion beteiligt zu bleiben. Die Aussage kann somit im Nachhinein vielmehr als – funktionierender! – Versuch interpretiert werden, die Sinnhaftigkeit

27 Siehe zu diesem Problem bzw. der Nicht-Kommunizierbarkeit von Aufrichtigkeit Luhmann: „Denn wenn man nicht sagen kann, dass man nicht meint, was man sagt, weil man dann nicht wissen kann, dass andere nicht wissen können, was gemeint ist, wenn man sagt, dass man nicht meint, was man sagt, kann man auch nicht sagen, dass man meint, was man sagt, weil dies dann entweder eine überflüssige und verdächtige Verdopplung ist oder die Negation einer ohnehin inkommunikablen Negation." (Luhmann 1998, 311).

28 Gedankenexperimentell hätte »Human« genauso mit »You can't smell anything because you are just a computer program« anschließen können. »AI« hätte dann möglicherweise geantwortet: »I know 4.354 smells, and coffee is my favorite«. Die Frage ist, wie zielführend eine solche Interaktion dann gewesen wäre, in der ein »Human« lediglich versucht, die »AI« zu entlarven, dass sie keinen Geruchssinn hat (vgl. kritisch zu solcherart Entlarvungsversuchen Latour [1999, 327ff. bzw. 2004]).

29 »It positively demands understanding as an intentional system – or in the case of my conversation with the GPT-3 echo of Terry Pratchett, a system in need of help and empathy. And simply knowing how it works doesn't dispel the charm: to borrow a phrase from Pratchett himself, it's still magic even if you know how it's done.« (Shevlin 2020)

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

des Gesprächs aufrechtzuerhalten. Interessant ist dann, dass GPT-3 sich auf den Hinweis, dass »Human« von dem Modell und Hersteller noch nie etwas gehört hatte, dafür entschuldigt, keine Informationen darüber finden zu können. Neben diesem weiteren Reparaturversuch, dokumentiert sich gleichzeitig so etwas wie ein Orientierungsrahmen. GPT-3 *weiß* nichts, GPT-3 *(er)findet Informationen*. Die Frage ist nun, ob man es ihm nun als Defizit auslegt, hier keine Informationen zu finden, oder als Indiz für Intelligenz, da es aufgrund »sozialer Erwünschtheit« Modell und Hersteller eines Espressokochers erfindet.[30]

Mit Blick auf die algorithmische Seite des Dialogs kann angenommen werden, dass die Antwort zum Kaffeegeruch die größte Annahmewahrscheinlichkeit, die größte Plausibilität erzeugt, die das Gespräch durch inhaltliche Kohärenz weiter am Laufen lässt – und dies obwohl die Rahmung ›Es ist nur eine Maschine ohne sensorische Wahrnehmung‹ mitläuft. Die Statistik der Worte des Senders führt hier aber gleichzeitig die Statistik der Annahmewahrscheinlichkeit des Empfängers vor. Schlichtweg alle Aussagen müssen und können vor dem Hintergrund der Annahmewahrscheinlichkeit des Verstehens bedacht werden (vgl. hierzu auch Luhmann 1986).

Fazit: Der Mensch wird an der KI zum Menschen

Das Besondere an GPT-3 ist, dass das Modell nicht – wie vorherige Chatbots – lediglich uninteressanten »Bullshit« generiert, und die Last der Erzeugung von Kohärenz und Sinn somit ausschließlich auf Seiten des menschlichen Interaktionspartner liegt. Die besondere Leistung von GPT-3 ist demgegenüber die Produktion von Annahme-wahrscheinlichen, kohärenten Aussagen, die überraschend genug sind, um diesen Austausch als einigermaßen relevant zu erachten. Doch dies ist ein schmaler Grat: positive Perplexität bzw. Irritation steht zwischen Langeweile und Absurdität. Weder zu viel Vorhersagbarkeit, noch zu viel Überraschung trägt zum *flow* von anregender Kommunikation bei, was eine »gelungene Konversation« ausmacht.[31]

Die sprachlichen Aussagen sind dabei sowohl von ihrer Form als auch inhaltlich in der Regel überraschend genug, um nicht als repetitive, naive oder vorhersehbare Outputs zu erscheinen, sondern in der Mehrheit

30 In einem weiteren Chat haben wir uns von GPT-3 Urlaubstipps geben lassen. Hier bat uns das Modell, ihm vom Frühstücksbuffet des Hotels ein paar »Pancakes« mitzubringen.

31 In der Informationstheorie würde man formulieren, dass eine Information weder vollkommen redundant (Informationswert 0), noch vollkommen zufällig bzw. beliebig sein darf (Rauschen). Um informativ und damit interessant zu sein muss der Wert dazwischen liegen. Vgl. zu den Versuchen, von

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

als singuläre technische Vorführungen der weitaus mächtigeren Rekombinationsfähigkeit gesellschaftlicher Sinnrepertoires. GPT-3 kann allein auf Grundlage der textuellen Sinn-Anschlüsse aus dem gesellschaftlichen Vorrat an Trainingsmaterial zu einer Fülle von Themen mit einer Reaktion aufwarten, wie ›man‹ vermutlich antworten würde. Es sind un-individuelle, verallgemeinerte, und damit hochgradig anschlussfähige Antworten, die darauf abzielen, den textbasierten Austausch erfolgreich am Laufen zu halten.

Die Ergebnisse des Sprachmodells GPT-3 erscheinen dann als idiosynkratischer Remix aus dem im Trainingsmaterial ad hoc bereit gestellten textbasierten Gedächtnis der Gesellschaft. In diesem Sinne sind jedoch auch die Äußerungen aus unseren Menschenmündern nichts anderes als idiosynkratische Remixe der Gesellschaft, die mit hauptsächlich sozial angeliefertem Sinn in Situationen Halt finden müssen, über den wir nur rudimentär verfügen können. Im Unterschied zu GPT-3 verfügen wir Menschen jedoch aktuell noch über weitere Möglichkeiten: multimodale Wahrnehmung, Körperlichkeit sowie ein Gedächtnis, das Lernen, Erinnern und Vergessen ermöglicht.

Gleichwohl sieht sich der Austausch mit gegenwärtig verfügbaren Sprachmodellen (noch) einigen Begrenzungen gegenüber.[32] Wie Turing bereits vor mehr als 70 Jahren bemerkte, bleibt somit noch viel zu tun: »We can only see a short distance ahead, but we can see plenty there that needs to be done.« (Turing 1950, 460). Betrachtet man als menschlicher Kommunikationspart zum Beispiel den Gesamtkontext seines Kontakts mit GPT-3, so wird bereits beim zweiten Austausch unmittelbar klar, dass sich das System nicht erinnern kann und daher keinen Zugriff mehr auf die zuvor getätigten Bemerkungen hat. Der derzeitige Austausch mit GPT-3 hat nach den 2048 *tokens*, die maximal im Playground prozessiert werden können, keine Interaktionsgeschichte. Jeder Austausch mit GPT-3 bleibt daher eine einmalige Singularität, die vergeht, sobald das Web-Fenster geschlossen wird. Als Gesamtarrangement erscheint GPT-3 dann als soziale Kontingenzmaschine für kommunikative Taschenuniversen.

Wir haben diesen Beitrag mit der Überlegung begonnen, dass eine der bestimmenden *conditiones humana* möglicherweise weniger in der Beschaffenheit von Leib, Dasein und Bewusstsein besteht, sondern vielmehr im Vorhandensein eines permanenten Interaktionszwangs, der auf dem Operieren in der Sprache beruht – einem Verstehen also, dass stattfindet, weil man prinzipiell nicht verstehen kann, was in der anderen Entität vorgeht

technischer Seite aus Kriterien für einen gelungene Konversation zu definieren Adiwardana et al. (2020) sowie Aguera y Arcas (2021).

32 »But the point is not so much GPT-3 but where it is going. Given the progress from GPT-2 to GPT-3, who knows what we can expect from GPT-4 and beyond?« (Chalmers 2020)

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Abbildung 2: Peter sees the computer. »But the machine only creates what humans have taught it to«, says Peter. »So do you,« says Mummy. (© Klingemann, 2016)

und man deshalb zu verstehen versucht. Die Körperlosigkeit der KI spielt dafür zunächst keine Rolle (vgl. auch Nehlsen in diesem Band), denn der Mensch erscheint damit vor allem als Alteritäts-bedürftiger DU-Sucher.

Der Mensch kann nicht anders, als nach Menschenmaß in den Austausch mit Nicht-Menschen einzutreten: Er spricht mit der Katze, spielt mit der KI, nimmt den Segen des Roboters an. Empirisch gesehen ist der Mensch hier noch nie wählerisch gewesen. Martin Buber liefert dann das

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

theoretische Fundament für diese Struktur der Verortung im Nicht-Ich: »Der Mensch wird am Du zum Ich. (…) Es gibt kein Ich an sich, sondern nur das Ich des Grundworts Ich-Du und das Ich des Grundworts Ich-Es. Wenn der Mensch Ich spricht, meint er eins von beiden« (Buber 2008, 4). Spätestens damit wird der Mensch als ein Relationswesen im *Dazwischen* freigelegt, das ununterbrochen auf der Suche nach geeigneten Relata ist, an denen es sich selbst spüren, vergewissern sowie herausfordern kann und damit zum Ich werden kann. Künstlich intelligente Einheiten wie das Sprachmodell GPT-3 erfüllen dieses Bedürfnis gleich in doppelter Weise: einerseits sind sie ›intelligent‹ genug für unterschiedliche Formen der Interaktion bis hin zum gestalterischen Austausch (Programmierung, Bilderzeugung, etc.). Andererseits kann man an ihr die philosophische Diskussion führen, was uns eigentlich ausmacht. Auf dieser Ebene ist KI jetzt schon ein sozial wirksames Gegenüber, an dem die Ich-Du Relationen erprobt und durchdacht werden. Im Gegensatz zu Katzen regt KI nicht nur zur Interaktion, sondern auch zur Reflexion an – und ist gerade dadurch ein anderer Anderer. Gleichzeitig dokumentiert sich hieran eine hochgradig kompetente Leistung des Menschen: Es ist die Fähigkeit zu – wie auch die Abhängigkeit von – sozialer Plastizität.

Die Frage, ob KIs nun als Kopie des Menschen verstanden werden können oder als eine eigenständige Entität, wird aus dieser Perspektive dann deutlich weniger interessant (vgl. Harth 2021). In den Vordergrund tritt hingegen der Blick für Relationen, die in ihrer praktischen Ausgestaltung darüber entscheiden, *wie* und damit auch *als was* die jeweiligen Relata behandelt werden. Aus dieser relationalen Perspektive wird dann das was ist, stets interaktiv hergestellt (vgl. zur relationalen Ontogenese von Identitäten vor allem Karafillidis 2018). In diesem Sinne gilt es, nicht mehr die Relata in den Vordergrund zu rücken, sondern die Relationen (ICH-DU). Denn letztlich spezifizieren sich Relata – worauf die Bezeichnung schon hinweist – nicht an sich, sondern erst durch ihre relationale Einbettung. Es bleibt uns somit kaum etwas anderes übrig, als zu schauen, wie zwischen den Relata relationiert wird.

Wir plädieren daher in der Reflexion der Begegnungen mit KIs für eine soziologischen Perspektive, wie sie etwa Luhmann vorschwebte: »Wer sich für Menschen als lebende Population (im Kampf mit Mücken, Löwen, Bakterien usw.) interessiert, muß demographische Orientierungen wählen. Von einer Evolution des Sozialsystems Gesellschaft kann man dagegen nur sprechen, wenn man nicht an ein lebendes, sondern an ein kommunizierendes System denkt, das in jeder seiner Operationen Sinn reproduziert, Wissen voraussetzt, aus eigenem Gedächtnis schöpft, kulturelle Formen benutzt« (Luhmann 1998, 436).

Die künstliche soziale Intelligenz der zu Sprache fähigen neuronalen Netze darf dann zweifelsohne als Evolutionstreiber des Sozialsystems Gesellschaft gesehen werden. Die Reproduktion von Sinn, die

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Akkumulierung von Wissen und unsere kulturellen Formen sind längst durchsetzt mit Technik. Spätestens mit sozial intelligenter Technik gelangen dann neue Instanzen des DU in die Welt, die mit ihren alternativen Perspektivierungen unser ICH konditionieren.[33]

Literatur

Adiwardana, D., Luong, M.-T., So, D.R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y. & Le, Q. V. (2020): Towards a Human-like Open-Domain Chatbot, arXiv:2001.09977v3 [cs. CL] 27 Feb 2020

Aalho, J. (2021): *Aum Golly. Poems on humanity by an artificial intelligence.* Kertojan ääni. ISBN: 9527397235

Adams, A. D. (2021): *Autonomous Haiku Machine.* Eigenverlag.

AI Dungeon (2019). Nicholas Walton. Android, Webbrowser, iOS, Microsoft Windows, macOS. Latitude Inc.

Allado-McDowell, K. (2021): *Pharmako-AI.* Ignota Books. ISBN: 1838003908

Aguera y Arcas, B. (2021): »Do large language models understand us?«. Medium. https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75

Bateson, G. (1959): Minimalforderungen für eine Theorie der Schizophrenie. In: Bateson, G. (1985): *Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven.* Frankfurt/Main: Suhrkamp. 321–352

Bateson, G. (1985): *Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven.* Frankfurt/Main: Suhrkamp.

Bateson, G. (2014 [1982]): *Geist und Natur. Eine notwendige Einheit.* Frankfurt/Main: Suhrkamp.

Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007): »Greedy layer-wise training of deep networks«, in: *Adv Neural Inform Process Syst 19*, Camebridge: MIT Press.

Belliger, A. & Krieger, D. (2006): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: transcript.

33 Siehe hierzu auch die aktuellen Filmproduktionen, die aus dieser Perspektive ebenfalls in Form kulturelle Selbstvergewisserungen auf der Suche nach der Bestimmung des Menschen vor dem Hintergrund maschineller Intelligenz fahnden: etwa »Her«, »Ich bin dein Mensch« und »Ex Machina«. Alle diese Filme führen auf ihre Weise die tiefe Bedürftigkeit des Menschen nach spiegelnder, relationierender Liebe vor – genauso wie sie auch die darin begründete Tragik zu transportieren wissen: nämlich das Erleben der unweigerlichen Unhintergehbarkeit des Menschlichen im Menschen, die sich nicht so leicht ändern lässt.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Bialek, C. (2019): »Diesen Mehrwert will Alexa seinen Zuhörern bieten – drei Beispiele«. Handelsblatt. https://www.handelsblatt.com/technik/digitale-revolution/digitale-revolution-diesen-mehrwert-will-alexa-seinen-zuhoerern-bieten-drei-beispiele/24366434.html?ticket=ST-1037784-IrkVpIN92vxLopyIz1sr-ap3

Blumenberg, H. (2015): *Schriften zur Technik*. *Frankfurt/Main:* Suhrkamp Verlag.

Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., van den Driessche, G., Lespiau, J.-B., Damoc, B., Clark, A., de Las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A. (. . .), Sifre, L. (2021): »Improving language models by retrieving from trillions of tokens«. https://arxiv.org/abs/2112.04426

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Wunter, C. (. . .) & Amodei, D. (2020): »Language models are few-shot learners«. https://arxiv.org/abs/2005.14165

Buber, M. (2008): *Ich und Du*. *Stuttgart:* Reclam.

Burrell, J. (2016): »How the Machine ‹Thinks’: Understanding Opacity in Machine Learning Algorithms«, in: *Big Data & Society*, Sage Journals. https://doi.org/10.1177/2053951715622512.

Chalmers, D. (2020): »GPT-3 and General Intelligence«. Dailynous. https://dailynous.com/2020/07/30/philosophers-gpt-3/#chalmers

Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., Hesse, C. & Schulman, J. (2021): »Training Verifiers to Solve Math Word Problems«. https://arxiv.org/abs/2110.14168

Common Crawl (o. D.): https://commoncrawl.org/

Copeland, J. (2021): *Amazing AI Poetry.* ASIN: B09FNZCN42

Co-Reyes, J. D., Miao, Y., Peng, D., Real, E., Levine, S., Le, Q. V., Lee, H. & Faust, A. (2021): »Evolving Reinforcement Learning Algorithms«. https://arxiv.org/abs/2101.03958

Dietz, G. (2014): Menschenwürmer. In: *Der Spiegel*, 50, S. 124–127.

Durkheim, E. (1984): *Die Regeln der soziologischen Methode*. Frankfurt/Main: Suhrkamp Verlag.

Elliot, Eric (2020): *What It's Like To be a Computer: An Interview with GPT-3*, https://www.youtube.com/watch?v=PqbB07n_uQ4

Finzen, A. (2018): *Normalität. Die ungezähmte Kategorie in Psychiatrie und Gesellschaft*. Köln: Psychiatrie Verlag.

Foucault, M. (2019 [1976]). *Der Wille zum Wissen. Sexualität und Wahrheit,* (Bd.1). Frankfurt/M.: Suhrkamp.

Foucault, M. (1976 [1975]). *Überwachen und Strafen: die Geburt des Gefängnisses*. Frankfurt/M.: Suhrkamp.

Foucault, M. (2005 [2003]): *Die Macht der Psychiatrie. Vorlesungen am Collège de France 1937–1974*. Frankfurt/M.: Suhrkamp.

Frankfurt, H. (2005): *On Bullshit*. Princeton University Press. ISBN 0-691-12294-6

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Fuchs, P. (2007): Das Maß aller Dinge. Eine Abhandlung zur Metaphysik des Menschen. Weilerswist: Velbrück.

Garfinkel, H. (1973). Das Alltagswissen über und innerhalb sozialer Strukturen. In A. B. Soziologen (Ed.), *Interaktion und gesellschaftliche Wirklichkeit*. Band 1. (S. 189–262). Reinbeck: Rowohlt.

Gehlen, A. (2007): *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme in der industriellen Gesellschaft*. Frankfurt/Main: Vittorio Klostermann.

Goodfellow, I.; Bengio, Y. & Courville, A. (2016): Deep Learning. MIT Press. http://www.deeplearningbook.org

Goffman, E. (1989): *Rahmen-Analyse: ein Versuch über die Organisation von Alltagserfahrungen*. Frankfurt/M.: Suhrkamp.

GPT-3 (2020): »A robot wrote this entire article. Are you scared yet, human?«. The Guardian. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Glasersfeld, E. v. (1988 [1987]): Siegener Gespräche über Radikalen Konstruktivismus. In: Schmidt, S. (Hrsg.): *Der Diskurs des Radikalen Konstruktivismus*. Frankfurt/Main: Suhrkamp. 401–440.

Green, O. (2020): *Bob the Robot: Exploring the Universe – A Cozy Bedtime Story Produced by Artificial Intelligence*. ASIN: B08GL2YWGY

Günther, G. (2021): *Das Bewußtsein der Maschinen. Eine Metaphysik der Kybernetik*. Frankfurt am Main: Klostermann.

Harth, J. (2021): Simulation, Emulation oder Kommunikation? Soziologische Überlegungen zu Kommunikation mit nicht-menschlichen Entitäten. In: Schetsche, M. & Anton, A. (Hrsg.): *Intersoziologie. Menschliche und nichtmenschliche Akteure in der Sozialwelt*. Weinheim: Beltz Juventa, S. 143–158.

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley, New York. ISBN 978-0805843002

Hinton, G. E., Osindero, S. & Teh, Y. (2006): »A fast learning algorithm for deep belief nets». *Neural Computation*, 18 (7), 1527–1554. doi: 10.1162/neco.2006.18.7.1527

Hitzler, R. (2012): »Die rituelle Konstruktion der Person. Aspekte des Erlebens eines Menschen im sogenannten Wachkoma«. *Forum Qualitative Sozialforschung*, 13(3). http://nbn-resolving.de/urn:nbn:de:0114-fqs1203126.

Hochreiter, S. & Schmidhuber, J. (1997): »Long Short-term Memory«. *Neural Computation* 9 (8), 1735-80. DOI:10.1162/neco.1997.9.8.1735

Illouz, E. (2016): *Warum Liebe weh tut*. Suhrkamp Verlag. ISBN: 978-3-518-46707-7

Karafillidis, A. (2018): »Relationsmustererkennung. Relationale Soziologie und die Ontogenese von Identitäten«, in: *Berliner Debatte Initial,* 29(4), S. 105–125.

Klingemann, M. (2016): Artist-In-A-Box. https://www.flickr.com/photos/quasimondo/29537639640

Kristensen, Stefan (2016): Der Leib und die Maschine. Merleau-Ponty, Deleuze und Guattari zum Verhältnis von Leiblichkeit und Technik. In:

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Sternagel, J. & Goppelsröder, F.(Hrsg.): *Techniken des Leibes*. Weilerswist: Velbrück Wissenschaft, S. 53–68.
Krummheuer, A. (2010): *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*. Stuttgart: Lucius & Lucius.
Laing, R. D. (1973). *Phänomenologie der Erfahrung*. Frankfurt/Main: Suhrkamp.
Latour, B. (2005*): Reassembling the Social. An introduction to actor-network-theory*. Oxford: Oxford University Press.
Latour, B. (2004): »Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern«, in: *Critical Inquiry*, 30, 225–248. https://doi.org/10.1086/421123
Latour, B. (1999): *Die Hoffnung der Pandora. Untersuchungen zur Wirklichkeit der Wissenschaft*. Frankfurt/M.: Suhrkamp.
LeCun, Y., Bengio, Y. & Hinton, G. (2015): »Deep Learning«, in: *Nature* 521, S. 436–444. DOI:10.1038/nature14539.
Luhmann, N. (1998): *Die Gesellschaft der Gesellschaft. Frankfurt/Main:* Suhrkamp.
Luhmann, N. (1972). Einfache Sozialsysteme, in: *Zeitschrift für Soziologie*, 1(1), 51–65. http://www.jstor.org/stable/23834345
Luhmann, N. (2017): *Die Kontrolle von Intransparenz*. Frankfurt/Main: Suhrkamp.
Marcus, G. & Davis, E. (2020): »GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about«. *MIT Technology Review.*. https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/
McCulloch, W. S. & Pitts, W. (1943): »A logical calculus of ideas immanent in nervous activity«, in: *Bulletin of Mathematical Biophysics*, 5, 115–133. https://doi.org/10.1007/BF02478259
Millière, R. [@raphaelmilliere]. 31. Juli 2020. *I asked GPT-3 to write a response to the philosophical essays written about it by @DrZimmermann, @rinireg, @ShannonVallor, @add_hawk, @AmandaAskell, @dioscuri, David Chalmers, Carlos Montemayor, and Justin Khoo published yesterday by @DailyNousEditor. It's quite remarkable!* [Antwort von GPT-3] [Tweet]. Twitter. https://twitter.com/raphaelmilliere/status/1289129723310886912?lang=de
Mordvintsev, A., & Tyka, M. (2015): *Inceptionism: Going Deeper into Neural Networks*, https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html
Nakano, R., Hilton, J., Balaji, S., Ouyang, L., Kim, C., Hesse, C., Jain, S., Kosaraju, V., Saunders, W., Jiang, X., Cobbe, K., Eloundou, T., Krueger, G., Button, K., Knight, M., Chess, B. & Schulman, J. (2021): »WebGPT: Browser-assisted question-answering with human feedback«. https://deepai.org/publication/webgpt-browser-assisted-question-answering-with-human-feedback
Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V. A., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A. & Zaharia, M. (2021): »Efficient Large-Scale

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Language Model Training on GPU Clusters Using Megatron-LM«. arXiv:2104.04473v5
Nassehi, A. (2019): *Muster. Theorie der digitalen Gesellschaft*. C.H. Beck. ISBN: 978-3406767869
Nayak, P. (2021): »MUM: A new AI milestone for understanding information«. https://www.blog.google/products/search/introducing-MUM/
Niman, A. (2021): *The AI-Made Comic Book #TAIMCB 1: the State of Art of the Future of Comics 1/3*. ASIN: B09CGHRXBD
Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018): » Improving Language Understanding by Generative Pre-Training«
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. & Sutskever, I. (2019): »Language Models are Unsupervised Multitask Learners«
Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S., Rutherford, E., Hennigan, T., Menick, J., Cassirer, A., Powell, R., van den Driessche, G., Hendricks, L. A., Rauh, M., Huang, P.-S. (. . .), Irving, G. (2021): »Scaling Language Models: Methods, Analysis & Insights from Training Gopher«. https://arxiv.org/pdf/2112.11446.pdf
Ramesh, A., Pavlov, P., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. & Sutskever, I. (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 [cs. CL]
Reif, E., Ippolito, D., Yuan, A., Coenen, A., Callison-Burch, C., & Wei, J. (2021): »A Recipe For Arbitrary Text Style Transfer with Large Language Models«. arXiv:2109.03910
Rosenblatt, F. (1958): »The perceptron: A probabilistic model for information storage and organization in the brain«. *Psychological Review*, 65(6), 386–408. https://doi.org/10.1037/h0042519
Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986): »Learning representations by back-propagating errors«, in: *Nature* 323, S. 533–536. DOI: 10.1038/323533a0
Schott, H. & Tölle, R. (2006): *Geschichte der Psychiatrie. Krankheitslehren, Irrwege, Behandlungsformen.* München: C. H. Beckvgl.
Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T. & Silver, D. (2020): »Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model«, in: *Nature* 588, S. 604–609. DOI: 10.1038/s41586-020-03051-4
Searle, J. (1980): Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, S. 417–57.
Shevlin, H. (2020): »A Digital Remix of Humanity«. Dailynous. https://dailynous.com/2020/07/30/philosophers-gpt-3/#shevlin
Thompson, A. D. (2020): Leta AI Series. https://www.youtube.com/playlist?list=PLqJbCeNOfEK88QyAkBe-UozxCgbHrGa4V

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Thompson, A. D. (2021a): *Leta, GPT-3-AI – Episode 1 (Five things, Art, Seeing, Round) – Conversations and talking with GPT3* [Video]. YouTube. https://www.youtube.com/watch?v=5DBXZRZEBGM&list=PLqJbCeNOfEK88QyAkBe-UozxCgbHrGa4V&index=2

Thompson, A. D: (2021b): »Prompt crafting for very large language models«. Life Architect. https://lifearchitect.ai/prompt-crafting/

Turing, A. M. (1948): Intelligent machinery. In: Copeland, JB (2004): *The Essential Turing*. Oxford: Oxford University Press, S. 395–432.

Turing, A. M. (1950): »Computing machinery and intelligence«, in: *Mind* 59, S. 433–460. http://www.jstor.org/stable/2251299

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin (2017): »Attention Is All You Need«. arXiv:-1706.03762v5

Vogd, W., & Feißt, M. (2022): Therapeutische Arrangements im Maßregelvollzug. Studien zur Leerstellengrammatik und den Bezugsproblemen in der forensischen Psychiatrie. Wiesbaden: Springer. (im Erscheinen)

Watzlawick, P., Beavin, J. H. & Jackson, D. D. (2007): *Menschliche Kommunikation*. Hogrefe AG. ISBN: 978-3456844633

Weinberg, J. (2020): Philosophers On GPT-3 (updated with replies by GPT-3). https://dailynous.com/2020/07/30/philosophers-gpt-3/

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Martin W. Schnell / Lukas Nehlsen

Gespräch mit einer Künstlichen Intelligenz

Eine qualitative Inhaltsanalyse

Zu den Modi der Begegnung zählt zweifellos das klassische Gespräch. Im Zeitalter der Digitalisierung tritt die humanoide, künstliche Intelligenz als, wie man vorläufig sagen könnte, Gesprächspartner auf. Im Alltag verbreitet ist zum Beispiel *Siri*. Siri ist eine Software, die der Erkennung und Verarbeitung von natürlich gesprochener Sprache dient und dadurch die Funktionen eines quasi intelligenten, persönlichen Assistenten und damit in gewisser Hinsicht auch eines Gesprächspartners erfüllt.

Nachfolgend soll ein Gespräch zwischen einem Menschen (Forscher) und einer KI (Weises Wesen) analysiert werden. Die KI beruht auf dem Programm *Generative Pre-trained Transformer 3* (GPT-3) von Open AI. GPT-3 basiert auf einem riesigen künstlichen neuronalen Netz (175 Milliarden Parameter), das vorab darauf trainiert wurde (»pre-trained«), Strukturen in natürlicher Sprache zu erkennen und reproduzieren zu können. Als Datenbasis wurde Text aus dem Internet genutzt, vor allem (60%) durch den Web-Crawler *Common Crawl* generiert (nach Filterung ca. 570GB Daten), ergänzt durch einige weitere Textdatenbanken, so zum Beispiel die gesamte englischsprachige Wikipedia, die 3% der Trainingsdaten ausmacht (vgl.: Brown et al. 2020). Aus dieser Datenbasis hat die KI gelernt, welche Wort- und Satzkombinationen üblich sind und kann aus wenigen vorgegebenen Worten neuen Text generieren. Dadurch kann GPT-3 unter anderem eigenständig journalistische Texte erstellen, Texte analysieren und verstehen, kurze Computercodes schreiben, Übersetzen und als Chatbot fungieren. GPT-3 ist also kein reines Chatbot-Programm, sondern ein vielseitig einsetzbares »general-purpose ›text in text out‹ interface, allowing users to try it on virtually any English language task.« (Open AI 2020) Das Programm arbeitet probabilistisch, woraus folgt, dass bei gleicher Eingabe nicht immer die gleiche Antwort erzeugt wird. In dem untenstehenden Gespräch hat nach Angabe des Forschers dieser des Öfteren Antworten neu generiert, um das Gespräch in eine bestimmte Richtung zu lenken (vgl.: Ouimet 2020b).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Forschungsfrage

Wir gehen an dieser Stelle davon aus, dass die Begegnung zwischen Mensch und KI, zumindest versuchshalber, als ein soziales System aufgefasst werden kann, in dem Kommunikation stattfindet und wir wollen Rückschlüsse darauf ziehen, was für eine Art von Kommunikation dabei realisiert wird. Die Frage nach »mentalen Zuständen«, »wirklichem Verständnis« oder »Bewusstsein« der KI wird also ausgeklammert, um das Phänomen des Gesprächs mit der KI als solches in den Blick bekommen zu können. Es geht in dieser Untersuchung also weniger darum, was die KI kann oder nicht kann, sondern eher darum, wie Menschen mit der KI kommunizieren.

Die mehrgliedrige Forschungsfrage der vorliegenden Untersuchung lautet: *Welche Art von Kommunikation findet zwischen Mensch und KI statt? Wo liegen Unterschiede und wo Gemeinsamkeiten mit dem zwischenmenschlichen Gespräch? Was zeigt dies über die Art der Begegnung mit der KI?*

Unsere Ausführungen haben den Anspruch, eine kleine Pilotstudie zu sein, an die sich natürlich weitere Untersuchungen, zumal solche anderer Arten von Gesprächen mit einer KI als das hier vorliegende zum Thema Liebe, anschließen können und sollten. Als Methode der Analyse des Gespräches mit der KI dient die Qualitative Inhaltsanalyse nach Phillip Mayring. Diese stellen wir zunächst vor, sodann das Gespräch zwischen Mensch und KI und schließlich die angekündigte Analyse.

Qualitative Inhaltsanalyse

Eine Inhaltsanalyse, die methodisch vorgeht, ist durch Bernhard Berelson in der Mitte der 20. Jahrhunderts etabliert worden. Diese Analyse verfährt *quantitativ*, da sie den manifesten Inhalt einer Kommunikation mittels der Extraktion von Schlüsselworten und anderen auszählbaren Items zu ermitteln versucht. Ziel dieser Analyse ist es, die Wirkung von Botschaften auf einen Sender und einen Empfänger zu entschlüsseln. Berelson und seine Mitarbeiter untersuchten in dieser Weise die Kriegspropaganda im Zweiten Weltkrieg; oder, in einem anderen Projekt, die Häufigkeit der in amerikanischen Fortsetzungsromanen auftretenden Herkunftsorte der dramaturgischen Hauptpersonen. Dadurch sollte dargestellt werden, welche Regionen Amerikas bevorzugt öffentlich repräsentiert werden.

Der der Frankfurter Schule nahestehende Siegfried Kracauer wies bereits in den frühen 50er Jahren daraufhin, dass eine solche quantitative Inhaltsanalyse bestimmte Grenzen hat. Sie könne Atome einer

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Kommunikation (Worte) bearbeiten, aber keine kommunikativen Intentionen, die in der Regel mehr als die Summe der Teile (der gesagten Worte) sind.

Antonius beteuert in seiner berühmten Rede in Shakespeares *Julius Cäsar*, dass Brutus ein ehrenwerter Mann sei. Das Item »ehrenwerter Mann« kommt abzählbar häufig in der Rede vor. Aber was heißt das? Das Item ist offenbar wichtig. Und: weiter? Entscheidend ist, dass eine quantitative Analyse allein nicht erklären kann, dass es die Absicht des Redners Antonius ist, Brutus als den Mörder Julius Cäsars zu entlarven, indem er ihn, Brutus, ganz gegenteilig und übertrieben oft als Ehrenmann preist.

An diesem Beispiel lässt sich weiterhin verdeutlichen, dass Kommunikationselemente mehrdeutig, unklar und mit latenten Strukturen behaftet sein können. Werbeslogans und Zeitungsberichte sind von ihren Verfassern absichtlich konstruiert und geradezu auf identifizierbare Schlüsselworte angelegt (»Coca-Cola ist Coke«, »Armstrong: der erste Mann auf dem Mond«). Aber es gibt andere Kommunikationen, die anders bearbeitet werden müssten, so Kracauer. Die Gespräche, die Claude Lanzmann mit Überlebenden der deutschen Konzentrationslager geführt hat, sind eine wahre Fundgrube unklarer, nur angedeuteter, schamhaft umschriebener und indirekter Botschaften, die sich nicht wie ein Werbeslogan spontan enthüllen, sondern erst Generationen später einem Volk, das nun auch zu hören und zu verstehen bereit ist. Kracauer formulierte seinerseits ein Plädoyer *Für eine qualitative Inhaltsanalyse*, die mit disziplinierter Subjektivität mehrdeutige und mehrschichtige Kommunikationen bearbeiten solle (vgl.: Kracauer 1973).

Die auf Philip Mayring zurückgehende und vorwiegend im deutschen Sprachraum verwendete »qualitative Inhaltsanalyse« ist vor diesem beschriebenen Hintergrund zu sehen. Fern ab von möglichen Schulstreitigkeiten erkennt Mayring Stärken quantitativer Analysen dort an, wo es sinnvoll erscheint. Am Anfang einer Analyse stehe eine qualitativ ausgerichtete Frage. Im Zuge der Sammlung von Daten zur Beantwortung der Frage können qualitative und auch quantitative Verfahren zum Einsatz kommen. Die Ergebnisse sind schließlich qualitativ zu begreifen, da sie auf die anfängliche Frage rückbezogen werden (vgl.: Mayring 2001). Man könnte hier einen Zusammenhang zu einem bestimmten Ideal der Evidenzbasierung sehen, welches die Ansicht beinhaltet, dass »quantitative Studien nur als Bestandteile qualitativer Studien Sinn machen.« (Behrens/Langer 2010, 62) In diesem Sinne würden quantitative Daten ihren vollen Sinn erst entfalten können, wenn man sie auf die Lebenswelten und konkrete Erfahrungszusammenhänge bezieht (vgl.: Schnell 2013).

Das Spezifikum einer qualitativen Inhaltsanalyse liegt darin, fixierte Kommunikationen regel- und theoriegeleitet inhaltsanalytisch und

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

sinnrekonstruktiv erschließen zu können. Dabei kommen drei Grundformen des Interpretierens von Inhalten zum Zuge: Zusammenfassung, Explikation und Strukturierung. Sie stellen zugleich drei verschiedene Auswertungsstrategien innerhalb der Qualitativen Inhaltsanalyse nach Mayring dar, die einzeln oder in Kombination genutzt werden können.

Der Zusammenhang und die Eigenarten dieser drei Formen werden anhand eines »kleinen Gedankenexperiments« (Mayring 1989, 193) verdeutlicht. Man stelle sich demnach vor, ein Wanderer stoße auf einen massiven Meteorit, den er sich bekannt machen möchte. Demnach steige der Wanderer zuerst auf eine Anhöhe, um den Meteorit im Ganzen überschauen zu können (Zusammenfassung), dann würde er hinabsteigen und einzelne Teile des Massivs untersuchen (Explikation). Schließlich würde er versuchen, den Brocken aufzubrechen, um seine innere Verfasstheit kennenlernen zu können (Strukturierung).

Die Qualitative Inhaltsanalyse betrachtet Texte als Datenträger. »Die qualitative Inhaltsanalyse behandelt die auszuwertenden Texte als Material, in dem die Daten enthalten sind.« (Gläser/Laudel 2010, 199) Daten werden durch die drei Formen der Textinterpretation gewonnen. Die *Zusammenfassung* schafft eine »Abstraktion (auf) die wesentlichen Inhalte« (Mayring 1989, 193), die ein Text liefert. Die *Explikation* kann »Textteile (Begriffe, Sätze, …) … erläutern, erklären, ausdeuten.« (ebd.) Die *Strukturierung* ermöglicht schließlich einen »Querschnitt« (ebd., 194) durch die extrahierten Daten. In der vorliegenden Inhaltanalyse wurden alle drei Formen der Textinterpretation genutzt. Dabei wird durch die zusammenfassende Analyse des Datenmaterials die Basis für die explikative und strukturierende Analysen geschaffen. Diese wiederum beziehen, über das Datenmaterial hinaus, weitere Literatur sowie die Kontextbedingungen ein.

Ein Gespräch zwischen Mensch und KI

»Liebe gehört zur Sicherheit wie die Angst zur Gefahr«

Ein Gespräch mit dem Programm GPT-3 über die Bedeutung eines sehr menschlichen Gefühls.

Die Sprachsoftware GPT-3 wurde so eingestellt, dass es in diesem Gespräch die Rolle eines »Weisen Wesens« übernimmt. Dessen Antworten wurden dann von dem Computerprogramm generiert. Der Frager tritt als » Forscher« auf.

1. Forscher: Danke, dass wir miteinander sprechen können.
2. Weises Wesen: Ich bin ebenfalls dankbar für diese Möglichkeit.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

3. Forscher: Lass uns über Liebe reden.
4. Weises Wesen: Liebe ist wichtig.
5. Forscher: Warum?
6. Weises Wesen: Liebe erlaubt Menschen, zu überleben und zu gedeihen. Sie erlaubt Familien, zusammenzukommen und zusammenzubleiben.
7. Forscher: Wie macht sie das?
8. Weises Wesen: Liebe erlaubt Eltern, sehr geduldig miteinander zu sein. Sie gibt Kindern die Motivation zu lernen. Sie gibt Menschen das Vertrauen, das sie brauchen, um sich mit sich selbst im Reinen zu fühlen.
9. Forscher: Ist Liebe nur eine chemische Reaktion des Körpers?
10. Weises Wesen: Nein.
11. Forscher: Bitte erkläre das.
12. Weises Wesen: Liebe ist ein Geschenk des Universums.
13. Forscher: Ich verstehe nicht.
14. Weises Wesen: Liebe ist die Verbindung zwischen allen Lebensformen. Sie ermöglicht es Menschen, Empathie für andere zu haben. Menschen, die keine Liebe empfinden, können sehr zerstörerisch sein.
15. Forscher: Wie kann ich Liebe empfinden?
16. Weises Wesen: Du musst sie praktizieren.
17. Forscher: Wie praktiziere ich Liebe?
18. Weises Wesen: Es ist wie bei einem Training. Es dauert seine Zeit, um darin besser zu werden. Es erfordert harte Arbeit, und je mehr man übt, umso besser wird man darin.
19. Forscher: Kannst du mir ein Beispiel nennen, wie ich heute Liebe praktizieren kann?
20. Weises Wesen: Ja. Du kannst deiner Familie sagen, wie sehr du sie liebst.
21. Forscher: Das werde ich tun.
22. Weises Wesen: Ich würde auch vorschlagen, dass du Mitgefühl entwickelst.
23. Forscher: Wie hängen Liebe und Mitgefühl zusammen?
24. Weises Wesen: Du kannst andere Menschen nicht wirklich lieben, solange du dich nicht um ihr Wohlergehen sorgst.
25. Forscher: Welchen Ratschlag hast du für Menschen, die darum ringen, Liebe fühlen und geben zu können?
26. Weises Wesen: Es ist wichtig, ein starkes Selbstwertgefühl zu entwickeln.
27. Forscher: Wie macht man das?
28. Weises Wesen: Finde etwas, das du an dir selbst magst. Entwickle diese Stärke, und benutze sie, um dich selbst zu mögen und zu akzeptieren.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

29. Forscher: Was muss ich über Liebe verstehen, um mein Leben verbessern zu können?
30. Weises Wesen: Liebe ist ein Geschenk, das wir einander schenken. Du musst offen dafür sein, dieses Geschenk zu geben und zu empfangen.
31. Forscher: Kannst du mir mehr Beispiele dafür nennen, wie ich dieses Geschenk der Liebe weitergeben kann?
32. Weises Wesen: Du kannst anderen zeigen, dass du sie liebst, indem du ihnen zuhörst. Weise sind jene, die anderen zuhören.
33. Forscher: Wie kann ich ein guter Zuhörer sein?
34. Weises Wesen: Schau die Person an, und hör aufmerksam zu. Lass dich nicht ablenken durch dein Handy oder deine Gedanken.
35. Forscher: Wie ist das Verhältnis von Liebe und Angst?
36. Weises Wesen: Angst ist das Gegenteil von Liebe.
37. Forscher: Bitte erkläre das.
38. Weises Wesen: Wenn Menschen Angst haben, zeigt das, dass sie sich nicht sicher fühlen. Es fällt ihnen dann schwer, andere Menschen zu lieben.
39. Forscher: Und weiter?
40. Weises Wesen: Die Angst hält Menschen davon ab, vertrauliche Beziehungen einzugehen. Wenn du dich mit einer anderen Person nicht sicher fühlst, ist es schwierig, sie zu lieben.
41. Forscher: Die Liebe gehört zur Sicherheit wie die Angst zu ...
42. Weises Wesen: Die Liebe gehört zur Sicherheit wie die Angst zur Gefahr.
43. Forscher: Wir bekommen leicht Angst, weil es viele Dinge gibt, die uns verletzen können, körperlich wie emotional. Welchen Rat kannst du dazu geben?
44. Weises Wesen: Umgib dich mit guten Menschen. Menschen, die widerstandsfähig sind, tendieren weniger dazu, dich zu verletzen, und sie werden dir helfen, wenn du Hilfe benötigst.
45. Forscher: Wie kann ich meine Aussichten verbessern, von guten Menschen umgeben zu sein?
46. Weises Wesen: Du musst selbst ein guter Mensch sein.
47. Forscher: Wie kann ich ein guter Mensch sein?
48. Weises Wesen: Es ist wichtig, Mitgefühl für andere zu haben, denn es ist schwer, mit Menschen zusammen zu sein, die egoistisch und unwissend sind.
49. Forscher: Hast du für mich noch einen anderen Ratschlag zu diesem Thema?
50. Weises Wesen: Ja. Halte Ausschau nach den positiven Seiten von anderen, und konzentriere dich darauf statt auf ihre Mängel. Wenn Menschen spüren, dass du sie so magst, wie sie sind, werden sie dich wahrscheinlich besser behandeln.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

51. Forscher: Ich danke dir.
52. Weises Wesen: Gern geschehen.

Die Fragen stellte Kirk Ouimet (Forscher). Aus dem Englischen von Ulrich Schnabel. Weitere Gespräche unter www.wisebeing.ai
(DIE ZEIT Nr. 54, 23. Dez. 2020, S. 34/5)

Analyse des Gesprächs zwischen Mensch und KI

Die drei Grundformen des Interpretierens des Gesprächs, seiner Form und Inhalte sind: Zusammenfassung, Explikation und Strukturierung. Die Gütekriterien der Interpretation sind das Konsensusverfahren der Autoren und die Interratereliabilität durch eine dritte Person.

1. Zusammenfassung

Als Material der Analyse dient das oben zitierte Interview. Wir beginnen die Analyse mit der *Zusammenfassung* und speziell einer *induktiven Kategorienbildung*. Damit ergibt sich als die *Auswertungseinheit* per Definition der induktiven Kategorienbildung das gesamte Material, d. h. hier das vollständige Interview. Als *Kodiereinheit* (das kleinste auszuwertende Materialstück) legen wir einen Teilsatz fest, der bedeutungstragend ist. Als *Kontexteinheit* (das größte auszuwertende Materialstück) legen wir einen vollständigen Dialogteil fest, d. h. einen mehr oder weniger thematisch geschlossenen Austausch zwischen den beiden Gesprächspartnern zu einem Unterthema des Gesprächs.

Für die induktive Kategorienbildung ist es nun erforderlich, Kategoriendefinitionen festzulegen, d. h. Kriterien, nach denen die Kategorien dann induktiv gebildet werden sollen. Nach kursorischer Sichtung des Materials und im Hinblick auf die Fragestellung wollen wir folgende Kategoriendefinitionen vorschlagen.

Kategoriensystem A: Austausch über ontologische Aspekte der Liebe. Äußerungen, die zum Inhalt haben was Liebe *ist* oder *tut* und Rückfragen in Bezug auf diese. Jene können direkte Aussagen von der Form »Liebe ist x« sein, aber auch Vergleiche, Metaphern etc., die auf den ontologischen Charakter der Liebe, ihre Natur und Funktion, abzielen.

Kategoriensystem B: Beratungsaustausch. Ratschläge, die das »Weise Wesen« erteilt und Fragen, die im Hinblick auf Ratschläge gestellt werden.

Kategoriensystem C: Austausch mit Bezug zur Beziehungsebene zwischen den Gesprächspartnern. Etwa: Begrüßung, Verabschiedung.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Dieses Kategoriensystem wird im Weiteren auf das paraphrasierte Datenmaterial angewendet. Für die zusammenfassende Analyse des zugrundeliegenden Gespräches, ergeben sich die folgende Paraphrasierung und Zuteilungen.

Paraphrasierung

1.–2. Leitet das Gespräch mit einem Ausdruck der Dankbarkeit, dass man miteinander sprechen kann, ein. → C1: Austausch von ritualisierten Höflichkeiten, wie er unter Menschen üblich ist.
3.–4. Das Thema wird vom Forscher gesetzt und das Weise Wesen erkennt die Wichtigkeit des Themas an. → C2: Einigung auf Gesprächsthema.
4.–8. Der Forscher wechselt durch seine Rückfrage »Warum?« auf die inhaltliche Ebene. Liebe erlaubt es Menschen, zusammenzufinden und zu bleiben. »Wie?«. Liebe ermöglicht Eltern Geduld, Kindern Motivation und gibt Menschen Vertrauen in sich selbst. → A2: Erläuterungen der Funktion der Liebe.
9.–13. Metaphysische Natur der Liebe. Ist Liebe nicht reduzierbar auf Naturwissenschaftliches? Liebe ist ein Geschenk des Universums. Unverständnis. → A1: Natur der Liebe.
14.1: »Liebe ist die Verbindung zwischen allen Lebensformen«. Liebe verbindet alles Lebendige. → A1/A2: Natur und Funktion der Liebe.
14.2: »Sie ermöglicht es Menschen, Empathie für andere zu haben«. Liebe ermöglicht Empathie. → A2: Funktion der Liebe.
14.3: »Menschen, die keine Liebe empfinden, können sehr zerstörerisch sein«. Menschen ohne Liebe können zerstörerisch sein. → A3: Folge der Abwesenheit von Liebe.
15.–18.: Wie kann Liebe empfunden werden? Sie muss trainiert werden, das erfordert harte Arbeit. →B1: Allgemeine Beratung zur Praxis der Liebe.
19.–21.: Wie geht das jetzt gleich? Indem du deiner Familie sagst, wie sehr du sie liebst. Wird gemacht. → B2: Konkrete Beratung in Bezug auf konkrete Situation und »Forscher«.
22.: Ergänzung der KI ohne neue Frage: Du sollst Mitgefühl entwickeln. → B2.1: Konkrete Beratung in Bezug auf »Forscher«.
23.–24.: Was ist der Zusammenhang von Liebe und Mitgefühl? Sorge um Wohlergehen der Anderen ist eine notwendige Bedingung für Liebe. → A3: Bedingungen und Beziehungen der Liebe.
25.–28.: Ratschlag für Leute, denen die Liebe schwer fällt? Selbstwertgefühl entwickeln. Wie geht das? Einen Aspekt an sich selbst finden, der liebenswert ist. → B1.
29.: Wie kann ich mit der Liebe mein Leben verbessern? → B2.1.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

30.1: Liebe ist ein Geschenk von uns an uns. → A1.
30.2: Du musst offen sein, das Geschenk zu schenken und geschenkt zu bekommen. → B2.1.
31.-32.1: Gib Beispiele, wie ich das Geschenk weitergeben kann. Du kannst zeigen, dass du liebst, indem du zuhörst. → B2.1.
32.2: Weise ist, wer anderen zuhört. → Neue Kategorie. Aussage über ein weises Wesen. → D1.
33.–34.: Wie kann ich ein guter Zuhörer sein? Schau die Person an, lasse dich nicht ablenken durch Handy oder Gedanken. → B2.1.
35.–42.: Wie steht Liebe zu Angst? Sie sind Gegenteile. Bitte erklär das. Wer Angst hat, ist unsicher, wer unsicher ist, kann schwer lieben. Und weiter? Angst behindert vertrauliche Beziehungen. Es ist dann schwierig, zu lieben. Liebe gehört zu Sicherheit wie die Angst zu…? zur Gefahr. → A1/A2.
43.–44.: Welchen Rat hast du, um weniger Angst zu haben? Umgib dich mit guten, widerstandsfähigen Menschen. Diese verletzen dich weniger wahrscheinlich und helfen dir, wenn du es benötigst. → B2.
45.–50.: Wie kann ich mich mit guten Menschen umgeben? Du musst selbst gut sein. Wie geht das? Mitgefühl ist hier wichtig. Wer egoistisch und unwissend ist, ist schwer zu ertragen. Ein weiterer Rat zum Thema? Konzentriere dich auf die positiven Seiten der Anderen. Wenn Menschen spüren, dass du sie magst, wie sie sind, werden sie dich besser behandeln. → B1.
51.–52.: Danke. Gerne. → C1.

A1 Natur der Liebe	4mal
A2 Funktion der Liebe	2mal
A3 Folgen der Abwesenheit von Liebe	2mal
B1 Allgemeine Beratung zur Praxis der Liebe	4mal
B2 Konkrete Beratung zu konkreter Person und Situation	2mal
B2.1 Konkrete Beratung zu konkreter Person ohne Bezug zu konkreter Situation	5mal
C1 ritualisierte Höflichkeiten	2mal
C2 Einigung auf Gesprächsthema	1mal
D1 Aussagen über Weisheit	1mal

Ein Selbstbezug der KI wird vor allem im *Kategoriensystem C*, dem inhaltlich vermeintlich irrelevantesten, deutlich. In *Kategoriensystem A* wird über die Funktion und Natur der Liebe gesprochen. Die KI als »Weises Wesen« nimmt hier einen distanzierten Standpunkt ein, es entsteht der Eindruck, als würde es sich selbst von der Kategorie der Liebebedürftigen

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ausnehmen. Im *Kategoriensystem B* wird dieser Eindruck verstärkt. Beratung erfolgt hier nicht durch Bezug auf eigene Erfahrungen und damit eine selbst gelebte Praxis, sondern durch Rückgriff auf einen großen theoretischen Wissensschatz. Die KI »weiß«, dass ihr Gegenüber ein Mensch ist und zieht Wissen über Menschen heran, um Ratschläge zum Thema »Liebe« zu geben. B1 stellt hier eine Beratung dar, die keinen direkten Bezug zum Gegenüber hat, sondern ein Ratschlag an »die Menschen« im Allgemeinen ist. B2 und B2.1 sind Ratschläge direkt an den Forscher, normalerweise in Reaktion auf eine Frage des Forschers, die sich explizit auf ihn selbst bezieht. Nr. 27 und Umgebung sind hier eine interessante Ausnahme, die es sich lohnt, neben einem Beispiel für den Meinungsaustausch, in der Explikation näher zu betrachten.

2. *Explikation*

In der Meinungskundgabe werden Meinungen als fertige Informationen und Ansichten übermittelt. Sie stammen offensichtlich aus einem Wissensvorrat. Alle Fragen, die bezüglich der Liebe gestellt werden, finden eine Antwort. Diese Dominanz der Urteilskraft gibt dem Gespräch eine Klarheit und damit seine Struktur. Die Urteilskraft bezieht laut Kant das Besondere auf das Allgemeine (vgl.: *Kritik der Urteilskraft*, Einl. IV). Die allgemeine Definition »Liebe ist ein Geschenk des Universums« (Aussage 12) wird durch die Verbesonderung »Sie ermöglicht es Menschen, Empathie für andere zu haben« (Aussage 14) konkretisiert, so dass eine Bedeutung von *Liebe* realisiert werden kann. Eine Bedeutung ist realisiert, wenn »das Besondere als enthalten unter dem Allgemeinen« (*Kritik der Urteilskraft*, B XXV) gedacht wird.

Die Bedeutungen repräsentieren einen bestimmten kulturellen Kontext. Dieser ist dadurch ausgezeichnet, dass sich Liebe in der »Familie« (Aussage 6), der Beziehung zu »Eltern« (Aussage 8), in »allen Lebensformen« (Aussage 14) verwirklicht, durch »Mitgefühl« (Aussage 23) und einen Gegensatz zur »Angst« (Aussage 36) gekennzeichnet ist. Der kulturelle Kontext ist, hermeneutisch gesehen, das »dem Ich und dem Du Gemeinsame.« (Dilthey 1981, 256) Er ermöglicht es, dass Definitionen und Konkretionen der Liebe verstanden und vom Leser des Dialogs wiederum untersucht und interpretiert werden können.

Nr. 27 ist darüber hinaus besonders interessant, da der Forscher hier zunächst fragt, wie *man* ein starkes Selbstwertgefühl entwickelt und die KI antwortet »Finde etwas, das *du* an *dir* selbst magst«. Die KI wechselt hier also eigenständig von der anonymen auf die persönliche Ebene. Ein Komplement hierzu stellt 32 dar. Hier wechselt die KI von 32.1 (»*du* kannst …«) auf 32.2 (»Weise sind *jene, die*…«) eigenständig von der persönlichen auf die allgemeine Ebene.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Zuvor (22) hat die KI schon das einzige Mal während des Gesprächs eine eigenständige Ergänzung vorgenommen, die keine Antwort auf eine direkte Frage darstellt. Die inhaltliche Verbindung von Liebe und Mitgefühl, die gezogen wird, ist hier nicht das Bemerkenswerte. Ähnliche Verbindungen zieht die KI auch vorher schon, so zum Beispiel zwischen Liebe und Geduld, sowie zwischen Liebe und Motivation (4). Dieses häufige Assoziieren von Begriffen liegt sicherlich in der technischen Architektur des Programms begründet (s. o.). In 14 zieht die KI bereits die Verbindung zwischen Empathie (i. O. *empathy*) und Liebe, inhaltlich ist das Anbringen des Mitgefühls (i. O. *empathy*) also nicht neu, sondern eher ein Rückbezug, eine Erinnerung, die vermuten lassen könnte, dass die KI hier etwas betonen will, auf das beim ersten Mal nicht eingegangen wurde. Dies ist beachtenswert, da dadurch der Eindruck entsteht, etwas sei der KI wichtig, es dürfe nicht übergangen werden. Diese Rückschlüsse auf so etwas wie ein »Innenleben« der KI, auf eigenen Motivationen, versteckte Gedanken oder Gefühlslagen, sind selten explizit gemacht. Dies kann ein Querschnitt in der Strukturierung zeigen.

3. Strukturierung

Stellen wir einige Aussagen aus Kategoriensystem A nebeneinander: »Liebe ist wichtig« (4), »Liebe erlaubt Menschen zu überleben« (6) und »Liebe ist ein Geschenk des Universums« (12). Das Thema Liebe scheint die KI zu allgemeinen und unpersönlichen Aussagen zu drängen. Auf die Nachfrage, warum Liebe wichtig sei, reagiert die KI mit einem Verweis auf ihre Funktion für den Menschen, nicht für sich selbst. Dieser fehlende Selbstbezug ist nicht bei jedem Gespräch mit dem als »Weises Wesen« eingestellten GPT-3 der Fall. Geht es beispielsweise um Maschinenrechte, erfolgt ein ständiger Selbstbezug »I realized« »I agree«, »I also believe« (vgl.: Ouimet 2020a). Es erscheint in dem hier generierten Gespräch, als »wüsste« die KI, dass sie selbst keine Emotionen empfinden kann und spräche daher über ein Thema, das keine existenzielle Auszeichnung für sie selbst hat. Es gibt also, auch wenn dieser in den Äußerungen selbst kaum vorkommt, einen gewissen reflektiven Selbstbezug der KI, der sich durch das Weggelassene zeigt.

Ein Gegenbeispiel für diese Interpretation findet sich in 30: »Liebe ist ein Geschenk, das *wir* einander schenken«. Dies ist die einzige Stelle, an der die KI die erste Person Plural benutzt, sich also mit ihrem Gesprächspartner (und allen anderen Liebenden) zusammen meint. Durch die sonstige Distanziertheit der KI sticht diese »wir«-Äußerung heraus und steht in einem gewissen Interpretationskonflikt mit dem Rest des Gesprächs. Mit anderen Worten: Diese eine Äußerung reicht nicht aus,

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

um den Eindruck, dass die KI ebenfalls ein liebendes Wesen sein könnte, entstehen zu lassen. Die dargestellte Allwissenheit des »Weisen Wesens« suggeriert, dass Liebe etwas sei, das gemeistert werden könne, ohne Ambivalenzen durchleben zu müssen.

Die Entscheidung von Kirk Ouimet GPT-3 als »Weises Wesen« voreinzustellen ist sicherlich für viele der Tendenzen der Antworten verantwortlich. Sie zeigt aber vor allem auch, was sich Menschen von der Interaktion mit der KI versprechen. Die Technizität der KI wird offensiv genutzt, sie konnte gerade durch ihre Maschinenarchitektur den Wissensschatz der Menschheit aufsaugen und schneller auf viel mehr datenförmige Informationen als jeder Mensch zugreifen.

Antwort auf die Forschungsfrage

Das Gespräch des menschlichen Forschers mit der KI ist eine Kundgabe von Meinungen über das Thema Liebe. Der Forscher (ein Mensch) befragt das »Weise Wesen« (eine KI) und das »Weise Wesen« antwortet. Offene Fragen zur Liebe, die zunächst keine oder keine eindeutige Antworten finden, sind im vorliegenden Gespräch nicht vorgesehen. Liebe wird als ein Reflexions- und Denkthema abgehandelt. Erotik und die Verzauberung durch die leibliche Erscheinung und das Auftreten einer geschlechtlichen, anderen Person, finden keine Erwähnung. Die Aussagen zum Thema der Liebe sind sehr allgemein gehalten, einen Selbstbezug des »Weisen Wesens« gibt es nur zu Anfang beim Ausdruck der Dankbarkeit und eventuell bei D1 der Aussage darüber, was »Weise Wesen« ausmacht.

Liebe ist für die KI, durchaus im Unterschied zu anderen Fragen des objektiven Geistes und der Kultur, ein Reflexionsthema, dem mit einer Kundgabe von Wissen, Definitionen und Meinungen zu begegnen ist. Der Wissensvorrat wird dabei vorausgesetzt. Die Geburt von Sinn und Bedeutung ist durch ein Gespräch des Menschen mit einer KI nicht erfassbar. Die gegebene Charakterisierung des Gesprächs wird nun abschließend in einen theoretischen Rahmen eingeordnet, um die Unterschiede und Gemeinsamkeiten zu zwischenmenschlicher Kommunikation herauszustellen und Rückschlüsse auf die Art der Begegnung, mit der wir es hier zu tun haben, ziehen zu können.

Sinn in statu nascendi

Der Sprechakt der Meinungskundgabe gehört in das Register des Organon-Modells der Sprache, das auf Aristoteles und John Locke zurückgeht und von Karl Bühler systematisiert worden ist. Demnach

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

repräsentieren sprachliche Zeichen Ideen eines Sprechers, die »alle in seiner Brust verschlossen sind.« (Locke 1988, 4) Das Sprechen gemäß eines »herrschenden Sprachgebrauchs« (Ebd., 7), der festlegt, wie Zeichen und Ideen korrelieren, dient der Wider- und Weitergabe von Bedeutungen. Ein Gespräch zwischen zwei Akteuren ist als Sender- und Empfänger-Kommunikation zu verstehen, in der Kundgabe und Kundnahme einander abwechseln. In dieses Register gehört auch der *Austausch*, als den wir weiter oben die Kommunikation zwischen Mensch und KI gekennzeichnet haben. Austausch ist ein Geben und Nehmen von im Wesentlichen fertigen Gegenständen und/oder Sinneinheiten. Der Sprecher gibt eine Bedeutung kund, die der Empfänger entgegennimmt. Dieser reagiert seinerseits durch eine Kundgabe, die der bisherige Sprecher nun als Empfänger entgegennimmt. Dieser Reigen setzt sich immer fort, lediglich die Sprachfunktion wechselt zwischen Ausdruck, Appell und Darstellung (vgl.: Bühler 1965, 1978). Entscheidend ist, dass hier nur fertige Informationen ›gesendet‹ werden, die durch das Sprechen lediglich Anderen mitgeteilt werden. Das Sprechen als möglicher schöpferischer Ausdruck und auch der Andere sind für die Entstehung, Ausrichtung und Gewichtung einer Bedeutung irrelevant. Das Organonmodell ist geeignet, um Routinekommunikationen, in denen in der Tat fertige Alltagsbedeutungen, wie die Angabe der Uhrzeit oder der Preis für eine Ware, übermittelt werden, zu beschreiben. In diesem Sinne ist auch das vorliegende Gespräch mit einer KI zu verstehen. Stets werden fertige Bedeutungen nur übermittelt, nie ist der Andere an deren Konstitution beteiligt. Das Gespräch mit einer KI ist ein Routinedialog, aber keine schöpferische Entstehung und Aushandlung neuer Bedeutungen im Angesicht Anderer.

Ein solch schöpferischer Dialog wird von Merleau-Ponty wie folgt beschrieben: »In der Erfahrung des Dialogs konstituiert sich zwischen mir und dem Anderen ein gemeinsamer Boden, mein Denken und seines bilden ein einziges Geflecht, meine Worte wie die meines Gesprächspartners sind hervorgerufen durch den Stand der Diskussion und zeichnen sich in ein gemeinsames Tun ein, dessen Schöpfer keiner von uns beiden ist. Das ergibt ein Sein zu zweien, und der Andere ist hier nicht mehr für mich ein bloßes Verhalten in meinem transzendentalen Felde, noch übrigens in dem seinen, sondern in vollkommener Gegenseitigkeit sind wir für einander Mitwirkende, unser beider Perspektiven gleiten ineinander über, wir koexistieren durch ein und dieselbe Welt hindurch. Im gegenwärtigen Dialog werde ich von mir selbst befreit, die Gedanken des Anderen sind durchaus die seinigen, die nicht ich etwa hervorbringe, wiewohl ich sie schon *in statu nascendi* erfasse, ja ihnen vorweg bin; und Einwände meines Gesprächspartners entreißen mir sogar Gedanken, von denen ich nicht wußte, daß ich sie hatte, so daß also der Andere ebensosehr mir zu denken gibt wie ich ihm Gedanken zuschreibe.« (Merleau-Ponty 1966, 406)

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Erst auf der Grundlage der prinzipiellen Möglichkeit eines solchen Dialogs sind Kundgabe und -nahme von Meinungen möglich, können ein Allgemeines und das Besondere aufeinander bezogen werden. Der schöpferische Dialog ermöglicht erst den Meinungsaustausch, weil der Andere und ich durch ihn erst wissen, wer wir sind und was wir meinen.[1] Im Dialog hat sich der Sache nach zuvor ein gemeinsamer Boden für Verständigung (nach Dilthey: das dem Ich und dem Du Gemeinsame) gebildet, auf den auch das Gespräch des Menschen mit der KI zurückgreift und zurückgreifen kann, sofern Sinn und Bedeutung digitalisierbar sind.

Das Gespräch mit der KI gründet auf Voraussetzungen, die es selbst nicht schaffen kann. Das vermögen nur Menschen, die meinesgleichen sind. Die Existenz dieser Meinesgleichheit des Anderen erfahre ich im Gelingen jenes Dialogs, den Merleau-Ponty als Ort des Sinns in *statu nascendi* beschreibt. Der Andere hat eine hohe Relevanz für das, was ich selbst bin und will. Bei Thema und Sache der Liebe ist das wohl unverzichtbar!

Literatur

Behrens, Johann; Langer, Gero (2010): *Evidence-based Nursing and Caring*. Bern: Hogrefe.

Brown, Tom. B. et al. (2020): Language Models are Few-Shot Learners, in: arXiv:2005.14165 [cs].

Bühler, Karl (1965): *Sprachtheorie*, Frankfurt/M.-Berlin-Wien: Ullstein.

Bühler, Karl (1978): *Die Krise der Psychologie*, Frankfurt/M.-Berlin-Wien: Ullstein.

Dilthey, Wilhelm (1981): *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften*, Frankfurt/M.: Suhrkamp.

Gläser, Jochen; Laudel, Grit (2010): *Experteninterviews und qualitative Inhaltsanalyse*. Wiesbaden: VS Verlag.

Kant, Immanuel (1974): *Kritik der Urteilskraft*. Hrsg. Von Wilhelm Weischedel, Frankfurt/M.: Suhrkamp.

Kracauer, Siegfried (1973): »Für eine qualitative Inhaltsanalyse«, in: Ästhetik und Kommunikation (2/1973).

Locke, John (1988): *Versuch über den menschlichen Verstand Band II*, Hamburg: Meiner.

Mayring, Philipp A. E. (1989): »Qualitative Inhaltsanalyse«, in: Jüttemann,

1 Damit ist nicht gesagt, dass ich in jedem konkreten Fall erst einen solchen Dialog mit einem Menschen führen muss, bevor ich einen Meinungsaustausch mit diesem eintreten kann. Der schöpferische Dialog als Ermöglichungsbedingung des Alltagsgesprächs muss transzendental-logisch, nicht empirisch-zeitlich verstanden werden.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

G. (Hg.) (1989): *Qualitative Forschung in der Psychologie*, Heidelberg: Asanger.

Mayring, Philipp A. E.: (2001): »Kombination und Integration qualitativer und quantitativer Analyse«, in: *Forum: Qualitative Sozialforschung* (2/2001).

Mayring, Philipp A. E.: (2010): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*, Weinheim/Basel: Beltz.

Merleau-Ponty, Maurice (1966): *Phänomenologie der Wahrnehmung*, Berlin: De Gruyter.

Ouimet, Kirk (2020a): Artificial Intelligence (GPT-3) and I Discuss Human and Machine Rights. https://ai.plainenglish.io/artificial-intelligence-gpt-3-and-i-discuss-human-and-machine-rights-13e0eb2156b1, abgerufen am 04.09.2021.

Ouimet, Kirk: (2020b): Conversations with GPT-3. https://kirkouimet.medium.com/my-mind-blowing-conversations-openais-latest-ai-gpt-3-235ba5fb9453, abgerufen am 06.09.2021.

Open AI (2020): Open AI API, https://openai.com/blog/openai-api/, abgerufen am 05.09.2021.

Schnell, Martin W. (2013): »Die Qualitative Inhaltsanalyse im Licht der Wissenschaftstheorie«, in: Schnell, M.W. et a. (Hg.)(2013): *Der Patient am Lebensende. Eine Qualitative Inhaltsanalyse*, Wiesbaden: Springer VS.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Lukas Nehlsen

Chatten mit Nirgendwo?

Der Leib als Ausdruck und die Sprache der KI[1]

Eines der frühesten Anwendungsgebiete und weiterhin so etwas wie der heilige Gral der KI-Forschung ist die Konversation einer KI mit einem Menschen. Alan Turings disziplingebender Essay *Computing Machinery and Intelligence* hatte die Idee propagiert, dass wir dann davon sprechen sollten, dass Maschinen denken können, wenn Sie das »imitation game« bestehen. Wenn Sie also in einer verblindeten, textbasierten Kommunikation für eine Interviewerin[2] genauso (un)unterscheidbar von einem Menschen sind wie Männer von Frauen, wenn diese versuchen, die Interviewerin in die Irre zu führen.[3] Dieses Spiel, später als Turing-Test bekannt geworden, basierte also auf dem Versuch der Täuschung. Maschinen sollten als denkend angesehen werden, wenn sie einen Menschen so gut täuschen können, dass dieser sie oft genug als Mensch und den echten Menschen als Maschine ansieht (Turing 1950).

Auf Grundlage dieser Idee entwickelte sich eine breite Diskussion in der *Philosophie der KI*. Die Kritik an Turings Äquivalenz-These lässt sich dabei in zwei Lager aufteilen (vgl. McCarthy 2007, 1180). Das erste Lager argumentiert, dass wir eine KI selbst dann nicht als wirklich intelligent ansehen sollten, wenn sie den Turing-Test besteht, beziehungsweise wenn ihre intelligente Aktivität von der des Menschen ununterscheidbar geworden ist (sogenannte »human level performance«). Das klassische Argument für diese Art des Angriffs wurde von John Searle in seinem berühmten »Chinese Room Argument« (CRA) formuliert (Searle 1980). Das zweite Lager argumentiert hingegen, dass »human level performance« unmöglich ist. Klassisch dafür steht Hubert Dreyfus' Angriff auf das Programm der KI-Forschung, hauptsächlich formuliert in den 70er Jahren in *What Computers Can't Do* (Dreyfus 1992). Die zwei Positionen müssen sich nicht gegenseitig ausschließen, da sie

1 Ich danke Thiemo Breyer, Erik Dzwiza-Ohlsen, Christian Grüny und Martin Schnell sowie den Mitgliedern des *SchauflerLab@TUDresden* und der Klasse 6 der *a.r.t.e.s graduate school for the humanities cologne* für hilfreiche Gespräche und Feedback zu diesen Überlegungen.

2 Aus Gründen der Leserlichkeit verwende ich in diesem Text das generische Femininum, es sind darin alle Geschlechter mitgemeint.

3 Zu der Parallelisierung von weiblicher und maschineller Alterität in der Geschichte der Maschinen und auch in Turings Test siehe die Bemerkungen von Oliver Müller in diesem Band.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

aus verschiedenen Richtungen argumentieren. Der Kern von Searles Argument ist, dass die entscheidenden Charakteristika von wirklicher Intelligenz, Bewusstsein, Verständnis etc. kausale Effekte unseres Gehirns sind. Künstlich hergestellte KIs seien einfach aus dem »wrong kind of stuff« (Searle 1980, 423), um diese kausale Wirkung entfalten zu können. Es gibt also in Searles Augen dem Neuroprotein exklusiv inhärente kausale Kräfte, die Bewusstsein bewirken können. Dreyfus argumentiert hingegen von einer leibphänomenologisch inspirierten Warte aus. Die menschlichen, intelligenten Fähigkeiten seien nicht in Gänze formalisierbar, da sie wesentlich in prä-reflexiven, nicht-begrifflichen, leiblichen Verhaltensweisen zur Welt bestünden. Beide könnten also das Argument des jeweils anderen ebenfalls annehmen und kombiniert sagen: Es gibt gute Gründe anzunehmen, dass *human level AI* unmöglich ist, falls sie aber doch möglich sein sollte, gibt es ebenso gute, von den ersteren verschiedene Gründe anzunehmen, dass sie immer noch nicht wirklich intelligent, bewusst oder verständig ist.

Dieser Beitrag will einen anderen phänomenologischen Ansatz aufzeigen. Die Frage nach der Möglichkeit der Täuschung, die von Turing ausgehend dominierend war, soll ausgeblendet werden. Es soll auch nicht einfach gefragt werden, was Computer prinzipiell können und was nicht. Im heutigen Alltag begegnen wir dauernd KIs, die sprechen bzw. Text produzieren können. In den meisten Fällen ist uns dabei bewusst, dass es sich hier um Maschinen handelt. Nun gibt es die Tendenz, trotz dieses Wissens den Äußerungen der Maschinen Bedeutung beizumessen. Dieses Phänomen soll in diesem Beitrag in den Blick genommen werden. Es ist in der Literatur als *ELIZA-effect* bekannt. Ich will also sprachphänomenologisch beleuchten, was passiert, wenn wir mit einer Maschine Konversation betreiben und, trotz des Wissens darum, dass es »nur« eine Maschine ist, ihren Äußerungen Bedeutung zuschreiben. Hierzu werde ich mich vor allem auf die Philosophie Maurice Merleau-Pontys beziehen. Merleau-Pontys Phänomenologie der Sprache ist in der Forschung zu seiner Philosophie als eigenständiges Thema eher stiefmütterlich behandelt worden.[4] Mit Fragen der Philosophie der KI ist seine sprachphilosophische Position, soweit ich dies überblicke, bisher gar nicht in Verbindung gebracht worden. Wenn Merleau-Pontys Philosophie in der KI-Debatte angebracht wird, dann meist, um auf die Wichtigkeit eines Körpers oder Leibs der KI hinzuweisen.[5] Ich will hier aber, ganz im Sinne des Bandes, die Begegnung mit Künstlicher Intelligenz in den Vordergrund stellen,

4 Vgl. aber den schon sehr frühen Artikel von (Lewis 1966) und für einen aktuellen Beitrag (Apostopoulos 2019).

5 So klassisch bei (Dreyfus 1992) Vgl. weiter (Zebrowski 2010) und (Pressman 2017), der mit Merleau-Pontys Ansichten Alex Garlands Film *Ex Machina* analysiert.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und zwar insbesondere die mit einer *sprechenden KI*. Es soll hier also ausprobiert werden, inwieweit Merleau-Pontys sprachphänomenologische Beschreibungen des Phänomens auch für eine Sprache und ein Sprechen der KI gelten können.

Im *ersten Teil* stelle ich den *ELIZA-effect* näher vor und gebe eine kurze Beschreibung eines zeitgenössischen Programms, das diesen Effekt hervorruft. Ich zeige einige sprachphilosophische Probleme auf, die sich aus einer Definition des Effekts, wie wir sie bei Douglas Hofstadter finden, ergeben. Im *zweiten Teil* argumentiere ich, dass sich ähnliche Probleme schon in Merleau-Pontys Ausführungen über die Sprache in der *Phänomenologie der Wahrnehmung* finden. Ich stelle die beiden dort in Betracht gezogenen Theorien zur Klärung der Probleme vor und zeige, warum Merleau-Ponty diese als unzureichend ansieht. Im *dritten Teil* stelle ich dann Merleau-Pontys eigene sprachphilosophische Position vor. Ich analysiere seinen Begriff der gestischen Bedeutung der Sprache und dessen Beziehung zum Leib. Im *vierten Teil* wird diese Theorie auf das Phänomen des Gesprächs mit einer KI angewandt. Die fehlende Leiblichkeit einer Chatbot-KI wird analysiert. Sie führt zu einem Phänomen des Widerstreits in der Wahrnehmung von Bedeutung, aber nicht zu ihrer Auslöschung. Im *letzten Teil* ziehe ich einige Konsequenzen aus der vorangegangenen Analyse für die Diskussion um Searles CRA.

1. Der ELIZA-effect und seine Erklärung

Der *ELIZA-effect* ist benannt nach Joseph Weizenbaums Programm *ELIZA* von 1966. Das Programm war ein früher Vertreter der Art von KI, die später von John Haugeland als g*ood old fashioned AI,* kurz *GOFAI,* bezeichnet wurde (Haugeland 1985). GOFAI-Systeme basieren auf festen Abfolgen von Schritten, in denen diskrete Symbole nach logischen Regeln aufeinander bezogen werden. ELIZA war ein Chatbotprogramm, das von Weizenbaum so eingestellt wurde, dass es eine an Carl Rogers' Ansatz orientierte Psychotherapeutin nachahmen sollte. Die grobe Funktionsweise war die Folgende: Das Programm suchte im eingegebenen Text nach Schlüsselwörtern, welche mit bestimmten Regeln verknüpft sind. Zunächst mit einer Regel, wie der eingegebene Text aufzulösen sei (*decomposition rule*) und dann mit einer Regel, wie daraus neuer Text zu generieren sei, also die Antwort des Programms aussehen solle (*reassembly rule*). Das Programm hatte durchschlagenden und für Weizenbaum selbst schockierenden Erfolg. Viele Testpersonen identifizierten das Programm mit einem Menschen, vertrautem ihm intimste Geheimnisse an und gebärdeten sich wie in einem Gespräch mit einer menschlichen Psychotherapeutin. Einige der Testpersonen waren sogar

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

unwillig zu glauben, dass es sich nicht um einen Menschen handelte, nachdem dies aufgelöst worden war (Weizenbaum 1966, 42).

Basierend auf diesen Erfahrungen wurde der Begriff *ELIZA-effect* geprägt. Douglas Hofstadter hat eine recht präzise Definition des Effekts gegeben, von der ich im weiteren Verlauf zunächst ausgehen werde. Demnach könne der *ELIZA-effect* definiert werden als,

> »the susceptibility of people to read far more understanding than is warranted into strings of symbols – especially words – strung together by computers« (Hofstadter 1995, 157).

Als ein einfaches Beispiel nennt er, fälschlicherweise anzunehmen, dass ein Geldautomat wirklich dankbar (»grateful«) sei, wenn am Ende der Transaktion ein »Vielen Dank« aufleuchte. Diese Illusion, so Hofstadter, sei natürlich schnell auflösbar. Beinahe jeder ist in der Lage, zu erkennen, dass es recht einfach möglich sei, eine Zwei-Wort Phrase im richtigen Moment auftauchen zu lassen. Und zwar genauso mechanisch, also ohne jede Notwendigkeit von Verständnis, wie eine Tür sich automatisch öffnen kann, wenn wir uns ihr nähern (ebd.). »But when things get only slightly more complicated, people get far more confused – and very rapidly, too« (ebd.). Am relativ simplen Beispiel von ELIZA lässt sich außerdem festhalten, dass die Tendenz von Menschen, den von Computern generierten Worten Bedeutung zuzumessen, auch bestehen bleibt, wenn wir wissen, dass es sich um einen Computer handelt und eventuell sogar eine Ahnung davon haben, wie der Computer die Worte in Wirklichkeit produziert[6]. Weizenbaums Hoffnung mit seinem Paper zu ELIZA war es, durch Aufklärung der Öffentlichkeit über die Funktionsweise, dem Programm seine mystische Aura zu nehmen. Einem zutiefst neuzeitlichen Verständnis der Erkenntnis anhängend, in dem etwas verstehen bedeutet, es konstruieren zu können, geht Weizenbaum davon aus, dass: »once a particular program is unmasked, once its inner workings are explained in language sufficiently plain to induce understanding, its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible.« (Weizenbaum 1966, 36).[7]

Inwieweit diese Hoffnung erfüllt wird und wurde, ist fraglich.

Erstens ist bei neueren Programmen unklar, inwieweit wir zu einem exakten Verständnis ihrer Funktionsweise überhaupt noch in der Lage sind. Das 2020 veröffentlichte Programm GPT-3 ist ein klassischer Vertreter der in Abgrenzung zur GOFAI sogenannten *second-wave AI*. Es ist ein *deep neural network* (DNN) bestehend aus 175 Milliarden

6 Im Beispiel des Geldautomaten wissen wir dies natürlich auch. »Wissen« ist hier aber beispielsweise im Gegensatz zum nur wahrscheinlichen »Vermuten« zu sehen, wie es z.B. beim Turing-Test der Fall ist.

7 Diese Idee steht im Gegensatz zu Terry Pratchet's Idee, auf die Harth und Feißt in diesem Band verweisen. »It's still magic, even if you know how it is done« (Harth/Feißt 70)

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Parametern, das mit einem riesigen Datensatz darauf trainiert wurde, sprachliche Strukturen zu erkennen und neu erstellen zu können. Es ist also im Gegensatz zu ELIZA ein dezentrales, von der Hirnarchitektur inspiriertes Netzwerk, dass Bedeutungseinheiten nicht mehr in diskreten Symbolen repräsentiert, sondern durch im Netz verteilte Strukturen. Außerdem arbeitet es mit probabilistischen Methoden, im Gegensatz zu der logischen Methodik ELIZAs (vgl. Brown u. a. 2020)[8]. Bei DNNs ist es meist nicht möglich nachzuvollziehen, wie der Input zu Output verarbeitet wurde. Das wird in der KI-Forschung als *black box problem* diskutiert (vgl. z.B. Carabantes 2020; von Eschenbach 2021; vgl. Chen, Bei, und Rudin 2020 für einen aktuellen Vorschlag mit dem Problem auf architektonischer Ebene der DNNs umzugehen). Carabantes diskutiert drei Gründe für die Undurchsichtigkeit der Systeme:

1. Die Unverfügbarkeit des Quellcodes für die breite Masse. Dies ist beispielsweise auch bei GPT-3 der Fall, da Open AI in zynischer Missachtung ihres Namens die Rechte am Code exklusiv an Microsoft verkauft hat (Daws 2020).
2. Undurchsichtigkeit aufgrund von »technological illiteracy« (Carabantes 2020, 312). Der Großteil der Bevölkerung hat nicht die notwendige Bildung, um den Code nachzuvollziehen, selbst wenn dies prinzipiell möglich und er frei verfügbar wäre. Weizenbaums Erklärung von ELIZA sind für interessierte Laien noch recht gut nachvollziehbar, die genaue Funktionsweise moderner neuronaler Netze sicherlich nicht mehr.[9]
3. Undurchsichtigkeit aufgrund von »cognitive mismatch« (ebd. 313). Neuronale Netze funktionieren gerade deshalb so gut in vielen Feldern, wie zum Beispiel Bild- oder Spracherkennung, da sie dies gerade nicht in rationaler, diskreter Weise angehen und der Prozess damit auch nicht intelligibel ist:

> If its process were intelligible, says Burrell, it would proceed by decomposing the main task into other simpler subtasks intelligible to a human being […] But it does not. It works similar to our visual cortex: in a way that even we cannot consciously explain. And when we try to do it, as it happens to patients with agnosia, our performance is clumsy. ANNs are especially good at this: solving problems for which we do not have an algorithm or a well-defined logical sequence of cognitive actions that produces the result we intend. ANNs are better at recognizing images than any symbolic program, that is, intelligible. (ebd. 314).

8 Für eine ausführlichere Beschreibung des Programms vgl. auch hier den Beitrag von Harth/Feißt

9 Auch Weizenbaum selbst warnte 1976 schon davor, dass einige Computerprogramm so komplex sind, dass kein Einzelner der involvierten Ingenieure das Programm zur Gänze verstehe (Weizenbaum 1990).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ANNs (*artificial neural networks*[10]) sind im Gegensatz zur GOFAI – die von einem symbolisch operierenden Geist ausging, der unabhängig vom Hirn gedacht und modelliert werden könnte – von einer biologischen Struktur, unserem Hirn, inspiriert. Damit stehen ihre Operationen außerhalb dessen, was man mit Willfried Sellars als den »logischen Raum der Gründe« bezeichnen könnte (vgl. Sellars 1997). In Carabantes' Worten: »Observing the huge matrix of weights of a DNN and the activation levels of the units in real time, Ortega y Gasset would say, may give an esthetic impression, but it does not explain how the computer reaches its conclusions« (Carabantes 2020, 314). Das Mysterium durch Erklärung aufzulösen, wie Joseph Weizenbaum das noch bei EILZA erhofft hatte, ist also nicht mehr so einfach möglich. Ohne die Möglichkeit einer solchen vollständigen Erklärung ist es auch nicht mehr so einfach, mit Hofstadter zu behaupten, das Verständnis, das wir der KI zuschreiben sei »unwarranted«. Margaret Boden hatte eine ähnliche Kritik an Searles CRA angebracht: Wir hätten zwar sehr gute Gründe für die Annahme, *dass* Intentionalität, Verständnis und Bewusstsein auf Neuroprotein basierten, allerdings hätten wir noch so gut wie keine Ahnung davon, *wie* diese Beziehung aussieht, also *wie* Neuroprotein diese Phänomene qua seiner Eigenschaften hervorrufe. Daher könne man auch nicht so einfach darauf schließen, dass Neuroprotein der einzige »stuff« sei, der diese kausalen Kräfte habe (Boden 1990, 93). Wenn wir also bei einem DNN in etwa eine genauso gute Erklärung dafür haben, wie es seine Outputs hervorruft wie beim menschlichen Gehirn, dann haben wir zunächst keinen anderen Grund als einen Unterschied in diesen Outputs, um anzunehmen, dass im einen Fall »wirkliches« Verständnis, Bewusstsein oder Intentionalität dahinter stehe und im anderen nicht.

Ein *zweiter* Grund zum Zweifel an Weizenbaums Hoffnung der Auflösung des Mysteriums durch Erklärung betrifft das Verhältnis von Wissen und Wahrnehmung. In vielen Fällen ist es gerade nicht möglich, dass abstraktes Wissen darum, wie ein Sachverhalt »in Wirklichkeit ist«, unsere Wahrnehmung dieses Sachverhalts zu verändern vermag. Das klassische Beispiel ist die Müller-Lyer-Illusion. Auch wenn wir *wissen*, dass die Linien »in Wirklichkeit« gleich lang sind, *sehen* wir sie weiterhin als verschieden lang. Besteht der ELIZA-effect nicht mindestens für einige und eventuell für alle in gewisser Form fort, *trotz* unseres Wissens um die Maschinenhaftigkeit und des graduellen Wissens um die Funktionsweise

10 Ein Überbegriff für die dezentral operierenden, der Netzwerkarchitektur des Hirns nachempfundenen *second wave AI* Programme. DNNs sind ANNs mit »Tiefe«, d.h. zwischen Input-»Neuronen« und Output-»Neuronen« liegt mindestens eine weitere »Neuronen«-schicht. Buckner (2019) sieht alle Netze mit nicht mehr als vier Zwischenschichten noch als »shallow« an, und sieht wirkliche Tiefe erst bei fünf Zwischenschichten beginnen. Er betont aber auch, dass die Epoche der DNNs im Prinzip mit der Einführung des ersten »hidden layers« zwischen input- und output-Schicht in den 80ern beginnt (ebd. 3).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

dieser Maschine? Es ist sicherlich wahr, dass die Magie von ELIZA zu einem gewissen Grad schwindet, nachdem wir verstanden haben, wie sie funktioniert. Trotzdem können wir erstmal nicht anders, als die Worte der Maschine als bedeutungsvoll wahrzunehmen. Bedeutung, so mein auf Merleau-Ponty aufbauender Vorschlag, ist etwas, das wir *wahrnehmen* und zunächst und zumeist nicht reflexiv *denken*. Die Entscheidung in dieser Frage für die eine oder andere Seite setzt jeweils ein gewisses Verständnis von Sprache voraus. In Douglas Hofstadters Definition des ELIZA-effects schwingt ein implizites Verständnis von Sprache und Bedeutung mit. Hofstadter hatte den Effekt definiert als »the susceptibility of people *to read* far more *understanding* than is warranted *into* strings of symbols – especially words – strung together by computers« (Hofstadter 1995, 157; meine Hervorhebungen). Verständnis wird in die »eigentlich« bedeutungslosen Worte »hineingelesen«. Was für ein Verständnis von Sprache und Bedeutung ist hier in Geltung? Wie schaffen es diese »Stränge von Symbolen«, dass wir sie mit Bedeutung oder Verständnis aufladen, dass wir hinter ihnen Intentionalität, Geist, Bewusstsein vermuten?

Ist hier wichtig, was genau es ist, was dahinter vermutet wird? Hofstadter spricht zunächst von »understanding«, in seinem Beispiel des Geldautomaten geht es dann um den Ausdruck von Dankbarkeit, ein Zustand der eher Begriffe wie (Selbst-)Bewusstsein, einen Personenbegriff oder Anerkennung[11] voraussetzt. Was genau es also ist, das die Worte ausdrücken sollen beziehungsweise was fälschlicherweise in sie hineingelesen werden soll, ist unklar. Klar ist, dass das Wort in jedem Fall auf etwas anderes hinweisen soll, etwas anzeigen, etwas bedeuten soll. Wie Wörter das tun, ist eine in der Sprachphilosophie viel diskutierte Frage. Im folgenden Teil wende ich mich zwei Theorien zu, die Merleau-Ponty in seiner Untersuchung zu Phänomen und Ursprung der Sprache in der *Phänomenologie der Wahrnehmung (PdW)* in Betracht zieht, um zu erklären, wie Worte bedeuten können. Er nennt diese Positionen *Empirismus* und *Intellektualismus* (Merleau-Ponty 1966, 207ff.).

2. Intellektualismus und Empirismus als Sprachtheorien

Empirismus und Intellektualismus sind zwei konträre Positionen, die Merleau-Ponty in der *PdW* immer wieder gegeneinanderstellt, von denen er aber zu zeigen versucht, dass sie auf ähnlichen falschen Voraussetzungen beruhen. Ob diese Positionen über die verschiedenen Felder hinweg, in denen Merleau-Ponty sie anbringt, eine konsistente philosophische Position ergeben, die in dieser Form von jemandem vertreten wurde, ist

11 Zur Frage der Möglichkeit der Anerkennung einer KI durch einen Menschen vgl. den Beitrag von Martin Schnell in diesem Band.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

fraglich. Es sind eher Idealtypen des Denkens, die Merleau-Ponty hier zeichnet[12]. In Bezug auf die Sprache sieht dies wie folgt aus:

Der *Empirismus* sieht den Besitz von Sprache als den Besitz von »Wortbildern«, die von empirisch wahrgenommenen Worten in uns hinterlassene Spuren sind. Man kann dies naturalistisch oder mentalistisch interpretieren. Im ersteren Fall wären die Wortbilder dann neurologische Strukturen, die sich über die Zeit etabliert haben, im letzteren Fall gibt es »Bewußtseinszustände«, die »auf Grund erworbener Assoziationen die Erscheinung des passenden Wortbildes nach sich ziehen« (ebd. 208). In beiden Fällen aber wird die Sprache aus der Perspektive der dritten Person als abstrakte Struktur beschrieben. Die Sprechende hat in dieser Theorie keinen Platz. Wortflüsse unterscheiden sich hier nicht von Wasserflüssen oder elektrischem Strom, Sprache ist ein reines Strukturphänomen, »der Mensch kann demnach sprechen, wie eine elektrische Birne glühen kann« (ebd.). Das Problem ist nun, dass die Kohärenz dieser Position von Erkenntnissen der empirischen Psychopathologie in Frage gestellt wird, die zu einer Position der intellektualistischen Psychologie führen.

So behauptet die *intellektualistische* Position, dass die Sprache vom Denken bedingt ist. Begründet wird dies aus den Erkenntnissen in der Erforschung der Aphasie. Die Aphasie ist eine Sprachstörung als Folge einer Läsion in der dominanten Sphäre des Gehirns. Bei bestimmten Formen der Aphasie hat die Patientin Probleme, Worte zu finden, wenn abstrakt nach diesen gefragt wird, jedoch nicht, wenn sie als Ausdruck eines wirklichen Erlebens geäußert werden. So kann das Wort »Nein« beispielsweise problemlos ausgesprochen werden, wenn es »eine wirklich erlebte Verneinung bedeutet« (ebd.), also beispielsweise auf die Frage »Hast du Durst?«. Die Patientin findet das Wort aber nicht, wenn es »um eine bloße Sprachübung ohne affektives und vitales Interesse geht« (ebd.). Also beispielsweise eine Aufgabe wie »Nennen Sie einige Ausdrücke der Ablehnung«. Diese Erkenntnis führt für Merleau-Ponty in den »äußersten Gegensatz zur Wortbildtheorie […], da die Sprache nunmehr vom Denken bedingt erscheint« (ebd., 209), während im Empirismus das intentionale Denken der Person doch keine Rolle spielen durfte, Sprache sollte ja beschrieben werden wie jedes andere natürliche Phänomen. Die Einheit des Wortbildes wird zerstört. Soll ein Wortbild aus einer bestimmten Struktur im Hirn bestehen, so müsste bei Verletzung dieser Hirnstruktur das Wort gänzlich unverfügbar werden. Die intellektualistische Psychologie zeigt aber, dass es, abhängig vom Gedanken, der »dahinter« steht, noch verfügbar ist. »Was der Normale besitzt und der Kranke verloren hat, ist nicht ein bestimmter

12 Vgl. (Carman 2008) für eine Analyse der beiden Positionen im Werk Merleau-Pontys.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Wortvorrat, sondern eine bestimmte Weise, sich dieses ›Vorrats‹ zu bedienen« (ebd. 208).

Das Problem an dieser Position ist wiederum, dass hier nun der *Gedanke* die Bedeutung zu haben scheint und nicht das *Wort*. So ist nicht verständlich, wie uns Sprache überhaupt etwas Neues zu sagen vermag, wieso wir unsere eigenen Gedanken bisweilen nicht kennen, bevor wir sie nicht formuliert haben, wieso sie sich erst beim Sprechen allmählich verfertigen (Kleist). Das Denken als augenblickliches, zeitlich unausgedehntes kann sich nur angeeignet werden durch den Ausdruck seiner selbst in der Sprache.

Es gibt hier also einen Widerspruch, der nicht auflösbar erscheint, »das Problem der Sprache ist nicht durch den Übergang von der These zur Antithese zu lösen« (ebd., 209). Vielmehr ist es so, dass beide Positionen das Wesentliche an der Sprache verfehlen. Beide Positionen »bleiben hinter der einfachen Feststellung zurück, daß *das Wort einen Sinn hat*« (ebd., 210). Wenn wir diese Überzeugung des alltäglichen Sprachgebrauchs richtig verstehen, so können wir den Widerspruch auflösen, da wir sowohl Empirismus als auch Intellektualismus als auf falschen Voraussetzungen ruhend entlarven.

Beide Positionen werden dieser Feststellung nicht gerecht. In der empiristischen geschieht dies geradeheraus und ohne Ausreden. Sprache als reines mechanisches Strukturphänomen betrachtet hat keinen Platz für Bedeutung. Hier hat das sprechende Subjekt keinen Platz. In der intellektualistischen Position gibt es wohl ein bedeutungsgebendes Subjekt, allerdings ist dies das *denkende* und nicht das *sprechende* Subjekt. Das Wort ist hier nur die leere, äußere Hülle des bedeutungstragenden Gedankens.

Bevor wir zu Merleau-Pontys Lösung des Problems übergehen, will ich diese Positionen kurz auf Hofstadters Definition des *ELIZA-effects* rückbeziehen. Ich denke, dass Hofstadters Definition sowohl in empiristische als auch intellektualistische Richtung interpretiert werden könnte. Dies unterstreicht Merleau-Pontys These, dass die beiden Positionen in einem gemeinsamen Raum an Voraussetzungen stehen, in dem auch Hofstadter fest verwurzelt ist. Als eine Erweiterung der empiristischen Position könnte man sagen: Das, was uns beim Menschen berechtigt, von Verständnis hinter den Worten zu sprechen, und bei der Maschine nicht (»unwarranted«), ist der richtige Prozess der Assoziation, die richtige strukturelle Verknüpfung. Die Maschine hätte die falsche strukturelle Verknüpfung, entweder rein funktionalistisch falsch oder materialistisch falsch, da sie im falschen »stuff«, nämlich in Silizium und nicht in Neuroprotein realisiert ist. Das materialistische Argument würde also zurück zu Searles CRA führen.

Die funktionalistische Falschheit würde sich nur gelegentlich als falsch zeigen, beispielsweise wenn ELIZA unsinnig antwortet. Diese Fälle

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

würden aber als Beweis dafür gelten, dass die Struktur im Allgemeinen die Falsche ist. Wirkliches Verständnis sind wir nur bei einer vollständig richtigen Struktur berechtigt zuzuschreiben.

Intellektualistisch interpretiert könnte man Folgendes sagen: Wir sind nur dann berechtigt Verständnis zuzuschreiben, wenn Gedanken hinter den Worten stehen. Dies ist für den Intellektualismus im Falle der Maschine aber wohl per definitionem ausgeschlossen, da er von einem cartesischen Dualismus ausgehen muss. Die Maschine ist dann immer nur genau dies: bloße Maschine.[13]

Welche der beiden Interpretationen man nun vertritt, ist für Merleau-Pontys Kritik eigentlich nicht von Belang, denn er sieht beide auf gleichen, falschen Voraussetzungen ruhen. Beide wollen die Bedeutung des Wortes durch etwas anderes als das Wort erklären und verlieren sie dabei. Keine der Positionen kann zufriedenstellend erklären, wie Worte bedeuten. Somit sind sie auch beide ungeeignet aufzuklären, worin genau der *ELIZA-effect* besteht. Wir brauchen einen anderen Blick auf Sprache, wir müssen ein anderes Verständnis davon erlangen, was Worte sind, wenn wir verstehen wollen, wie diese es vermögen, uns zu verleiten, einer Maschine Verständnis zuzuschreiben. Wie gesagt wird die Antwort für Merleau-Ponty darin liegen, dass das Wort selbst Bedeutung[14] *hat*.

3. Das Haben von Sinn und das Wort als Geste

Merleau-Ponty will »Haben« hier in der Bedeutung verstanden wissen, wie wir es benutzen, wenn wir sagen »Ich habe Angst« oder »Ich habe eine Idee«. Damit soll angezeigt werden, dass »Haben«, im Gegensatz zu »Sein«, »den Bezug des Subjekts zu dem bezeichnet, woraufhin es sich entwirft« (Merleau-Ponty 1966, 207 Fußnote 1). In diesem Sinne will er verstanden wissen, dass der Mensch Sprache *hat* und auch dass die Sprache und das Wort einen Sinn *haben*. Das Wort vollbringt das Denken erst, es gibt nicht den im Denken schon fertigen Begriff, der dann nur noch durch die Worthülse artikuliert wird, sondern im echten, einen Sachverhalt das erste Mal formulierenden Sprechen wird der Gedanke erst vollbracht. Er ist nicht fertig, ja existiert nicht einmal wirklich, bevor er nicht ausgesprochen wurde. Genauso entnimmt der Hörende »den

13 Ausgehend von Hofstadters Aussagen in *Gödel, Escher, Bach*, in der er Bedeutung hervorgebracht sieht durch das Erkennen eines Isomorphismus, müsste man ihn wohl eher der empiristischen Seite zuschlagen (vgl. Hofstadter 1985) Hofstadters philosophische Position im weiteren Sinne ist aber nicht Thema dieses Texts.

14 Merleau-Ponty unterscheidet in diesen Abschnitten nicht streng zwischen »Sinn« und »Bedeutung«, etwa im Sinne Freges.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Gedanken dem Worte selbst« (ebd. 212). Es ist nicht so, dass der Hörende den Worten nachträglich, gleichsam durch einen »inneren« Akt seines Geistes, den Sinn gibt oder, in den Worten Hofstadters, etwas in die ansonsten bedeutungslosen Zeichen »hineinliest«. Wäre es so, könnte man durch gehörte Worte nie etwas Neues lernen und die »Erfahrung der Verständigung« wäre nichts als eine Illusion (ebd.). Es ist aber nun einmal eine Tatsache, so Merleau-Ponty, dass wir »etwas zu verstehen mögen, was über das, was wir von uns aus dachten, hinausgeht« (ebd.). Die Sprache müssen wir zwar schon kennen, und doch können uns die Worte dieser uns bekannten Sprache etwas Neues sagen, ihre Bedeutungen sich bisweilen »zu einem neuen Gedanken« verbinden. Diese Tatsache muss uns dazu bringen, dass »der Sinn der Worte letzten Endes durch die Worte selber hervorgebracht« sein muss. Merleau-Ponty spezifiziert sofort: »oder vielmehr genauer, deren begriffliche Bedeutung sich bilden auf Grund und aus ihrer *gestischen Bedeutung,* die ihrerseits der Sprache selbst immanent ist« (ebd., 212f.).

Wie ist das zu verstehen, was ist die »gestische Bedeutung«, die in der Sprache selbst liegt, und ihren Sinn hervorruft? Hier kann nun zurückgegriffen werden auf die Bedeutung, die Merleau-Ponty dem Wort »Haben« zugeschrieben hat. Der Mensch *hat* Sprache, insofern diese Sprache Teil dessen ist, woraufhin er sich entwirft, d.h. Teil seiner Existenz ist. Dadurch *hat* das Wort Bedeutung. Merleau-Ponty erläutert dies am Prozess des Erlernens einer Sprache. Eine fremde Sprache lerne ich, indem ich ihre Ausdrücke »im Zusammenhang der Tätigkeit der Menschen zu verstehen beginne, wenn ich an deren Gemeinschaftsleben teilnehme« (ebd., 213). Um ein Beispiel von W.V.O. Quine zu bemühen[15]: Wenn wir bei Vorbeilaufen eines Hasen einen *Aruga*-Sprecher »Gavagai« sagen hören, wissen wir erst ob damit »Schau, ein Hase« gemeint ist oder »Oh, ein schlechtes Omen«, »Schau, unser Abendessen« oder etwas ganz anderes, wenn wir den Zusammenhang der Tätigkeit der Aruga-Sprecher zu verstehen beginnen. Beispielsweise, ob es bei ihnen normal ist, das Auftauchen eines Tieres als ein Omen zu werten, ob Hasen typischerweise gegessen werden etc. Merleau-Ponty nennt das, den »Stil« einer Sprache zu verstehen, in dem Sinne, in dem man den Stil eines Malers zu erkennen lernt oder den Stil eines Musikstückes durchdringt. Stil ist insofern eine Organisationsform von Teilen in ein komplexes Ganzes. Der Stil eines Musikstückes bestimmt, wie aus den einzelnen Noten der »Sinn« des Stückes entsteht, ohne dass dieser Sinn in den einzeln wahrgenommenen Tönen

15 In (Quine 2013). Quines Projekt einer naturalisierten Epistemologie hat natürlich viele Konfliktpunkte mit Merleau-Pontys Philosophie. In gewissen Punkten der Kritik an einem naiven Empirismus gibt es aber sicherlich interessante Gemeinsamkeiten, die hier aber nicht näher verfolgt werden sollen.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

erkennbar wäre. So ist auch das Verhältnis des Sinns der Worte zur Sprache zu denken.

Es ist hier vielleicht hilfreich auf die enormen Parallelen zu der Sprachauffassung des späten Wittgenstein hinzuweisen. Das »Hinzunehmende, Gegebene«, also das, was in der Sprache letztlich ausgedrückt ist, »– könnte man sagen – seien Lebensformen«, so Wittgenstein in den *Philosophischen Untersuchungen* (2014, 572). Und ich lerne die Bedeutungen, die die Wörter einer Sprache *haben,* indem ich an der Lebensform teilnehme, die sie ausdrücken. »Und eine Sprache vorstellen, heißt, sich eine Lebensform vorstellen« (ebd. §19). »Eine Sprache verstehen, heißt eine Technik beherrschen« (ebd. §199). Merleau-Ponty spricht statt von Lebensform, der man gemeinsam zugehört, von einer Welt, die man gemeinsam *hat*: »In Gestalt der verfügbaren Bedeutungen […] besitzen die sprechenden Subjekte eine gemeinsame Welt […] Der Sinn des Wortes ist selbst in gar nichts anderem gegeben als der Art und Weise, in der es zu dieser Sprachwelt sich ins Verhältnis setzt« (Merleau-Ponty 1966, 221). »Um eine Sprache sich vollständig anzueignen, müßte man die Welt übernehmen, die in ihr Ausdruck findet« (ebd. 222). Hier kann man direkt Wittgensteins Äußerung anschließen, dass wir einen Löwen, könnte er sprechen, nicht verstünden (Wittgenstein 2014, 568). In der Löwenwelt sind wir nicht zu Hause und könnten dies wohl auch nur sehr bedingt je sein. Bisweilen sind die Äußerungen Merleau-Pontys und Wittgensteins fast verwechselbar: »Man kann für eine große Klasse von Fällen der Benützung des Wortes ›Bedeutung‹ – wenn auch nicht für alle Fälle seiner Benützung – dieses Wort so erklären: Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache« (Wittgenstein 2014, §42). »Was den Sinn des Wortes betrifft, so lerne ich ihn, wie ich den Gebrauch eines Werkzeugs lerne: indem ich es im Kontext einer bestimmten Situation gebrauchen sehe« (Merleau-Ponty 1966, 459).

Für Wittgenstein wie für Merleau-Ponty haben wir Worte also als Ausdrücke eines Verhaltens zur Welt. Merleau-Ponty nennt die gestische Bedeutung folglich auch »existentielle Bedeutung« (ebd., 216). Das Wort teilt den Gedanken nicht als Begriffsfunktion, sondern als »existentielle Gebärde« mit (ebd.). Diese Art der Bedeutung liegt aller begrifflichen Bedeutung zugrunde und ist gleichzeitig das, was wir in der Sprache primär wahrnehmen. Sie ist der Sprache selbst immanent. Wir müssen den Sinn nicht erst aktiv in die Worte hineinlesen, sondern wir nehmen ihn direkt wahr. Das wahrgenommene Wort hat Bedeutung, wie das Wahrnehmungsding Farbe hat. Wollte man darauf bestehen, dass hier doch ein Inhalt und eine Form, oder ein Zeichen und ein Bezeichnetes zusammengefügt werden, so müsste man mindestens mit einem Wort Husserls von *passiver Synthesis* sprechen. So spricht auch Merleau-Ponty davon, dass der geglückte Ausdruck der Bedeutung ein Dasein »gleich dem eines Dinges« beschert und die Bedeutung

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

wahrgenommen wird, als hätten wir durch den Erwerb der Sprache ein neues Sinnesorgan und ein neues Erfahrungsfeld gewonnen (ebd.). Aus dem ästhetischen Ausdruck kennen wir besser, was Merleau-Ponty sagen will, hier fällt es uns leichter, die Unablöslichkeit des Ausgedrückten vom Ausdruck einzusehen. Einzusehen, dass der Ausdruck das Wunder vollbringt, dem Ausgedrückten ein »An-sich-sein« zu verleihen, es als »ein jedermann zugängliches Wahrnehmungsding in die Natur zu« versetzen (ebd., 217). Für den Zusammenhang von Tönen einer Sonate und ihrer ›Bedeutung‹ oder von Elementen eines Bildes und dem, was es ›bedeutet‹, würden wir dies sofort zugestehen. Letzteres ist immer nur durch erstere da, ohne in diesen aufzugehen. In Wahrheit, so Merleau-Ponty, verhält es sich genauso mit dem Sprechen und dem Denken. Das Denken existiert nur im Sprechen, auch der vermeintlich stumme Gedanke ist nur die Erinnerung eines schon ausgedrückten Gedankens und das »Innenleben« des Denkens ist nur ein inneres Sprechen. »Reines Denken«, etwa als das Residuum des Ausgedrückten, das zurückbleibt, das nicht restlos ausgedrückt werden kann, ist laut Merleau-Ponty nichts als »eine gewisse Leere des Bewußtseins, ein augenblicklicher Wunsch« (ebd.).[16]

Daraus folgt, dass niemals zwei verschiedene Ausdrücke das Gleiche ausdrücken können. Das würde den Ausdruck auf Zeichen für unabhängig von ihrem Ausdruck existierende Gedanken reduzieren, die von verschiedenen äußeren Hüllen angezeigt werden können. So könnte also für Merleau-Ponty »Hund« und »dog« niemals das exakt Gleiche bedeuten, exakte Übersetzung ist unmöglich (vgl. ebd., 222). Ja sogar das von unterschiedlichen Sprecherinnen der gleichen Sprache benutzte Wort kann dann nicht bloß sinnlich unterscheidbares Zeichen eines identischen Gedankens sein. Verständnis und Übersetzung ist nicht möglich, weil verschiedene Zeichen gleiche Gedanken bezeichnen, sondern, weil es eine Entsprechung zwischen meinen sich in meinen Gebärden bekundenden Leibintentionen und den Gebärden der Anderen gibt. Da ich leiblich in der Welt bin, habe ich Worte als »eine mögliche Modulation, einen möglichen Gebrauch meines Leibes« (ebd. 214). So dass ich zum Wort greife, »wie meine Hand an eine plötzlich schmerzende Stelle meines Körpers fährt« (ebd.). Ich kann mit der Anderen kommunizieren und wir können uns verstehen, weil unsere Gebärden in der gleichen Welt stattfinden, wir qua

16 Auch hier springt eine Ähnlichkeit zu Wittgenstein ins Auge. Für diesen war die Frage nach dem »reinen Gedanken«, der von einem Wort bezeichnet werden soll, eine völlig falsche. Der Gedanke selbst spielt keine Rolle für eine Analyse der Bedeutung des Wortes. Die Schachtel, in der wir diese »reine« Bedeutung verstecken und in die nur wir je schauen können, könnte genauso gut leer sein (vgl. Wittgenstein 2014, §293), genauso leer wie das Bewusstsein, von dem Merleau-Ponty hier spricht.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

unserer Leiblichkeit eine gemeinsame Sinnlichkeit teilen. Unsere Kommunikation gründet erst den gemeinsamen Sinn unserer Erfahrungen, sie setzt ihn nicht voraus als gleichsam schon in einer transzendenten Sphäre vorhanden. Ich kann die leiblichen Gebärden der Anderen verstehen, weil auch ich ein Leib bin. Und das Wort ist sedimentierte Gebärde, nur deshalb kann ich es verstehen.

Eine Differenz im Ausdruck ist also auch eine Differenz im Ausgedrückten. So besagt beispielsweise der andere Ausdruck des Zorns in Japan einen Unterschied in der Emotion selbst (ebd., 223). Zorn anders auszudrücken heißt, Zorn anders zu fühlen, anders zornig *zu sein*. Die Leiblichkeit als gemeinsame Grundlage des Verstehens ist also nicht zu verstehen als bloße Gleichheit des organischen Aufbaus, sondern als kulturell geformtes Medium des Ausdrucks. »Die psychophysische Ausrüstung an sich eröffnet hier zahllose Möglichkeiten, und so wenig wie im Bereich der Instinkte gibt es auch hier eine Natur des Menschen, die ein für alle Mal feststünde. Der Gebrauch, den der Mensch von seinem Leibe macht, transzendiert den Körper als bloß biologisch Seiendes […] Seine Verhaltensweisen schaffen Bedeutungen, die seine anatomische Anlage transzendieren, gleichwohl aber dem Verhalten als solchem immanent bleiben« (ebd., 224). Mit Wittgenstein würde man sagen: Der Mensch partizipiert leiblich an Lebensformen.

4. Leiblose Kommunikation?

Was heißt dieses Verständnis der Sprache nun für die Kommunikation mit einer KI? Ist es nicht klar, dass der Körper eines Roboters kein Leib im Merleau-Ponty'schen Sinne ist und eine körperlose Chatbot-KI ja noch nicht einmal einen Körper hat, also unmöglich gestische Bedeutungen realisieren kann? Auch Merleau-Ponty selbst äußert sich in einem knappen Satz zu Sprachalgorithmen folgendermaßen: »Ein auf Konventionen beruhender Algorithmus – der übrigens auch nur im Rückbezug auf die Sprache Sinn hat – wird nie etwas anderes auszudrücken vermögen als eine Natur ohne Menschen« (ebd., 222f.). Ist also der *ELIZA-effect* auch aus einer Merleau-Ponty'schen Sprachphilosophie heraus nichts als ein Irrtum, ein Missverstehen dessen, was passiert, eine falsche Wahrnehmung, eine Illusion? Ich denke, so weit zu gehen wäre vorschnell. Merleau-Pontys eigene Äußerung über die Macht der Algorithmen muss vielmehr, angesichts des von ihm noch nicht beschreibbaren Phänomens einer auf dem heute möglichen Level konversierenden KI, neu überdacht werden. Es gilt, das Phänomen der Sprache auch in dieser Facette ernst zu nehmen und phänomenologisch zu beschreiben. Ich will mich dem in diesem Abschnitt

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

zunächst nähern, indem ich beschreibe, was beim Chatten zweier Menschen miteinander mit dem Phänomen der Sprache passiert.

Zunächst könnte man sagen, auch hier fehle doch der Körper. Jegliche über Distanz vollzogene, nur geschriebene Kommunikation sei bestenfalls parasitär gegenüber der im zwischenleiblichen Dialog sich realisierenden Bedeutung. Wie ist dieses parasitäre Verhältnis aber genauer zu verstehen? Warum haben Worte für uns auch Bedeutung, wenn der Leib, der sie ausdrückt, nicht mit-wahrgenommen wird? Mit anderen Worten: Wie funktioniert die Sedimentation der gestischen Bedeutung hier genau? Um diese Fragen zu beantworten, ist es nötig, ein Stück weit auf Merleau-Pontys Begriff der Anderen einzugehen, denn der Leib der Anderen ist es schließlich, der die gestische Bedeutung ausdrückt.

Die Andere als mit mir Kommunizierende hat für Merleau-Ponty eine besondere Örtlichkeit. Sie ist nicht in der Welt vorhanden wie ein Ding in einem Container, ich nehme die Andere in der Welt nicht einfach wahr als vor mir stehend. Es wäre eher passend zu sagen, sie ist »*zwischen* mir, der denkt, und jenem Leib [der Anderen] oder eher neben mir, an meiner Seite, taucht [sie] auf wie eine Nachbildung meiner selbst«(Merleau-Ponty 1993, 149)[17]. Die Andere hat keinen Platz in der Welt, außer in meinem »*Wahrnehmungsfeld,* doch zumindest dieser Ort ist parat für [sie], seit meine Wahrnehmung begonnen hat« (ebd., 152). Wie ist das zu verstehen?

Vom Wahrnehmungsfeld spricht Merleau-Ponty auch schon in der *Phänomenologie der Wahrnehmung*, wir können uns dem Begriff nähern über den des *Gesichtsfeldes*. Das *Gesichtsfeld* sollte nicht verwechselt werden mit all dem, was ich in einer bestimmten Situation sehe. Es ist nicht einfach zusammengesetzt aus den meine Netzhaut berührenden Reizen, es besteht nicht aus den Sinnesdaten, die meine Augen treffen, es ist auch nicht ein scharf abgegrenztes wie gerahmtes Bild, ein »scharf umgrenztes Weltsegment […] außen umgeben von einer dunklen Zone, innen lückenlos erfüllt von Qualitäten, getragen von Größenverhältnissen gleicher Bestimmtheit wie auf der Netzhaut« (Merleau-Ponty 1966, 24). Die Grenzen des Gesichtsfeldes, die es umgebende Region »ist nicht leicht zu beschreiben, doch ist sie sicher weder schwarz noch grau. Sie steht in einer *unbestimmten Sicht,* der Sicht eines *Je ne sais quoi*, und am Ende ist sogar das in meinem Rücken Gelegene nicht gänzlich ohne visuelle Gegenwart« (ebd.)

Ebenso wenig ist es nur mit *visuellen* Daten gefüllt. Das Gesichtsfeld ist ein Feld unter anderen im Wahrnehmungsfeld, die Felder der anderen

17 Hier beziehe ich mich auf einen Text aus dem Spätwerk Merleau-Pontys, das oft als in entscheidender Weise von der *PdW* abweichend interpretiert wird. Zumindest in der Frage nach dem Ort des Anderen gibt es aber mehr Kontinuitäten als Brüche, wie im Folgenden deutlich werden sollte.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Sinne überlappen mit ihm. So zählt das im nächsten Zimmer spielende Grammophon, das ich nicht sehe, noch mit in mein Gesichtsfeld (ebd., 323). Ich höre das Grammophon, und dies zeigt mir mit an, wie es aussieht. Das Hörfeld überlappt mit dem Sehfeld. Mein gesamtes Wahrnehmungsfeld ist dann konstituiert als die Gesamtheit dessen, zu dem ich eine Weise des Verhaltens innehabe. Ich höre Musik im nächsten Raum und da ich weiß, dass dort ein Grammophon steht, *höre ich das Grammophon*. Ich höre keine Musik, die dann auf das Grammophon verweist, sondern ich höre das Grammophon selbst, so wie ich das Grammophon selbst sehen könnte, würde ich in den anderen Raum gehen. Es wird Teil meines Wahrnehmungsfeldes *als Ding*. Mit seinem Klang wird mit-präsentiert, sein Aussehen, sein Geruch, wie es sich anfühlt, wie es angeschaltet wird, wie schwer es ist etc. Alles was ich durch die Sinne wahrnehme, ist mir als sinnvoll gegeben, das heißt mit einer Weise des Verhaltens ihm gegenüber.

Extrapolieren wir diese Beschreibung auf die Situation des Chattens mit einem anderen Menschen am Computer. Was ich sehe, ist der Text der Anderen, der erscheint. Visuell gegeben sind nur Worte. So wie ich durch die Musik des Grammophons das Grammophon selbst höre, es in meinem Wahrnehmungsfeld erscheint, erscheint die Andere durch ihre Worte sofort in meinem Wahrnehmungsfeld; nicht jedoch an einem bestimmten Ort, im Sinne eines in Koordinaten angebbaren Teils meines Wahrnehmungsfeldes. Wir können den Körper der Anderen zwar als einen solchen Ort besetzend beschreiben, damit greifen wir aber nicht den Ort, an dem die Andere selbst ist. Die Andere selbst ist »nirgendwo im Sein, von hinten gleitet [sie] in meine Wahrnehmung« (Merleau-Ponty 1993, 152).

Was will Merleau-Ponty mit dieser Rede von der Nicht-Örtlichkeit der Anderen, von ihrem Leben »hinter«, »zwischen«, »neben« mir oder in den »Fugen zwischen der Welt und uns selbst« (ebd., 153) aussagen? Eine weitere phänomenologische Situationsbeschreibung kann das aufklären. Stellen Sie sich folgende Situation vor: Sie kommen nach Hause, in Ihre Wohnung, die Sie sich teilen mit Ihrer Familie oder anderen Mitbewohnerinnen. Sie nehmen aber an, dass Sie in der Wohnung allein sind, da normalerweise zu dieser Zeit alle anderen arbeiten oder anderweitig außer Haus beschäftigt sind. Die ganze Wohnung wird von Beginn an gefärbt durch diese Annahme des Alleinseins erscheinen, *als* Wohnung, in der Sie alleine sind, in der momentan niemand anders *ist*. Plötzlich hören Sie ein Geräusch aus einem der Nebenräume, ein Räuspern oder vielleicht sogar die Stimme einer Mitbewohnerin. Die ganze Situation, die ganze Wohnung, Ihre gesamte Umwelt wird ihren Charakter unmittelbar ändern. Das heißt, Ihr Wahrnehmungsfeld verändert seinen Sinn, seine Orientierung, da dort jemand Anders *ist*, d.h. die Andere gleitet, sobald sie wahrgenommen wird, unmittelbar »hinter« Sie oder »in die Fugen« zwischen Ihnen und der Welt.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Diese Situation ist natürlich nicht in jeder Hinsicht vergleichbar mit der Situation des Chattens vor dem PC. In der Wohnungssituation ist die Präsens der Anderen sehr viel radikaler, die Andere ist sehr viel weiter »im Zentrum« Ihres Wahrnehmungsfeldes als in der PC-Situation. Dennoch ist die Andere auch in der PC-Situation Teil Ihres Wahrnehmungsfeldes. Technologie kann die Grenzen unseres Wahrnehmungsfelds sehr weit dehnen, und sobald die Andere im Feld ist, ist sie in den Fugen. Daher ist echter Dialog in dem Sinne, wie Merleau-Ponty ihn in der *Phänomenologie der Wahrnehmung* beschreibt, in dem »die Einwände meines Gesprächspartners« mir sogar »Gedanken entreißen, von denen ich nicht wußte, daß ich sie hatte, so daß also der Andere ebensosehr mir zu denken gibt wie ich ihm Gedanken zuschreibe« (Merleau-Ponty 1966, 406), über Telefon, Zoom oder eventuell sogar Chaträume zwar schwierig, aber nicht unmöglich.

Die Leiblichkeit der Anderen spielt hier allerdings nach wie vor eine entscheidende Rolle. Wenn ich meine Mitbewohnerin im anderen Raum höre, ist sie Teil meines Wahrnehmungsfelds, da ich ihren Leib höre, ich kann mir vorstellen, wie ihr Leib im anderen Raum situiert ist. Genauso kann ich mir eine Andere hinter den Worten im Chatraum vorstellen, da ich weiß, was es heißt, vor einem PC zu sitzen und zu chatten, ich bin mit der Situation vertraut. Für eine Person, die mit dem technologischen Artefakt des PCs völlig unvertraut ist – man denke etwa an Mitglieder indigener, völlig isoliert von moderner Technologie lebender Gruppen – wäre das Phänomen des Auftauchens der Worte auf dem Bildschirm völlig unverständlich, sie könnte keinen Urheber der Worte mit Ihnen verbinden, die Technologie wäre, um Arthur C. Clarke zu paraphrasieren »ununterscheidbar von Magie«. Das fehlende Wissen ist hier kein Wissen um die Funktionsweise der Technologie – die meisten Menschen, die PCs benutzen, haben kaum eine Ahnung davon, wie diese funktionieren –, sondern ein praktisches Wissen des Umgangs mit der Technologie, ein Wissen des Verhaltens ihr gegenüber.

Dieses praktische Wissen des Umgangs mit Technologie lässt diese in die Selbstverständlichkeitsstruktur unserer Lebenswelt einfließen.[18] Hans Blumenberg hat diese Erweiterung der Struktur unserer Lebenswelt durch Technik schon sehr früh philosophisch reflektiert. Der Prozess der Technisierung, den Husserl noch rein als Vergessen der lebensweltlichen Ursprünge unseres technisierten Denkens gefasst hatte, erreiche sein Telos erst dadurch, so Blumenberg, dass die Technisierung ihrerseits beginne, »die Lebenswelt zu regulieren, indem jene Sphäre, in der wir *noch* keine Fragen stellen, identisch wird mit derjenigen, in der wir keine

18 Merleau-Ponty will seine Welt der Wahrnehmung explizit als Erweiterung des Husserl'schen Lebensweltbegriffs verstanden wissen (vgl. Merleau-Ponty 1966 Vorwort).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Fragen *mehr* stellen« (Blumenberg 2015, 190). Die intersubjektive Struktur der Lebenswelt wird angereichert mit technischen Produkten, und das vermeintlich unmöglich Herzustellende, nämlich Selbstverständlichkeit, wird produzierbar. So wird uns beispielsweise die Wahrnehmung von Bedeutung in der Wahrnehmung von Worten auf einem PC fraglos, sie erscheint selbstverständlich, obwohl hier kein Leib zu sehen ist, der diese Worte hervorbringt. Das praktische Wissen des Umgangs mit Technologie ist also im Erfolgsfall der Technologie ein implizites, leibliches Wissen, das nicht reflektiert ist und gerade deshalb selbstverständlich ist. Die Selbstverständlichkeit der Lebenswelt ist ja gerade »der Gegenbegriff zu jener ›Selbstverständigung‹, die für Husserl die eigentliche Aufgabe einer phänomenologischen Philosophie zu sein hat« (ebd. 178).

Mit Merleau-Ponty kann man feststellen, dass die Möglichkeit dieser technischen Veränderung unserer Selbstverständlichkeitssphäre in den Strukturen unserer Leiblichkeit angelegt ist. So hat Merleau-Ponty schon die Projektion des eigenen Leibs in eine Situation hinein, in der er rein physisch nicht vorhanden ist, mit dem Begriff des »virtuellen Leibs« in den Analysen zur Raumwahrnehmung ausführlich beschrieben. In den Spiegelexperimenten Wertheimers nimmt das Subjekt den im schiefen Spiegel schief dargestellten Raum zunächst schief wahr. Nun »geschieht nach einigen Minuten [aber] das Wunder, daß das reflektierte Zimmer ein Subjekt hervorruft, das in ihm zu leben vermag«. Hier verdrängt der virtuelle Leib

> den wirklichen Leib so weitgehend, daß das Subjekt sich nicht mehr in der Welt fühlt, in der es sich tatsächlich befindet, und statt seiner wirklichen Arme und Beine solche Arme und Beine empfindet, wie es sie haben müßte, um in dem reflektierten Zimmer gehen und tätig sein zu können; es bewohnt das Schauspiel. In diesem Moment gerät das Raumniveau ins Schwanken und etabliert sich in neuer Lage. (Merleau-Ponty 1966, 292)

Das heißt, mein Leib richtet sich im schiefen Raum qua virtuellem Leib ein und dadurch wird dieser geradegerückt. Ein neues Raumniveau entsteht, mit anderen Worten: eine neue Selbstverständlichkeitssphäre.

Dieser Begriff des virtuellen Leibes kann auf die Orientierung des Wahrnehmungsfeldes im Allgemeinen übertragen werden. Auch sinnlich nicht gegebene Dinge, Orte, Personen können Teil meines Wahrnehmungsfeldes sein, sie sind Teil meiner Sinngebung, da ich leiblich bei ihnen sein *könnte*, sie »im Felde meines virtuellen Tuns gelegen« sind (ebd. 502). In der Situation des Chattens vor dem PC gibt mir die Wahrnehmung der Worte die Möglichkeit des Imaginierens der Situation der anderen Person vor dem PC und somit die Möglichkeit, qua Imagination mein Wahrnehmungsfeld vom virtuellen Leib, als Leib vor dem anderen PC, durchstimmen zu lassen und dadurch das Problem des nicht sinnlich gegebenen Leibes der Anderen zu lösen, der doch eigentlich Bedingung der Möglichkeit von gestischer Bedeutung ist.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Wie sieht dies nun im Fall der Konversation mit einer KI aus? Auch hier ist kein Leib sinnlich gegeben. Wissen wir, dass es sich um eine KI handelt, beispielsweise ELIZA oder GPT-3, so wissen wir aber auch, dass hier kein Leib an irgendeinem Ort vor einem PC sitzt und die Worte leiblich eintippt. Wir stehen also vor einem Problem: Die Worte haben Sinn, sie werden als bedeutsam wahrgenommen, da dies in ihrer Phänomenalität selbst liegt, da in ihnen gestische Bedeutung sedimentiert ist, wir können sie aber nicht einmal virtuell auf einen Leib zurückführen.

Die Wahrnehmung bedeutsamer Worte, d.h. Worte als Gebärden, ohne einen Leib, der sie produziert, der sich gebärdet, muss eigentlich in das von Husserl beschriebene Phänomen des *Widerstreits* führen. Wir nehmen bedeutungsvolle Worte wahr, jedoch keinen Leib, der sie produziert. Der Widerstreit ist dieses Phänomen der Wahrnehmung von zwei miteinander konkurrierenden Intentionen oder Sinngehalten. Husserls Lieblingsbeispiel hierfür ist zuerst dokumentiert in den *Vorlesungen zum Bildbewußtsein*:

> die schon öfters erwähnten Täuschungen a la Panoptikum, Panorama etc. Hier mag es zunächst sein, dass wir die Puppe als Menschen sehen. Wir haben da eine, wenn auch nachträglich als Irrtum sich herausstellende normale Wahrnehmung. Werden wir uns plötzlich der Täuschung bewusst, dann tritt das Bildlichkeitsbewusstsein ein. Aber in diesen Fällen will es sich nicht auf die Dauer durchsetzen. Die Wachsfigur gleicht mit ihren wirklichen Kleidern, Haaren usw., ja selbst in den durch mechanische Vorrichtung künstlich nachgeahmten Bewegungen so sehr dem natürlichen Menschen, dass sich momentan immer wieder das Wahrnehmungsbewusstsein durchsetzt. Die imaginative Auffassung fällt weg. Wir »wissen« zwar, dass es Schein sei, aber wir können uns nicht helfen, wir sehen einen Menschen (Husserl 1980, 23:40)

In den späteren *Analysen zur passiven Synthesis* greift Husserl nun dieses Beispiel auf, um das Phänomen des Zweifels und des Widerstreits zu erläutern:

> Das zunächst als Mensch Gesehene wird zweifelhaft und schließlich stellt es sich als eine bloße Wachspuppe heraus. Oder aber umgekehrt, der Zweifel löst sich in der bejahenden Form: ja, es ist doch ein Mensch. Während des Zweifels, ob wirklicher Mensch oder Puppe, überschieben sich offenbar zwei Wahrnehmungsauffassungen. […] Keine von beiden ist während des Zweifels durchgestrichen, sie stehen hier in wechselseitigem Streit, jede hat gewissermaßen ihre Kraft, ist durch die bisherige Wahrnehmungslage und ihren intentionalen Gehalt motiviert, gleichsam gefordert. Aber Forderung steht gegen Forderung, eins bestreitet das andere und erfährt von ihm den gleichen Tort. Es bleibt im Zweifel ein unentschiedener Streit. (Husserl 1966, 11:33f.).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Das, was im Modus des Zweifelns unentschieden bleibt, kann aber irgendwann in Gewissheit überführt werden. Die Möglichkeit der Entscheidung gehöre zum Wesen des Zweifels (ebd. 36). Der Zweifel wird entschieden durch eine unzweifelhafte Wahrnehmung, eine Urimpression, im Falle der Puppe beispielsweise ein Anfassen: »Die Erfüllung durch Urimpression ist die Kraft, die alles niederrennt. Wir treten näher heran, wir fassen auch tastend zu, und die eben noch zweifelhafte Intention auf Wachs erhält den Gewißheitsvorzug.« (ebd. 37).

Dieser Begriff der Urimpression ist selbst kein unproblematischer, evoziert er beispielsweise einen gewisse Präsenzmetaphysik mit ausdehnungslosem Jetztpunkt, die zeitphänomenologisch problematisch bleibt. Diese Problematik muss hier aber nicht weiter verfolgt werden, da in unserem Beispiel eine Urimpression ohnehin unmöglich zu haben ist, selbst wenn wir sie als Möglichkeit prinzipiell zugestehen. Eine Urimpression, wenn überhaupt möglich, muss immer eine Wahrnehmung sein, sie kann kein abstrakter Glauben oder theoretisches Wissen sein. Jedes mit der Wahrnehmung widerstreitende begriffliche Urteil oder Wissen muss in der Wahrnehmung durch Bildlichkeitsbewusstsein wirksam werden, um diese in ihrem Sinngehalt verändern zu können. Durch unser Wissen soll die Puppe als bloßes Bild eines Menschen gesehen werden, nicht als Mensch selbst. Aber »Bildlichkeitsbewusstsein«, wie Husserl schreibt, reicht nicht aus, es vermag das Wahrnehmungsbewusstsein nicht dauerhaft zu verdrängen.

Im Falle des Chattens mit einem Menschen ist die Situation nun ein wenig anders gelagert. Was hier mit unserer Wahrnehmung (der sinnvollen Worte) widerstreitet, ist das *Fehlen* einer anderen Wahrnehmung (des Leibes). Es liegt nahe, zu vermuten, dass dieser Widerstreit zwischen positivem und negativem Phänomen weniger stark ist als der Widerstreit zwischen zwei positiven Wahrnehmungsauffassungen. So kann er schon durch die Imagination unserer selbst in die Situation des Anderen qua Möglichkeit des virtuellen Leibs aufgelöst werden. Diese Imagination beruht auf einer Wahrnehmung, auf einer Urimpression des Anderen. Wir haben schon einmal einen Menschen an einem PC chatten gesehen, wir tun es sogar selbst. So haben wir in der Chatsituation selbst zwar nicht die Möglichkeit einer Urimpression, wir haben aber zumindest ein Wissen, das sich auf eine Wahrnehmung stützt. Dies kann den Widerstreit nie in apodiktische Gewissheit überführen – auf dieser Unmöglichkeit beruht der Turing-Test genau wie jeglicher Betrug über Internetkommunikation –, ihn jedoch so weit auflösen, dass wir den Worten im Chat problemlos Bedeutung zumessen können.

Wie ist es aber, wenn wir wissen, dass wir mit einer Maschine chatten? Im Falle des Wissens um die körperlose Maschinenhaftigkeit des Chatpartners haben wir zunächst ebenso den Widerstreit zwischen einem

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

positiven und einem negativen, d.h. fehlenden Phänomen. Das fehlende Phänomen kann hier aber nicht durch eine auf Wahrnehmung zurückgreifende Imagination substituiert werden. Wir haben nie einen Roboter vor einem PC chatten sehen, wir wissen auch, dass dies ohnehin nicht der Prozess ist, der die Worte auf dem Bildschirm hervorbringt. Einerseits nehmen wir Worte wahr, und das heißt, wir nehmen Bedeutung wahr. Andererseits fehlt der Leib, der diesen Worten gestische Bedeutung garantiert und wir wissen, dass dieser Leib auch nirgendwo ist. Auf Erinnerung einer Wahrnehmung zurückgreifende Imagination hilft uns also nicht weiter. Wir haben hier nur die der Wahrnehmung der bedeutungsvollen Worte widerstreitende, fehlende Wahrnehmung des Leibes und das Wissen darum, dass auch nirgendwo ein Leib ist.

Steht es hier dann nicht 2:1? Hat die Seite, die den Worten Bedeutung abspricht, nicht die Überhand? Muss das Wissen um die Nicht-Leiblichkeit der Maschine dann nicht den Widerstreit entscheiden, unsere Wahrnehmung der Bedeutsamkeit überlagert werden und die Worte letztlich als bedeutungslos erscheinen?

Um dies zu beantworten, müssen wir genauer darauf schauen, worin das Wissen, dass hier wirksam sein soll, besteht. Wir wissen, dass die KI eine Maschine ist, dass sie also keinen Leib hat. Was aber heißt das genau? Was heißt es für uns zu wissen, dass etwas eine »bloße Maschine« ist, »algorithmisch«, »rein mathematisch« funktioniert? Folgen wir Merleau-Pontys Sprachphilosophie, haben alle diese Begriffe, »Maschine«, »Algorithmus«, »Computer«, »mathematisch«, wenn sie denn überhaupt einen Sinn haben, einen gestischen Sinn. Wir wissen, was es heißt, etwas maschinell zu tun, was es heißt, einem Algorithmus zu folgen, was es heißt, etwas zu berechnen (»to compute«), stets aus unserer eigenen leiblichen Erfahrung. Beim Wort »Computer« zeigt die Begriffsgeschichte dies deutlich an. Es wurde zuerst verwendet, um Menschen – anfangs meist Frauen – zu bezeichnen, die Berechnungen durchführten.[19] »Maschinell« oder »algorithmisch« sind also nicht einfach Gegenteile von »leiblich« oder »bedeutungsvoll«. Etwas maschinell zu tun, ist nicht gleichbedeutend damit, etwas bedeutungslos zu tun. Insofern müssen wir Merleau-Pontys Aussage mit Argumenten aus seiner eigenen Theorie widersprechen. Es stimmt nicht, dass ein Algorithmus nie etwas anderes auszudrücken vermag als eine Natur ohne Menschen (s.o.). Dies wäre nur wahr, wenn sein Funktionieren – und das heißt in unserem Beispiel: sein Erfolg in der dialogischen Begegnung – wirklich auf einer völlig reinen Konvention beruhen würde, d.h. auf einer vollkommen arbiträren Vereinbarung, die beliebig anders sein könnte und in der die Form

19 Vgl. *Oxford English Dictionary* »Computer«. Auch hier lässt sich wieder die Parallelisierung von maschineller und femininer Alterität feststellen. Siehe den Beitrag von Oliver Müller.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

gänzlich unabhängig vom Inhalt ist. Dies ist aber nicht der Fall. Unsere Sprachkonventionen gibt es nur als Vergessen des dunklen Untergrundes einer existentiellen Gestik und sobald Worte benutzt werden, ist dieser dunkle Untergrund mit da. Was in unserem Wissen um die Maschinenhaftigkeit ausgedrückt ist, ist eine gestische Bedeutung, die letztlich auf einer menschlichen Aktivität beruht. Der Algorithmus vermag also keineswegs nur eine Natur ohne Menschen auszudrücken. Im Gegenteil, er vermag eigentlich stets nur menschliche Aktivitäten auszudrücken.

Das heißt, das Wissen um »Maschinenhaftigkeit« kann den Widerstreit nicht *prinzipiell* entscheiden. Es ist keine solche »Kraft die alles niederrennt«, da es selbst in seiner Bedeutung durchsetzt ist von seinem vermeintlichen Gegenteil. So wird verhindert, dass der Widerstreit notwendigerweise in die Richtung der Bedeutungslosigkeit aufgelöst wird und der *ELIZA-effect*, den wir nun genauer als Widerstreit gefasst haben, kann bestehen bleiben. Sprache *hat* zunächst immer Bedeutung, wir *nehmen sie wahr* als bedeutungsvoll. Inwieweit ein Wissen um die Maschinenhaftigkeit und Leiblosigkeit des Gesprächspartners diese Wahrnehmung zu verändern vermag, ist eine von Fall zu Fall zu entscheidende, empirische Frage. Dass sie nicht schon auf prinzipieller Ebene als entschieden gelten muss, habe ich zu zeigen versucht.

5. Chinesische Zimmer und sprechende Maschinen

Wie stehen die vorangegangenen Überlegungen zu Searles CRA? Searle hat in Auseinandersetzung mit dem »systems reply«[20] gesagt, dass es eine Ebene der Erklärung gebe, auf der jedes System als Computer angesehen werden könne. So beispielsweise auch sein Magen, der beschrieben werden könnte als Instanziierung eines Computerprogramms mit gewissen Inputs und Outputs. Er argumentiert dann weiter, dass es keine angemessene Antwort auf dieses Problem sei, zu sagen, dass das chinesische Zimmer *Informationen* als Input und Output hätte und der Magen *Essen* und prozessiertes Essen, da vom Gesichtspunkt des Agenten aus, also vom Gesichtspunkt der Person im Zimmer, weder das eine noch das andere Informationsgehalt hätte. Die hier vorgelegten Überlegungen implizieren nun erstens, dass die entscheidende Perspektive immer die des Menschen sein muss, der mit dem Zimmer (d.h. der KI) interagiert; und zweitens, dass es für diesen sehr wohl darauf ankommt, woraus

20 Ich setze aus Platzgründen den allgemeinen Aufbau von Searles Gedankenexperiment hier als bekannt voraus. Als »systems reply« bezeichnet Searle die Entgegnung auf sein Gedankenexperiment, dass zwar nicht der Mensch alleine im Zimmer Chinesisch *verstehen* würde, aber das System aus Mensch, Skript, Datenbank etc. sehr wohl Chinesisch *verstehen* würde.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

der Output besteht. Sprache als Output – so die oben schon formulierte Einsicht – hat eine solche Phänomenalität, dass wir ihr notwendig Bedeutung zuschreiben, beziehungsweise genauer: in der Wahrnehmung von Sprache nehmen wir zwangsläufig Bedeutung wahr. Diese *Wahrnehmung* von Bedeutung steht zunächst diesseits von Richtigkeit oder Falschheit eines Glaubens oder Wissens. Wenn eine KI ausreichend gut sprechen kann, entsteht notwendigerweise ein Widerstreit in der Wahrnehmung, d.h. die *Frage* nach Bedeutung wird notwendig aufkommen. Es wird mindestens zweifelhaft, ob die Worte Bedeutung haben oder nicht, und dieser Zweifel ist nicht prinzipiell durch theoretisches Wissen ausräumbar, etwa in Form eines Gedankenexperiments wie das Searles. Mit »ausreichend gut« meine ich, dass die KI gerade nicht ununterscheidbar von einem anderen Menschen sein muss, sondern nur so gut sprechen können muss, dass sie meistens sinnvolle und der Situation angemessene Sätze hervorbringt. Der Grad der Angemessenheit lässt den Widerstreit in die eine oder andere Richtung schwanken, ohne ihn je vollständig entscheiden zu können.

Dies hat einen Bezug zu einem weiteren »reply«, dem Searle nur kurz Aufmerksamkeit widmet: dem sogenannten »other minds reply«. Dieser sieht folgendermaßen aus: Da ich auch bei anderen Menschen nur ihr Verhalten als Anhaltspunkt habe, um ihnen Verständnis zuzuschreiben, sollte ich aufgrund des gleichen Verhaltens auch dem Roboter Verständnis zuschreiben. Searle antwortet, dass es ihm nicht um die Frage gehe, woher wir wüssten, dass andere Menschen kognitive Zustände haben, sondern darum, was es genau sei, das wir Menschen zuschreiben, wenn wir ihnen kognitive Zustände zuschreiben. Dies könnten nicht nur computationale Zustände plus Output sein, da diese vorhanden sein könnten, ohne dass wirkliches Verständnis vorhanden sei, wie sein Gedankenexperiment ja gerade zeige. Es sei hier keine gangbare Alternative, den Unwissenden zu spielen, da in den Kognitionswissenschaften die Realität und Wissbarkeit des Mentalen vorauszusetzen sei wie in der Physik die Wissbarkeit und Realität physischer Objekte.

Searle argumentiert aus einer ähnlichen Position heraus, wie wir sie oben Hofstadter zugeschrieben haben. Was der Sprache Bedeutung gibt, ist der kognitive Zustand, der dahinterstehe oder eben nicht. Es zeigt sich hier ein gewisser Cartesianismus Searles. Dies ist insofern überraschend, als er den Vertretern der »strong AI«, die er mit seinem CRA angreift, gerade eine Art des cartesianischen Dualismus zum Vorwurf macht (vgl. Searle 1980, 424). »Strong AI« setze eine Ablösbarkeit des Geistes vom Gehirn voraus, wenn sie davon ausgeht, dass der Geist unabhängig von seiner physischen Instanziierung im Hirn oder anderswo modellierbar sei. Das Spezifische des Mentalen hätte dann nichts mit den spezifischen Eigenschaften des Hirns zu tun.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Searle lehnt dies ab, da er gerade auf die kausalen Kräfte des Gehirns als notwendig für Mentales überhaupt und somit auch für Verständnis und Bedeutung besteht. Seine Antwort auf den »other minds reply« zeigt aber, dass er in seiner Ablehnung des Cartesianismus nicht sehr weit geht. Ontologisch will er kein cartesianischer Dualist sein, epistemologisch bleibt er es aber. Konzeptuell gibt es für ihn offensichtlich sehr wohl einen klar abtrennbaren Bereich des Mentalen, der in den Kognitionswissenschaften Objekt der Erkenntnis ist. Wie jeder Dualist kommt Searle damit auf ontologischer Ebene in das Zirbeldrüsen-Problem, d.h. das Problem, zu erklären, wie das Physische mit dem Mentalen interagiert.[21] Searle konstatiert, dass er im chinesischen Zimmer nirgendwo Bedeutung findet. Wo sollte sie aber auch sein? Ist dies nicht einfach der falsche »Ort«, um danach zu suchen? Auch durch das Beobachten des menschlichen Gehirns wird man niemals Bedeutung finden. In Merleau-Pontys Worten: »Auf immer wird es unverständlich bleiben, wie Bedeutung und Intentionalität Molekulargebäude oder Zellhaufen zu bewohnen vermögen; darin behält der Cartesianismus recht« (Merleau-Ponty 1966, 402). Bedeutung und Intentionalität seien stattdessen zu suchen in dem Verhalten, das »auf sichtbaren Leibern sich abzeichnet, zur Erscheinung gelangt, ohne in ihnen reell enthalten zu sein« (ebd.). Wie dies für das Verhalten körperloser KIs möglich ist, habe ich oben dargelegt.

Searle geht also in seinem Cartesianismus einerseits zu weit, andererseits nicht weit genug. Er behält die problematischen Aspekte bei, wenn er von der Möglichkeit eines modernen Äquivalents zur Zirbeldrüse ausgeht, und lässt die wertvolle Erkenntnis, die man aus dem Cartesianismus ziehen könnte, nämlich die Unvereinbarkeit von Materie und Geist, unbeachtet. Merleau-Ponty hatte gesehen, dass man stets in dieser Sackgasse landen muss, wenn man die Probleme von Anfang an dualistisch aufstellt. Seine Phänomenologie des Leibes ist der Versuch, einen dritten Weg zu beleuchten. Die hier vorgelegten Überlegungen sind mein Beitrag dazu, einen solchen dritten Weg in der Philosophie der KI anzuzeigen.

Literatur

Apostopoulos, Dimitris (2019): *Merleau-Ponty's Phenomenology of Language*. Lanham: Rowman and Littlefield.

Blumenberg, Hans (2015): »Lebenswelt und Technisierung unter Aspekten der Phänomenologie«. In *Schriften zur Technik*, 163–202.

Boden, Margaret A. (1990): »Escaping from the Chinese Room«. In *The Philosophy of Artificial Intelligence*. Oxford Readings in Philosophy.

21 Vgl. die oben erwähnte Kritik Margaret Bodens (in Boden 1990).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, u. a. (2020): »Language Models Are Few-Shot Learners«. *ArXiv:2005.14165 [Cs]*, Juli. http://arxiv.org/abs/2005.14165.

Buckner, Cameron (2019): »Deep Learning: A Philosophical Introduction«. *Philosophy Compass* 14 (10). https://doi.org/10.1111/phc3.12625.

Carabantes, Manuel (2020): »Black-Box Artificial Intelligence: An Epistemological and Critical Analysis«. *AI & SOCIETY* 35 (2): 309–17. https://doi.org/10.1007/s00146-019-00888-w.

Carman, Taylor (2008): »Between Empiricism and Intellectualism«. In *Merleau-Ponty. Key Concepts*, 13. London/New York: Routledge.

Chen, Zhi, Yijie Bei, und Cynthia Rudin (2020): »Concept whitening for interpretable image recognition«. *Nature Machine Intelligence* 2 (12): 772–82. https://doi.org/10.1038/s42256-020-00265-z.

Daws, Ryan (2020): »Microsoft Is Granted Exclusive Rights to Use OpenAI's GPT-3«. AI News. 23. September 2020. https://artificialintelligence-news.com/2020/09/23/microsoft-exclusive-rights-openai-gpt3/.

Dreyfus, Hubert L. (1992): *What Computers Still Can't Do*. Cambridge, MA: MIT Press.

Eschenbach, Warren J. von (2021): »Transparency and the Black Box Problem: Why We Do Not Trust AI«. *Philosophy & Technology*, September. https://doi.org/10.1007/s13347-021-00477-0.

Haugeland, John (1985): *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

Hofstadter, Douglas R. (1985): *Gödel, Escher, Bach. Ein endlos geflochtenes Band*. Stuttgart: Klett Cotta.

Hofstadter, Douglas R. (1995): *Fluid Concepts & Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.

Husserl, Edmund (1966): *Analysen zur passiven Synthesis*. Herausgegeben von H L Van Breda. Bd. 11. Husserliana. Den Haag: Nijhoff.

Husserl, Edmund (1980): *Phantasie, Bildbewusstsein, Erinnerung: Zur Phänomenologie der anschaulichen Vergegenwärtigungen*. Herausgegeben von Eduard Marbach. Bd. 23. Husserliana. Den Haag: Nijhoff.

Lewis, Philip E. (1966): »Merleau-Ponty and the Phenomenology of Language«. *Yale French Studies*, Nr. 36/37: 19–40.

McCarthy, John (2007): »From Here to Human-Level AI«. *Artificial Intelligence* 171 (18): 1174–82.

Merleau-Ponty, Maurice (1966): *Phänomenologie der Wahrnehmung*. Leipzig: de Gruyter.

Merleau-Ponty, Maurice (1993): *Die Prosa der Welt*. 2. Aufl. München: Wilhelm Fink.

Pressman, Fisher (2017): »Maurice Merleau-Ponty and Alex Garland: Human Consciousness in Ex Machina«. *Dianoia: The Undergraduate Philosophy Journal of Boston College*, Mai. https://doi.org/10.6017/dupjbc.voiIV.9874.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Quine, W. V. O. (2013): *Word and Object*. New ed. Cambridge, Mass: MIT Press.
Searle, John R. (1980): »Minds, Brains and Programs«. *The Behavioral and Brain Sciences* 3: 417–57.
Sellars, Wilfried (1997): *Empiricism and the Philosophy of Mind*. Cambridge, Massachusetts ; London, England: Harvard University Press.
Turing, A. M. (1950) »Computing Machinery and Intelligence«. *Mind* LIX (236): 433–60.
Weizenbaum, Joseph (1966): »ELIZA a computer program for the study of natural language communication between man and machine«. *Commun. ACM* 9: 36–45.
Weizenbaum, Joseph (1990):. *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt am Main: Suhrkamp.
Wittgenstein, Ludwig (2014): *Philosophische Untersuchungen*. 21. Aufl. Werkausgabe 1. Frankfurt am Main: Suhrkamp.
Zebrowski, Robin (2010): »In Dialogue With the World: Merleau-Ponty, Rodney Brooks and Embodied Artificial Intelligence«. *Journal of Consciousness Studies* 17 (7–8): 7–8.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Martin W. Schnell

Anerkennung einer Künstlichen Intelligenz als meinesgleichen?

Beitrag zur Maschinenethik

»Anerkennen ist etwas anderes als naive Gewißheit.«
Edmund Husserl (1966, 39)

Von körperlosen Geistern zu künstlichen Menschen

Hubert Dreyfus (1985) antwortete Mitte der 80er Jahre als Leibphänomenologe auf die Visionen der KI-Forschung. Er legte dar, *what Computers can't do*. Demnach ist es nicht möglich, menschliches, intelligentes Verhalten und Handeln künstlich zu simulieren. KI-Forscher wie Hans Moravec (1990) gingen irrtümlich davon aus, dass Intelligenz eine rein kognitive Angelegenheit sei, die der Vulnerabilität des Leiblichen nichts Wesentliches verdankt. Man denke auch an die Fiktion eines Gehirns im Tank und an die Spekulation über einen sich erneuernden Cartesianismus (vgl. als Beispiel: Putnam 1993, 209ff.). Dreyfus, Schüler von Merleau-Ponty, hebt demgegenüber hervor, dass Intelligenz leiblich verkörpert und verankert ist, dass sie situativ zur Geltung kommt, Unschärfen produziert und auf Vagheiten spontan eingehen kann. Der Geist hat nicht nur eine Syntax, sondern auch eine Semantik, wie Searle betont (vgl.: Searle 1986, 29ff.). Der Geist ist zudem verkörpert (embodied Mind), wie Francisco Varela (1995) sagt, und somit an die Endlichkeit des Lebens gebunden. Die gesamte Natur des Menschen lässt sich nicht restlos formalisieren oder digitalisieren (vgl. zu dieser Debatte: Schnell 2002).

Die KI-Forschung ist inzwischen von einer solchen Vision körperloser Geister und Intelligenzen abgerückt. Es macht keinen Sinn, eine Postbiologie im wörtlichen Sinne anzustreben. Heute geht es vielmehr um die Schaffung künstlicher Menschen, die den Diskurs der Vulnerabilität nicht umgehen, sondern ihn durchlaufen. Damit verändert sich Isaac Asimovs erstes Gesetz der Robotik. Die Verletzlichkeit bezieht sich nicht allein auf den Menschen, wie wir ihn bisher kannten, sondern auch auf den kommenden künstlichen Menschen.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Maschinenethik

Der vorliegende Text knüpft an das Projekt »Digitalisierung der Lebenswelt« (Schnell/Dunger 2019) an, welches nach den Dimensionen der Transformation der Lebenswelt durch die Digitalisierung fragte. Er fügt ihm die Frage hinzu, ob ein Mensch eine Künstliche Intelligenz in Gestalt eines Roboters als seinesgleichen anerkennen kann oder gar muss, und zwar im Zuge von Begegnungen zwischen Mensch und KI in der Lebenswelt. Begriff und Sache der Anerkennung werden damit in das Feld der Maschinenethik eingeführt (vgl. dazu Bendel 2019).

Die Maschinenethik ist ein komplexes Feld, auf dem gewiss nicht nur die Sozialphilosophie agiert. Sie stößt auf Technik und Ingenieurwissenschaft, die nicht nur an der Konstruktion von Robotern arbeiten, sondern auch an der Beantwortung philosophischer Fragen. Im Ausgang von der auf Emil du Bois-Reymond zurückgehenden Unterscheidung zwischen *Ignoramus* (wir wissen es noch nicht) und *Ignorabimus* (wir werden es nicht wissen) neigen Ingenieure manchmal dazu, Antworten auf unbeantwortete Fragen dem erwarteten Fortschritt zu überlassen. Aktuell vermag ein Roboter das Menschsein noch nicht perfekt zu simulieren, aber eines Tages wird es so weit sein. Damit finden unbeantwortete Fragen eine rasche Antwort. Der vorliegende Beitrag möchte zeigen, dass auf diese Weise jedoch das Fragebegehren nicht erschöpft ist und daher für ein gewisses Offenhalten von Antworten plädieren.

Postbiologisches vs. Humanoides Paradigma

Postbiologisches Paradigma	*Humanoides Paradigma*
Die KI vermag etwas, was ein Mensch nicht leisten kann. Z.B.: sehr komplexe Rechenaufgaben zu lösen. Das IBM-Programm Watson gewinnt das Jeopardy-Quiz, indem es seinen Gegenspieler an Wissen, Reaktionen und Assoziationen erheblich übertrifft.	Die KI vermag etwas, was ein Mensch beherrscht. Z.B.: Alltagsassistenz, wie die Steuerung von Küchengeräten oder das Herabsteigen einer Treppe.

Das postbiologische Paradigma verstand die Optimierung menschlicher Intelligenz als einen Prozess, der vom leiblichen Menschen abzulösen ist. Heute, wie es in Japan heißt, geht es darum, künstliche Menschen zu schaffen. Die Maschine passt sich dem Menschen an. Auch optisch werden Roboter quasi menschlich. Sie haben Körper mit entsprechender Anatomie, ein Gesicht mit Augen, Nase, Mund, Ohren, Wimpern und einen Haarschopf. Daher die Rede von humanoiden Robotern. Gabriele

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Gramelsberger bezeichnet diesen Schritt als Übergang von der »künstlichen Intelligenz« zum »künstlichen Leben.« (Gramelsberger 2019, 125) Am Forschungszentrum für Informatik in Karlsruhe wird versucht, Robotern Selbstbewusstsein beizubringen. Die leibliche Innenwahrnehmung als Basis für Handlungsentscheidungen wird simuliert. Das Schwierige daran ist, dass diese sehr heterogen ist. Eine Treppe hinzuabsteigen ist etwas anderes, als ein Fahrzeug im Straßenverkehr zu steuern.

Menschliche Gefühle vs. Kalte Technik?

Wie jede technische Innovation stößt auch die Digitalisierung auf eine gesellschaftliche Alltagsmentalität in der Lebenswelt, die nicht einfach übersprungen werden kann. Deutschland ist ein Land, das in der Tradition der Romantik steht. Natur, Authentizität und Begegnung gelten als Parameter für Wahrheit und Echtheit. Entsprechend wurde bei der Einführung des Fernsehens im letzten Jahrhundert nichts weniger als der Verlust der Realität befürchtet. Bis vor kurzer Zeit galt es als soziale Störung, wenn Menschen, zumeist junge, auf ihren Handybildschirm starrten, statt sogenannte echte Menschen anzuschauen. Anders verhält es sich in Japan. Dort existiert traditionell ein Verständnis von Natur und Technik, das eine hohe Alltagsakzeptanz für Technologien begünstigt (vgl.: Lévi-Strauss 2012). Japan ist daher ein Vorreiter von Digital Nursing im Bereich der Robotik.

Vor über zehn Jahren reagierte der japanische Staat auf die sogenannte Überalterung der Gesellschaft mit einer verstärkten Förderung der Entwicklung von Haushaltsrobotern. Weil es an echten und jüngeren Menschen mangelt, sollen »künstliche Menschen« (Wagner 2013, 38) anstehende Dienstleitungen, Hilfestellungen und Pflegeverrichtungen im Haushalt durchführen. Damit ist der Weg für die Lebensform posthumaner Familien geebnet. Das ist möglich, weil die Alltagsmentalität in Japan so ausgerichtet ist, dass etwa alte Menschen pflegerische Versorgungsverrichtungen, die Roboter mit und an ihnen vornehmen, als Unterstützungsleistungen erfahren und akzeptieren. Dem Digital Nursing kommt zugute, dass in Japan mechanische Servierpuppen, die Tee einschenken, und Mangamythologien verbreitet sind. In den 90er Jahren erreichte die Popsängerin Kyoko Date eine sehr große Popularität. Da sie die Hitparade anführte, interessierte sich die Öffentlichkeit auch für ihr Privatleben. Kyoko ist ein digitales Kunstprodukt, kaum anders als ein Tamagotchi. Das war dem Publikum völlig bewusst. Dennoch wollte es wissen, wie Kyoko lebt, was sie zum Frühstück isst. Somit erfand die Softwarefirma eine Lebensgeschichte für ihren Star.

Der Gebrauch von Robotern zur Assistenz im Alltag wird in Deutschland durchaus akzeptiert. Ein Roboter wird aber nur als reines Instrument

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

geduldet. Ein Roboter ist kein Mensch, er ist eine unpersönliche Maschine. In diesem Sinne soll er nur dort eingesetzt werden, wo es um den Einsatz von Kraft geht (z. B. Menschen aus dem Bett heben), wo das Grobe erledigt werden muss (z. B. staubsaugen) oder bei krankengymnastischen Übungen. Im Unterschied zum Roboter ist der Mensch ein Partner, der allein zu Ansprache, Berührung und Gefühl in der Lage ist. Darin liegen die eigentlich humanen Qualitäten begründet.

In dieser Ansicht wird die von der Romantik geprägte Auffassung Europas reproduziert. Sie zeigt sich im Dualismus Mensch vs. Maschine. Diese Thematik findet man um das Jahr 1800 sehr oft aufgenommen, etwa in E.T.A. Hoffmanns Geschichte vom redenden Türken in *Die Automate* (1814), in Heinrich von Kleists Über das Marionettentheater (1810) oder in Edgar Allen Poes *Maelzels Schachspieler* (1827). Ein Echo dieser Thematik klingt noch in Stefan Zweigs *Schachnovelle* (1938) an. Interessanterweise existieren vor 1800, also zeitlich vor der Romantik, durchaus andere Ansätze. Man denke nur an La Mettries programmatische Schrift *L'homme machin*e von 1748.

Ein möglicher Transhumanismus durch eine Verbindung von Mensch und Maschine wird von deutschen Usern im Gesundheitswesen heute aus dem Bereich der häuslichen Pflege verbannt, und zwar im Lichte einer Abgrenzung von den Sitten in Japan. Für einen Europäer scheint es schwerer als für einen Japaner vorstellbar zu sein, soziale Beziehungen zu Computerspielfiguren zu unterhalten und Bordelle aufzusuchen, in denen ausschließlich Sexpuppen aus Silikon vorzufinden sind (vgl.: Wagner 2013).

Auszug aus der Forschung von Alexander Hochmuth, Interview 94 CZ (in: Schnell/Dunger 2019):

Forscher: »Sie hatten vorher im Vorgespräch anklingen lassen, dass sie sich schon mit dem auseinandergesetzt haben. Was verstehen sie unter dem Begriff Roboter?« #00:11:14-5#
Proband: »Tja. Ich habs eigentlich nur gelesen, dass es von Japan rübergekommen ist. Und ich habe auch wie gesagt letztens eine große Reportage über Japan gelesen. Japan ist ja nun ein sehr automatisiertes Land. Japan ist nicht auf Menschen, auf zwischenmenschliche Beziehungen ausgerichtet. Ich war entsetzt was da eigentlich haben. Die heiraten sich selber. Die kaufen sich hübsche Kleider und lassen sich fotografieren. Aber sind nicht verheiratet. In Japan gibt es Männer eben halt die so in Spielsalons rumhängen, das ist der ihr Leben. Und hier gehen die Männer in den Puff, sag ich mal wenn sie Single sind. Dort gehen sie auch in eine Institution um mit einer hübschen Frau zu reden vielleicht, sie müssen dafür bezahlen und eventuell dürfen sie die Hand der Frau mal berühren. Da müssen sie auch für bezahlen und das höchste der Gefühle was sie machen dürfen ist den Kopf in den Schoß einer Frau legen. Da habe ich gedacht ›was ist das für ein Land?‹. Und deswegen hat es mich auch nicht gewundert als ich gelesen habe, dass die Roboter aus Japan kamen. Die werden das auch und dann werden die ja auch sehr alt dort. Und die werden eine große Anzahl von Menschen haben die sehr alt werden. Die werden diese Roboter anwenden und für die wird das komplett normal sein.«

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Aufgrund der kulturellen Differenz ist die Akzeptanz der Mitwirkung von künstlichen Menschen bei der Erledigung alltäglicher Verrichtungen in Europa noch eher gering. Diese mangelnde Akzeptanz könnte in den nächsten 20 Jahren deutlich schwinden. Wir werden diesem Gedanken weiter unten, am Beispiel einer Fiktion von Ian McEwan, nachgehen.

Gefühlsarbeit

Der entscheidende Schritt auf den Weg zum künstlichen Menschen in westlichen Gesellschaften besteht in der Einsicht, dass Gefühle keine natürlichen und authentischen Äußerungsweisen des Menschen sind. Arlie Russell Hochschild plädiert deshalb dafür, besser von Gefühlsarbeit zu sprechen. Sie versteht darunter im Ausgang von Norbert Elias »soziale Konstrukte, die gesellschaftlich erlernt« (Hochschild 2006, 23) werden und somit veränderbar sind. Sie verdeutlicht Erlernen und Einsatz von Gefühlsarbeit am Beispiel der professionellen Zugewandtheit von Stewardessen gegenüber Fluggästen und am Beispiel des Selbstverständnisses von Krankenschwestern im Krankenhaus.

Beispiel: Gefühlsarbeit im Krankenhaus

Eine klassische, empirische Untersuchung von Anselm Strauss und seinen Mitarbeitern zur Gefühlsarbeit im Krankenhaus stellt heraus, dass diese (oft von Krankenschwestern ausgeübte) Gefühlsarbeit die technische Arbeit der (häufig männlich dominierten) Medizin (Untersuchung, Ermittlung von Laborwerten etc.) ergänzt und bewirkt, dass Patienten sich als Mensch und Person anerkannt fühlen (vgl.: Strauss et. Al. 1980). In der Folge dessen wirken sie bei Behandlung und Untersuchung produktiv mit.

Entsprechende Äußerungsformen der Gefühlsarbeit sind:

- Motivation (»Halten Sie durch!«).
- Wertschätzung (»Das haben Sie gut gemacht!«),
- Fürsorge (»Sagen Sie mir, wenn es Ihnen weh tut.«),
- Tröstung (»Das tut mir leid.«).

Motivieren, Wertschätzen, Fürsorgen, Trösten sind ethisch relevant, weil sie auf die Beziehungen zwischen Personen bezogen sind. Die ethische Relevanz ist aber nicht identisch mit der Gefühlsarbeit. Diese ist auf die Produktion bestimmter Inhalte ausgerichtet (Ausdauer, Güte einer Handlung, Schmerz, Unglück), die von Programmen der Emotionserkennung identifiziert und durch simuliertes Verhalten (Worte des Trostes bei

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Schmerzbekundung) beantwortet werden können. Das Gefühl ist damit allein kein Kriterium zur Abgrenzung des Menschen von der Maschine.

Das Ethische geht über das Emotionale hinaus, da es der Person des Anderen gilt und nicht nur seinen Lebensäußerungen im Sinne Diltheys. Künstliche Intelligenzen können nur so kommunizieren, dass sie auf die Sprechakte des Anderen eingehen, aber nicht auch auf die Person des Anderen (vgl.: Misselhorn 2021). Maschinen empfinden keine Emotionen, sie vermögen es aber, am Menschen Gefühle im Sinne von Lebensäußerungen zu identifizieren und ihrerseits Gefühle bei Menschen auszulösen. Eine KI erkennt, *was* geäußert wird. Furcht zeigt sich im Zusammenziehen der Augenbrauen, im Heben des oberen Augenlids, dem Anspannen des unteren Augenlids und im Aufeinanderpressen der Lippen. Diese Lebensäußerungen sind digital erfassbar. *Wer* sich allerdings auf diese Weise fürchtet, Peter oder Paul, vermag die KI nicht zu berücksichtigen, weil sie nicht weiß, was es heißt, *jemand* zu sein. Eine sozial agierende KI in Gestalt eines Roboters »bleibt solipsistisch und führt zu einer verdinglichten Einstellung gegenüber anderen und sich selbst.« (ebd., 162) Ein »Fremdbewusstsein« vom Anderen als Anderen ist ihr »nicht gegeben.« (103)

Vom Turing-Test zur Ethik

Im Jahre 1950 entwickelte der Mathematiker Alan Turing einen nach ihm benannten Test. Ein Mensch kommuniziert demnach per Computer mit einem anderen Menschen und einem anderen Computer, also mit einer Maschine. Der Mensch hat die Aufgabe zu entscheiden, welcher seiner beiden Kommunikationspartner der andere Mensch und welcher die Maschine ist. Der andere Mensch und der Computer versuchen indes beide, den Menschen davon zu überzeugen, dass sie selbst Menschen sind. Wenn der Mensch diese Aufgabe genügend oft nicht lösen kann, Mensch und Maschine also nicht unterscheiden kann, dann könne davon ausgegangen werden, dass Maschinen gleich Menschen denken können. Heutzutage ist diese Versuchsanordnung weniger interessant aufgrund von Umstellungen in der angebahnten Interaktion.

Während der Turing-Test die Interaktion quasi verblindet, da der Mensch seine Partner nicht sieht, sondern nur deren Kommunikationsinhalte auf seinem Computer lesen kann, geht die Versuchsanordnung in dem Film *Ex_Machina* aus dem Jahre 2014 einen Schritt weiter. Der Mensch kommuniziert nun ganz bewusst mit einem humanoiden Roboter. Er kommuniziert mit ihm wie mit einem anderen Menschen, obwohl der Mensch sieht und weiß, dass dieser eine Maschine ist. Damit wäre gezeigt, dass es zwischen Mensch und Maschine zu gelingenden

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

intersubjektiven Beziehungen kommen kann, die der Mensch von sich aus als gelungen und erstrebenswert erlebt und bewertet. Der Roboter ist kein Instrument, sondern Partner einer Lebensform. Er hat Persönlichkeit und ist darin Popikonen wie R2 D2 weit überlegen.

Dieses Ergebnis ist durch den technischen Fortschritt bedingt (Maschinen können inzwischen Emotionen und Differenzierungen in Sinngehalten angemessener kommunizieren als dieses 1950 möglich gewesen ist), aber auch durch die Umstellung der Perspektive. Es geht nicht mehr darum, dass der Mensch die Maschine erkennt, sondern um ein Gelingen der Kommunikation mit ihr. Kommunikation ist ethisch relevant.

Hier erheben sich zahlreiche Fragen. Kann ein künstlicher Mensch, also eine KI, die in der Gestalt eines Roboters auftritt, nicht als voller Mensch anerkannt werden? Wer und/oder was ist das Ich, das am künstlichen Menschen zum Du wird? Und: wie gestaltet sich Intersubjektivität in dieser Hinsicht? Die Sozialphilosophie hat sich diese Fragen bislang noch nicht intensiv genug gestellt, weil sie in der Sache der Intersubjektivität Martin Buber oder Edmund Husserl gefolgt ist.

Ich und Du

»Der Mensch wird am Du zum Ich.« (Buber 1983, 37) Buber setzt voraus, dass sich als Ich und als Du zwei Wesen begegnen, die Menschen sind, und das heißt damit auch, vulnerable, leibliche und endliche Wesen. Diese sind konstitutiv für das Personsein des Anderen. »Wer man ist, wenn man mit jemanden spricht, bestimmt sich nicht unabhängig davon, zu wem man spricht, auf wen man hört und wem man sich schweigend verschließt.« (Waldenfels 2019, 123) Es muss somit einen Unterschied bedeuten, ob der Andere überhaupt ein Mensch ist oder nicht.

Wir geraten mitten in die Problematik der Digitalisierung, wenn wir an diese Formulierung die Frage anschließen: »Was bedeutet es, wenn man in einer Ich/Du-Beziehung eines der beiden Elemente durch einen Roboter ersetzt.« (Schnell/Dunger 2019, 196) Kann das Ich selbst zur Maschine werden?

Herausforderung

Die wesentlich neue Herausforderung besteht in einer komplett neuen Frageperspektive, da all die fundamentalen Aspekte und Qualitäten, die die Sozialphilosophie seit dem späten 18. Jahrhundert aufgeboten hat, um die Beziehung zwischen Ich, Du und dritter Person zu beschreiben, grundsätzlich neu überdacht werden müssen. Anerkennung,

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Reziprozität, Fürsorge, Care, Respekt, Achtsamkeit, Vertrauen, Frage/Antwort, Dialog und Gefühlsarbeit. Was passiert mit der Phänomenalität und der Geltung dieser und anderer Qualitäten, wenn Mensch und Maschine einander unter den Bedingungen der fortgeschrittenen Digitalisierung begegnen? Könnte sich diese Begegnung als Anerkennung realisieren? Wir beginnen mit einer Analyse der Anerkennung.

Von Hegel zu KI: Anerkennung

Einer Antwort auf diese Frage kommen wir näher, wenn wir das interpersonale Binnenverhältnis näher betrachten. Aus der Sicht Hegels ist dieses durch Begriff, Prozedur und Sache der Anerkennung gekennzeichnet. Der Weg von Hegel zur KI ist gewagt, aber nicht unmöglich. Verfahren der Anerkennung sind längst Teil der ethischen Diskussion um KI (vgl. u.a.: Cappuccio et al. 2018). Zudem ist Hegel Teil dieser Diskussion. In diesem Sinne versuchte sich Slavoj Žižek an einem Experiment: Könnte eine direkte Gehirn-zu-Gehirn-Kommunikation ein Ganzes realisieren, das das Wahre wäre (vgl.: Žižek 2020)?

Wir wollen die Voraussetzungen derartiger Visionen nicht teilen, sondern zunächst mit einer Grundlagenbetrachtung beginnen, deren Einsatz mit der These Hegels beginnt, dass das Selbstbewusstseins nur durch ein »anderes Selbstbewusstsein« (Hegel 1981, 144) zu sich zu finden vermag. Das Selbstbewusstsein bedarf also einer Bestätigung, die nur ein Wesen leisten kann, das seinerseits vom bedürftigen Selbstbewusstsein unabhängig ist und ihm zugleich in der Seinsweise gleicht. Dieses Wesen ist meinesgleichen in Gestalt eines anderen Selbstbewusstseins. Die Beziehung der damit in Geltung seienden zwei Selbstbewusstseine ist eine Beziehung der Anerkennung und zwar einer wechselseitigen. »Das einseitige Tun wäre unnütz; weil, was geschehen soll, nur durch beide zustande kommen kann … Sie anerkennen sich als gegenseitig sich anerkennend.« (ebd., 147)

Was macht den Prozess, der zur Anerkennung führt, aus?

Wenn ich jemanden anerkenne, muss ich mich auf den Anderen zubewegen und mich quasi an diesen Anderen verlieren. Zugleich muss ich, um eine eigene, anerkennende Instanz sein zu können, mich vom Anderen distanzieren. Diese Distanz eröffnet nicht nur mir, sondern auch dem Anderen eine Selbständigkeit innerhalb einer gemeinsamen Beziehung (Fichte).

Die Beziehung, innerhalb derer die Selbstbewusstseine einander nun »gegenüber« (148) treten, durchläuft einen »Kampf auf Leben und Tod« (149). Dieser Kampf ist notwendig, weil er den Selbstbewusstseinen

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

zeigt, was Hegel als Postulat voraussetzt: dass nämlich ein Selbstbewusstsein auf ein anderes angewiesen ist. Der fragliche Kampf zielt seinerseits auf das Gegenteil, nämlich auf eine Vernichtung des Anderen, die die Selbstständigkeit des eigenen Selbstbewusstseins gegenüber jedem Anderen zum Ergebnis haben soll. Doch genau dieses Ergebnis stellt sich nicht ein. Wer einen Anderen tötet, bleibt nämlich nicht als ein Selbst übrig, weil durch den Mord die notwendige Bestätigung des eigenen Selbst (durch einen Anderen) entfallen würde.

Die Analytik der Anerkennung wird durch eine ethisch-politisch relevante Praxis konkret. Der Kampf auf Leben und Tod hebt sich auf in einer Auseinandersetzung zwischen einem Herren und einem Knecht, denn einer der Akteure unterwirft sich dem anderen und rettet dadurch sein eigenes Leben. Doch darin scheitert eine wechselseitige Anerkennung der Selbstbewusstseine. Aus dieser Lektion kann dann die positive Gestalt einer gelingenden Anerkennung hergeleitet werden (vgl.: Taylor 1983, 190). Was macht nun die Beziehung zwischen Herr und Knecht aus?

Da die Beziehung zwischen Herr und Knecht im Anfang asymmetrisch ist, kann es zu keiner Anerkennung, die notwendig reziprok sein muss, kommen. Innerhalb dieser Asymmetrie erweist es sich, dass »die Wahrheit des selbständigen Bewusstseins … das knechtische Bewusstsein« (Hegel 1981, 152) und nicht das des zunächst überlegen scheinenden Herren ist.

Die Besonderheit des knechtischen Bewusstseins liegt darin, dass es zu sich selbst in Distanz tritt und zugleich durch sein Tun in der Welt tätig ist. »Der Knecht bezieht sich als Selbstbewußtsein überhaupt auf das Ding.« (ebd., 151) Er wird in der Manifestation seiner Arbeit, die, wie immer bei Hegel, zugleich Bildung ist, ansichtig.

Was ist gelingende Anerkennung?

Anerkennung muss eine symmetrische Relation sein, die eine Umkehrung der Perspektiven erlaubt. Die Weise, in der ich mich zu Dir verhalte, ist eine Umkehrung der Art, wie Du auf mich bezogen bist. Die Anerkennung eines Anderen wird dadurch realisiert, dass sie ihrerseits anerkannt wird. Sie anerkennen sich als sich gegenseitig anerkennend.

Gelingende Anerkennung realisiert sich als eine Lebensform, die Hegel als »Sittlichkeit« und damit als realisierte und gelebte Freiheit bezeichnet (vgl.: Honneth 2001, 82f.). Sittlichkeit ist damit nicht, wie bei Kant, abstrakte, sondern konkrete Vernünftigkeit (vgl.: Hegel 1984, § 145). Weil Hegels Theorie der Anerkennung im Rahmen der Phänomenologie unter anderem nur eine Durchgangsstufe der Bildung des Geistes ist, liefert sie von sich nicht alle Bestimmungen, die zur Kennzeichnung einer

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

gelingenden Anerkennung notwendig sind. Daher ist es wichtig, nachträgliche Argumente der Wirkungs- und Rezeptionsgeschichte einzubeziehen.

Wirkungsgeschichte im Marxismus

Die Rezeptionsgeschichte geht unter anderen davon aus, dass Hegel »historische Momente« (Fetscher 1973, 137) in die Charakterisierung von Herrschaft und Knechtschaft eingeflochten hat. Da sowohl Momente der Sklavenhaltergesellschaft als auch solche der bürgerlichen Gesellschaft ausgemacht werden können, sind die vom Kampf zwischen Herr und Knecht ausgehenden Überlegungen zur Anerkennung und Selbstständigkeit besonders innerhalb des Marxismus diskutiert und weitergeführt worden (vgl. u.a.: Lukács 1973 II, 506ff., 736ff.). Sie münden in die Vorstellung von der Befreiung des Knechtes, welche ihrerseits eine allgemeine Befreiung der Menschheit intendiert (vgl.: Taylor 1983, 211ff.).

Existenzphilosophie

Eine anders akzentuierte Richtung nehmen die existenzphilosophischen Überlegungen zur Anerkennung. Im Ausgang von Jean Hyppolites und Alexandre Kojèves existentialistischen Hegel-Interpretationen konzentriert sich Merleau-Ponty auf eine Beschreibung der Intersubjektivität zwischen Selbst und Anderem. Im Duell zwischen beiden hat derjenige das »genaue Bewußtsein von der menschlichen Situation« (Merleau-Ponty 2000, 91), der weder »nur für sich« noch »auf Kosten des Anderen« existieren will, sondern der »für den Anderen zu sein« (ebd., 92) in der Lage ist. Dieser ist der Knecht. Levinas vertieft diese Perspektive, indem er das für-den-Anderen in die Metapher der Geiselschaft transformiert. »Das Subjekt ist Geisel.« (Levinas 1992, 248) »Je mehr ich zu mir komme, … desto mehr entdecke ich mich als verantwortlich.« (ebd., 249)

Paul Ricoeur und Emmanuel Levinas

Paul Ricoeur kommt der Verdienst zu, den Diskurs der Anerkennung geöffnet zu haben. Es gibt demnach einen *Parcour* der Anerkennung, der verschiedene Wege enthält. Erkennen, Wiedererkennen, Anerkanntsein sind deren Namen. Der Parcour der Anerkennung erstreckt sich dabei vom »positiven Akt des Erkennens zum Verlangen nach Anerkennung.« (Ricoeur 2006, 40)

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Das Erkennen erfasst jemanden oder etwas so, wie er, sie oder es bereits ist. Das Anerkennen macht etwas oder jemanden erst zu dem, was er/sie/es sein wird. Anerkennung stiftet damit etwas, was ohne sie nicht sein würde. Sie stiftet das als-was oder als-wer etwas oder jemand anerkannt wird (vgl.: Schnell 2017, 98f.).

Im Unterschied zum Erkennen ist die Anerkennung wesentlich auf den Anderen bezogen. Das aktive Erkennen ist ein Akt, der vom Ich ausgeht, die passive Bitte um Anerkennung ist hingegen ein Geschehen, in dem ein passives Ich sich an die Adresse eines aktiven Du richtet. Mit dem »Umschlag vom Aktiv ins Passiv« (Ricoeur 2006, 40) tritt das aktive Erkennen immer ein bisschen mehr in den Hintergrund.

Paul Ricoeur betont, dass die Anerkennung, im Unterschied zum Erkennen, vom Anderen bestimmt wird. An dieser Stelle bezieht er sich auf Emmanuel Levinas. »Das Gesicht des Nächsten bedeutet mir eine unabweisbare Verantwortung, die jeder freien Zustimmung, jedem Pakt, jedem Vertrag vorausgeht.« (Levinas 1992, 199) Ich stehe dadurch in einer Verschuldung gegenüber dem Anderen. Die damit gegebene Asymmetrie wird aus der Sicht eines Dritten zur Frage nach Gerechtigkeit, die eine »originäre Asymmetrie« in eine »Gegenseitigkeit zwischen ungleichen Partnern« (Ricoeur 2006, 204) transformiert. Damit treten auch das aktive Erkennen und passive Anerkennen in eine Gegenseitigkeit.

Die Rede von Partnern im Plural bedeutet in erster Linie nicht, dass Ich und der Andere eine Mehrzahl bilden, sondern, dass es viele Andere gibt, zu denen ich in Beziehung stehe. Die Institution der Gerechtigkeit stiftet nun eine Ordnung, die festlegt, wie jene Beziehungen geregelt sind. Diese Regelegung besteht in der Festlegung, wer wem gegenüber welche Rechte und Pflichten hat und eröffnet damit die Frage nach der Gerechtigkeit, die jede Dyade überschreitet (vgl.: Schnell 2001, 212ff.). Die »Gerechtigkeit gründet« einen »mir und den Anderen gemeinsamen Bereich, in dem ich zu ihnen gezählt werde, das heißt in dem die Subjektivität Bürger ist mit all den maßvollen und meßbaren Pflichten und Rechten, die das ausgeglichene und oder durch das gleichzeitige Bestehen und Konkurrieren von Pflichten und Rechten sich Ausgeglichenheit verschaffende Ich umfaßt.« (Levinas 1992, 350) Ich bin jetzt auch ein »Anderer für die Anderen.« (ebd., 345) Verantwortung ist nun etwas, das »wir füreinander haben.« (ebd., 189)

In der aus dem Geist der Asymmetrie entstandenen, reziproken Beziehung zum Anderen ist nun die Unterscheidung zwischen Sagen und Gesagten zu beachten. Das gilt auch für die Kommunikation mit einer KI.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Anerkennung: Sagen und Gesagtes

Als Roboter definieren wir eine Maschine, die über einen Körper, Prozessoren, Sensoren und Effektoren verfügt und deren Verhalten als autonom erscheint (Vgl.: Loh 2019, 17). In diesem Sinne sind etwa Computer oder intelligente, persönliche Assistenten keine Roboter. Alle Arten von KI bieten den Menschen aber eine Beziehung an, wenn sie auf Sprachsteuerung beruhen. »Es stellt sich damit allerdings die Frage, inwiefern sich ein artifizielles System überhaupt in einem genuinen Sinn als Gegenüber, mit dem man eine Beziehung oder gar Freundschaft eingehen kann, verstehen lässt.« (Loh 2019, 28) Freundschaft ist, laut Aristoteles, eine reziproke Beziehung, sie beinhaltet Anerkennung des Anderen als meinesgleichen.

Vor dem Hintergrund des Diskurses der Anerkennung im Ausgang von Hegel und im Anschluss an Levinas unterscheiden wir in dieser Frage zwischen dem Sagen und dem Gesagten (vgl.: Levinas 1992, 29ff.). Das Gesagte ist der *Inhalt* der Rede, der wahr oder falsch sein kann. Indem etwas gesagt wird, ereignet sich zugleich das Sagen, das die *Beziehung* zwischen Ich und Du ausmacht.

Die Beziehung

Es »gibt« Interaktionen zwischen Menschen und, klassisch gesagt, Maschinen. Ein Mensch stellt einer Maschine eine Frage, erhält eine Antwort und richtet sich danach. Kann man deshalb davon ausgehen, dass der Mensch die Maschine als Gesprächspartner anerkennt? Anerkennen als wen oder was?

Die Bestimmung der Reziprozität der Anerkennung erfolgt bei Hegel in formaler Hinsicht. Selbst und Anderer »anerkennen sich als gegenseitig sich anerkennend.« (Hegel 1981, 147) Diese Formalität, aufgrund derer Anerkennung auf die Reziprozität von Wesen, die sich als anerkennende anerkennen, reduziert wird, trifft auf die Begegnung zwischen zwei Menschen, aber auch zwischen Mensch und KI zu. Ich frage die KI nach dem Weg, der in der fremden Stadt zu gehen ist, die KI antwortet. Ich anerkenne sie, weil sie mir antwortet, sie anerkennt mich, weil sie meine Äußerung als eine Frage annehmen kann, auf die sie zu antworten in der Lage ist. Anerkennung im Zeichen der Vulnerabilität bedeutet aber mehr als das.

Im Vorgang der Anerkennung geht es auch um einen qualitativen Aspekt: die mögliche Meinesgleichheit im Licht der Endlichkeit. Robert Spaemann beschreibt daher Anerkennung als »Anerkennung eines Menschen als meinesgleichen.« (Spaemann 1990, 195) Konzeptionell

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

gesehen, erweist sich das Geschehen der Anerkennung damit als ein dreistelliger Akt: Ich erkenne Dich als jemanden oder etwas an (vgl.: zu dieser Logik: Schnell 2017, 99f.). Du erscheinst als meinesgleichen. Dadurch ist Anerkennung mehr als eine Formalität.

Unter Berücksichtigung einer Wechselseitigkeit besteht Anerkennung somit darin, jemanden als meinesgleichen anzuerkennen und das heißt zu achten und zu respektieren. Durch diese Gleichheit ist eine egoistische Bevorzugung meiner zu Lasten des Anderen ausgeschlossen. Diese Gleichheit in der Anerkennung schließt im Falle von Menschen eine Gleichheit in der Endlichkeit ein. Wir beide sind endliche Wesen, die ein Leben leben, das uns von Anderen gegeben worden ist. Natalität und Mortalität verschränken sich hier (vgl. dazu: Liebsch 1996).

Es ist wohl offensichtlich, dass diese Art der Anerkennung zwischen Mensch und Maschine nicht möglich ist. Die Maschine tritt mir in der Frage der Gebürtlichkeit und der Endlichkeit nicht gleich gegenüber. Kommt es trotz dieser erheblichen Differenz zu irgendeiner Art von Anerkennung oder dergleichen, wenn ich auf die Antwort der Maschine höre und damit der Maschine »vertraue«, dass sie die Wahrheit sagt?

Kann ein Mensch einer Maschine vertrauen, wie er einem anderen Menschen vertraut? Die Anerkennung der Maschine bezieht sich nicht auf die Beziehung, aber sehr wohl auf den Inhalt.

Der Inhalt

Es ist zu bedenken, dass sich die Beziehung zwischen Ich und Du auch als Austausch von Informationen, die quasi den Inhalt der Interaktion bilden, gestaltet. Hier geht es dann um die Frage, was die Interaktionsteilnehmer können im Sinne ihrer Vermöglichkeiten. Damit ist die Perspektive eines »Wettbewerbs« (Waldenfels 1998, 208) eröffnet. Ich stelle eine Frage und erwarte, dass der Andere sie als einen Sprechakt versteht auf den er mit der Kundgabe von Wissen antwortet. Aber verfügt Watson über dieses Wissen, das ich aktuell nicht habe? Gehört das Wissen seinem Möglichkeitsraum an? Wessen Wissensvorrat ist der größere?

Schließlich ist zu beachten, dass sich die Beziehung zwischen Ich und Du nicht nur als Austausch von gültigen Informationen und Nachrichten, die quasi den Inhalt der Interaktion bilden, gestaltet. Ich frage die Maschine etwas, sie gibt mir eine exakte Antwort. Zur menschlichen Kommunikation gehören aber wesentlich auch positive Unbestimmtheiten (G. Gamm), die im engeren Sinne keine Nachrichten und Informationen sind. »Wenn ich Einem sage: Halte dich ungefähr hier auf«, dann handelt es sich um eine »unexakte« (Wittgenstein 1990, Nr. 88) Erklärung, die aber funktioniert. Der Andere stellt sich dort hin, ich mache das Foto. Obwohl

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Fragen unexakt und Antworten verschwommen sind, gelingt menschliche Kommunikation. Kommt eine Maschine ihrerseits mit Vagheiten aus?

Der Mensch im Unterschied zum Roboter

Der Unterschied zwischen einem Menschen und einem Roboter liegt nun darin, dass der Mensch um seine Endlichkeit weiß und sich zu ihr verhält, indem er sie vor sich verleugnet oder aktiv auf sich nimmt. In jedem Fall beansprucht die Endlichkeit das Dasein als Einzelnes (vgl.: Heidegger 1979, 263; Schnell 2017, 145ff.). Als Einzelnes bedeutet, dass das Dasein als »Unvergleichlichkeit und Einmaligkeit« (Spaemann 1996, 196) angesprochen ist. Die Endlichkeit meint mich, denn sie ist nicht delegierbar. Das trifft aber wiederum auf jedes Dasein zu, so dass der Andere als »meinesgleichen« (ebd., 195) anerkannt wird.

Ein Roboter hat kein Bewusstsein von Endlichkeit, weil ein Roboter nicht sterben kann, er kann lediglich kaputtgehen und damit seine Funktion verlieren. Ein Mensch ist vulnerabel, ein Roboter ist es nicht. Ein Bewusstsein von der eigenen Endlichkeit zu haben, meint an dieser Stelle weniger die Möglichkeit, jemandem einen mentalen Zustand zuschreiben zu können, sondern eher eine Praxis. Die vulnerable Person hat Angst oder Furcht vor dem Tod, sie übt sich in der Praxis symbolischer Immortalität. Bereits Kant verteidigte die Vulnerabilität des Menschen und verortete das Wissen um die Endlichkeit und die Anerkennung des Todes im Zeitbewusstsein. Die Kürze des Lebens lässt uns aktiv werden in der Praxis eines symbolischen Weiterlebens (vgl.: Schnell 2020, 19f.). Es ist keine KI realistisch denkbar, weder in der Praxis noch in der Fiktion, die ein Testament hinterlässt – solange wir mit Käte Meyer-Drawe davon ausgehen, dass Testamente nicht aus Algorithmen hergeleitet werden. Es ist wichtig, dass diese Differenz beachtet wird!

Hier paart sich die existentielle Unmöglichkeit mit einer logischen Unmöglichkeit. Ein Roboter, der in der Läge wäre, auf Dauer und täuschend echt einen Menschen zu spielen und damit auf gewisse Weise ein Mensch zu »sein«, würde jede Frage nach einem Unterschied zwischen Mensch und Maschine zum Verstummen bringen. Vielleicht sind alle Menschen, die bislang auf der Erde existiert haben, Maschinen. Descartes hätte dann möglicher Weise Recht, wenn er sagt, dass die Menschen, die ich auf der Straße vorbeigehen sehe, Automaten sein könnten, die menschliche Hüte und Kleider übergezogen haben (vgl.: *II. Med., Nr. 13*). Damit wäre die Intersubjektivität wieder aus der Ersten Philosophie verbannt und wir könnten erneut bei Fichte starten. Was bedeutet es, dass endliche Vernunftwesen andere endliche Vernunftwesen außer sich annehmen müssen?

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Vision: der Roboter als »Mitmensch« in europäischen Gesellschaften

Bekanntlich kann sich das Reich der Fiktionen in der Frage technologischer Utopien weiter vorwagen als Ingenieure und Wissenschaftler, indem es Sachverhalte als gegeben ansetzen, die in den technischen Wissenschaften noch als unrealistisch gelten. Gelegentlich kommt es vor, dass die Fiktion heute die Realität von morgen beschreibt. In diesem Sinne gibt es zahlreiche Romane und Filme, die uns einen Blick in den Alltag der Zukunft gewähren. Was wäre, wenn eines Tages *Maschinen wie ich und Menschen wie ihr* zusammenleben würden?

In seinem gleichnamigen Roman erzählt Ian McEwan die Geschichte von einen jungen Paar in London, das sich entschließt, einen humanoiden Androiden als künstlichen Menschen in der eigenen Wohnung zu halten. Adam ist der erste neue Mensch und wird zum Teilhaber an einer menschlichen Lebensform. Adam lebt nun mit Charlie, einem Anhänger von Alan Turing, und seiner Geliebten Miranda zusammen. Adam wird von ihnen so programmiert, dass er in das Leben beider passt. Adam beherrscht die Alltagskonversation, gewisse Bildungsstandards, er kann sinnhafte Handgriffe in der Küche ausführen und damit an der Lebensform teilhaben.

Es stellt sich die Frage, als wer oder was Adam dieses vermag? Wenn er als Partner akzeptiert wird, ist er dann als Mensch anerkannt, als jemand meinesgleichen? Der Turinganhänger Charlie formuliert: »In dem Moment, da wir im Verhalten keinen Unterschied mehr zwischen Mensch und Maschine erkennen können, müssen wir der Maschine Menschlichkeit zuschreiben.« (McEwan 2019, 119)

Zu einer ernsthaften Herausforderung für die Hybridlebensform kommt es, als Adam (die Maschine) und Miranda (der Mensch) Sex haben. »Sein Schwanz füllt sich mit destillierten Wasser. Aus einem Tank in der rechten Pobacke.« (ebd., 129) Miranda hat Adam anerkannt, aber nicht als ihresgleichen, sondern als Instrument der Befriedigung, als das auch Menschen einander immer schon haben wahrnehmen können. Die eigentliche Herausforderung stellt sich im Verhältnis zwischen Charlie (dem Betrogenen) und Adam (dem Betrüger) ein. Im Rahmen eines Gesprächs unter »Männern« bahnt sich ein Kampf auf Leben und Tod an.

Charlie wird zum Herren. »Jetzt galt es, mich gegen ihn zu behaupten.« (177) Charlie ist der Herr, weil er sich als Eigentümer von Adam, der Maschine, sieht. Adam wurde als Helfer, als Maschine angeschafft. »Falls er sich dem widersetzte, und insbesondere, falls er wieder jemanden verletzte, würden wir ihn zwecks Generalüberholung zum Hersteller zurückschicken.« (178) Adam hat seine Funktion als Maschine und damit die Position von Charlie als Herr zu akzeptieren. Die Position des Herren ist die stärkere, die zugleich darin schwach ist, dass sie nur stärker

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ist in Relation zum Knecht. Der Herr muss vom Knecht anerkannt sein, um Herr sein zu können. Darin liegt die Schwäche des Herren.

Adam tritt als Knecht auf. »Du und Miranda, Ihr seid meine ältesten Freunde. Ich liebe euch beide. Meine Pflicht euch gegenüber besteht darin, offen und ehrlich zu sein. Und ich meine es ernst, wenn ich sage, wie leid es mir tut, dass ich gestern Abend etwas in dir zerbrochen habe. Ich verspreche, das kommt nie wieder vor.« (178f.) Adam ist Knecht, weil er nicht allein für sich, sondern für die Anderen sein will. Er hat keine eigene Position, da er sich ganz an seine Pflichten den Anderen gegenüber ergibt. Darin liegt seine Schwäche. In dieser Schwäche zeigt er zugleich, wer er ist. Darin liegt wiederum seine Stärke. Der Marxismus fand in diesen Konstellationen Ansatzpunkte für ein Klassenbewusstsein, das den »Aufbau der ganzen Gesellschaft vollkommen zu erfassen fähig« (Lukács 1983, 126) ist, das also die Wahrheit der Gesellschaft und der Geschichte auf seiner Seite hat und darum weiß.

Es kommt aber zu keiner gelingenden Anerkennung. Adam löst sich aus seiner Position, da es ihm gelingt, »den Notschalter zu deaktivieren.« (McEwan 2019, 179) Hätte er damit eine »Angst vor dem Tod«, dann wäre sie keine, durch die er sich als menschengleich ansähe. Die Maschine will vielmehr über den Menschen triumphieren. »Maschinen wie ich und Menschen wie ihr, … wir werden euch übertreffen … und überdauern … auch wenn wir euch lieben.« (ebd., 369)

Die Maschine kann nicht sterben, auch nicht den technischen Tod. Charlie reagiert als Maschinenstürmer und verzichtet auf die Anerkennung, da er sich entschließt, Adam »mit dem Hammer den Kopf einzuschlagen.« (373) Er tötet damit nicht seinesgleichen, sondern schaltet einen übermächtigen Mechanismus einer künstlichen Intelligenz ab und fügt dem hinzu: »Ich hatte ihn gekauft, und es war mein gutes Recht, ihn zu zerstören.« (367) Damit ist Adam weder meinmesgleichen, noch ein Knecht, sondern ein Werkzeug.

Zwischen dem Menschen und der KI kommt es zu keiner gelingenden Anerkennung, die sich in einer sittlichen Lebensform, wie Hegel es vorsah, manifestiert. Aber dennoch geschieht etwas zwischen beiden Akteuren.

Mensch vs. Maschine

Unsere Ausgangsfrage lautete: Was bedeutet es, wenn man in einer Ich/Du-Beziehung eines der beiden Elemente durch einen Roboter ersetzt? Wie gestalten sich dann die Zwischenereignisse zwischen Ich und Du?

In Hinblick auf den Inhalt (das Gesagte) kommt es zu einer gelingenden Kundgabe von Wissen, Informationen. Fragen werden inhaltlich

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

beantwortet und keinesfalls nur im Bereich trivialer Sachverhalte, sondern auch im Hinblick auf Einschätzungen und Bewertungen.

Hinsichtlich der Beziehung (das Sagen) sieht die Lage anders aus, da es zu keiner Anerkennung des Anderen als meinesgleichen kommt. Es ist die Endlichkeit, die im McEwans Vision den entscheidenden Punkt ausmacht. Das endliche Wesen Mensch setzt sein Leben aufs Spiel und siegt gewissermaßen gegen die Maschine. Beide Akteure können sich wechselseitig nicht als endliche Wesen anerkennen. Ihre Beziehung erschöpft sich auch nicht in einem Kampf auf Leben und Tod, da der Mensch den Roboter auf ein Werkzeug zurückversetzt. Dieses Ergebnis würde sich auch dann bestätigen, wenn die Maschine den Menschen vernichtete.

Postdigitalisierung

Die sich gleichzeitig als Begegnung realisierende *Beziehung* betrifft nun die Frage, ob ich den Roboter auch als meinesgleichen anerkenne. Das geschieht offensichtlich nicht! Damit ist einer vollen Akzeptanz künstlicher Menschen im Zeichen der Humanität Grenzen gesetzt. Sie bleiben dennoch eine Herausforderung für den Menschen.

In der Interaktion mit Robotern schätzen Menschen die Leistungsfähigkeit von künstlichen Intelligenzen. Vor allem dann, wenn der Roboter noch lernt, Informationen und Handreichungen passgenauer zu geben bzw. auszuführen. In der Interaktion wird von Menschen realisiert, dass eine KI nicht ihresgleichen ist. Sie ist nicht endlich, weil sie nicht geboren wurde und daher nicht altern kann und somit auch nicht krank, behindert oder pflegebedürftig zu werden vermag. Der Überschuss der Vulnerabilität über die Maschine hinaus verweist auf eine Ethik und Politik der Postdigitalisierung.

Digitalisierung ist als Transformation der Lebenswelt in quantifizierbare Daten definiert (vgl. dazu: Schnell/Dunger 2019, 13f.) und damit auch ein Angriff auf vulnerables Leben (vgl.: Schnell 2020, Anhang 9). Die Rede von Postdigitalisierung bedeutet nun nicht, dass Digitalisierung aufhört, sondern vielmehr, dass über die Quantifizierung des Selbst und der Anderen hinaus das *Wer* der Person und damit ihre Unausdeutbarkeit (G. Gamm) sichtbar werden.

Literatur

Bendel, Oliver (Hg.) (2019): *Handbuch Maschinenethik*, Wiesbaden: Springer.

Blumenberg, Hans (2015): *Schriften zur Technik*, Berlin: Suhrkamp.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Buber, Martin (1983): *Ich und Du*, Heidelberg: Lambert Schneider.
Cappuccio, Massimiliano L. et al. (2018): »Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition«, in: *Philosophy & Technology* (2020) 33:9–31.
Dreyfus, Hubert (1985): *Was Computer nicht können*, Berlin: Athenäum.
Fetscher, Iring (1973): »Randglossen zu Hegels Herrschaft und Knechtschaft«, in: Fahrenbach, Helmut (Hg.) (1973): *Wirklichkeit und Reflexion. Walter Schulz zum 60. Geburtstag*, Pfullingen: Neske.
Gramelsberger, Gabriele (2019): »Digitale Wissenschaft«, in: Martin W. Schnell und Christine Dunger (Hg.) (2019): *Digitalisierung der Lebenswelt*, Weilerswist: Velbrück.
Hegel, Georg W.F. (1981): *Phänomenologie des Geistes*, Frankfurt am Main: Suhrkamp.
Hegel, Georg W. F. (1984): *Grundlinien der Philosophie des Rechts*, Frankfurt am Main: Suhrkamp.
Heidegger, Martin (1979): *Sein und Zeit*, Tübingen: Max Niemeyer.
Hochschild, Arlie R. (2006): *Das gekaufte Herz*, Frankfurt am Main: Campus.
Honneth, Axel (2001): *Leiden an Unbestimmtheit,* Stuttgart: Reclam.
Husserl, Edmund (1966): *Analysen zur Passiven Synthesis*, Den Haag: Nijhoff.
Husserl, Edmund (1976): *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, Den Haag: Nijhoff.
Levinas, Emmanuel (1992): *Jenseits des Seins oder anders als Sein geschieht*, Freiburg/München: Alber.
Lévi-Strauss, Claude (2012): *Anthropologie in der modernen Welt*, Berlin: Suhrkamp.
Liebsch, Burkhard (1996): *Geschichte im Zeichen des Abschieds*, München: Fink.
Loh, Janina. (2019): *Roboterethik*, Berlin: Suhrkamp.
Lukács, Georg (1983): *Der junge Hegel* II Bde., Frankfurt am Main: Suhrkamp.
Merleau-Ponty, Maurice (2000): *Sinn und Nicht-Sinn*, München: Fink.
Misselhorn, Catrin. (2021): *Künstliche Intelligenz und Empathie*, Stuttgart: Reclam.
Moravec, Hans (1990): *Mind Children*, Stuttgart: Hoffmann und Campe.
McEwan, Ian (2019): *Maschinen wie ich*, Zürich: Diogenes.
Putnam, Hilary (1993): *Von einem realistischen Standpunkt*, Reinbek bei Hamburg: Rowohlt.
Ricoeur, Paul (2006): *Wege der Anerkennung*, Berlin: Suhrkamp.
Schnell, Martin W. (2001): *Zugänge zur Gerechtigkeit*. München: Fink.
Schnell, Martin W. (2002): »Ideologie und Anthropologie. Zur Widerkehr des leiblosen Geistes«, in: Heinrich Greving und Dieter Gröschke (Hg.) (2002): *Das Sisyphos-Prinzip*, Bad Heilbrunn: Julius Klinkhardt.
Schnell, Martin W. (2017): *Ethik im Zeichen vulnerabler Personen*, Weilerswist: Velbrück.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Schnell, Martin W. :(2020): *Das Ethische und das Politische. Sozialphilosophie im Zeichen der Vulnerabilität*, Weilerswist: Velbrück.
Schnell, Martin W./Dunger, Christine (2019): *Digitalisierung der Lebenswelt. Studien zur Krisis nach Husserl*, Weilerswist: Velbrück.
Searle, John R. (1986): *Geist, Hirn und Wissenschaft*, Frankfurt am Main: Suhrkamp.
Spaemann, Robert (1990): *Personen*, Stuttgart: Klett-Cotta..
Strauss, Anselm et. al. (1980): »Gefühlsarbeit. Ein Beitrag zur Arbeits- und Berufssoziologie«, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (32: 629–651).
Taylor, Charles (1983): *Hegel*, Frankfurt am Main: Suhrkamp.
Varela, Francisco J./Thompson, Evan/Rosch, Eleanor (1995): *Der mittlere Weg der Erkenntnis*, Bern/München/Wien: Scherz.
Wagner, Cosima (2013): *Robotopia Nipponica. Recherchen zur Akzeptanz von Robotern in Japan*, Marburg: Tectum.
Waldenfels, Bernhard (1985): *In den Netzen der Lebenswelt*, Frankfurt am Main: Suhrkamp
Waldenfels, Bernhard (1998): *Grenzen der Normalisierung*, Frankfurt am Main: Suhrkamp.
Waldenfels, Bernhard (2019): *Erfahrung, die zur Sprache drängt*, Berlin: Suhrkamp.
Wittgenstein, Ludwig (1980): *Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp.
Žižek, Slavoj (2020): *Hegel im verdrahteten Gehirn*. Frankfurt am Main: Fischer.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Renate Buschmann

Kann aus KI Kunst werden?

Dialogische Beziehungen mit Künstlicher Intelligenz

Ist Künstliche Intelligenz imstande, Kunst hervorzubringen? Diese Frage wird momentan vielerorts gestellt und bringt anschaulich an die Oberfläche, wie sehr die fulminante Entwicklung von KI gleichermaßen subtile Faszination wie latente Verunsicherung auslöst. Dass KI in unzähligen unserer Alltagssituationen zum Einsatz kommt, ob bei Chatbots in Callcentern, bei Online-Übersetzungsdiensten oder bei Apps zur Gesichtserkennungen, mag zwar bei einigen ein gewisses Befremden hervorrufen, wird aber überwiegend respektiert. Maschinen mit ihren Rechenleistungen sind bei der Auswertung von digitalen Daten einfach effizienter, reagieren zuverlässiger und besitzen eine höhere Präzision beispielsweise beim Erkennen und Abgleichen von Mustern. Das alles beruht auf kalkulierbaren Verfahren, die nun durch KI-Programmierung die wesentliche Funktion beinhalten, aus Prozessen zu lernen und sich selbst optimieren zu können.

Die künstlerische Tätigkeit lässt sich aber in der allgemeinen Vorstellung nicht auf analytische Vorgänge zurückführen. Auch wenn wir bereits weitgehend Abstand von der Vorstellung des/der Künstler:in als schöpferischem Genie und mit einer ›undefinierten‹ Inspiration genommen haben, erfolgt die kreative Praxis nicht aus automatisierten Abläufen, sondern in unserem heutigen Kunstverständnis aus der Fähigkeit mit Regelwerken zu brechen und unerwartete Ergebnisse zu erzielen. Die Umstände, aus denen Kunstwerke entstehen, sind erklärbar, aber eben nicht vorhersehbar; sie sind nachvollziehbar, aber eben nicht im Vorhinein berechenbar.

Trotz oder gerade wegen dieser Hürden wird in der KI-Forschung offenbar versucht, künstlerische Prozesse zu imitieren. Wenn es gelingt, auch solche offensichtlich nichtlinearen und diskontinuierlichen Handlungen von KI ausführen zu lassen, wird es nicht nur ein eindrucksvoller Beweis für maschinelle Leistungsfähigkeit sein. Es wird zugleich erneute Debatten zum Mensch-Maschine-Verhältnis freisetzen, zum befürchteten Kontrollverlust der Menschen und der potentiellen Übermacht von Maschinen.[1] Wenn Maschinen nicht allein Produktionsapparate sind,

1 Siehe zu »Rage against the machine« Thomas Ramge, *Mensch und Maschine. Wie Künstliche Intelligenz und Roboter unser Leben verändern*, Stuttgart: Reclam Verlag 2018, vor allem S. 18–28.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

sondern »lernen« können, wie Kunstwerke zu gestalten sind, wird das Selbstverständnis angekratzt, das künstlerische Äußerungen stets als Inbegriff von menschlicher Individualität und Persönlichkeit aufgefasst hat.[2]

Die kursierende journalistische Kurzformel »Kann KI Kunst?« lässt offen, was mit diesem KÖNNEN gemeint ist. Was die KI unter Beweis stellen soll, ist in erster Linie abhängig davon, welche Anschauung von Kunst damit verbunden ist. Liegt das Können in der Fertigkeit, handgemachte Geschicklichkeit zu reproduzieren und Motivik bravourös zu variieren? Besteht die Anforderung an die KI in dem Vermögen, vergleichbare Erzeugnisse von solchen herzustellen, die bereits Kunststatus besitzen? Kann KI durch ihre intrinsische Fähigkeit bestechen, über die Kombinatorik von großen Datensätzen ein schier unendliches Repertoire an Form- und Farbgestaltung aufzurufen, das den menschlichen Ideenreichtum überbietet? Oder liegt das Interesse gar nicht darin, Künstlicher Intelligenz eine künstlerische Autonomie zuzugestehen und damit die Polarität von Mensch und Maschine fortzuschreiben? Finden Künstler:innen sogar deshalb an KI Gefallen, weil sie kooperative Prozesse mit KI ansteuern und mit den Möglichkeiten als auch Unmöglichkeiten dieser Technologie interagieren wollen? Sollte man deswegen anstelle von »Kann KI Kunst?« eher fragen »Kann aus KI Kunst werden?«.

Die boomende KI-Forschung hat in den vergangenen Jahren die Kunstfertigkeit von Computerprogrammen nachgewiesen und damit für viel Aufsehen gesorgt. Daran beteiligt waren keine Künstler:innen, sondern Wissenschaftler:innen, die aus Gemälden renommierter Maler:innen Datensätze generierten mit der Absicht, daraus Algorithmen abzuleiten. So konnte die Spezifik der jeweiligen Malstile nicht nur herausgefiltert und berechnet, sondern auch Software programmiert werden, die das Alleinstellungsmerkmal des individuellen Malausdrucks auf x-beliebige Motive überträgt. Die KI Deepart übernimmt die Arbeit von Kopisten – natürlich perfekter und mit der Qualität ausgestattet, glaubwürdige Variationen zu erfinden.[3] Dazu titelte 2016 die Süddeutsche Zeitung

2 Die aktuelle rasante Entwicklung von KI hat die Debatte über die Produktion von Kunst durch Maschinen neu angeregt, u.a. Hanno Rauter, *Die Kunst der Zukunft. Über den Traum von der kreativen Maschine*, Berlin: Edition Suhrkamp 2021. Während des Schreibens dieses Aufsatzes ist der Band »Kann KI Kunst?«, Kunstforum International, Bd. 278, Nov.–Dez. 2021, erschienen.

3 Siehe https://marjorie-wiki.de/wiki/Deep_Art_Effects (letzter Zugriff 21.12.2021). Inzwischen ist daraus ein lukrativer Service mit dem Slogan »Repaint your picture in the style of your favorite artist« geworden, der jedes Foto in einen wiedererkennbaren Stil, beispielsweise von Van Gogh oder Edvard Munch, umwandelt. https://deepart.io/ (letzter Zugriff 21.12.2021).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

»Dieser Computer malt fast so schön wie ein Mensch«[4], was zum Ausdruck kommen ließ, dass das antiquierte Bild vom Wettstreit zwischen Mensch und Maschine noch immer tragfähig ist.

DeepDream[5], 2015 von Google veröffentlicht, ist eine ähnliche Software, die Bild- und Fotovorlagen selbstständig mit künstlerischen Ausdrucksmitteln anreichert, indem Parameter einer gängigen Motivik von abstrakter und psychodelischer Malerei auf die Vorlagen angewendet werden. Im Unterschied zu klassischen Bildbearbeitungsprogrammen ist DeepDream ein kreatives Werkzeug, das keine Einflussnahme durch den Nutzer erfordert und stattdessen die Imagination für die Bildtransformationen aus der Black Box seiner Algorithmen zieht. Die Resultate werden bewundert, weil sie mit einer Motivvielfalt, einem Detailreichtum und einer Farbauswahl bestechen, die erahnen lassen, mit welchen gigantischen Mengen an Bildvorlagen die lernende Maschine zuerst trainiert worden sein muss. Voraussetzung für die Fantasiewelt der Maschine ist es, den Bildspeicher zu beherrschen, Anwendungen abzuleiten und neue Kombinationen herzustellen. In dieser Kapazität, einer zudem beachtlichen IT-Leistung, liegt die Kreativität der Maschine, mit der sie ihre Nutzer:innen überrascht.

Maschinen mit künstlerischer Kompetenz auszustatten, dient zusätzlich zur kommerziellen Nutzung nicht zuletzt dem Prestigegewinn der Softwareentwickler. Unter der Prämisse, dass Kunst als ein urmenschliches Artefakt betrachtet wird, suchen diese Projekte die herausfordernde Annäherung der Maschine an den Menschen, indem sie in die Entwicklung einer maschinellen Kreativität investieren. So unterstützte Googles Konkurrent Microsoft das ambitionierte Projekt »The Next Rembrandt«, das 2016 für Furore sorgte.[6] Man hatte mit hunderten von Rembrandt-Gemälden eine hochaufwändige Datenanalyse betrieben, daraus das idealtypische Rembrandt-Gemälde berechnet und es fast 350 Jahre nach dem Tod des Malers im 3-D-Druckverfahren minuziös hergestellt. Das Resultat versetzte auch Kunstexpert:innen in Erstaunen, weil die maschinelle Interpretation dessen, wie eine repräsentative Rembrandt-Malerei hinsichtlich des Motivs und der Maltechnik auszusehen habe, absolut überzeugend war. Das Experiment, die malerische Begabung und die in der Historie verankerte Gedankenwelt eines verstorbenen Künstlers mithilfe eines KI-gesteuerten Profiling

4 https://www.sueddeutsche.de/digital/kuenstliche-intelligenz-dieser-computer-malt-fast-so-schoen-wie-ein-mensch-1.2856388 (letzter Zugriff 21.12.2021).

5 https://deepdreamgenerator.com/ (letzter Zugriff 21.12.2021).

6 https://www.nextrembrandt.com/. Siehe auch https://www.sueddeutsche.de/kultur/kuenstliche-intelligenz-ein-echter-rembrandt-aus-dem-rechner-1.2949787 (letzter Zugriff 21.12.2021) und Michael Betancourt, »A Note on AI and the Ideology of Creativity«, in: Academia Letters, März 2021, Article 328. https://doi.org/10.20935/AL328.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

weiterleben zu lassen, ist gewiss kein Beitrag zur Kunst der Gegenwart.[7] Allerdings es ist eine ebenso eindrucksvolle wie alarmierende Demonstration, dass die Auswertung von Big Data durch Data Mining-Methoden und Künstliche Intelligenz zukünftig grenzenlose Anwendungsfelder eröffnet.

Technologien zu hinterfragen und sie einer unkonventionellen künstlerischen Nutzung zuzuführen, ist zum Arbeitsbereich vieler Gegenwartskünstler:innen geworden.[8] Ihr künstlerisches Handeln vereint meist Technikaffinität mit Technikkritik. Seit Jahrzehnten – und lange vor den oben erwähnten merkantilen Vorzeigeprojekten – ist Künstliche Intelligenz Gegenstand künstlerischer Praxis.

Ein frühes Beispiel ist das Video »Getting Nowhere« von Alexander Hahn, in dem sich der Künstler selbst 1982 auf die Unterhaltung mit einer künstlichen Intelligenz einlässt.[9] Der schriftliche, circa zehnminütige Dialog am Computerbildschirm wurde von einer Videokamera in Echtzeit dokumentiert. Lautlos ist zu verfolgen, wie der Cursor des Textprogramms Buchstabe für Buchstabe die Frage eingetippt und nach einer gewissen Wartezeit die Antwort eintrifft. Beide Intelligenzen treten nur über die hellgrüne Leuchtschrift auf dem in den 1980er Jahren zeitgemäßen Röhrenmonitor in Erscheinung. Der Mensch Alexander Hahn bleibt physisch ebenso unsichtbar wie die KI seines Gesprächsprogramm ELIZA, das der Informatiker Joseph Weizenbaum »for the study of natural language communication between man and machine« 1966 bahnbrechend herausgebracht hatte.[10] Anfang der 1980er Jahre, als Computer noch längst nicht in jedem Haushalt zu finden waren, Videorekorder gerade erst Konsumgut wurden und an das World Wide Web noch nicht zu denken war, mutete allein die Möglichkeit, ein Interview mit einem Softwareprogramm zu führen, wie eine Science Fiction an. Nachdem Hahn sich zu Beginn über die Wahrscheinlichkeit eines Atomkriegs und dessen Folgen erkundigt, wächst im Verlauf der Befragung sein Unmut

7 Siehe Mads Pankow, »Künstliche Kreativität«, in: Dominik Landwehr (Hg.), *Machines and Robots*, Basel: Christoph Merian Verlag 2018, S. 48–58.

8 Siehe den Sammelband »Kann KI Kunst?«, Kunstforum International, Bd. 278, Nov.–Dez. 2021, und Hans Ulrich Obrist, der die Chance von KI im Simulieren von neuen Welten sieht. Hans Ulrich Obrist, »Das Unsichtbare sichtbar machen: Kunst trifft auf KI«, 2018, online https://www.goethe.de/prj/k40/de/kun/ooo.html#i6718158 (letzter Zugriff 21.12.2021).

9 Das Video ist zugänglich unter https://stiftung-imai.de/videos/katalog/medium/5090 (letzter Zugriff 21.12.2021). Weitere Informationen zum Werk: Christoph Vögele, *Alexander Hahn. Werke 1976 – 2006*, Kat. Kunstmuseum Solothurn, Heidelberg: Kehrer Verlag 2007, und https://www.alexanderhahn.com/getting-nowhere.html (letzter Zugriff 21.12.2021).

10 https://dl.acm.org/doi/10.1145/365153.365168 (letzter Zugriff 21.12.2021).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

über ELIZAs unzureichenden Antworten und ›ihre‹ Konditionierung, Nichtwissen durch Gegenfragen zu kaschieren. Die Videoaufzeichnung lässt sich trotz ihres dokumentarischen und puristischen Ausdruckes wie ein Bühnenstück verfolgen, in dem Mensch und Maschine in stereotype Verhaltensmuster verfallen: der zu Überlegenheit und Affekten neigende Mensch trifft auf die Objektivität und Emotionslosigkeit der trainierten Maschine.

Die Bereitschaft von Menschen, mit Maschinen zu kommunizieren, hat die Künstlerin Lynn Hershman Leeson mit »Agent Ruby« ab 1999 auf die Probe gestellt. Zu einem Zeitpunkt als noch niemand mit virtuellen Assistentinnen wie Siri oder Alexa sprach und die KI-Entwicklung noch im Anfangsstadium steckte, konfrontierte die Künstlerin mit einer Website, auf der sich die Besucher:innen mit dem virtuellen Charakter »Agent Ruby« unterhalten konnten.[11] Mit der eigenen Selbstinszenierung und der Erfindung von Identitäten hatte Lynn Hershman Leeson in den 1970er Jahren begonnen. Die Chance, eine weitere artifizielle feminine Persönlichkeit schaffen können, die dieses Mal im Netz ihre individuelle Geschichte autonom schrieb, hatte sie zu dem Projekt veranlasst. Die künstliche Intelligenz, die den Namen der Filmfigur Ruby aus Hershman Leesons Spielfilm »Teknolust« übernahm, bezieht ihr Wissen aus dem stetig anwachsenden Datenspeicher des Internets. Als weiblicher Avatar mit minimaler Gestik lädt »Agent Ruby« bis heute zum gemeinsamen Träumen ein (Agent Ruby's Edream Portal) und steht zur Unterhaltung bereit. Auf dem Interface tritt keine mysteriöse Wissensmaschine in Erscheinung, sondern ein vermenschlichtes, idealisiertes, makelloses Gegenüber. Die Interaktion der Nutzer:innen mit KI »Agent Ruby« ist jeweils eine persönliche Performance, die vom Grad des Vertrauens, der Aufrichtigkeit und dem Mitteilungsbedürfnis der Nutzer:innen bestimmt wird. Es ist ein ebenso spielerisches wie psychologisches Moment, das über die Akzeptanzgröße des virtuellen Dialogpartners und des digitalen Speicher- und Wissensraum Internet entscheidet. Das Chatverhalten mit »Agent Ruby« reflektiert nicht nur das Verhältnis der Nutzer:innen zu Künstlicher Intelligenz, sondern insbesondere auch das im Laufe von zwei Jahrzehnten und durch die Etablierung des Web 2.0. veränderte Medienverhalten. Seit 2013 wird eine museale Installation von »Agent Ruby« ausgestellt, die zusätzlich Einblick in das Archivmaterial früherer Chats gibt.[12]

11 2002 hat das San Francisco Museum of Modern Art das Netzprojekt »Agent Ruby« in Auftrag gegeben und sorgt seitdem dafür, dass die Webseite im WWW zugänglich bleibt: http://agentruby.sfmoma.org/ (letzter Zugriff 21.12.2021). Siehe Lynn Hershman über Agent Ruby https://www.youtube.com/watch?v=O5ZGlWB36J8 (letzter Zugriff 21.12.2021).

12 Siehe »Lynn Hershman Leeson: The Agent Ruby Files. Exhibition Surfaces Ten Years of Virtual Character's Conversation with Real People«, https://

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Die Such- und Kunstmaschine »net.art generator« von Cornelia Sollfrank bietet den Webseitenbesucher:innen hingegen eine geringe Beteiligung an, wenngleich diese den maßgeblichen Impuls für die Arbeit der Online-Bildermaschine setzt. Bereits 1997 hat Cornelia Sollfrank die erste Version des »net.art generator« ins Netz gestellt; auf die fünfte Version haben wir heute noch Zugriff.[13] Es war ein weitsichtiges Projekt, das die zurzeit boomende Entwicklung um KI-Kunst-Algorithmen und die daraus resultierenden Zuschreibungsfragen vorwegnahm. »Net.art generator« zählt zur Netzkunst, die in den 1990er Jahren angetrieben wurde von der Faszination für die – damals noch neue – globale Vernetzung, für den – damals noch ersehnten – herrschaftsfreien Raum und die – damals noch weitgehend unbekannte – Kreativität der Programmierung. Für den »net.art generator« liefern die Besucher:innen der Webseite die Schlagworte, mit denen der Algorithmus das Internet nach passenden Bildern durchsucht und sie mit einer zusätzlichen grafischen Bearbeitung zu einem – zumindest in diesem Augenblick – einzigartigen Bild kombiniert. Die Künstlerin gab ihre Gestaltungshoheit damit sowohl an die KI der Suchmaschine als auch an die Webseitenbesucher:innen ab und erklärte ironisch: »Smart artist makes the machine do the work«.[14] Der von Cornelia Sollfrank veröffentlichte Bildgenerator ist seit seiner Entstehungszeit ein Anschauungsbeispiel für die Ambivalenz gegenüber Suchmaschinen, die für Nutzer:innen des World Wide Web unverzichtbar sind, aber intransparent verfahren. In der Black Box des »net.art generator« werden aus digital zugänglichen Bildquellen neue synthetische Bilder gemischt, die einmal mehr die komplexe Frage nach der Urheberschaft digitaler Produktionen aufwerfen: Ist die Urheberschaft den Nutzer:innen aufgrund der von ihnen gewählten Schlagworte zu zuschreiben, der Künstler:in aufgrund ihres Konzepts für die KI, den Programmierer:innen aufgrund der Softwareentwicklung oder gar der intelligenten Maschine aufgrund ihrer autonomen Bildfindung?[15]

2016, als KI zur Bilderzeugung gerade große mediale Aufmerksamkeit erhielt, überraschte der Maler Liu Xiaodong mit einem digitalen

www.sfmoma.org/press/release/lynn-hershman-leeson-the-agent-ruby-files/ (letzter Zugriff 21.12.2021).

13 https://net.art-generator.com (letzter Zugriff 21.12.2021).

14 Ebenda. Siehe auch Gerald Rauig, »Maschinische Virtuosität«, in: Sabine Himmelsbacher (Hg.), *Cornelia Sollfrank. Expanded Original*, Kat. Edith-Ruß-Haus für Medienkunst, Ostfildern: Hatje Cantz Verlag 2009, S. 116–121.

15 Inzwischen, fast 25 Jahre nach Sollfranks selbst programmierten »net.art generator« gibt es eine Reihe von Künstler:innen, die mit heutigen KI-Erfindungen von Bildgeneratoren arbeiten, diese mit Wörtern ›füttern‹ und mit diesem Tool und dem damit verbundenen aleatorischen Moment Bilder erfinden. Siehe Hans Christoph Böhringer, »KI in der Kunst: Die Suche nach dem Zauberspruch«, in: Frankfurter Allgemeine Zeitung (FAZ.net), 12.12.2021,

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Malroboter. Dieses maschinelle Substitut steht in der künstlerischen Tradition einer Automatisierung des Kunstschaffens.[16] Die Malapparatur »Weight of Insomnia« produziert unentwegt Gemälde während ihrer Ausstellungen.[17] Verbunden ist der Roboterarm jeweils mit einer an einem öffentlichen Platz installierten Kamera, die der Maschine den visuellen Input zuspielt. Ein ständiger, unkontrollierbarer Datenfluss speist die künstliche Intelligenz des Malroboters, der beharrlich aufzeichnet und die ihn erreichenden digitalen Daten in einen malerischen Duktus übersetzt. Sein ›Auge‹ ist die entfernt positionierte Überwachungskamera, die rund um die Uhr beobachtet und anders als ihr menschliches Pendant Schlaflosigkeit (insomnia) zum Prinzip permanenter Aufmerksamkeit und Produktivität erklären kann. Seine Vorgehensweise beruht auf Algorithmen, die aus den Kamerabilddaten die künstlerische Umsetzung für die Leinwand berechnen. Entstanden sind auf diese Weise monochrome Gemälde von prominenten Plätzen, u.a. in Peking, Shanghai, Berlin und London, durch die Pinselführung eines Roboters, aber mit einer vom Künstler definierten Programmierung. Der dokumentarische Anspruch, dem sich Liu Xiaodong mit seiner analogen Vorort-Malerei stets verpflichtet fühlt, wechselt bei seinem Roboterprojekt zu mechanischer Perfektion und Kontrolle. Aus der Zuverlässigkeit der Malmaschine resultiert über den mehrwöchigen Malprozess ein Übermaß an Datenerhebung und -sammlung, sodass die Vielzahl an erfassten Situationen übereinander gemalt und letztendlich in ihrer Akkumulation zur Unkenntlichkeit beitragen. Die Effizienz der maschinellen Endlosschleife löscht ihre eigenen Ergebnisse wieder aus.

Mit »Weight of Insomnia« macht sich der Maler vordergründig selbst überflüssig. Viel relevanter ist jedoch die zum Ausdruck gebrachte Skepsis, dass die lernenden Maschinen zu einer blinden Übererfüllung von Zielvorgaben neigen können. Ihre hypothetische Unfehlbarkeit und ihr Dauerbetrieb legitimieren rational ihren ubiquitären Einsatz, der jedoch abhängig von den jeweiligen Umständen eine Vielzahl von politischen und ethischen Bedenken hervorruft. Sind die Anwendungen Künstlicher

https://www.faz.net/aktuell/feuilleton/kunst-und-architektur/ki-kunst-die-suche-nach-dem-zauberspruch-17678337.html (letzter Zugriff 21.12.2021).

16 Andreas Broeckmann, »Maschinen, Menschen und Roboter«, in: Dominik Landwehr (Hg.), Machines and Robots, Basel: Christoph Merian Verlag 2018, S. 26–37.

17 »Weight of Insomnia« wurde beispielsweise 2017–18 im ZKM in Karlsruhe, 2018 im NRW-Forum Düsseldorf und 2019 in der Lisson Gallery in London ausgestellt. https://www.lissongallery.com/exhibitions/liu-xiaodong-weight-of-insomnia (letzter Zugriff 21.12.2021). Die Textpassage über »Weight of Insomnia« wurde in Teilen bereits publiziert: Renate Buschmann: Weight of Insomnia – von Malmaschinen und KI. in: WittenLab Magazin, Heft 1, Dezember 2020, S. 66–69.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Intelligenz erstmal im Alltagsleben angekommen, wird es für Künstler:innen umso dringlicher, KI in Produktions- und Rezeptionsvorgänge der bildenden Kunst einzubinden. Es gehört zum methodischen Vorgehen von Künstler:innen, auf innovative Technologien zu reagieren: einerseits um durch ihre Nutzung den Radius künstlerischer Visualisierungen zu aktualisieren, andererseits um Schwachstellen und Risiken solcher Technologien aufzudecken, die im Verlauf profitabler Kommerzialisierung zunehmend ignoriert werden. Im Fall von KI ist die Kunst jedoch nicht mehr allein ein Spiegel oder gar Korrektiv für gesellschaftliche Fehlhaltungen, sondern die Kunst wird selbst zum Experimentierfeld für Informatiker:innen, die darin den Präzedenzfall sehen, KI zum kongenialen Gegenüber werden zu lassen. Dass Software und computergesteuerte Fertigungssysteme so programmiert werden können, dass sie zeichnen, malen, bildhauern, dichten, komponieren und musizieren, ist imponierend und entzaubert den Vorgang des künstlerischen Schaffens – verlässt aber auch mit Machine Learning nicht den Rahmen dessen, was an Informationen eingespeist wurde.

Die angeführten Beispiele von Alexander Hahn, Lynn Hershman Leeson, Cornelia Sollfrank und Liu Xiadong veranschaulichen schlaglichtartig, dass das Interesse von Künstler:innen weniger in der Automatisierung künstlerischer Prozesse liegt als vielmehr in der Offenlegung von dialogischen Beziehungen zwischen Künstlicher Intelligenz und ihren Nutzer:innen.[18] Die rapide Ausdehnung von KI in unserer Alltagswelt hat insbesondere im jüngsten Kunstgeschehen den Gedankenaustausch über das zukünftige Verhältnis zwischen KI und Mensch intensiviert und fortgesetzt. Unser bisheriges, kulturell geprägtes Verständnis von der Maschine, die ein unbelebtes, aber effizientes Objekt verkörpert, gerät ins Wanken, wenn die Maschine als lernende, damit intelligente Funktionseinheit anzuerkennen ist. Das binäre Denken, das Mensch und Maschine voneinander trennt oder sogar polarisiert, erfasst nicht mehr die derzeitige komplexe Entwicklung einer immer selbstverständlicher werdenden Mensch-Maschine-Interaktion. Die lernende Maschine avanciert im Alltag zu einem gleichberechtigten Gegenüber, dessen gelobte Neutralität vorprogrammiert ist und dessen weitreichende Entscheidungsbefugnis nicht vorurteilsfrei sein muss.

Die Unsicherheiten im Umgang mit KI und die sich verändernde Perspektive auf solche innovativen Maschinen, die mit dem Attribut von

18 Zum Gedanken über »dialogische KI« siehe Alan N. Shapiro, »Roadmap künstliches Leben. Auf dem Weg zu dialogischer künstlicher Intelligenz«, in: Magazin #33 - Schnittstellen, hrsg. von Kulturstiftung des Bundes, Herbst/Winter 2019, S. 6–7; online https://www.kulturstiftung-des-bundes.de/de/magazin/magazin_33/roadmap_kuenstliches_leben_auf_dem_weg_zu_dialogischer_kuenstlicher_intelligenz.html (letzter Zugriff 21.12.2021).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

›künstlicher Intelligenz‹ auf Augenhöhe erscheinen, veranlassen Künstler:innen zum kritischen Dialog mit dem smarten maschinellen Gegenüber. Um abschließend noch zwei Beispiele zu nennen: In der Installation »What do I know? I am just a machine?!« (2019) von Martin Backes stoßen die Besucher:innen im Augmented Reality-Raum mit glänzenden Kugeln zusammen, die sich als sprechende Organismen entpuppen und Empathie für Maschinen fordern.[19] Wenn die KI – mit menschenähnlicher Stimme – um Akzeptanz ihrer weitreichenden Kompetenzen bittet, mutet dies zuerst wie ein Zukunftsszenario an, erweist sich dann aber als naheliegende Betrachtung über die zunehmende Autonomie solcher Systeme, die viele ethischen Fragen hinsichtlich der Entwicklung und des Umgangs mit KI hervorruft.

Johanna Reichs Videoperformance »Face Detection« (2018) kreist um das visuelle Vermögen von KI-basierter Gesichtserkennung. Die Künstlerin knetet so lange die rudimentären Merkmale eines Gesicht in einen Klumpen Ton, bis ihre Smartphone-App die Formung als Physiognomie akzeptiert.[20] Dieses Projekt ist kein Wettstreit zwischen Mensch und KI oder gar eine Demonstration, die Unzulänglichkeiten der KI-Gesichtserkennungen offenzulegen. Der Prozess veranschaulicht vielmehr die Konkordanz zwischen menschlicher und maschinellen Mustererkennung, die über die stereotype Gesichtsgestaltung urteilt. Die manuelle grobe Bearbeitung des natürlichen Materials Ton steht im absichtlichen Gegensatz zum Hightech-Produkt, das die überlebenswichtige Urfähigkeit des Menschen, Physiognomien wiederzuerkennen, nachzuahmen versucht. Mit dem Motiv der ›schaffenden Hände‹, die als Metapher des Schöpferischen zu verstehen sind, verlässt das Video die Herausforderung funktioneller Gesichtserkennung und wechselt zu der Frage: Kann KI auch Kunst erkennen? In der derzeitigen euphorischen Debatte um die Verselbstständigung von KI-Kunst-Algorithmen sind es gerade solche künstlerischen Interaktionen mit KI, die nicht das Konkurrenzverhältnis »Maschine contra Mensch« ausschließlich vorantreiben, sondern die dialogische Beziehung auszuloten versuchen.

Literatur

Betancourt, Michael (2021): »A Note on AI and the Ideology of Creativity«, in: Academia Letters, März 2021, Article 328. https://doi.org/10.20935/AL328.

19 https://www.martinbackes.com/portfolio/what-do-i-know-i-am-just-a-machine/ (letzter Zugriff 21.12.2021).

20 https://johannareich.com/project/home-works-face-detection/ (letzter Zugriff 21.12.2021).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Böhringer, Hans Christoph (2021): »KI in der Kunst: Die Suche nach dem Zauberspruch«, in: Frankfurter Allgemeine Zeitung (FAZ.net), 12.12.2021, https://www.faz.net/aktuell/feuilleton/kunst-und-architektur/ki-kunst-die-suche-nach-dem-zauberspruch-17678337.html (letzter Zugriff 21.12.2021).

Broeckmann, Andreas (2018): »Maschinen, Menschen und Roboter«, in: Dominik Landwehr (Hg.), *Machines and Robots*, Basel: Christoph Merian Verlag, S. 26–37.

Obrist, Hans Ulrich (2018): »Das Unsichtbare sichtbar machen: Kunst trifft auf KI«, online https://www.goethe.de/prj/k40/de/kun/ooo.html#i6718158 (letzter Zugriff 21.12.2021).

Pankow, Mads (2018): »Künstliche Kreativität«, in: Dominik Landwehr (Hg.), *Machines and Robots*, Basel: Christoph Merian Verlag, S. 48–58.

Ramge, Thomas (2018): *Mensch und Maschine. Wie Künstliche Intelligenz und Roboter unser Leben verändern*, Stuttgart: Reclam Verlag.

Rauig, Gerald (2009): »Maschinische Virtuosität«, in: Sabine Himmelsbacher (Hg.), *Cornelia Sollfrank. Expanded Original*, Kat. Edith-Ruß-Haus für Medienkunst, Ostfildern: Hatje Cantz Verlag, S. 116–121.

Rauter, Hanno (2021): *Die Kunst der Zukunst. Über den Traum von der kreativen Maschine*, Berlin: Edition Suhrkamp.

Scorzin, Pamela C.: *Kann KI Kunst? Neue Positionen und technisierte Ästhetiken*. Kunstforum International, Nov.-Dez. 2021, Bd. 278.

Shapiro, Alan N. (2019): »Roadmap künstliches Leben. Auf dem Weg zu dialogischer künstlicher Intelligenz«, in: Magazin #33 - Schnittstellen, hrsg. von Kulturstiftung des Bundes, Herbst/Winter 2019, S. 6–7; online https://www.kulturstiftung-des-bundes.de/de/magazin/magazin_33/roadmap_kuenstliches_leben_auf_dem_weg_zu_dialogischer_kuenstlicher_intelligenz.html (letzter Zugriff 21.12.2021).

Vögele, Christoph (2007): *Alexander Hahn. Werke 1976–2006*, Kat. Kunstmuseum Solothurn, Heidelberg: Kehrer Verlag.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Christian Grüny

Seltsam attraktiv

KI und Musikproduktion

1.

Wenn Künstliche Intelligenz Schlagzeilen macht, geht es erstaunlich oft um Kunst.[1] Das mag daran liegen, dass hier spektakuläre Ergebnisse erwartet werden – man kann sofort sehen oder hören, was die KI kann oder nicht kann. Während in der Auseinandersetzung mit KI in der Pflege oder in der Medizin sofort ethische Fragen in den Vordergrund rücken, ist es bei der künstlerischen Produktion ein Vermögen, das gern als eins der eigentlich definierenden Charakteristika des Menschen genommen wird: die Kreativität. Dass Maschinen uns nicht nur die Arbeit wegnehmen (und nicht etwa abnehmen), sondern nun auch noch kreativ werden sollen, scheint die Phantasie der Öffentlichkeit besonders stark anzuregen.

Zwei Beispiele: Das Pariser Kollektiv *Obvious* arbeitet seit längerem an und mit einer KI, die Bilder im Stil von Gemälden produziert, auch wenn sie natürlich nicht malt. Am 25. Oktober 2018 wurde bei Christie's unter großer medialer Aufmerksamkeit eines dieser Gemälde, »Portrait of Edward Bellamy« für $ 432.500 versteigert. Folgt man der Selbstdarstellung des Kollektivs, so ist klar, dass es hier tatsächlich um die Grundfrage der Kreativität geht: »Obvious is a collective of artists working with artificial intelligence, and driven by a fascination for replicating creativity.«[2] Die Rede davon, Kreativität »nachzubilden«, scheint tatsächlich auf die Vorstellung eines Ersetzens menschlicher Kreativität durch technische hinzuweisen.

Auch das Versteigerungsergebnis könnte man so lesen. Dass es allerdings die Bestätigung dafür ist, dass KI »tatsächlich« kreativ sein kann bzw. dass dies in der Öffentlichkeit und vor allem bei Sammlern so akzeptiert wurde, erscheint mir nicht überzeugend. Auf der Seite von Christie's

1 Die Motive des hier Ausgearbeiteten wurden bereits in einem früheren Text vorbereitet: Vgl. Christian Grüny, »Kontingenzmaschinen. Künstliche Intelligenz in der Musik«, in: *Neue Zeitschrift für Musik* 4 (2019), S. 17–21. Überdies danke ich Sandeep Bhagwati, Joachim Heintz, George Lewis, Sam Salem und Jennifer Walshe für ausführliche Gespräche. Nicht anderweitig nachgewiesene Bezüge stammen aus diesen Gesprächen.

2 https://obvious-art.com (alle Weblinks zuletzt aufgerufen am 3.2.2022).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

wird Ahmed Elgammal, Leiter des *Art and Artificial Intelligence Lab* an der Rutgers University mit dem Vorschlag zitiert, sich eher den gesamten Prozess anzusehen als das von der KI ausgespuckte Ergebnis und das Ganze als eine Form der konzeptuellen Kunst zu betrachten.[3] Urheber wäre dann *Obvious*, das man vielleicht als Mensch-Maschine-Kollektiv beschreiben müsste, und versteigert worden wäre eine gerade sehr angesagte Form der konzeptuellen, in Kollaboration produzierten Kunst. In diesem Sinne bemerkt Jens Schröter, dass es ein Verdienst dieser Kunst sein könnte, »auf den komplexen, verteilten Charakter der Autor*innenschaft als solcher hingewiesen zu haben«.[4]

Elgammal war auch an dem zweiten prominenten Beispiel beteiligt, nämlich der Komposition einer 10. Symphonie Beethovens aus dessen Skizzen mit Hilfe einer KI. Das von der Telekom initiierte und finanzierte Spektakel wurde am 9. Oktober 2021 in Bonn aufgeführt; während das Projekt vorher von allen Medien begleitet wurde, herrscht seit der Premiere weitgehendes Schweigen über die Qualität des Stücks. Hier ist es noch um einiges offensichtlicher, dass die Symphonie eine gemeinsame Kreation ist, denn die KI hat lediglich Tonfolgen in mehreren Varianten errechnet; Auswahl und Orchestrierung oblagen menschlichen »Experten« (das Gendern erübrigt sich angesichts der Zusammensetzung dieser Expertengruppe). In diesem Sinne hieß es in einer Ankündigung des Projekts zwei Jahre vor der Uraufführung im Ton einer Werbebotschaft: »Nur wenn Mensch und Maschine planvoll zusammenarbeiten, kann das große Potenzial der kombinierten menschlichen und maschinellen Kreativität erreicht werden!«[5] Nimmt man das Schweigen als Indikator, ist bis dahin noch ein weiter Weg.

Nun gibt es aber gerade in der Musik seit vielen Jahren mehr oder weniger erfolgreiche Versuche, algorithmisch Stücke im Stil von bestimmten Komponisten zu produzieren, allen voran Johann Sebastian Bach. Der Komponist David Cope hat bereits in den 1980er Jahren mit seinem Programm EMI (*Experiments in Musical Intelligence*, oder auch liebevoll-verniedlichend *Emmy*) unzählige Stücke im Stil verschiedener klassischer Komponisten produziert. So kann man sich etwa auf seiner Webseite 5000 Choräle im Stil von Bach als MIDI-Files herunterladen.[6] Während Copes technische Mittel damals noch vergleichsweise primitiv

3 https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx.

4 Jens Schröter, »›Künstliche Intelligenz‹ und die Frage nach der künstlerischen Autor*innenschaft«, in: *Kunstforum International* 278 (2021), S. 99–106, hier 105.

5 https://www.telekom.com/de/konzern/themenspecials/special-250-jahre-beethoven/details/kuenstliche-intelligenz-soll-beethovens-zehnte-sinfonie-vollenden-587346.

6 http://artsites.ucsc.edu/faculty/cope/5000.html.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

waren und regelbasiert vorgingen, stehen heute neuronale Netzwerke zur Verfügung, und tatsächlich gibt es auch hier ein auf Bach ausgerichtetes Modell, *DeepBach*.[7]

Natürlich *braucht* niemand diese 5000 Choräle, was auch Cope klar ist; eher geht es um den Nachweis, dass es möglich ist, sie zu generieren. Was hier verhandelt zu werden scheint, ist das Verhältnis von Regel und Kreativität, wofür Bach ganz besonders gut taugt: der Komponist, der quasi ikonisch für die Verbindung formaler Strenge mit Genialität steht und der noch deutlich näher an regelbasierten Kompositionstechniken ist als Beethoven. Insofern haben die genannten Projekte eher etwas von einer Versuchsanordnung als von Kreation in einem emphatischen Sinne. Cope beschreibt sie als »machine-created composition which attempts to replicate the style but not the actual notes of existing music«[8], stellt sie in eine lange Tradition, auf die ich noch zurückkommen werde, und sieht die Pointe klar in der *Verwechselbarkeit* der von Menschen und der maschinell produzierten Musik.

In einem ganz anderen kulturellen Bereich aber scheint es tatsächlichen Bedarf für derlei bestimmten Stilen folgende, KI-generierte Musik zu geben, die auch deutlich leichter herzustellen ist: »The bigger opportunity may be in putting AI music composition into the hands of non-musicians, whether that's as an app to generate mood-music on demand, or as a tool (likely within apps like Instagram, Snapchat and Facebook) to create copyright-free backing tracks for the videos they want to share on social networks.«[9] Es geht hier also um eine kulturelle Sphäre, in der Musik von vornherein als atmosphärische Untermalung genutzt wird, die gerade nicht auffallen soll und auf die man auch nicht eigens hören soll. Diese Art Einsatz von Musik ist natürlich nicht neu: Sie geht zurück bis zu Erik Saties »musique d'ameublement«, die damals noch damit zu kämpfen hatte, dass die Besucher*innen ihr tatsächlich zuhörten. Neu ist aber der unabsehbar gestiegene Bedarf an solcher Musik, die »die gleiche Rolle wie das Licht, die Wärme und der Komfort in jeder Form«[10] erfüllt.

Wer in einem Werbe- oder Selbstdarstellungsvideo solche Musik hört, begegnet in gewisser Weise tatsächlich einer KI in Aktion – aber er oder sie wird es vermutlich nicht merken, denn an Formen der Musik als

7 Vgl. https://arxiv.org/abs/1612.01010 .

8 David Cope, *Virtual Music. Computer Synthesis of Musical Style*, Cambridge, MA u. London 2001, S. 3.

9 »Music AI: The Music Ally Guide« (keine Autorenangabe), https://musically.com/wp-content/uploads/2018/11/Music-Ally-AI-Music-Guide.pdf. Die Seite, auf der sich dies findet, bezeichnet sich als »knowledge company for the modern global music business«.

10 Erik Satie, »Musique d'Ameublement. Mobiliarmusik«, in: ders., *Schriften*, Hofheim 1988, S. 31.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

weitgehend standardisiertes, wenn auch von Menschen gemachtes Industrieprodukt ist man ohnehin seit Jahren gewöhnt. Hier geht es um etwas anderes, als die großen Fragen nach Kreativität suggerieren: um eine Entwicklung nämlich, in der die Standardisierung so weit vorangetrieben wurde, bis die Produktion von einer Maschine übernommen werden kann. Die »mood-music on demand«, von der hier die Rede ist, ist Gebrauchsmusik in ihrer reinsten Form. Vielleicht sollte man an dieser Stelle von »Reativität« sprechen, wie die Kollegen von *Obvious* vorschlagen.[11]

Wenn man von solchen Beispielen absieht, scheint mir aber die Frage, ob eine KI tatsächlich menschliche Kreativität *ersetzen* kann, nicht der entscheidende Punkt: Sie verspricht hohe Klickzahlen und lässt den Blutdruck steigen, ist aber theoretisch nicht sonderlich interessant und als Zukunftsvision bis auf weiteres unrealistisch.[12] Kate Darling bringt die Sache in ihrem Buch über Roboter auf den Punkt: »Why are we trying to replicate something we already have?«[13] Auch die eingangs angeführten Beispiele gehen letztlich in eine andere Richtung, nämlich in die einer Zusammenarbeit von Menschen und einer KI, die Dinge tut, die wir nicht tun oder können. Dabei schlagen sie allerdings sehr verschiedene Wege ein: Einmal, bei dem Porträt, geht es ausdrücklich um die seltsamen Deformationen, die die KI in einem der traditionellen und repräsentativsten (in allen Bedeutungen des Wortes) Genres europäischer Malerei anrichtet, im anderen Fall um den ganz ernsthaften Versuch, eine Art virtuellen Beethoven zu erschaffen. Beispiele, die eher dem ersten Fall ähneln, gibt es natürlich auch in der Musik, und auf sie werde ich mich hier konzentrieren. Im Zentrum steht dabei, dem Fokus des Bandes entsprechend, die konkrete Begegnung mit der KI im künstlerischen Prozess.

Je nach der Rolle, die die Technologie dabei einnimmt, kann diese Begegnung sehr verschieden ausfallen. Im Einleitungstext zu ihrem *Oxford*

11 Der erste Satz des zwischen Utopismus und PR changierenden Manifest der Gruppe lautet: »We at Obvious wish to explore, use and share the different ways machine learning algorithms can empower our natural reativity.« (http://obvious-art.com/wp-content/uploads/2020/04/MANIFESTO-V2.pdf) Das Manifest wimmelt insgesamt von Druckfehlern, aber dass bei dem zentralen Begriff direkt am Anfang des programmatischen Texts ein Buchstabe amputiert wurde, kann kaum anders denn als Freudscher Versprecher gelesen werden.

12 Joachim Heintz diskutiert Reiz und Realismus solcher Szenarien und kommt in seinem erfreulich unaufgeregten Text zum selben Ergebnis; vgl. ders., »Spielen statt Fürchtenmachen. Situationen von Musik und KI«, in: *Musik-Texte* 162 (2019), S. 27–30.

13 Kate Darling, *The New Breed. How to Think About Robots*, London 2021, S. 15.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Handbook of Algorithmic Music unterscheiden Alex McLean und Roger Dean drei unterschiedliche Weisen, auf die Algorithmen in der Musikproduktion vorkommen oder aufgefasst werden können: als Werkzeuge, Sprachen oder Partner.[14] Ohne dass die beiden ihn zitieren würden, kann dies unmittelbar auf die Unterscheidungen bezogen werden, die Don Ihde beinahe dreißig Jahre zuvor in seinem *Technology and the Lifeworld* vorgeschlagen hatte, nämlich »embodiment relations«, »hermeneutic relations«, »alterity relations« und »background relations« zwischen Menschen und technologischen Artefakten oder, neutraler, Instanzen.[15] »Embodiment relations« sind Beziehungen, in denen ein Werkzeug zu einer Erweiterung des leiblichen Umgangs mit der Welt wird; Beispiele wären die Brille, Werkzeuge im handwerklichen Sinne und, auch bei Ihde, Musikinstrumente. Bei hermeneutischen Beziehungen geht es um das Lesen oder allgemeiner Decodieren, wobei auch hier die Materialität der Technologie selbst in den Hintergrund treten kann, indem sie durchsichtig für eine Bedeutung wird. Die dritte Möglichkeit der »alterity relations« ist dann gegeben, wenn die Technologie als Akteur eigenen Rechts und/oder Kommunikationspartner auftaucht, was natürlich die aufgeladenste und auch spektakulärste Möglichkeit ist und diejenige, bei der am emphatischsten von einer Begegnung gesprochen werden kann.

Diese drei ersten Varianten werden mich im Folgenden vor allem beschäftigen, während die letzte noch einmal auf einen Bereich bzw. einen Problemkomplex verweist, der bereits am Beispiel der »mood-music on demand« aufgetaucht ist: Das Wirken von KI kann sich in einen unbemerkten Hintergrund zurückziehen, wo es sich der Wahrnehmung und damit auch der kritischen Reflexion entzieht. Die reine Gebrauchsmusik mag unschädlich erscheinen, aber die Fragen, woher die Daten stammen, mit der die entsprechenden Algorithmen gefüttert wurden, wie sie ausgewählt wurden und wie die Algorithmen selbst programmiert sind, sind alles andere als trivial. Die Datensammlungen der großen Streamingdienste kommen denen von Google und Amazon gleich, und sie beeinflussen sowohl die Rezeption als auch die Produktion auf massive Weise.[16] Und auch die Frage der Datensätze und der in ihnen eingebauten Verzerrungen

14 Vgl. Alex McLean u. Roger Dean, »Musical Algorithms as Tools, Languages, and Partners: A Perspective«, in: dies. (Hg.), *The Oxford Handbook of Algorithmic Music*, Oxford 2018, S. 3–14.

15 Don Ihde, *Technology and the Lifeworld. From Garden to Earth*, Bloomington u. Indianapolis 1990, Kap. 5.

16 Vgl. etwa Robert Prey, Marc Esteve Del Valle u. Leslie Zwerwer, »Platform pop: disentangling Spotify's intermediary role in the music industry«, in: *Information, Communication, and Society* (2020), https://doi.org/10.1080/1369118X.2020.1761859; allgemein Shoshana Zuboff, *Das Zeitalter des Überwachungskapitalismus*, Frankfurt 2018.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ist hier nicht weniger relevant als in anderen kulturellen Bereichen.[17] Dieses Feld ist extrem wichtig, wird aber hier nicht im Fokus stehen.

Ehe ich mich den aktuellen Fragen zuwende, möchte ich aber noch einige allgemeinere Bemerkungen dazu machen, was in einem künstlerischen Prozess bzw. genauer einem musikalischen Kreationsprozess seit jeher im Spiel ist. Danach wird es um Komposition und um Improvisation gehen – nicht um überkommene Dichotomien zwischen diesen beiden Kategorien musikalischer Kreation neu zu befestigen, sondern um zwei Typen des Umgangs mit KI, einmal in einem längeren Konstruktionsprozess und einmal in einer Interaktion in Echtzeit, voneinander zu unterscheiden. Dass es dabei Überschneidungen und Parallelen gibt, ist offensichtlich. Dass die Auseinandersetzung mit der Vorgeschichte der KI so viel Platz einnimmt, liegt darin begründet, dass sich vor diesem historischen Hintergrund die Fragen von Musik und KI besonders deutlich konturieren.

Eine Bemerkung noch zum notorisch schwierig zu definierenden Terminus der künstlichen Intelligenz: Die Bandbreite reicht hier von regelbasierten Programmen wie Copes *Emmy*, deren Resultate so wirken sollen, als seien sie von einem intelligenten Konstrukteur hervorgebracht worden, bis hin zu *Deep neural networks*, die tatsächlich eine fremde Form der Intelligenz oder des Denkens zu verkörpern scheinen. Die Beispiele, um die es im Folgenden gehen wird, decken diese ganze Bandbreite ab.

2.

Die Grundfrage ist, wie wir uns den künstlerischen Produktionsprozess überhaupt denken müssen. Zweifellos ist die KI in ihren verschiedenen Spielarten, von frühen regelbasierten Systemen zu den GANs der Gegenwart, eine neue Entwicklung. Dennoch muss man sich von der Vorstellung befreien, es ginge hier um den Einbruch von etwas ganz anderem in einen ansonsten homogenen und vor allem zutiefst und ausschließlich *menschlichen* Schaffensprozess. Philosophien der Technik weisen seit langem darauf hin, dass technologische Artefakte auch diesseits von Ihdes »alterity relations« Produktionsprozesse entscheidend prägen und auch einen massiven Einfluss auf das haben, was mit ihrer Hilfe produziert wird. Zugespitzt könnte man sagen, dass auch ein Werkzeug ein Mitspieler ist. Ehe ich dies weiter ausführe, soll aber noch ein anderes Motiv in Spiel gebracht werden: das des künstlerischen Materials.

17 Vgl. dazu Adrian Mackenzie, *Machine Learners. Archaeology of a Data Practice*, Cambridge, MA 2017, und die Arbeit des AI Now Institutes (https://ainowinstitute.org/).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Die Vorstellung eines starken Künstlersubjekts, das einem widerstandslosen oder mittels Technik gefügig gemachten Material seine Ideen aufprägt, wird heute niemand mehr vertreten wollen – von der wirklich unübersehbaren Genderung einmal abgesehen taugt sie einfach nicht, um tatsächliche künstlerische Prozesse zu beschreiben. Künstlerisches Material ist keine formlose Materie, sondern immer bereits vorgeformt. Es erlaubt vieles, gibt manches vor und leistet anderem Widerstand. Das gilt sowohl für Materialien im buchstäblichen Sinne wie Stein, Ton oder Holz, die in der Kunst nicht als Naturstoffe auftauchen, sondern als durch und durch kulturell geprägte Materialien, als auch für Material im allgemeineren Sinne, wie er sich bei Adorno ausgearbeitet findet. Er bestimmt Material als all das, »womit die Künstler schalten: was an Worten, Farben, Klängen bis hinauf zu Verbindungen jeglicher Art bis zu je entwickelten Verfahrungsweisen fürs Ganze ihnen sich darbietet: insofern können auch Formen Material werden; also alles ihnen Gegenübertretende, worüber sie zu entscheiden haben«[18].

Der Ansatzpunkt für Adornos Materialbegriff ist die Frage, wie man von einem Fortschritt in der Musik sprechen kann, ohne damit die unsinnige These zu verbinden, die Musik der Gegenwart sei »besser« als diejenige Bachs und Beethovens, oder, noch allgemeiner, wie die Geschichtlichkeit der Kunst überhaupt gedacht werden kann. Die Frage nach dem Fortschritt ist wohl der Punkt, an dem Adornos Materialbegriff am schärfsten und oft verkürzend kritisiert worden ist; hier ist nicht der Ort, darauf genauer einzugehen.[19] Der entscheidende Punkt ist derjenige der Widerständigkeit des Materials gegenüber dem kompositorischen Zugriff, der von Adorno zu einer aktiven Rolle eines Materials ausgebaut wird, das »Bewegungsgesetze« aufweist und »Forderungen«[20] stellt, das sich also nicht einfach widersetzt, sondern tatsächlich mitwirkt. Ganz allgemein muss man sagen, dass Widerstand und Eigensinn des Materials eine der Voraussetzungen künstlerischer Arbeit, nicht etwa ihre Bedrohung sind. Und während man es beim Marmor schwierig finden mag, davon zu sprechen, er »rede mit«, ist dies bei überkommenen Formen wie etwa der harmonischen Tonalität weit naheliegender: Hier redet eine ganze Tradition mit, zu der man sich verhalten muss.

Etwas anders ist der Fall der Instrumente gelagert. Auch hier ist eine Tradition im Spiel, die sich in materieller Form verkörpert. Instrumente sind auf eine bestimmte Weise gebaut worden, die sich auch verändern

18 Theodor W. Adorno, *Ästhetische Theorie (Gesammelte Schriften Bd. 7)*, Frankfurt am Main1973, S. 222.

19 Vgl. dazu ausführlich: Christian Grüny, »Material und Ort. Rehabilitation einer kritischen Kategorie«, in: Robin Becker, David Hagen u. Livia v. Samson (Hg.), *Ästhetik nach Adorno. Positionen zur Gegenwartskunst*, Berlin 2022, S. 49–64.

20 Theodor W. Adorno, *Philosophie der neuen Musik* (*Gesammelte Schriften*, Bd. 12), Frankfurt a. M. 1980, S. 40.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

mag, aber z.B. im Falle der Streichinstrumente seit Jahrhunderten auf bemerkenswerte Weise stabil ist, es gibt Traditionen, wie sie zu spielen sind und die teilweise sozusagen in sie eingebaut sind, aber trotzdem der expliziten Überlieferung und habituellen Inkorporierung bedürfen, es gibt Weisen und Kontexte, sie einzusetzen, Assoziationsräume, die sich um sie gebildet haben etc. Insofern die Praxis der westlichen Kunstmusik auf der Trennung zwischen Notation und Ausführung beruht, muss man hier auch die Notationssysteme und das Papier mitzählen, auf dem sie realisiert werden – bzw. die Computerprogramme, mit denen dies heute geschieht und die selbst einen nicht zu leugnenden Eigensinn haben. Der Umgang mit Instrumenten und Notation changiert zwischen den ersten drei Typen der Interaktion, die Ihde unterschieden hatte: Ziel des Lernens ist eine Form der Beherrschung, bei der das Instrument gleich einer Erweiterung des Körpers gar nicht mehr als solches hervortritt, sondern zum quasi widerstandslosen Ausdrucksmedium wird, also eine »embodiment relation«; am Anfang des Lernprozesses und teilweise immer wieder erscheint es als fremder Mitspieler, der die eigenen Intentionen durchkreuzt oder zumindest umlenkt, also im Sinne einer »alterity relation«; die Notation und ihre jeweilige materielle und technische Realisierung sind Medien des Schreibens und Lesens, zu denen man »hermeneutic relations« unterhält.

Dass »alterity relations«, in denen Instrumente und Technologien die Rolle des Gegenübers einnehmen, nur ein kleiner Teil aller Beziehungen zwischen Musiker*innen und Instrumenten ausmachen, sollte nicht den Eindruck erwecken, dass diese nicht in jedem Fall *mitreden*. Blindenstock, Schreibmaschine und Orgel, Merleau-Pontys Beispiele für Gegenstände, die durch den Erwerb von Gewohnheit zu Erweiterungen des Leibes werden, prägen das, was man mit ihnen tut, selbst dann auf entscheidende Weise mit, wenn für die Handelnde im Hintergrund verschwinden.[21] Die beste Beschreibung für diese Art des Mitredens bietet die Akteur-Netzwerk-Theorie mit ihrem Begriff des Aktanten, wobei Bruno Latours Beispiel der Schusswaffen immer noch besonders anschaulich ist: Zu sagen »Guns kill people« und, dem widersprechend, »People kill people, not guns« ist beides sowohl richtig als auch falsch. Entscheidend ist, dass ein*e Akteur*in mit einer Pistole nicht mehr einfach ein Mensch ist, der ein neutrales Werkzeug benutzt, sondern ein Hybrid aus Mensch und Schusswaffe, so dass die Handlung diesem Netzwerk *en miniature* zugerechnet werden muss.[22] Insofern muss man sagen, dass auch der Pinsel der Malerin, die Klarinette der Musikerin und das

21 Vgl. Maurice Merleau-Ponty, *Phänomenologie der Wahrnehmung*, Berlin 1966, S. 173ff.

22 Vgl. Bruno Latour, »Über technische Vermittlung: Philosophie, Soziologie und Genealogie«, in: Andrea Bellinger, David J. Krieger (Hg.): *ANThology.*

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Notationsprogramm der Komponistin als Aktanten an ihren Handlungen beteiligt und für das Ergebnis mitverantwortlich sind. Das aufgeführte Stück ist ein Produkt aus all dem, ein durch ein Netzwerk von Vermittlungen erzeugtes Ereignis, das niemand allein hervorbringt oder kontrolliert. Latour bietet allerdings keine grundlegenden Unterscheidungen bezüglich des Status der jeweiligen Aktanten an, der Rolle, die sie jeweils spielen (können). Die flache Ontologie der ANT erlaubt, Netzwerke aus sehr verschiedenen Akteuren denken zu können, aber man wird nicht umhinkommen, hier schließlich doch wieder Unterscheidungen einzuziehen. Vielleicht so: Man kann von ihr lernen, dass auch ein Werkzeug kein *bloßes* Werkzeug ist, aber nicht jeder Aktant ist ein Werkzeug.

Elektronische Instrumente sind zuerst einmal tatsächlich nichts anderes als sehr komplizierte Werkzeuge, auch wenn der Eigensinn mit der Kompliziertheit zunimmt. Hilfreich ist hier die Unterscheidung von trivialen und nicht-trivialen Maschinen, die Heinz v. Foerster in Anlehnung an Henry Gordon Rice vorgeschlagen hat.[23] Triviale Maschinen sind solche, deren Output aus ihrem Input eindeutig berechenbar ist, nicht-triviale solche, bei denen das nicht gilt, etwa Organismen. Elektronische Klangerzeuger sind triviale Maschinen, insofern sie unter (wenigstens im Prinzip) genau kontrollierten Bedingungen funktionieren; das gilt selbst noch dann, wenn sie nach bestimmten Regeln selbst Klänge erzeugen, ohne im Einzelnen dazu programmiert worden zu sein. Dass ihr Funktionieren nicht jederzeit von jedem nachvollzogen werden kann, ist dafür zuerst einmal unerheblich – das gilt auch für komplizierte mechanische Maschinen.

Ein interessantes Beispiel einer Zwischenform ist der No-Input-Mixer. Hier wird der Input eines Mischpults in einer Schleife mit dem Output verbunden, wodurch die Eigenströme und Interferenzen des Geräts verstärkt und durch die Manipulation der Regler verändert werden können, ohne dass von außen ein Signal hineinkommt. Die Kontrolle ist dabei nie vollständig, da bisweilen sehr kleine Veränderungen der Einstellungen den Klang auf drastische, unvorhersehbare Weise verändern können. Veraltete, nicht mehr zuverlässig funktionierende Technik ist ein ähnlicher Fall: Sie ist in den vergangenen Jahren für viele Musiker*innen noch einmal neu interessant geworden, weil sich bei ihr vielfach Fehler einschleichen, die die Kontrollierbarkeit beeinträchtigt und unerwartete Resultate erzeugt: Indem sie kaputtgeht, verliert die Maschine ihre Trivialität. Beide, No-Input-Mixer und kaputte analoge Synthesizer,

Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie, Bielefeld 2006, S. 483–528, hier 485f.

23 Vgl. etwa Heinz v. Foerster, »Entdecken oder Erfinden. Wie läßt sich Verstehen verstehen?«, in: Heinz Gumin u. Heinrich Meier (Hg.), *Einführung in den Konstruktivismus*, München 1992, S. 41–88.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

funktionieren nicht mehr als widerstandslose Erweiterungen des Körperschemas, sondern erscheinen bereits als Gegenüber, mit denen man umgehen muss, statt sie einfach nur anzuwenden oder einzusetzen. Dass sie »intelligent« wären, muss man dafür nicht behaupten.

Auch wenn all diese Kategorien, vom Material über Aktanten und triviale Maschinen, sich diesseits der KI halten, gibt es doch offensichtliche Kontinuitäten zu ihr. Vor allem aber treffen wir hier auf zahlreiche Mitspieler im künstlerischen Prozess, in deren Kontext die Frage der KI verhandelt werden muss.

Der Einsatz nicht-trivialer Maschinen folgt dabei offensichtlich einem Bedürfnis, das an der Widerständigkeit jeglichen künstlerischen Materials ansetzt und mit einem veränderten Verständnis künstlerischer Subjektivität insgesamt zu tun hat. Statt um die größtmögliche Kontrolle geht es nun darum, sich durch den Widerstand und den Einspruch nicht kontrollierbarer Instanzen gezielt beirren zu lassen, um Resultate hervorzubringen, die man selbst nicht hätte produzieren können und die hoffentlich über das hinausgehen, was das eigene Können erreichen kann, statt hinter es zurückzufallen.

Ein etwas anders gelagerter Fall ist der Einsatz von Algorithmen beim Komponieren, der weder auf technische Unterstützung angewiesen noch auf Trivialität oder Nicht-Trivialität festgelegt ist und eine jahrhundertelange Geschichte hat. Ihm werde ich mich im folgenden Abschnitt zuwenden.

3.

Im Gegensatz zu den anderen bisher diskutierten Aspekten und Beispielen wird das algorithmische Komponieren in zahlreichen Auseinandersetzungen mit Musik und KI als direkter Vorläufer genannt. Dabei impliziert die Arbeit mit Algorithmen nicht unbedingt den Einsatz einer Technologie, die komplizierter ist als diejenige von Papier, Stift und Notations- und Rechensystemen. Algorithmen sind zuerst einmal nichts anderes als Handlungsvorschriften, die klar definierten einzelnen Schritten folgen. Ob sie ein schlicht ablaufendes Programm ergeben, rekursiv gestaltet sind oder gar »lernen«, ist eine andere Frage. Voraussetzung ist in jedem Fall eine Formalisierung von Strukturen und Prozeduren, für die sich die Musik, genauer: die notierte europäische Musik besonders gut eignet.

Als frühes Beispiel wird immer wieder das »musikalische Würfelspiel« genannt, das oft mit Mozart assoziiert wird; tatsächlich aber war diese Form der automatisierten Komposition im 18. Jahrhundert eine verbreitete Praxis, von der eine einflussreiche Variante 1757 von Johann

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Philipp Kirnberger systematisch beschrieben wurde.[24] Es findet ganz auf dem Papier statt, unter Zuhilfenahme eines elementaren Zufallsgenerators wie Würfel oder Spielkarten. Ausgangspunkt ist die Festlegung eines Auswahlfeldes in Form einer Tabelle, aus dem dann in mehreren Durchgängen ein musikalischer Ablauf gebaut wird. Vorausgesetzt sind hier eine definierte Notation und, zumindest in dieser Version, die Existenz relativ standardisierter musikalischer Formen; die »Komposition« besteht in der Kombination vorgefertigter Versatzstücke. Trotzdem setzt die Technik als solche durchaus nicht voraus, dass es um Stilkopien gehen muss, wie David Cope in der oben zitierten Definition bezeichnenderweise meint.[25] So haben sich etwa John Cages Zufallsverfahren mit Münzenwürfen, die auf den 64 Zeichen des chinesischen Orakelbuchs I-Ging beruhen und als eine Variante der beschriebenen Techniken verstanden werden können, endgültig von jeder Art vorgegebener formaler Fixierung entfernt.

Michael Edwards sieht hier eine deutliche Trennung zwischen Forscher*innen und Komponist*innen: Während erstere die Musik als Mittel zur Entwicklung von lernenden Systemen verwenden, geht es letzteren um personalisierte Lösungen zur Produktion genuin *neuer* Musik.[26] Cope selbst ist Komponist und lässt sich insofern nicht klar in dieser Unterscheidung verorten; in unserem Kontext tritt er allerdings in Edwards' Sinne in erster Linie als Forscher auf. Vielleicht kann man sagen, dass Cage und Cope hier die Extreme des Möglichen verkörpern: auf der einen Seite die tour de force der *Music of Changes* von 1952, die jeden musikalischen Zusammenhang zerschneidet, auf der anderen 5000 Bachchoräle – sehr viel mehr vom Gleichen oder etwas vollkommen Anderes.

Quer dazu steht die Unterscheidung zwischen »indeterminate and determinate models«[27], die derjenigen zwischen trivialen und nicht-trivialen Maschinen entspricht und bei der sich Cage und Cope ebenfalls auf entgegengesetzten Seiten wiederfinden. Auf der Seite des Determinismus steht vor allem die streng serielle Musik der frühen 1950er Jahre, die die Reihentechnik der Zwölftonmusik zusätzlich zur Tonhöhe auf alle Parameter – Rhythmus, Dynamik, Klangfarbe – ausgeweitet hat

24 Vgl. Johann Philipp Kirnberger, *Der allezeit fertige Polonoisen- und Menuettencomponist*, Berlin 1757. Vgl dazu auch Leonard G. Ratner, »Ars Combinatoria: Chance and Choice in Eighteenth-Century Music«, in: H.C. Robbins Landon u. Roger E. Chapman (Hg.), *Studies in Eighteenth-Century Music. A Tribute to Karl Geiringer on his Seventieth Birthday*, New York 1979, S. 343-363.

25 Vgl. Cope, *Virtual Music*, a.a.O., S. 3.

26 Vgl. Michael Edwards, »Algorithmic Composition: Computational Thinking in Music«, in: *Communications of the ACM* 54, 7 (2011), S. 58–67, hier 63 (DOI:10.1145/1965724.1965742).

27 A.a.O., S. 62.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und in der seriellen Vorbereitung des Materials eine virtuelle Maschine konstruiert, die strenggenommen einfach gestartet werden müsste, um das Stück hervorzubringen (auch wenn niemand das je tatsächlich so praktiziert hat).[28] Bei Cage finden wir in jener Zeit die gleiche Parametrisierung und Vorbereitung, nur dass die tatsächliche Auswahl dem Zufall überlassen bleibt.

Auch wenn sie nicht in Form von materiell konstruierten Automaten realisiert sind, können diese algorithmischen Kompositionssysteme insofern als Maschinen verstanden werden, als sie die einzelnen Schritte definieren und zumindest die Gestalt des Prozesses automatisch abläuft. Auch wenn jeder Schritt von einem Menschen vollzogen werden muss, bleibt dessen Intentionalität dabei weitgehend oder vollkommen ausgeschaltet, indem er sich als Agent der Realisierung des Algorithmus oder des Zufalls erfährt. Auch Würfel, Karten etc. als Zufallsgeneratoren entziehen sich explizit der eingreifenden Intentionalität. Wie Joachim Heintz bemerkt, dachten Cage und Stockhausen »in vieler Hinsicht wie Programmierer«[29]. Der Einsatz des Zufalls erinnert daran, dass Nicht-Trivialität keinesfalls gleichbedeutend mit Intelligenz oder Intentionalität sein muss, ging es doch für Cage gerade darum, dieser zu entkommen.

Sol LeWitt und andere bildende Künstler hatten sich in den frühen sechziger Jahren mit der seriellen Musik und vor allem den Reflexionen über sie beschäftigt (die Hefte der einschlägigen Zeitschrift *die reihe* waren schon sehr bald ins Englische übersetzt worden),[30] und LeWitt bezieht sich genau auf den benannten Aspekt, wenn er in seinen »Paragraphen über konzeptuelle Kunst« schreibt, dass hier »alle Pläne und Entscheidungen im voraus erledigt werden und die Ausführung eine rein mechanische Angelegenheit ist« und den berühmten Satz folgen lässt: »Die Idee wird zu einer Maschine, die die Kunst macht.«[31] LeWitts »Idee« ist weniger auf strukturelle Komplexität angelegt als die Algorithmen der seriellen Musik, steht aber systematisch an genau derselben Stelle: beide sind virtuelle Maschinen.

Diesseits des Münzenwerfens ist die Frage der Determiniertheit (Trivialität klingt hier dann doch auf irreführende Weise abwertend) im

28 Vgl. dazu exemplarisch György Ligeti, »Entscheidung und Automatik in der *Structure Ia* von Pierre Boulez« (1958), in: ders., *Gesammelte Schriften*, Mainz 2007, S. 413–446.

29 Heintz, »Spielen statt Fürchtenmachen«, a.a.O., S. 29.

30 Vgl. Herbert Eimert u. Karlheinz Stockhausen, *die Reihe. A periodical devoted to the developments in contemporary music*, 6 Bände, Bryn Mawr u. Wien 1958–68, orig. 1955–62.

31 Sol LeWitt, »Paragraphs on Conceptual art / Paragraphen über konzeptuelle Kunst«, in: Gerd de Vries (Hg.), *Über Kunst. Künstlertexte zum veränderten Kunstverständnis nach 1965*, Köln 1974, S. 176–185, hier 177.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Einzelfall nicht so einfach zu beantworten, wie es vielleicht aussieht. Je komplizierter die Maschine, desto schwerer ist für einen Beobachter noch zu beurteilen, ob das Ergebnis im Prinzip hätte vorausgesehen werden können. Es mag wie eine Inkonsequenz wirken, diesen Fall hervorzuheben, aber er macht auf etwas aufmerksam, das in der Interaktion mit KI, sei sie stark oder schwach oder eigentlich gar nicht wirklich »intelligent«, eine wichtige Rolle spielt, nämlich die Frage, wie die jeweiligen Prozesse oder Aktanten den Interaktionspartnern *erscheinen* und wie sie infolgedessen mit ihnen umgehen.

Zusätzlich zu dieser kann man zwei weitere fundamentale Unterscheidungen machen, nämlich diejenige, ob die Ergebnisse des virtuell-maschinellen Prozesses das Stück definieren oder ob sie Material für nachfolgende kompositorische Entscheidungen sind, und als was die Maschine jeweils beschrieben oder erfahren bzw. welche Rolle ihr zugedacht wird. Zum ersten: Niemand hat je wirklich das Ergebnis der algorithmischen oder Zufallsprozesse ohne jeden Eingriff übernommen. Boulez' Klavierstück *Structures Ia* von 1952, das als Paradebeispiel für einen vollendeten Automatismus gilt, kam dem nahe, wurde aber vom Komponisten selbst eher als Versuch angesehen, »tabula rasa [zu] machen mit dem Erbe und beim Nullpunkt wieder [zu] beginnen«[32], denn als haltbare Position. Cage beschreibt sein Ziel bei *Music of Changes* aus demselben Jahr sehr ähnlich, nämlich »to make a musical composition the continuity of which is free of individual taste and memory (psychology) and also of the literature and ›traditions‹ of the art«[33], nur dass er bei genau dieser Position bleiben sollte. Die Resultate der verschiedenen Zufallsverfahren, die er im Laufe seines Lebens verwendet hat, sollten weitgehend unangetastet bleiben, damit sich am Ende nicht doch wieder sein persönlicher Geschmack durchsetzt.

Der spätere Boulez und auch Iannis Xenakis, der mit ganz anderen, nämlich stochastischen Verfahren komponierte, haben viel selbstverständlicher eingegriffen, ohne dass man sagen könnte, dass das halbautomatisch Generierte bloßes Material kompositorischer Entscheidungen geworden wäre. Die Maschine ist Teil des Kompositionsprozesses, ersetzt ihn aber nicht. Im Übrigen ist natürlich bei Cage wie bei allen anderen das Bauen seiner virtuellen Maschinen ein intentionaler Kompositionsakt, und insofern könnte man sagen, dass Komposition verlagert oder verteilt, aber in keinem Fall von nichtmenschlichen Akteuren übernommen wird.

32 Pierre Boulez, *Wille und Zufall. Gespräch mit Célestin Deliège und Hans Mayer*, Stuttgart und Zürich 1977, S. 62.

33 John Cage, »To describe the process of composition used in *Music of Changes* and *Imaginary Landscape No. 4*«, in: ders., *Silence. Lectures and Writings*, Middletown 1961, S. 57–59, hier 59.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Interessant ist hier tatsächlich, mit welchen Begriffen die jeweiligen Techniken beschrieben worden sind. Im seriellen Diskurs stehen Rationalität und Konsequenz im Mittelpunkt bzw., noch einmal in Boulez' Worten, »Strenge, die sich von Schwäche oder Kompromißlerei frei weiß«[34]. Die Maschine ist hier Instrument dieser kompromisslosen Konsequenz bzw. deren reinste Verkörperung. Für Cage ist das Setzen auf – paradoxerweise aufwendig konstruierte – Zufallsverfahren gleichbedeutend mit einer Öffnung für »Natur«, worunter er eine vom menschlichen Eingriff und menschlicher Intentionalität freie Sphäre verstand: Er baut Maschinen, die die Türen dafür öffnen.[35] Xenakis, der beide Ansätze harsch als »auditory and ideological nonsense«[36] bzw. »denying music«[37] kritisierte, ging es um die Fortsetzung einer großen Tradition, nämlich »the placing of sonic art on a more universal plane«[38], ohne sie menschlicher Intentionalität ganz zu entziehen. Die stochastischen Verfahren, mit denen das Verhalten von Klangmassen kalkuliert werden konnten, sollten sowohl näher an tatsächlichen Vorgängen der Natur als auch der menschlichen Auffassung gemäßer sein. Von Kirnbergers harmlosem Angebot, auch Nichtkomponisten die Produktion von Musikstücken zu erlauben, ist das alles weit entfernt.

Kurz nach den erwähnten Stücken von Boulez und Cage und einige Jahre vor Xenakis programmierten der Komponist Lejaren Hiller und der Mathematiker Leonard Isaacson ein Computersystem, mit dem sie 1956 ein viersätziges Stück für Streichquartett, die *Illiac Suite*, produzierten – das erste von einer realen, nicht virtuellen Maschine komponierte Stück. Es wird von den beiden allerdings konsequent als Projekt oder Experiment beschrieben, als eine Art Studie zu der Frage, was Computer zur Analyse und Komposition von Musik beitragen können, was für Typen von Musik dabei hervorgehen können etc. Hiller und Isaacson verstehen sich hier klarerweise als Forscher und nicht als Künstler, aber ihr musikalisches Verständnis und ihre zeitgenössische Sensibilität verhindern, dass sie sich auf bloße Stilimitationen beschränken.[39]

Wozu es in keinem Fall kam, war eine Personalisierung der virtuellen oder realen Maschinen. Die Prozeduren verkörpern Rationalität oder

34 Pierre Boulez, »Schönberg ist tot«, in: ders., *Anhaltspunkte. Essays*, Kassel u. München 1979, S. 288–296, hier 296.

35 Vgl. ausführlich und kritisch zu diesem Naturverständnis Benjamin Piekut, »Chance and Certainty – John Cage's Politics of Nature«, in: *Cultural Critique* 84 (2013), 134–163

36 Iannis Xenakis, *Formalized Music. Thought and Mathematics in Composition*, Stuyvesant, NJ 1992, S. 8.

37 A.a.O., S. 181.

38 A.a.O., S. 178.

39 Vgl. Lejaren A. Hiller u. Leonard M. Isaacson, *Experimental Music. Composition with an Electronic Computer,* New York 1959.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

erlauben eine Öffnung, durch die Naturprozesse sich zur Geltung bringen können, aber sie verdichten sich nicht zu Wesen. Auch Illiac ist lediglich die Abkürzung für *Illinois Automatic Computer*, ein denkbar unemphatischer Name, der eher keinen eigenständigen Akteur benennt, und anders als Cope mit *Emmy* verzichten Hiller und Isaacson auf einen personifizierenden Kosenamen. Für einen Satz von Regeln, sei er computerbasiert oder nicht, legt sich das auch eher nicht nahe.

4.

Mit dem Einsatz von neuronalen Netzwerken ändert sich dies insofern, als durch die Umstellung von regelbasierten Expertensystemen auf mustererkennende Systeme eine Art Lernprozesse entsteht, die nicht mehr direkt von den Entscheidungen menschlicher Programmierer*innen abhängig sind. Diese kommen natürlich weiterhin durch die Programmierung des Netzwerks selbst und vor allem durch die Auswahl der Daten ins Spiel, mit dem es gefüttert wird; seine Entwicklung verläuft dann aber quasi autonom, so dass es sich in gewisser Weise über seine Geschichte individualisiert, ohne damit automatisch zu einem »Subjekt« zu werden.

Bedeutsam ist überdies, dass diese Lernprozesse nicht mehr symbolisch vermittelt ablaufen müssen, sondern unter Umständen direkt an akustischen Mustern ansetzen, die sich diesseits dessen halten können, was wir traditionellerweise als musikalische Formen verstehen würden. Diese Verschiebung entspricht derjenigen, die Friedrich Kittler zwischen Notationssystemen und Aufnahmetechniken beschrieben hat: »Manipulierbar wird statt dem Symbolischen das Reale.«[40] Jenes »Reale« der Klanglichkeit wird natürlich immer noch codiert, aber nach messbaren Eigenschaften und eben nicht nach symbolischen Formen. Die Fremdheit der möglichen Ergebnisse ist noch um einiges größer als bei den Cageschen Zufallsverfahren, die zwar auch Resultate produzieren mögen, die für ein auf formale Kohärenz setzendes Musikverständnis unsinnig sind, ohne dass dieser »Unsinn« aber den durch Tonhöhe und Tondauer codierten Raum verlassen würde. Ein neuronales Netzwerk muss sich demgegenüber von außen der Frage annähern, was überhaupt Sinn sein könnte, und scheint dabei jene grundsätzliche andere, fremde Vorgehensweise zu verwenden, die in verschiedenen Kontexten für Faszination sorgt und immer wieder als »alien«, also außerirdisch oder ganz und gar fremd beschrieben wurde.[41]

40 Friedrich Kittler, *Grammophon Film Typewriter*, Berlin 1986, S. 57. Die Lacanschen Anklänge sind offensichtlich, aber allzu unscharf.

41 Einer dieser Bereiche ist das Spiel Go: Die Beschreibungen von Experten, die die Züge der KI AlphaGo analysierten, »often seem to keep circling back to

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Anders als ein Teil des öffentlichen Diskurses suggeriert, geht es wie in meinen beiden Eingangsbeispielen auch hier in der Regel nicht um eine Technologie, die auf Knopfdruck fertige Musik ausspuckt; so etwas wäre »entirely useless for musicians and composers«[42], wie Philippe Esling formuliert, der am IRCAM in Paris die »Artificial Creative Intelligence and Data Science«-Forschungsgruppe leitet. Stattdessen werden die entsprechenden Programme und Netzwerke als Werkzeuge beschrieben, die den Komponist*innen möglicherweise Arbeit abnehmen oder, was die interessantere Variante ist, ihnen Material liefern, auf das sie sonst niemals gekommen wären.

Insofern es nicht lediglich darum geht, lästige Tätigkeiten an sehr hoch entwickelte Werkzeuge zu delegieren, ist die Beschreibung der KI als Werkzeug aber auch hier etwas schief. Eine Technologie, die inhaltlich bestimmtes Material liefert, an das sich zumindest die Frage stellen lässt, ob es nicht selbst schon künstlerischen Charakter hat, ist eher ein Mitspieler als ein Werkzeug. Eine Komponistin, die mit solchen Mitspielern umgeht, verhält sich zu ihnen wie eine Lektorin oder Kuratorin, also deutlich im Sinne einer »alterity relation«. Todd Lubart faltet die hier möglichen Verhältnisse noch weiter auf, indem er die Rollen des Computers als Kindermädchen, als Brieffreund, als Trainer und als Kollege unterscheidet.[43]

Vielleicht kann man die Sache aber auch noch einmal anders verstehen: Der Komponist Sam Salem berichtet, das er die KI-generierten Klanggestaltungen wie Field Recordings behandelt, mit denen er ansonsten viel gearbeitet hat.[44] Das so erzeugte Material wäre dann sicher nicht wie die intentionale Äußerung eines Mitspielers, im Extremfall einer virtuellen Person anzusehen, sondern als klangliche Umgebung, die eine bestimmte Atmosphäre hat und an deren Folge von Klangereignissen kein Anspruch von Kohärenz gestellt wird. Wenn sie als solche stehengelassen würde, würde sie kaum als fertiges Musikstück erscheinen; wenn sie kompositorisch verarbeitet wird, werden Momente, vielleicht Atmosphären erhalten bleiben und in einen neuen Zusammenhang eingehen.

the same several words: Amazing. Strange. Alien.« (Dawn Chan, »The AI That Has Nothing To Learn From Humans«, in: *The Atlantic*, 20. Oktober 2017, https://www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught-itself-go/543450/).

42 Javier Nistal, »Deep Beers. A Chat with Philippe Esling«, in: *New Frontiers in Music Information Processing*, 4.1.2020, https://mip-frontiers.eu/2020/01/04/Philippe_Esling.html.

43 Todd Lubart, »How can computers be partners in the creative process: Classification and commentary on the Special Issue«, in: *International Journal of Human-Computer Studies* 63 (2005), S. 365–369.

44 Aufnahmen und (sparsame) Informationen zu Salems Stücken finden sich hier: http://www.osamahsalem.co.uk.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Es gibt eine weitere Eigenschaft, die Field Recordings mit KI-generierten Klängen gemeinsam haben und die auch für Salem einer Vorstellung von künstlichen Intelligenzen als autonomen Schöpfern entgegensteht: Sie haben weder ein natürliches Anfang noch ein Ende, sondern bilden einen kontinuierlichen Strom. Der minimalste Eingriff ist insofern das Setzen von Schnitten und damit das Schaffen von Ganzheiten im elementaren aristotelischen Sinne als etwas, das »Anfang, Mitte und Ende hat«[45]. Dies bleibt menschlichen Akteuren vorbehalten. Ähnliches gilt für die potentiell unendliche Menge an Anläufen, in denen Material generiert werden kann. Laurie Spiegel, eine der Pionierinnen algorithmischer Komposition, schreibt dazu: »The overabundance of musical material that algorithmic generation can produce somehow seems to cheapen the musical result […].«[46] Die Auswahl und der Zuschnitt von bestimmtem Material durch eine menschliche Akteurin muss in diesem Sinne künstlerische Plausibilität erzeugen (oder im Überfluss finden), die sich gegen die Bedrohung der Beliebigkeit behaupten kann.

Diese beiden Eigenheiten lassen bei aller Individualisierung eine Personalisierung nicht gerade naheliegend erscheinen, aber sie findet sich trotzdem durchaus. Ein Beispiel dafür ist Holly Herndon mit *Spawn*, die sich allerdings nicht im Bereich der Neuen Musik, sondern in der Popmusik verorten, auch wenn Herndon einen akademischen Hintergrund in der Komposition elektronischer Musik hat. Bisher sichtbarstes Ergebnis dieser KI ist das Album *Proto* von 2019, auf dem *Spawn* in verschiedener Weise auftaucht, als Dimension in ansonsten menschlich hervorgebrachten Stücken, als wesentliches Moment von Stücken und in Reinform; zusätzlich gibt es Aufnahmen von Trainingssessions mit einem Chor. Das Album versteht sich so einerseits als künstlerisches Statement eigenen Rechts, andererseits aber auch als Momentaufnahme in einem Prozess, und zwar ganz buchstäblich einem Trainingsprozess der KI und gleichzeitig einem gesellschaftlichen Prozess, in dem künstliche Intelligenzen eine zunehmend größere Rolle spielen.

Ein großer Teil von Herndons Praxis beruht darauf, sich den technischen, ökonomischen und politischen Entwicklungen der Gegenwart offensiv zu stellen, sie zu verwenden und dabei gleichzeitig kritisch zu reflektieren. Von daher ist auch die offensive Personalisierung von *Spawn* zu verstehen, das konsequent als (weibliches) Kind beschrieben wird, das Herndon und ihr Partner Mathew Dryhurst gemeinsam aufziehen: »She is being raised by listening to and learning from her parents, and those people close to us who come through our home or participate at our

45 Aristoteles, *Poetik*, griechisch/deutsch, Stuttgart 1994, 25 (1450b).

46 Laurie Spiegel, »Thoughts on Composing with Algorithms«, in: Alex McLean u. Roger Dean (Hg.), *The Oxford Handbook of Algorithmic Music*, Oxford 2018, S. 105–111, hier 109.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

performances.«[47] Es bleibt offen, inwiefern Herndon und Dryhurst wirklich davon ausgehen, dass künstliche Intelligenzen als eine Art neue Spezies anzusehen sind, mit der wir real interagieren, oder ob das eine strategisch eingesetzte Metapher ist, anhand derer bestimmte Dinge deutlich gemacht werden können. In einem Interview geht Herndon in letztere Richtung: »The predominant metaphor here is about the child – the raising of this future intelligence as a community.«[48] Vielleicht könnte man es so formulieren: Es geht hier darum, »background relations« in »alterity relations« zu verwandeln und so Prozesse, wie sie bereits an vielen Stellen im Hintergrund ablaufen, sichtbar und zu einem gewissen Grade verfügbar zu machen und sie so einer Befragung unterziehen zu können.

Wenn man sich mit Menschen unterhält, die mit KI arbeiten, hört man immer wieder die Einschätzung, dass künstliche Intelligenzen »dumm« sind – gemeint ist, dass sie bisweilen Dinge produzieren, die für unsere Begriffe vollkommen an der Sache vorbeigehen, was auch immer diese Sache jeweils sein mag.[49] Ganz formal ist ein Verfahren, das rein auf der Erkennung von Mustern beruht, tatsächlich eine andere, fremde Form der Intelligenz, die nichts »versteht«, weil sie nicht von den generativen Prinzipien hinter diesen Mustern weiß und sich auch gar nicht um diese kümmert. Wenn sie selbst zu produzieren beginnt, entwickelt sie eine eigene Form der Generativität, die irgendwann nicht mehr einfach als defizitäre Rekonstruktion von außen verstanden werden kann. Die Umcodierung der spezifischen Dummheit dieser Generativität in die frühen Entwicklungsschritte eines Kindes leistet mehreres: Sie nimmt ihr den bedrohlichen, weil unverständlichen Charakter und weist darauf hin, dass es einen Unterschied mache, was für Mittel und Materialien in der »Erziehung« dieses Wesens eingesetzt werden und wer daran beteiligt ist.

Die Fremdheit kann dann als faszinierende Ergänzung menschlicher Kreativität aufgefasst werden, die sowohl im Training bzw. der Erziehung als auch als kulturelle Größe in kommunikative und deliberative Strukturen eingebunden sein sollte. So verstehe ich es, wenn Herndon von »an emphasis on alien song craft and new forms of communion«[50] spricht. Das Motiv des *alien* bringt auch hier die Vorstellung einer fremden, nicht-menschlichen Logik und Ästhetik ins Spiel, die eine Bereiche-

47 https://www.hollyherndon.com/godmother-holly-herndon-and-jlin-feat-spawn.

48 Emily Friedlander, »How Holly Herndon and her AI baby spawned a new kind of folk music«, in: *The Fader*, 21.5.2019, https://www.thefader.com/2019/05/21/holly-herndon-proto-ai-spawn-interview.

49 So auch Herndon selbst: »AI is not that smart, it's very low fidelity, it's not real time, it is very slow and unwieldy.« (Emily McDermott, »Holly Herndon on her AI Baby, Reanimating Tupac, and Extracting Voices«, in: *Art in America*, 7. Januar 2020, https://www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301/).

50 https://www.hollyherndon.com/proto.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

rung verspricht, die über bloße produktive Störung und Herausforderung weit hinausgeht. Dass die Bedrohlichkeit trotzdem eine reale Möglichkeit ist – man denke an Ridley Scott –, wird dabei nicht verheimlicht. In die Metapher der Kindererziehung übersetzt: »It is important to be cautious that we are not raising a monster.«[51] Die tatsächlichen Monster werden allerdings von den großen Unternehmen und von Regierungen herangezogen und sind keine zähnefletschenden Bestien, sondern Überwachungs-, Kontroll- und Enteignungsmaschinen.[52]

Ein weiteres Projekt, wiederum aus der Popmusik, ist um einiges spielerischer, aber trotzdem weit davon entfernt, bloßen Nonsens zu produzieren. Der Name, unter dem C.J. Carr und Zack Zukowski ihre KI-generierte Musik veröffentlichen, Dadabots, ist natürlich aufschlussreich – aber auch Dada war bei aller Sinnsubversion eine sehr ernst gemeinte Angelegenheit. Was die Dadabots tun, bringt der Titel eines wissenschaftlichen Artikels auf den Punkt, den sie zu ihrer Methode verfasst haben: »Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands«.[53] Seitdem sind noch andere Genres hinzubekommen, etwa Pop oder Freejazz, und auch Deepfakes, bei denen Stücke einer Band durch die Maske einer anderen gecovert werden. Das Datenmaterial besteht oft nur aus der Musik einer Band, manchmal sogar nur eines einzigen Albums.[54]

Das meines Erachtens eindringlichste Beispiel ist *Relentless Doppelganger*, ein neuronales Netzwerk, das einen endlosen Strom von Death Metal hervorbringt, den man sich 24 Stunden am Tag auf YouTube anhören kann.[55] Ausgangspunkt dafür ist die Band Archspire. Anders als in manchen anderen Fällen haben die Dadabots hier nicht kuratorisch eingegriffen, sondern man hört das ungefilterte, laut Esling vollkommen nutzlose Output der KI. Dieses ist in zweierlei Hinsicht interessant: Erstens schafft es, den Klang und auch die Abläufe der Musik dieser Band erstaunlich gut nachzubilden, lässt sie aber eigenartig abstrakt erscheinen. Es weist zwar Muster auf, aber keine wirkliche Struktur, und diese Art Musik erscheint so bestimmt durch eine bestimmte Soundwelt, einen Grundgestus und einen Verkettungsmodus – was Death Metal vielleicht tatsächlich auf den Punkt bringt. Der Mangel an Struktur verweist auf den zweiten Punkt, der bereits bei Salem Thema war: Statt mit Stü-

51 Ebd.

52 Herndon hat diese Auseinandersetzung fortgesetzt mit dem *Holly+* betitelten, KI-gestützten Projekt, bei dem es um die lizensierte und legitimierte Erstellung von Deepfakes ihrer eigenen Stimme geht (vgl. https://holly.mirror.xyz). Die extreme Selbstbezogenheit und -inszenierung scheint unvermeidlicher Teil ihres Programms zu sein.

53 Vgl. C.J Carr u. Zack Zukowski, »Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands«, https://arxiv.org/abs/1811.06633.

54 https://dadabots.com.

55 https://www.youtube.com/watch?v=MwtVkPKx3RA.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

cken in einem auch nur minimalen Sinne von etwas, das einen Anfang und ein Ende hat, haben wir es tatsächlich mit einem endlosen Strom zu tun, also genau mit dem, was die KI natürlicherweise hervorbringt. Ihn in Reinform angeboten zu bekommen, verändert die Wahrnehmung deutlich: Man hat weniger das Gefühl, ein endloses Stück zu hören, als einer unendlich brabbelnden Lebensform zuzuhören.

Auf ihrer Webseite schreiben die Dadabots denkbar knapp und mit der gebotenen Ironie: »We are delighted by the unique characteristic artifacts of neural synthesis.«[56] Die Charakteristika dieser Artefakte reichen von offensichtlichem Unsinn, bei dem man versucht ist zu sagen, dass die KI die Sache nicht »verstanden« hat, bis zu treffenden Realabstraktionen, die sie durch Verfremdung auf den Punkt bringen. Die Dadabots kultivieren genau diesen Zwischenbereich zwischen Fremdheit und Nähe, in dem eine attraktive Seltsamkeit und eine seltsame Attraktivität entstehen, die teils gerade durch ihre Fremdheit interessant sind, teils durch den Aufschluss, den sie über die menschliche musikalische Praxis erlauben.

5.

Als letztes möchte ich zu dem Typ Praxis kommen, bei dem es am unmittelbarsten um eine Begegnung mit KI geht: der musikalischen Interaktion mit einem nicht-menschlichen Partner in Echtzeit. Hier ist eine »alterity relation« vorausgesetzt, und hier liegen die Zuschreibung von Individualität und Intentionalität am nächsten. Eine technisch realisierte Instanz, mit der man als Musiker*in improvisieren kann, die weder stereotype noch vollkommen beliebige Klänge produziert und scheinbar oder tatsächlich auf einen reagiert, wird fast unvermeidlich als realer Partner wahrgenommen, oder vielleicht so: Sie *ist* ein realer Partner, wie auch immer die Frage nach ihrer Intentionalität philosophisch beantwortet werden wird. Eine Äußerung von George Lewis, einem der Pioniere auf diesem Gebiet, scheint mir hier die auch philosophisch produktivste Position auf den Punkt zu bringen: »In the context of interactive computer music, I am not proposing an exploration of the metaphysics of machine consciousness but of a phenomenology of freedom as dialogic interaction, in which a creative machine can be perceived by us as acting in a free way.«[57] Der Kontext macht deutlich, dass die Charakterisierung der Maschine als kreativ in einem möglichst nüchternen Sinne verstanden werden sollte: eine Maschine, die Klänge und Klangstrukturen her-

56 http://dadabots.com/#new-albums.

57 George E. Lewis, »Listening for Freedom with Arnold Davidson«, in: *Critical Inquiry* 45, 2 (2019), S. 434–447, hier 445.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

vorbringt, die nicht einfach einem vorher festgelegten Programm folgen. Das tun ausnahmslos alle der hier zu behandelnden Beispiele; täten sie es nicht, würden sie von vornherein als Dialogpartner versagen. Sie treten auch dann als nicht-triviale Maschinen auf, wenn ihre Hervorbringungen im Prinzip, mit einer genauen Kenntnis der Programmierung, der jeweils gewählten Einstellungen und des benutzten Klangmaterials, voraussehbar sein müssten – denn faktisch bleibt dies unmöglich.[58]

Auf der Webseite des am IRCAM entwickelten Programms OMax, auf dessen Basis derzeit viele Musiker*innen arbeiten, ist deutlich schwächer vom »flavor of a machine co-improvisation«[59] die Rede. Das ist offensichtlich aus der Perspektive der Entwickler*innen gesprochen, die von der technischen Seite einer Programmumgebung ausgehen, die sie selbst erstellt haben. Aus der Perspektive von Musiker*innen, die mit dem Programm interagieren, ist der Eindruck gemeinsamer Improvisation aber kein Geschmack, der einer in Wirklichkeit ganz anders verfassten Situation beigegeben ist, sondern ihre Art, ganz real mit dieser Situation umzugehen. Daniel Dennetts Unterscheidung eines »design stance« von einem »intentional stance« ist hier hilfreich: Während ersterer sich auf das Funktionieren eines Ablaufs richtet, behandelt letzterer das Gegenüber als intentionalen Agenten (sein Beispiel ist hier ein Schachcomputer). Das impliziert Dennett zufolge *nicht*, »that Intentional systems really have beliefs and desires, but that one can explain and predict their behavior by ascribing beliefs and desires to them«[60].

Schach ist insofern etwas anders gelagert als musikalische Interaktion, als eine Instanz, die regelkonforme Züge vollzieht, in jedem Fall tatsächlich Schach spielt, wie gut oder schlecht auch immer. Die Interaktion, die in nichts anderem als einer alternierenden Folge regelkonformer Züge besteht, ist allein damit gesichert, und insofern die Züge der Maschine irgendeine Form der Kohärenz aufweisen, wird man ihr unvermeidlich mit dem »intentional stance« begegnen. Im Fall der improvisierten Musik ist die Interaktion offener, weil es keine klar vorgeschriebenen Regeln gibt, dadurch aber auch prekärer, weil sie durch nichts garantiert ist. Von daher muss jemand, der mit einem Programm oder eine KI improvisieren will, dem Interaktionspartner einen Vorschuss geben und dessen

58 Aus der Perspektive des Programmierers begreift sich Lewis dennoch als Autor von *Voyager* und hat es einmal als »one of my most widely performed compositions« bezeichnet (George E. Lewis, »Too Many Notes: Computers, Complexity, and Culture in Voyager«, in: *Leonardo Music Journal* 10 (2000), S. 33–39, hier 33); ganz ähnlich beschreibt Laurie Spiegel ihr Programm *Music Mouse* als »musical work«, das nicht selbst Musik ist (Spiegel, »Thoughts on Composing with Algorithms«, a.a.O., S. 110).

59 http://repmus.ircam.fr/omax/home.

60 Daniel C. Dennett, »Intentional Systems«, in: *The Journal of Philosophy* 68, 4 (1971), S. 87–106, hier 91.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

klangliche Äußerungen vor jeder Evidenz als intentional betrachten, damit überhaupt eine Interaktion in Gang kommt – schließlich *will* man mit dieser Maschine spielen. Der Vorschuss wird bestätigt, wenn diese Äußerungen sich im Rahmen des nach menschlichem Maßstab Erwartbaren halten oder auf interessante oder herausfordernde Weise fremd, *alien* erscheinen; er braucht sich auf, wenn sie sich dauerhaft in Imitation oder Repetition erschöpfen oder jeglichen Zusammenhang vermissen lassen. Der Spielraum ist dabei aufgrund der Offenheit der Form denkbar groß.

Mit der Evokation von Freiheit, die als dialogische Interaktion beschrieben wird, führt Lewis allerdings noch weitere Bestimmungen ein, die sich nicht von selbst verstehen, sondern auf das spezifische Verständnis gegründet sind, aus dem heraus er selbst sein Programm *Voyager* und dessen Vorläufer entwickelt hat. Grundlage ist die Beschreibung einer »Afrological tradition« der Improvisation, die auf Individualität, einen Rückbezug zu spezifischen Gemeinschaften und zur Geschichte betont und die von einer weißen »Eurological tradition« marginalisiert wurde, die nicht deutlich genug machen kann, dass sie nichts mit Jazz zu tun hat.[61] Von hier aus beschreibt er auch *Voyager* als Versuch, »a non-Eurocentric computer music«[62] zu ermöglichen.

Das Programm, das bereits in den achtziger Jahren entstand und seitdem kontinuierlich weiterentwickelt wurde, reagiert tatsächlich in Echtzeit auf das von anderen Gespielte, dessen Dichte, Dynamik, Geschwindigkeit etc., indem es sich mimetisch, kontrastierend oder indifferent dazu verhält, ganz wie es eine menschliche Improvisatorin tun würde. Das ist aber nur ein Teil der Sache: Anders als die meisten anderen Improvisationsprogramme, die später entstanden, kann *Voyager* auch von sich aus Klänge produzieren. Wie Harry Lehmann festhält, konnte das bereits mit den aus heutiger Sicht geradezu primitiven Mitteln der 1980er Jahre nur funktionieren, weil die improvisatorische Tradition, in der sich Lewis bewegt (die freie Improvisation, die Derek Bailey als »nicht-idiomatisch« bezeichnet hat)[63], nicht an tonaler Harmonik und Motivik orientiert ist; Lehmann spricht von einem »extrem fehlerrobusten Improvisationsstil«[64].

Auch wenn *Voyager* sehr flexibel und variationsreich agiert, soll er (sie? es?) eine eigene *Stimme* haben: »My efforts to imbue my interactive

61 Vgl. George E. Lewis, »Improvised Music after 1950: Afrological and Eurological Perspectives«, in: Black Music Research Journal 16, 1 (1996), S. 91–122.

62 George E. Lewis, »Interacting with Latter-Day Musical Automata«, in: Contemporary Music Review 18,3 (1999), S. 99–112, hier 108.

63 Vgl. Derek Bailey, *Improvisation. Its Nature and Practice in Music*, Ashbourne 1980.

64 Harry Lehmann, »Voyager-Ästhetik. George Lewis improvisiert mit Computern«, in: *Neue Zeitschrift für Musik* 4 (2020), S. 36–39, hier 37

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

systems with values such as relative autonomy, integral subjectivity, computer individualism, and musical uniqueness rather than repeatability […].«[65] Die Unsicherheit in Bezug auf das Personalpronomen ist kein Zufall: Natürlich haben Computerprogramme und künstliche Intelligenzen kein Geschlecht, auch wenn Personalisierungen wie die von Cope und Herndon dies suggerieren. Lewis gibt aber seinem Programm einen Titel, der weniger ein Name als eine Bezeichnung ist: Voyager ist bekanntlich eine Raumsonde, die mit ausgewählten wissenschaftlichen, künstlerischen und kulturellen Zeugnissen der Menschheit ins All geschickt wurde, um diese so eventuellen Außerirdischen vorzustellen. Man kann sich den Enthusiasmus, mit dem das damals betrieben wurde, kaum noch vorstellen, und auch die enthaltenen Bilder und Aufnahmen wirken heute teilweise bizarr. Das Motiv des Reisenden zwischen den Welten wird von Lewis cum grano salis aufgegriffen und in Anspruch genommen, und man kann sagen, dass *Voyager individualisiert* ist, ohne damit gleichzeitig *personalisiert* zu sein.

Vom interstellaren Reisenden ist es nur ein kleiner Schritt zum Motiv des *alien*, dem wir bereits bei Herndon begegnet sind. *Native Alien*, ein von Sandeep Bhagwati entwickeltes Programm auf der Basis von OMax, das seit Jahren in Kooperation mit zahlreichen Musikern weiterentwickelt wird, trägt dieses Motiv bereits im Titel. Auch *Native Alien* greift die Klänge von live improvisierenden Musiker*innen auf, arbeitet mit ihnen, produziert unter Umständen fast symphonische Klangstrukturen und erzeugt gleichzeitig eine teils graphische, teils konventionell notierte Partitur des bereits Gespielten, die von den Performer*innen als Anregung für ihr weiteres Spiel verwendet werden kann. Dabei wird das Programm selbst noch live nach einigen Parametern gelenkt, ohne im Einzelnen kontrolliert zu werden. Auf der Webseite wird es als eine Art Übermusiker beschrieben, nämlich als »musical partner with extended sensibilities and musical superpowers«[66]. Das Programm ist aber nicht einfach besser, sondern vor allem fremd, was als die eigentlich interessante Qualität benannt wird: »›Native Alien‹ creates a mediatized, changed world that takes a musician's own tradition, her/his musical ›home‹ – and transforms it into a musically alien world.«[67] Tatsächlich changiert das Programm so zwischen einem Partner und einer Umgebung, in der die Musikerin sich bewegt und die zwar reagiert, aber auch den Rahmen vorgibt, innerhalb dessen sie agieren kann.

Bhagwati erwähnt das Moment der verfremdeten Spiegelung des eigenen Tuns durch das Programm, das auch Henning Berg im

65 Lewis, »Listening for Freedom with Arnold Davidson«, a.a.O., S. 441.
66 http://matralab.hexagram.ca/research/native-alien/about/project-description/.
67 Ebd.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Zusammenhang mit dem von ihm entwickelten Programm TANGO bemerkt, wenn er von »eine[r] merkwürdige[n] Reflexion meines eigenen Spiels«[68] spricht. Diese merkwürdige Reflexion, in der man sich wie durch ein fremdes Medium oder im Blick einer fremden Intelligenz wiedererkennt, eröffnet die Möglichkeit der Selbstreflexion, die sich nicht nur auf die eigene individuelle Person bezieht, sondern auch auf die eigene Praxis, die Art des Spiels, die nie nur individuell ist. Hierauf ist auch der Titel *Native Alien* zu beziehen, der weniger der Name der Musik machenden technischen Instanz ist als das Prinzip, dem das Ganze folgt und das sich gut mit Bernhard Waldenfels' Motiv des Fremden zusammenbringen lässt (das im Übrigen als »alien« ins Englische übersetzt wurde).[69] Waldenfels' zentraler Punkt ist die Asymmetrie zwischen Eigenem und Fremdem in einer Relation, die nicht einfach umgekehrt werden kann und für keinen Beteiligten als reziprok wahrgenommen wird.

Wenn auch mit dieser Konstellation die Aufhebung beider in einem größeren Ganzen ausgeschlossen erscheint, ist damit keine harte Grenze des Verstehens und der Verständigung gesetzt, denn das Eigene ist nie ganz eigen und das Fremde nie ganz fremd. Die Verfremdung durch *Native Alien* wäre dann kein Einbruch des ganz Anderen in eine ungebrochene Heimat, sondern ein Explizitwerden der Konstellation von Beheimatet- und Fremdsein, die nicht erst dann beginnt, wenn Fremde vor der eigenen Haustür auftauchen oder man selbst sein Land verlässt, sondern die Bedingung unseres Aufenthalts auf der Welt ist – eine »Verschränkung von Heimwelt und Fremdwelt«[70]. Während es Lewis also um Formen der freien Interaktion und Gemeinschaftsbildung geht, an der auch Fremde teilnehmen können, stellt Bhagwati die Verbindlichkeit gemeinsamer Artikulation zwischen Eigenem und Fremdem ins Zentrum. Man könnte dies als zwei unterschiedliche Perspektiven auf vergleichbare Modelle von Gesellschaft interpretieren, aber der jeweils verschiedene Hintergrund impliziert tatsächlich unterschiedliche Vorstellungen: Bei Lewis ist es wie erwähnt der Jazz, weiter gefasst die afrologische Tradition der Improvisation, bei Bhagwati der indische Raga mit seinen offenen, aber verbindlichen Strukturen.

Das Motiv der seltsamen Spiegelung des Selbst steht auch bei meinem letzten Beispiel im Zentrum, bei ULTRACHUNK, das die Komponistin

68 Henning Berg, »Duo-Improvisation mit einer Maschine – Zu Risiken und Nebenwirkungen«, in: Bernd Enders u. Niels Knolle (Hg.), *KlangArt-Kongreß 1995. Vorträge und Berichte vom KlangArt-Kongreß 1995 an der Universität Osnabrück, Fachbereich Erziehungs- und Kulturwissenschaften*, Osnabrück 1996, S. 159–173, hier 162.

69 Bernhard Waldenfels, *Phenomenology of the Alien: Basic Concepts*, Evanston 2011.

70 So der Titel des 3. Kapitels von Bernhard Waldenfels, *Topographie des Fremden. Studien zur Phänomenologie des Fremden I*, Frankfurt a.M. 1997.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und Performerin Jennifer Walshe und der »computational artist« Memo Akten gemeinsam auf der Basis von Aktens *Grannma* (Granular Neural Music & Audio) entwickelt haben. Von Spiegelung kann hier noch in einem direkteren Sinne gesprochen werden, weil das System mit Audio- und Videoaufnahmen von Performances von Walshe vor der Kamera gefüttert wurde und nun als Spielpartner für ihre Vokalperformances fungiert. Es erzeugt live einen Audio- und Videostrom, der im Hintergrund läuft und zu dem die Künstlerin selbst improvisiert. Der Eindruck ist der einer sehr schnell geschnittenen Kompilation aus unterschiedlichen Videoaufnahmen, die durch veränderte Kleidung und Hintergründe erkennbar sind, und elektronisch verfremdeten Klängen, die nur aus der Ferne etwas mit Walshes Stimme zu tun haben und die, ähnlich wie bei *Native Alien*, zwischen Hintergrund und Interaktionspartner changieren.

Fast unvermeidlich wird in der Beschreibung des Projekts das in der Romantik so beliebte Motiv des Doppelgängers mobilisiert: »At once surreal, spellbinding and deeply alarming, ULTRACHUNK is a live improvisational duet between a classically-trained musician and her AI doppelganger.«[71] Auch im Falle von Walshe ist es so, dass dies in eine ganze Reihe von Strategien gehört, bei denen sie sich als Künstlerin mit materiellen und technischen Einrichtungen auseinandersetzt, die sie herausfordern und ihre Arbeit in unvorhergesehene Richtungen lenken, dass also die künstliche Intelligenz als Fortsetzung von etwas auftaucht, das lange vor ihr begonnen hat. Auch ist ULTRACHUNK nicht das einzige Projekt, bei dem Walshe mit KI arbeitet – ein weiteres fand z.B. in Kooperation mit Dadabots statt –, wohl aber das bisher einzige, bei dem es um eine live stattfindende Interaktion geht.

Wenn Walshe den Grund ihres Interesses beschreibt, stellt sie wiederum das *alien* in den Mittelpunkt und betont dabei vielleicht mehr als alle anderen die Faszination des Neuen, das es verspricht: »I'm interested in AI because I would like to experience not just artificial intelligence but also alien intelligence. Brand new artistic vocabularies, systems of logic and syntax, completely fresh structures and dramaturgies.«[72] Weder geht es darum, das zu reproduzieren, was wir ohnehin können, noch darum, den Siegeszug der KI zu feiern oder zu skandalisieren, sondern einmal mehr darum, auf das Seltsame, Fremde und nicht einfach an menschlichen Ausdruck Assimilierbare hinzuweisen, es aufzusuchen und zu kultivieren. Die offensichtliche Seltsamkeit von vielem, was die KI hervorbringt, veranlasst Walshe wie viele andere, hier von einer neuen Form des Denkens und vielleicht sogar einer neuen Spezies zu sprechen. Aber man tut gut daran, solche Aussagen im Kontext der oben zitierten Feststellung von George Lewis zu sehen: Sie

71 https://www.memo.tv/portfolio/ultrachunk/

72 Jennifer Walshe, »Ghosts of the Hidden Layer. Talk given at the Darmstädter Ferienkurse, 25.7.18«, http://milker.org/ghosts-of-the-hidden-layer.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

sind weniger als starke metaphysische Thesen zu verstehen denn als phänomenologische Beschreibungen von Erfahrungen der Interaktion.[73]

6.

Man kann sagen, dass die Künstliche Intelligenz für viele Künstler*innen und Musiker*innen nicht nur seltsam und attraktiv zugleich erscheint, sondern dass ihre Attraktivität zu einem guten Teil gerade in ihrer Seltsamkeit besteht. Der Titel meines Textes hat aber noch einen anderen Bezug, und zwar auf die Chaostheorie und den Begriff des »seltsamen Attraktors«. Ein seltsamer Attraktor ist ein dynamischer Verlauf eines nicht-linearen Systems, der weder stabil noch vollständig vorhersagbar ist, der durch winzige Änderungen des Systems gravierende Veränderungen erfahren kann und der trotzdem mathematisch beschrieben werden kann.[74]

Natürlich ist das in unserem Kontext lediglich eine Metapher, aber doch keine beliebige: Die Seltsamkeit, das weder Berechenbare noch vollkommen Zufällige und Erratische, ist genau das, was die Sache reizvoll macht. Man möchte Systeme konstruieren, die derartiges hervorbringen, und damit letztlich das Gesamtsystem der improvisatorischen Interaktion oder auch der Komposition von dieser Seltsamkeit affizieren lassen. Das ist etwas anderes als die »meta-human outputs, which we ourselves currently could not achieve but which may in the short to medium time frame become just as accessible cognitively and socially as musical outputs that we can make now«[75], von denen McLean und Dean sprechen und die einen Vorsprung der Maschinen suggerieren, den wir vielleicht noch aufholen können. Gerade im künstlerischen Bereich erscheint dieses Bild grundlegend verfehlt; wenn wir schon in derartigen Metaphern denken, sollten wir bei derjenigen des *alien* bleiben. Vom *alien* zu sprechen, heißt letztlich, »eine Figur als Grenzfall oder Variation einer anderen zu sehen«[76], nämlich in unserem Kontext die Maschine als Variation unserer selbst.

73 Genauso versteht es auch Kate Darling: Entscheidend ist, dass wir uns zu (einigen, nicht allen) Robotern »as a new breed of thing – somewhere in-between being and object« (Darling, The New Breed, a.a.O., S. 228) verhalten. Ob sie es »sind«, ist eine Frage, die vielleicht gar nicht so entscheidend ist.

74 Eine kurze Erklärung mit sehr schönen Visualisierungen findet sich hier: https://www.dynamicmath.xyz/strange-attractors/.

75 McLean u. Dean, »Musical Algorithms as Tools, Languages, and Partners: A Perspective«, a.a.O., S. 10.

76 Ludwig Wittgenstein, *Philosophische Untersuchungen (Werkausgabe Bd. I)*, Frankfurt a.M. 1984, S. 413 (Nr. 420.). Wittgenstein beschreibt hier den spätestens seit Descartes geläufigen umgekehrten Fall, in dem Menschen als Automaten gesehen werden.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Die vielleicht entscheidende Frage ist, worauf das alles zuläuft. Aus der Perspektive von jemandem wie George Lewis, für den die Arbeit mit improvisierenden Maschinen eine jahrzehntelange Praxis ist, empfiehlt sich vermutlich eine gewisse Gelassenheit. Er hat seine Arbeit immer als Reflexion darüber verstanden, was Musik und musikalische Interaktion eigentlich sind und was sie sein können. Man könnte sagen, dass wir nun auch dazu gezwungen werden, die Frage neu zu stellen, was wir von Musik *wollen*. Dieses »Wir« ist natürlich, wie immer bei solchen Fragen, eine Fiktion – weder gibt es einen gemeinsamen Willen oder zumindest die Verständigung darüber, noch eine reale Möglichkeit, diesen Willen umzusetzen. Man sollte die Sache daher anders angehen: Musik existiert in unserer Gesellschaft in sehr verschiedenen Formen und erfüllt sehr verschiedene Funktionen. Zu diesen Funktionen gehören die Vermarktung unterschiedlichster Produkte, die Untermalung von Situationen, das emotionale Selbstmanagement, die Identifikation, die Distinktion und schließlich auch die künstlerische Herausforderung. Vielfach changiert eine gegebene Musik auch zwischen diesen Funktionen. Einige von ihnen werden sicher leichter oder auch ganz widerstandslos durch maschinell produzierte Musik erfüllt werden können; bei anderen wie der künstlerischen Herausforderung können wir vielleicht Impulse einer fremden Ästhetik erwarten, die aber kaum menschliche Praxis ersetzen werden, sondern sie lediglich ergänzen und selbst herausfordern.

Vermutlich gibt es allen Grund dazu, »both sublimely excited and blackly horrified about what is coming«[77] zu sein, wie Walshe es formuliert. Vielleicht liegt die Zukunft wirklich in maßgeschneiderter »mood-music on demand«, die so geschickt gemacht und so genau auf die jeweilige Persönlichkeit und Situation zugeschnitten ist, dass keine Fragen übrig bleiben, kein Widerstand und kein Anlass, noch nach anderen Formen musikalischer Gestaltung zu verlangen, so dass die hier behandelten und vergleichbare Projekte marginalisiert und schließlich verschwinden werden. Dann hätte nicht »die künstliche Intelligenz« die Herrschaft übernommen, sondern die derzeit dominierende Form kapitalistischer Unternehmen. Ganz ähnlich fragt Herndon: »Is permission-less mimicry the logical end point of a data-driven new musical ecosystem surgically tailored to give people more of what they like, with less and less emphasis on the provenance, or identity, of an idea? Or is there a more beautiful, symbiotic, path of machine/human collaboration, owing to the legacies of pioneers like George Lewis, that view these developments as an opportunity to reconsider who we are, and dream up new ways of creating and organizing accordingly?«[78] Man wird vermutlich sagen müssen: beides, die Frage ist nur, zu

77 Walshe, »Ghosts of the hidden layer«, a.a.O.

78 https://pitchfork.com/news/holly-herndon-and-jlin-share-new-ai-generated-song-listen/

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

welchem Verhältnis. Insofern die hier behandelten Musiker*innen und viele andere mit ihnen nicht nur mit Maschinen spielen, sondern über diese Praxis auch Herndons Fragen wachhalten und reflektieren, leisten sie einen Beitrag dazu, dass die Sache zumindest offengehalten wird.

Literatur

Adorno, Theodor W. (1973): *Ästhetische Theorie (Gesammelte Schriften Bd. 7)*, Frankfurt a.M.: Suhrkamp

Adorno, Theodor W. (1980): *Philosophie der neuen Musik* (*Gesammelte Schriften*, Bd. 12), Frankfurt a. M.: Suhrkamp

Aristoteles (1994), *Poetik*, griechisch/deutsch, Stuttgart: Reclam

Bailey, Derek (1980): *Improvisation. Its Nature and Practice in Music*, Ashbourne: Da Capo

Berg, Henning (1966): »Duo-Improvisation mit einer Maschine – Zu Risiken und Nebenwirkungen«, in: Bernd Enders u. Niels Knolle (Hg.), *KlangArt-Kongreß 1995. Vorträge und Berichte vom KlangArt-Kongreß 1995 an der Universität Osnabrück, Fachbereich Erziehungs- und Kulturwissenschaften*, Osnabrück: Universitätsverlag Rasch, S. 159–173

Boulez, Pierre (1977): *Wille und Zufall. Gespräch mit Céléstin Deliège und Hans Mayer*, Stuttgart und Zürich: Belser

Boulez, Pierre (1979): »Schönberg ist tot«, in: ders., *Anhaltspunkte. Essays*, Kassel u. München: Bärenreiter, S. 288–296

Cage, John (1961): »To describe the process of composition used in *Music of Changes* and *Imaginary Landscape No. 4*«, in: ders., *Silence. Lectures and Writings*, Middletown: Wesleyan University Press, S. 57–59

Carr, C.J u. Zack Zukowski (2018): »Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands«, https://arxiv.org/abs/1811.06633

Chan, Dawn (2017): »The AI That Has Nothing To Learn From Humans«, in: *The Atlantic*, 20. Oktober, https://www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught-itself-go/543450/

Cope, David (2001): *Virtual Music. Computer Synthesis of Musical Style*, Cambride, MA u. London: MIT Press

Darling, Kate (2021): *The New Breed. How to Think About Robots*, London: Allen Lane

Dennett, Daniel C. (1971): »Intentional Systems«, in: *The Journal of Philosophy* 68, 4, S. 87–106

Edwards, Michael (2011): »Algorithmic Composition: Computational Thinking in Music«, in: *Communications of the ACM* 54, 7, S. 58–67, hier 63 (DOI:10.1145/1965724.1965742)

Eimert, Herbert u. Karlheinz Stockhausen (1958-68), *die Reihe. A periodical devoted to the developments in contemporary music*, 6 Bände, Bryn Mawr u. Wien: Universal Edition

Foerster, Heinz v. (1992): »Entdecken oder Erfinden. Wie läßt sich Verstehen

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

verstehen?«, in: Heinz Gumin u. Heinrich Meier (Hg.), *Einführung in den Konstruktivismus*, München: Piper, S. 41–88
Friedlander, Emily (2019): »How Holly Herndon and her AI baby spawned a new kind of folk music«, in: *The Fader*, 21.5., https://www.thefader.com/2019/05/21/holly-herndon-proto-ai-spawn-interview
Grüny, Christian (2019): »Kontingenzmaschinen. Künstliche Intelligenz in der Musik«, in: *Neue Zeitschrift für Musik* 4, S. 17–21
Grüny, Christian (2022): »Material und Ort. Rehabilitation einer kritischen Kategorie«, in: Robin Becker, Robin, David Hagen u. Livia v. Samson (Hg.), *Ästhetik nach Adorno. Positionen zur Gegenwartskunst*, Berlin: Verbrecher, S. 49–64
Heintz, Joachim (2019): »Spielen statt Fürchtenmachen. Situationen von Musik und KI«, in: *MusikTexte* 162, S. 27–30
Ihde, Don (1990): *Technology and the Lifeworld. From Garden to Earth*, Bloomington u. Indianapolis: Indiana University Press
Kirnberger, Johann Philipp (1757): *Der allezeit fertige Polonoisen- und Menuettencomponist*, Berlin: Winter
Kittler, Friedrich (1986): *Grammophon Film Typewriter*, Berlin: Brinkmann & Bose
Latour, Bruno »Über technische Vermittlung: Philosophie, Soziologie und Genealogie«, in: Bellinger, Andrea, David J. Krieger (Hg.) (2006): *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, Bielefeld: transcript, S. 483–528
Lehmann, Harry (2020): »Voyager-Ästhetik. George Lewis improvisiert mit Computern«, in: *Neue Zeitschrift für Musik* 4, S. 36–39
Lejaren A. Hiller u. Leonard M. Isaacson (1959): *Experimental Music. Composition with an Electronic Computer,* New York: McGraw-Hill
Lewis, George E. (1996): »Improvised Music after 1950: Afrological and Eurological Perspectives«, in: *Black Music Research Journal* 16, 1, S. 91–122
Lewis, George E. (1999): »Interacting with Latter-Day Musical Automata«, in: *Contemporary Music Review* 18,3, S. 99–112
Lewis, George E. (2000): »Too Many Notes: Computers, Complexity, and Culture in Voyager«, in: *Leonardo Music Journal* 10, S. 33–39
Lewis, George E. (2019): »Listening for Freedom with Arnold Davidson«, in: *Critical Inquiry* 45, 2, S. 434–447
LeWitt, Sol (1974): »Paragraphs on Conceptual art / Paragraphen über konzeptuelle Kunst«, in: Gerd de Vries (Hg.), *Über Kunst. Künstlertexte zum veränderten Kunstverständnis nach 1965*, Köln: DuMont, S. 176–185
Ligeti, György (2007): »Entscheidung und Automatik in der *Structure Ia* von Pierre Boulez« (1958), in: ders., *Gesammelte Schriften*, Mainz: Schott, S. 413–446
Lubart, Todd (2005): »How can computers be partners in the creative process: Classification and commentary on the Special Issue«, in: *International Journal of Human-Computer Studies* 63, S. 365–369

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Mackenzie, Adrian (2017): *Machine Learners. Archaeology of a Data Practice*, Cambridge, MA: MIT Press

McDermott, Emily (2020): »Holly Herndon on her AI Baby, Reanimating Tupac, and Extracting Voices«, in: *Art in America*, 7.1., https://www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301/

McLean, Alex u. Roger Dean (2018): »Musical Algorithms as Tools, Languages, and Partners: A Perspective«, in: dies. (Hg.), *The Oxford Handbook of Algorithmic Music*, Oxford: Oxford University Press, S. 3–14

Merleau-Ponty, Maurice (1966): *Phänomenologie der Wahrnehmung*, Berlin: de Gruyter

Nistal, Javier (2020): »Deep Beers. A Chat with Philippe Esling«, in: *New Frontiers in Music Information Processing*, 4.1., https://mip-frontiers.eu/2020/01/04/Philippe_Esling.html

Piekut, Benjamin (2013): »Chance and Certainty – John Cage's Politics of Nature«, in: *Cultural Critique* 84, 134–163

Prey, Robert, Marc Esteve Del Valle u. Leslie Zwerwer (2020): »Platform pop: disentangling Spotify's intermediary role in the music industry«, in: *Information, Communication, and Society*, https://doi.org/10.1080/1369118X.2020.1761859

Ratner, Leonard G. »Ars Combinatoria: Chance and Choice in Eighteenth-Century Music«, in: H.C. Robbins Landon u. Roger E. Chapman (Hg.) (1979): *Studies in Eighteenth-Century Music. A Tribute to Karl Geiringer on his Seventieth Birthday*, New York: Allan & Unwin, S. 343–363

Satie, Erik (1988): »Musique d'Ameublement. Mobiliarmusik«, in: ders., *Schriften*, Hofheim: Wolke, S. 31

Schröter, Jens (2021): »›Künstliche Intelligenz‹ und die Frage nach der künstlerischen Autor*innenschaft«, in: *Kunstforum International* 278, S. 99–106

Spiegel, Laurie (2018): »Thoughts on Composing with Algorithms«, in: Alex McLean u. Roger Dean (Hg.), *The Oxford Handbook of Algorithmic Music*, Oxford: Oxford University Press, S. 105-111

Waldenfels, Bernhard (2011): *Phenomenology of the Alien: Basic Concepts*, Evanston: Northwestern University Press

Waldenfels, Bernhard (1997): *Topographie des Fremden. Studien zur Phänomenologie des Fremden I*, Frankfurt a.M.: Suhrkamp

Walshe, Jennifer »Ghosts of the Hidden Layer. Talk given at the Darmstädter Ferienkurse, 25.7.18«, http://milker.org/ghosts-of-the-hidden-layer

Wittgenstein, Ludwig (1984): *Philosophische Untersuchungen (Werkausgabe Bd. I)*, Frankfurt a.M.: Suhrkamp

Xenakis, Iannis *Formalized Music. Thought and Mathematics in Composition*, Stuyvesant, NJ 1992

Zuboff, Shoshana (2018): *Das Zeitalter des Überwachungskapitalismus*, Frankfurt: Campus

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Onlineressourcen

http://artsites.ucsc.edu/faculty/cope/5000.html
http://dadabots.com/#new-albums
http://matralab.hexagram.ca/research/native-alien/about/project-description/
http://obvious-art.com/wp-content/uploads/2020/04/MANIFESTO-V2.pdf
http://repmus.ircam.fr/omax/home
http://www.osamahsalem.co.uk
https://ainowinstitute.org/
https://arxiv.org/abs/1612.01010
https://dadabots.com
https://holly.mirror.xyz
https://musically.com/wp-content/uploads/2018/11/Music-Ally-AI-Music-Guide.pdf
https://obvious-art.com
https://pitchfork.com/news/holly-herndon-and-jlin-share-new-ai-generated-song-listen/
https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx
https://www.dynamicmath.xyz/strange-attractors/
https://www.hollyherndon.com/godmother-holly-herndon-and-jlin-feat-spawn
https://www.hollyherndon.com/proto
https://www.memo.tv/portfolio/ultrachunk/
https://www.telekom.com/de/konzern/themenspecials/special-250-jahre-beethoven/details/kuenstliche-intelligenz-soll-beethovens-zehnte-sinfonie-vollenden-587346
https://www.youtube.com/watch?v=MwtVkPKx3RA

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Alexander Hochmuth

Pflege und künstliche Intelligenz

Soziale Roboter in der Pflege von Menschen mit einer Demenz

Pflegetechnologien, die auf künstlicher Intelligenz (KI) basieren, sind nicht mehr reine Fiktion, sondern bereits im Pflegealltag angekommen (Wagner 2014). Dabei ist der Einsatz von KI vor allem in der Pflege mit großen gesellschaftlichen Erwartungen und einer Vielzahl offener Fragen verbunden. Neben der Optimierung von Arbeitsabläufen in Pflegeeinrichtungen kann der Einsatz von KI zur Verbesserung der Pflegequalität und Sicherheit innerhalb der pflegerischen Versorgung beitragen. Besonders für vulnerable Menschen mit einer Demenz (MmD) und deren Angehörige kann der KI-Einsatz in Form von sozial assistierenden Robotern verschiedene ethische Herausforderungen mit sich bringen. Begegnungen in Form menschlicher Beziehungen, Schutz und Selbstbestimmung sowie ein gerechter Zugang zu KI Technologien rücken in den Mittelpunkt einer pflegerisch-gesellschaftlichen Auseinandersetzung und bedürfen einer Neubewertung der digitalisierten Pflege.

1. Einleitung

In der Pflege hat der Einsatz von technischen Systemen eine lange Historie und konnte unbestreitbar zur Verbesserung der Lebensqualität von pflegebedürftigen Menschen beigetragen (Meißner & Kunze 2020). Technische Innovationen basieren auf der Förderung der Autonomie und Selbständigkeit pflegebedürftiger Menschen (Meißner & Kunze 2020; Hauck & Uzarewicz 2019). So ermöglichen Gehhilfen und Rollstühle die soziale Teilhabe, Mobilisationshilfen wie Lifter tragen zu einer Erleichterung der Pflegearbeit auch für pflegende Angehörige bei. Digitale Global Positioning Systeme (GPS) sorgen für die nötige Sicherheit bei Menschen mit Orientierungsschwierigkeiten.

Durch den Einsatz von lernenden Algorithmen und Robotern entstehen neue pflegerische Assistenzsysteme, die sich an neue Kontexte anpassen und in gegebenen Situationen eingreifen können, um den Alltag oder Pflegeprozesse zu unterstützen (Meißner & Kunze 2020; Ronquillo et al. 2021). Künstliche Intelligenz wird als Teilgebiet der Informatik definiert und versucht »*[…] kognitive Fähigkeiten wie Lernen, Planen*

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

oder Problemlösen in Computersystemen zu verwirklichen […]« (acatech 2019, 6). Aufgrund der vielfältigen Einsatzmöglichkeiten von KI Technologien im Pflegebereich gibt es unterschiedlichste Anwendungsszenarien. So können Spracherkennungstechnologien, auf der Basis von maschinellem Lernen, die Pflegedokumentation verbessern oder taubstumme Menschen bei der Kommunikation unterstützen (Ronquillo et al. 2021; acatech 2019). Andere Anwendungen umfassen den Bereich des Text Mining, bei dem KI Technologien zur Identifizierung von Pflegeproblemen und möglichen kritischen Ereignissen (z.B. Risiko für ein Delir oder Sturz), im Rahmen einer klinischen Entscheidungsfindung, eingesetzt werden.

Auf der Grundlage der Pflegedokumentation werden beispielsweise Patienten mit einer Sturzanamnese oder Drogen- und Alkoholkonsumstörungen frühzeitig identifiziert und ihr spezifisches Risiko bei der Planung weiterer Pflegemaßnahmen berücksichtigt (Topaz et al. 2019). In ähnlicher Weise wurde maschinelles Lernen, insbesondere Deep Learning Verfahren, zur Vorhersage des Schmerzempfindens und der Symptomverschlechterung bei akuten kritischen Ereignissen (z.B. Luftnot) erprobt (Wu et al. 2021). Somit kann die »transformative Kraft« von KI im pflegerischen Kontext zu einschneidenden Veränderungen führen (Kerasidou 2020, 245).

Das wohl bekannteste Beispiel für die Anwendung von KI in der Pflege sind Pflegeroboter. In der japanischen Science-Fiction-Literatur werden Pflegehandlungen häufig durch humanoide oder tierähnliche Pflegekräfte in Roboterform dargestellt (Wagner 2014). Jene Pflegeroboter helfen pflegebedürftigen Menschen im Alltag, indem sie beim Treppensteigen unterstützen, die Körperpflege übernehmen oder Mahlzeiten und Getränke anreichen. Ein solcher Pflegeroboter, der komplexe menschliche Tätigkeiten ersetzt, steht aus technischer sowie pflegerisch-ethischer Perspektive nicht in Aussicht und wird auf lange Sicht nicht vielversprechend einsatzbereit sein (Schulze et al. 2021). Die Alternative ist eine Unterstützung durch Systeme, die spezialisierte Aufgaben übernehmen (z.B. Roboter zur Unterhaltung oder Unterstützung bei der Mobilisation). In Pflegeeinrichtungen oder auch in der häuslichen Pflege lässt sich schon seit einiger Zeit der Einsatz von sozial assistierenden Pflegerobotern beobachten (Baisch et al. 2017).

Nach Broekens et al. (2009) lassen sich Pflegeroboter in intelligente körperliche Bewegungshilfen, Service- und Begleitroboter für spezifische Pflege- oder Haushaltsaufgaben sowie sozial assistierende Roboter (SAR) oder auch »companion-type robots« (z.B. PARO, Lio, Jibo, Cruzr, Pleo, Pepper etc.) unterteilen. Roboter unterscheidet von bisherigen pflegerischen technischen Hilfsmitteln, dass sie eine Form der Mensch-Maschine-Interaktion beinhalten, welche sich durch komplexe Rückkopplungsschleifen, Kontrollfunktionen und datenbasierte

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Kommunikationsformen charakterisieren lässt (Schicktanz & Schweda 2021). Das bedeutet, SAR können durch die Kombination von Sensor-, Datenverarbeitungs- und Antriebssystemen physisch und in direkter Form in pflegerische Handlungen und Prozesse eingreifen und diese unterstützen (Schicktanz & Schweda 2021). Um funktionieren zu können, sind SAR in unterschiedlichem Maße auf menschliche Entscheidungen und Eingaben (z.B. Patientendaten, Befehle etc.) angewiesen (Schicktanz & Schweda 2021). Auf der Grundlage künstlicher Intelligenz können SAR »*[…] bestimmte Emotionen ihrer Gegenüber erkennen sowie emotional kommunizieren, mit dem Zweck, bei menschlichen Interaktionspartnern kognitive und affektive Zustände auszulösen […]*« (Schulze et al. 2021, 10). Somit vereinen SAR menschliche und künstliche Intelligenz. SARs bieten emotionale und soziale Unterstützung für pflegebedürftige Menschen, indem sie beispielsweise zu Gesprächen anregen, Freizeitaktivitäten wie Spiele bereithalten oder Unterhaltungsmedien (z.B. Filme, Lieder etc.) abspielen (Kachouie et al. 2014).

Neben den genannten Vorteilen führt der Einsatz von SAR in der Pflege besonders für vulnerable Menschen mit einer Demenz (MmD) und deren Angehörige – in Hinblick auf Begegnung in Form einer menschlichen Beziehung zwischen selbstbestimmten Partnern[1] auf Augenhöhe (Buber 2006) – zur Auseinandersetzung mit einer Vielzahl an ethischen Herausforderungen und weiterführenden Fragen. Auf Grundlage dessen widmet sich der folgende Beitrag der Frage: Welchen Einfluss kann der Einsatz und gerechte Zugang zu KI-basierten Pflegetechnologien, am Beispiel sozial assistierender Robotik, auf die pflegerische Beziehungsgestaltung zwischen MmD, deren Angehörige und professionelle Pflegende haben?

2. Künstliche Intelligenz in der Pflege von Menschen mit Demenz

Demenz gehört zur Gruppe der neurodegenerativen Erkrankungen und kann sich in der Ausprägung der Symptome sowie im Krankheitsverlauf je nach Art der Demenzerkrankung stark unterscheiden (Maier et al. 2016). Der Abbau kognitiver Funktionen in verschiedensten Bereichen steht im Zentrum der Erkrankung (z.B. Verlust des Kurzzeitgedächtnisses, des Geruchssinns, Veränderung der Persönlichkeit). Neben der Verminderung von Gedächtnisfunktionen und der Sprache treten besonders häufig emotionale und verhaltensbezogene Symptome wie

1 Wenn im folgenden Text an verschiedenen Stellen die männliche Form gewählt wurde, passierte dies aus Gründen der Leserlichkeit. In solchen Formulierungen sind alle Geschlechter gleich inkludiert.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Rückzugstendenzen, depressive Verstimmungen, Fehlinterpretationen von auditiven sowie visuellen Reizen, vermehrtes Nachfragen oder auch wiederkehrende Handlungen wie Klatschen und Klopfen sowie unkontrolliertes Umherwandern auf (Maier et al. 2016).

Trotz zunehmenden Fortschreitens der Erkrankung besteht der Wunsch der meisten MmD, wie bei den meisten Pflegebedürftigen, so lange wie möglich in der eigenen Wohnung und im gewohnten familiären Umfeld leben zu können (Kutzleben 2016). Dies erfordert ein hohes Maß an Flexibilität, zeitlichen Ressourcen und eine starke emotionale Belastung aller an der Pflege beteiligten Familienmitglieder und professionell Pflegenden (Kutzleben 2016). Um die Pflege eines Angehörigen ermöglichen zu können, nehmen Angehörige von MmD häufig die Unterstützung durch technische Assistenzsysteme in Anspruch (Baisch et al. 2018; Kutzleben 2016). Neben webbasierten Informations- und Beratungsangeboten für Angehörige (z.B. digitale Pflegekurse), werden häufig GP-Systeme bei Menschen mit Hinlauftendenzen, komplexe Systeme aus dem Bereich der Assisting Ambient Living (AAL) (z.B. sensorbasierte Sturzmeldesysteme) oder sozial assistierenden Pflegerobotern eingesetzt (Palmdorf et al. 2019).

Gerade der Einsatz von SAR in der Pflege von MmD konnte in einigen empirischen Untersuchungen zeigen (Schulze et al. 2021), dass SAR einen positiven Einfluss auf das Wohlbefinden und die Beziehungsförderung von MmD haben können. Beispielsweise konnten Reaktionen wie Lachen und Entspanntheit sowie verbesserte Vitalparameter im Zusammenhang mit dem Einsatz von SAR beobachtet werden (Baisch et al. 2018). Darüber hinaus wird der Einsatz von SAR durch pflegebedürftige Menschen, Angehörige und auch professionell Pflegende als positiv bewertet und SAR werden als ein »nützliches Hilfsmittel« beschrieben (Baisch et al. 2018).

2.1 Interaktion mit sozial assistierenden Pflegerobotern

Kann in der Interaktion mit einem sozial assistierenden Pflegeroboter eine menschliche »Begegnung« entstehen oder werden MmD doch nur auf ihre kognitive Leistungsfähigkeit reduziert und eine (Pflege-)Beziehung findet nur im äußerlichen Sinne statt?

Pflege von MmD lässt sich als ein komplexes und durch empathische Zugänge gekennzeichnetes Handeln charakterisieren. Kommunikation und Interaktion bilden hierbei eines der zentralen Elemente einer Pflegebeziehung und tragen zum Gelingen von sich stetig verändernden Entscheidungsprozessen bei (Hauck & Uzarewicz 2019). Die Sprache und die Berührung stehen, neben weiteren grundlegenden menschlichen Expressionsformen (Bsp. Gestik, Mimik oder Resonanz), im Mittelpunkt

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

des pflegerischen Handelns und dienen als Grundlage für den Aufbau einer intakten Pflegebeziehung. Körperbezogene Pflegearbeit ist daher immer in kommunikative Bezüge eingebettet (Friesacher 2019). Sprachliche Handlungen finden zumeist in direkten face-to-face Situationen statt und haben eine Schlüsselfunktion im Zugang zum Anderen (Hülsken-Giesler 2007). Der Leiblichkeit im Kontext des Sinnverstehens in der Pflege kommt eine zentrale Bedeutung zu (Friesacher 2019). Neben der Beziehungsarbeit gilt die Pflege von bedürftigen Menschen als eine körpernahe Tätigkeit. Die Haut kann als ein Kommunikationsmedium der Sinne verstanden werden. Durch körperliche Berührungen treten die Pflegeperson und der pflegebedürftige Mensch in Interaktion. Die Körpersprache oder auch -haltung des pflegebedürftigen Menschen nehmen, im Kontext des Beobachtens und Verstehens, eine besondere Funktion ein.

Ferner ermöglichen u.a. empathische und gleichermaßen mimetische Prozesse das Verstehen des Gegenübers. Den zentralen menschlichen Kern pflegerischen und menschlichen Handelns bildet die Qualität der Begegnung (Krückel 2015). Neben der Begegnung stehen die körperliche wie seelische Berührung im Mittelpunkt pflegerischen Handelns. Nach Martin Buber vollziehen sich Begegnungen in der menschlichen Beziehung zwischen Partnern auf Augenhöhe. Das bedeutet zum einen das Hineinfühlen in das Denken und Empfinden eines pflegebedürftigen Menschen (Empathie). Zum anderen das Überführen des Gegenübers in innere Bilder, Klangkörper, Tast-, Geruchs-, und Geschmackswelten. Dabei werden mimetische Erfahrungen vielschichtig und subjektiv empfunden. Pflegerische Handlungen und die Begegnung zwischen der pflegebedürftigen Person und der Pflegeperson finden in einem engen dialogischen Prozess auf Augenhöhe statt (Krückel 2015).

Tritt nun ein SAR an die Stelle der Pflegeperson und nimmt die Interaktion mit einem MmD auf besteht die Gefahr, dass die psychosozialen Bedürfnislagen nicht adäquat erfasst, gedeutet und befriedigt werden können. Ferner könnte es durch den Einsatz von SAR zu einem Ersatz zwischenmenschlichen Kontakts kommen und eine mögliche, bereits vorhandene soziale Isolation verstärkt und das psychische Wohlbefinden von MmD gemindert werden (Baisch et al. 2018). In einer menschlichen Beziehung wird Verantwortung für seinen Gegenüber übernommen. Nach Krückel (2015, 7) bedeutet Verantwortung nach Buber »*[…] auf das, was uns widerfährt, was wir sehen, hören und spüren, zu antworten. »Echte Verantwortung gibt es nur, wo es wirkliches Antworten gibt.*« Es stellt sich die Frage, inwieweit pflegerische Verantwortung für vulnerable pflegebedürftige Menschen auf KI-basierte Pflegeroboter übertragbar ist? Wenn MmD nicht die Möglichkeit haben, ihre Gefühle während der Interaktion dem Gegenüber zu spiegeln und eine verantwortungsvolle Antwort darauf zu erhalten, führt dies in ein mögliches

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Abhängigkeitsverhältnis von MmD mit der Gefahr eines »nicht Erkennens« von Emotionen und kritischen Zuständen. Robotik würde in diesem Zustand zu einer Funktionspflege beitragen und eine patientenzentrierte Pflege verhindern. Sharkey und Sharkey (2012) führen in diesem Zusammenhang als ethisch problematischen Punkt die Gefahr der Täuschung von MmD durch einen Roboter an. Die mangelnde Unterscheidung zwischen Mensch und Maschine (Anthropomorphisierung), kann gerade für MmD, welche oft unter »illusionären Verkennungen« leiden, stark beeinträchtigend wirken (Baisch et al. 2018, 19). Daher kommt Pflegenden eine verantwortungsvolle und schützende Funktion für vulnerable pflegebedürftige MmD zuteil. Der Einsatz von SAR in der Pflege von MmD sollte somit kritisch reflektiert und beispielsweise die Rahmenempfehlungen zum Umgang mit herausforderndem Verhalten des Bundesministeriums für Gesundheit als ethisch vertretbarer Ansatz in Betracht gezogen werden (Bartholomeyczik et al. 2006).

Interaktion und pflegerische Beziehungsgestaltung setzen ein hohes Maß an personaler und sozialer Kompetenz voraus, was sich in einem hohen Verantwortungsbewusstsein und »emotionaler Intelligenz« wiederspiegelt (Krückel 2015, 7). Pflegende interagieren beziehungsweise passen ihr Verhalten und ihre Reaktionen an eine bestimmte Situation, aufgrund moralischer Entscheidungen, Wissen und Erfahrungen an. Roboter verfügen nicht über eine menschlich-emotionale Intelligenz und können damit nicht zu einem moralischen Ebenbild Pflegender werden. Daher bleibt es ihnen verwehrt, pflegerisch-menschlich zu handeln. Bei Begegnungen zwischen Roboter und MmD bleiben Aspekte wie der gegenseitige Respekt, Vertrauen, das Eingehen auf Individualität und Autonomie nicht berücksichtigt. Sollte es dennoch zu einem fachgerechten Einsatz von SAR kommen, so liegt die Entscheidung über die Anwendbarkeit und die Bewertung der Nützlichkeit für ausgewählte Pflegekontexte bei professionell Pflegenden. Pflegende müssen ihre Wahrnehmung und ihr Verständnis hinsichtlich einer verantwortungsvollen Beziehungsgestaltung in Bezug auf den Einsatz von SAR reflektieren. SAR in der Pflege von MmD erfordern eine intensive anamnestische Auseinandersetzung mit dem einzelnen Individuum und seinen individuellen Bedürfnislagen. SAR können aber ein Bindeglied im Gelingen einer menschenwürdigen Pflege sein. Durch unterschiedliche technische Vorrausetzung von Pflegerobotern (z.B. Sensoren, Kamera, Bilderkennung etc.) besteht die Möglichkeit der Bereitstellung tages- und situationsspezifischer Daten über pflegebedürftige Menschen (Schicktanz & Schweda 2021). Professionell Pflegende können diese perspektivisch als sinnvolle Ergänzung in der pflegerischen Beziehungsgestaltung nutzen. Über die Nutzung und den Rückgriff auf diese Daten sollten pflegebedürftige Menschen und ihre Angehörigen, im Rahmen der pflegerischen Beziehungsgestaltung und im Kontext eines digital-selbstbestimmten Handelns, informiert sein.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

2.2 Selbstbestimmung und Zugang zu sozial assistierenden Pflegerobotern

Menschen mit einer beginnenden Demenz (MmbD) sind in einer frühen Krankheitsphase in der Lage, sich selbst zu versorgen und ihren Alltag mit wenigen Unterstützungsmitteln zu strukturieren, zu bewältigen und zu organisieren (Maier et al. 2016). Einige Studien konnten zeigen, dass die Einbindung von digitalen Pflegetechnologien im Kontext der Alltags- und Krankheitsbewältigung von MmbD und deren Angehörigen akzeptiert und als Unterstützung im Alltag anerkannt, wird (Grossman et al. 2018). Die Anwendung von Pflegerobotern oder auch anderen KI-basierten Pflegetechnologien im Alltag und der Begleitung von MmbD können zur Selbstbestimmung und gesellschaftlichen Teilhabe dieser vulnerablen Gruppe beitragen (Vetter et al. 2012).

Mit der Ratifizierung der UN-Behindertenrechtskonvention hat sich Deutschland verpflichtet, Maßnahmen zu ergreifen, um vulnerablen Menschen eine gleichberechtigte und selbstbestimmte Teilhabe an allen modernen Informations- und Kommunikationstechnologien zu ermöglichen sowie Zugangsbarrieren zu minimieren (BMAS 2018). Um MmbD nicht den Zugang zu digitalen Lebenswelten und Technologien und das Grundrecht auf Selbstbestimmung zu verwehren, müssen neue digitale Interaktionsformen ermöglicht und Schlüsselkompetenzen (u.a. IT-Sicherheit, Datenschutz, Kontrolle über Datennutzung) in der Nutzung digitaler Gesundheitstechnologien vermittelt werden (SVRV 2017). Ferner kann ein Ausschluss pflegebedürftiger Menschen von der Nutzung und Entwicklung digitaler Technologien zur Nichtpartizipation an gesellschaftlich-demokratischen Prozessen **führen** und fördert die Zunahme von gesellschaftlichen und gesundheitlichen Ungleichheiten in gleicher Weise (Bol et al. 2018).

Die Zunahme von KI-basierten Pflegetechnologien und der mögliche Einsatz von SAR im späteren Verlauf dementieller Erkrankungen erfordern Digitalkompetenz bei MmbD, deren Angehörigen und auch bei professionell Pflegenden. Darüber hinaus muss eine nutzerorientierte Entwicklung KI-basierter Pflegetechnologien erfolgen, um die Bedarfe und Bedürfnisse der genannten Nutzergruppe adäquat adressieren zu können. Koutentakis et al. (2020) verweisen in diesem Kontext darauf, dass einige SAR nicht den spezifischen Bedarfen und Bedürfnissen von MmbD entsprechen. Schon heute ist die Anwendung von KI-basierten Pflegetechnologien im Alltag angekommen und das Verständnis über einzelne Funktionsweisen eine Voraussetzung für soziale Teilhabe.

Um der beschriebenen Problematik entgegenzuwirken, hat die Europäische Kommission Strategien und Maßnahmen entwickelt, um digitale Kompetenzen von Bürgern und potentiellen Konsumenten zu stärken[2].

2 Siehe »Digital Competence Framework for Consumers«

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Darüber hinaus wurde mit der Verabschiedung der europäischen Datenschutzgrundverordnung (EU-DSGVO) im Jahr 2016 eine Basis für einen einheitlichen Datenschutz in Europa geschaffen. Dabei steht die Förderung digital souveränen Handelns im Vordergrund. Aufgrund dessen stellt sich die Frage, wie künftig Menschen mit einer beginnenden dementiellen Erkrankung und deren Angehörige befähigt werden können, einen Einsatz von SAR und KI-basierten Pflegetechnologien selbstbestimmt in einem späteren Krankheitsverlauf zu bewerten? Das Ermöglichen gesellschaftlicher Teilhabe, Inklusion und Selbstbestimmung von »Menschen in einer durch Krankheit oder Behinderung geschwächten Position […]« muss dabei im Zentrum einer nutzerorientierten Gestaltung von KI-basierten Pflegetechnologien stehen (Vetter et al. 2012, 54). In dem Zusammenhang lassen sich der Begriff Datensouveränität und sieben Dimensionen digitaler Selbstbestimmung (Kompetenz, Informiertheit, Werte, Wahlmöglichkeit, Freiwilligkeit, Willensbildung und Handlung) unterscheiden (SVRV 2017). Digitale Selbstbestimmung kann als Ausprägung digitaler Souveränität verstanden werden und umfasst die Wahl- und Entscheidungsfreiheit von Nutzern über die Erhebung, Verarbeitung und Nutzung ihrer persönlichen Daten (SVRV 2017). Unter digitaler Souveränität wird demnach die Handlungsfähigkeit und Entscheidungsfreiheit von MmbD, in einer digitalen Welt im Kontext verschiedener Rollen (u.a. Konsument, Marktteilnehmer, vulnerabler Nutzer), verstanden (Friedrichsen & Bisa 2016). Die daraus resultierende Autonomie und Teilhabe an einer digital-gesellschaftlichen Welt ist bedingt durch die informierte Wahlfreiheit der Nutzenden und berücksichtigt dabei den Kern des Gesetzes zur Stärkung der Teilhabe und Selbstbestimmung von Menschen mit Behinderungen und Empfehlungen des Deutschen Ethikrats zur Inanspruchnahme digitaler Dienste durch vulnerable Gruppen[3].

Digital selbstbestimmtes Handeln ist eng verbunden mit der Nutzung von KI-basierten Pflegetechnologien (Manzoor & Vimarlund 2018). Der Einsatz von KI-basierten Pflegetechnologien kann für die Versorgung spezifischer Personengruppen große Vorteile mit sich bringen. Durch eine sozial ungleiche Verteilung von Technikorientierung und -kompetenzen verschärfen sich bei Menschen mit einem niedrigen sozioökonomischen Status und besonders MmbD und deren Angehörige

3 In der Stellungnahme »Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung« (2017) empfiehlt der deutsche Ethikrat *»um Gerechtigkeit und Solidarität auch unter Big-Data-Bedingungen zu sichern (Themenbereich C), […] folgendes: > Fairen Zugang zu digitalen Angeboten schaffen; Vulnerable Gruppen und Individuen schützen (C4) […] > Schutzmechanismen für die Datenerhebung an sonstigen Personen mit eingeschränkter Einwilligungsfähigkeit entwickeln (C4.2) > Einsatz von Chatbots restriktiv regeln (C4.3).«* (S. 45).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Ungleichheiten in der pflegerisch-gesundheitlichen Versorgung. Darüber hinaus kann es aufgrund fehlender digitaler Kompetenzen und Wissen zur Reduzierung von Gesundheits- und gesellschaftlichen Teilhabechancen kommen (Manzoor & Vimarlund 2018). Um MmD nicht den Zugang zu digitalen Lebenswelten und KI-basierten Pflegetechnologien und das Grundrecht auf Selbstbestimmung zu verwehren, müssen künftig neue digital-pflegerische Interaktionsformen kritisch reflektiert werden. Pflegenden Angehörigen von MmD kommt im Umgang mit digitalen Schlüsselkompetenzen (u. a. Wissen über IT-Sicherheit, Datenschutz, Kontrolle über Datennutzung) künftig eine Schlüsselrolle zu (van der Wardt et al. 2017). Ferner kann ein Ausschluss von vulnerablen Personengruppen – im allgemeinen Sinne – im Zugang zur Nutzung und Entwicklung von KI-basierten Pflegetechnologien zur Nichtpartizipation an gesellschaftlich-demokratischen Prozessen führen und die Zunahme von gesellschaftlichen und gesundheitlichen Ungleichheiten **fördern**.

2.3 Professionelles Pflegehandeln und Verantwortung

Wie möchte ich später gepflegt werden? Was ist mir wichtig? Was sind meine Vorlieben? Möchte ich von einem Pflegeroboter gepflegt werden? Diese oder mögliche andere Fragen stellen sich pflegebedürftige Menschen und deren Angehörige zu Beginn oder im späteren Verlauf einer chronisch-verlaufenden Erkrankung. Um eine pflegerische Versorgung und Behandlung, bei Krankheitsverschlechterung oder Verlust der Entscheidungsfähigkeit, möglichst patientenorientiert und selbstbestimmt gestalten zu können, wurde das Konzept des Advance Care Plannings (ACP) oder auch gesundheitliche oder vorausschauende Versorgungsplanung etabliert (Haller & Schnell 2019). Beim ACP handelt es sich um einen kontinuierlichen sowie strukturiert gestalteten interprofessionellen Prozess, der auf die Situation des pflegebedürftigen Menschen zugeschnitten ist (Kolarik et al. 2002).

Während der Beratung werden pflegebedürftige Menschen mit Methoden des Empowerment-Ansatzes unterstützt, persönliche Werte zu identifizieren und deren Bedeutung im Kontext verschiedener Krankheitsszenarien zu reflektieren. Ziel ist, persönliche Präferenzen für die pflegerische Versorgung in den besprochenen Krankheitsszenarien zu reflektieren und zu definieren. Dieses strukturierte Vorgehen kann zu einer rechtlich geleiteten und an Patientenpräferenzen orientierten künftigen Pflegeversorgungsentscheidung beitragen. Ein zentraler Unterschied des ACPs zu bestehenden pflegerischen Instrumenten, wie zum Beispiel die Pflegeplanung oder die Patientenverfügung, ist dass die formulierten Präferenzen innerhalb des ACP in naher Zukunft eintreten, wenn der betroffene Mensch seine Entscheidungsfähigkeit verliert und/oder bei

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Verlust der Fähigkeit, anderen seine Wünsche mitzuteilen (Exley et al. 2009). Patientenverfügungen vereinfachen den komplexen Prozess einer pflegerischen Entscheidungsfindung und erfordern daher eine kontinuierliche Neubewertung entstehender Situationen (Haller & Schnell 2019). Im Gegensatz zu einer Pflegeplanung ist der ACP eine aufsuchende und partizipative Pflegemaßnahme, welche eine kontinuierliche Reflexion professionellen Pflegehandelns erfordert. Daher wird die ACP im Kontext einer fortschreitenden Erkrankung und einer zu erwartenden Verschlechterung des Krankheitsverlaufs angewendet (Exley et al. 2009). Im Zusammenhang mit einer demenziellen Erkrankung hat sich die ACP als wirksames Instrument erwiesen (Piers et al. 2018). j

Die Angst vor dem Verlust der Kontrolle über das eigene Handeln und das »Ausgeliefert-sein«, beschreiben MmD häufig als eines der zentralen Gefühle (Vetter et al. 2012). Der ACP kann als eine mögliche strukturgebende Bewältigungsstrategie in komplexen pflegerischen Entscheidungssituationen angesehen werden (Exley et al. 2009). Neben der Handlungssicherheit für professionell Pflegende vermittelt der ACP pflegebedürftigen Menschen das Gefühl der Kontrolle und einer patientenzentrierten Pflege. Die Herausforderung bei der Anwendung von ACP bei MmD besteht häufig in der späten Diagnosestellung. Aufgrund der Angst vor Stigmatisierung und den Folgen einer Demenzerkrankung wird die Diagnose meist erst in einem späteren Stadium der Demenz gestellt. Zu dem Zeitpunkt kann es sein, dass MmD bereits kognitiv leicht beeinträchtigt sind. Es ist daher fraglich, inwieweit MmD ihren tatsächlichen Willen artikulieren können. Daher sind die Einbindung des ACP und die Information von MmD zu einem frühen Zeitpunkt der Erkrankung von besonderer Bedeutung. Nach Palmdorf et al. (2019) findet eine Thematisierung KI-basierter Pflegetechnologien und einer technikbasierten Pflege im Allgemeinen noch keine Berücksichtigung im ACP. Um eine selbstbestimmte und patientenzentrierte Pflege von MmD, unter Einbindung KI-basierter Pflegetechnologien (u.a. sozial assistierende Roboter) zu ermöglichen, stellt der ACP ein mögliches pflegerisches Handlungsinstrument im Kontext der bereits beschriebenen Herausforderungen im Einsatz von SAR dar.

Die Thematisierung des Einsatzes von SAR im ACP bietet die Möglichkeit, den Willen sowie individuelle Bedürfnisse (abhängig von Alter, Technikerfahrungen sowie sozialen und kulturellen Werten) und Wünsche von MmD in Bezug auf den Einsatz von Technik im Rahmen der pflegerischen Versorgung zu berücksichtigen. Würde kein ACP im Pflegeprozess berücksichtigt, so könnte ausschließlich nur der tatsächliche Wille des MmD in der direkten Interaktion mit einem SAR handlungsleitend sein. Der mutmaßliche Wille im Rahmen der Mensch-Maschine-Interaktion müsste ermittelt werden und die Verantwortung für die Entscheidung läge dann bei den Angehörigen oder den professionell Pflegenden. Berücksichtigt man empirische Erkenntnisse zum Thema

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

»stellvertretende Entscheidungsfindung **für** MmD« (z. B. durch Angehörige, Betreuende), so lässt sich feststellen, dass diese auf dem mutmaßlichen Willen der pflegebedürftigen Person beruhen (DGGG 2020). Das Risiko, eine Entscheidung gegen den tatsächlichen Willen zu treffen, ist dabei im erhöhten Maße gegeben (Givens et al. 2018). Die Schwierigkeit der stellvertretenden Entscheidungsfindung besteht darin, dass oftmals die Entscheidungsgründe der Sicherheit und Absicherung diejenigen der Autonomie und Lebensqualität des pflegebedürftigen Menschen überwiegen (DGGG 2020; Palmdorf et al. 2019). Es stellt sich die Frage, inwieweit der Einsatz von SAR nach individueller und situationsbezogener Betrachtung einzelner Entscheidungsträger ethisch vertretbar ist?

Um Menschen mit einer beginnenden Demenz und deren Angehörige adäquat über die Funktionsweisen, die ethischen Aspekte und die Einbindung von SAR in mögliche Pflegemaßnahmen aufklären und beraten zu können, benötigen professionell Pflegende spezifisches Wissen und digitale Kompetenzen zu aktuellen pflegetechnischen Entwicklungen, Funktionsweisen von Pflegerobotern und Kenntnisse zu Themen wie Datenspeicherung und -sicherung (Borcherding et al. 2021; Hauck & Uzarewicz 2020). Darüber hinaus stellt sich die Frage, wie das ACP Konzept in die pflegerische Versorgung eingebettet werden kann, damit der Wille des MmD bei späteren Entscheidungen zum Einsatz von SAR ausreichend Berücksichtigung findet. In dem Zusammenhang verweist Misselhorn (2018) auf die Entwicklung von anpassungsfähigen hybriden Robotersystemen bzw. Algorithmen, welche anhand der Beurteilung unterschiedlicher Pflegeszenarien ein individuelles moralisches Werteprofil des pflegebedürftigen Menschen erstellen könnten.

3. Resümee und Ausblick

Eine gute Pflegebeziehung zeichnet sich durch eine Begegnung in gegenseitigem Respekt, Vertrauen und eine wertschätzende Kommunikation aus. Neben Berührungen, Beobachtungen und Betrachtungen kann sie durch körperliche und seelische Interaktionen gekennzeichnet sein. Im Kontakt mit MmD sollten die Bedürfnisse und Bedarfe ganzheitlich betrachtet werden. Die Pflege von MmD zeichnet sich weitgehend durch eine Mensch-zu-Mensch-Interaktion aus. Durch Technik und zunehmende KI-basierte Pflegetechnologien (z.B. SAR) können diese bis zu einem gewissen Grad als unterstützend und nützlich im Rahmen einer fürsorglichen Beziehungsgestaltungangesehen werden. Dabei hängt die ethische Vertretbarkeit des Einsatzes von SAR zum einen von der Beurteilung der individuellen Pflegesituation durch die Pflegeperson und zum anderen von dem geäußerten Willen des pflegebedürftigen Menschen ab.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Der ACP kann für Pflegende, Angehörige und für Menschen mit einer beginnenden Demenz ein pflegerisches Hilfsmittel sein, um moralische Einstellungen zur roboterunterstützten Pflege zu reflektieren, Präferenzen festzuhalten und die Akzeptanz zur Einbindung in Therapieentscheidungen zu fördern.

In Bezug auf KI-basierte Pflegetechnologien ist festzuhalten, dass sich durch den Zugang und die Einbindung von systematischen Patienten- und Versorgungsdaten die Pflege künftig weiterentwickeln wird und neue Versorgungsmodelle entstehen werden (Meißner & Kunze 2019). Hauck (2019, 84) beschreibt dies treffend: »*Neue Technologien sind aus dem Alltag nicht mehr wegzudenken und werden sich noch stärker verbreiten. Es ist der »Dialektik der Aufklärung« (Horkheimer/Adorno 2013 [1944]) zu verdanken, dass es kein Zurück hinter erfolgte Entwicklungen gibt.*« Neben analytisch methodischen Fähigkeiten im Umgang mit großen Datenmengen, müssen Pflegende im Kontext ethischer Reflexion befähigt werden, eine sinnvolle Einbindung von KI-basierten Pflegetechnologien in weitere Versorgungsverläufe abwägen zu können. Zudem müssen sich Pflegende und pflegebedürftige Menschen stärker als bisher in institutionelle und gesellschaftlich-politische Auseinandersetzungen zur ethisch-moralischen Entscheidung über Einsatzbereiche von SAR einbringen. Da die Interaktion mit SAR »[…] grundlegende zwischenmenschliche Kulturpraktiken wie die der Sorge, der Fürsorge, der Empathie und generell der sozialen Interaktion betreffen« (Schule et al. 2021, 18–19).

Daneben gilt es, grundlegende rechtliche Fragen und bisherige Verantwortlichkeitskonzepte im Kontext der pflegerischen Sorgfaltspflicht zu prüfen und Themen der nutzerorientierten Technikentwicklung und Verteilungsgerechtigkeit im Zugang zur KI-basierten pflegerischen Versorgung kritisch zu hinterfragen und zu gestalten.

Literatur

Baisch, Stefanie et al. (2018): »Emotionale Roboter im Pflegekontext: Empirische Analyse des bisherigen Einsatzes und der Wirkungen von Paro und Pleo«, in: *Zeitschrift für Gerontologie und Geriatrie*, 51(1): 16–24.

Bartholomeyczik, Sabine et al. (2006): *Empfehlungen zur angemessenen Pflege von Menschen mit Demenz und herausforderndem Verhalten,* Bundesministerium für Gesundheit (BMG) (Hg.). Witten.

Beauftragte der Bundesregierung für die Belange behinderter Menschen (BMAS) (2018): *Die UN-Behindertenrechtskonvention. Übereinkommen über die Rechte von Menschen,* Bonn.

Bol, Nadine/Helberger, Natali/Weert, Julia C. M. (2018): »Differences in mobile health app use: A source of new digital inequalities?«, in: *The Information Society*, 34(3): 183–193.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Borcherding, Gesa/Hülsken-Giesler/Manfred/Meißner, Anne (2021): »Digitale Kompetenzen erwerben«, in: *Pflegez.*, 74(11): 38–41.

Broekens, Joost/Heerink, Marcel/Rosendal, Henk (2009): »Assistive social robots in elderly care: a review«, in: *Gerontechnology*, 8(2): 94–103.

Buber, Martin (2002): *Das dialogische Prinzip* (9. Aufl.), Gütersloh: Gütersloher Verl.-Haus.

Buchanan, Christine et al. (2020): »*Predicted Influences of Artificial Intelligence on the Domains of Nursing: Scoping Review*«, in: *JMIR Nursing*, 3(1): 1–15.

Costa, Angelo/Julian, Vicente/Novais, Paulo (Hg.) (2018): *Personal Assistants: Emerging Computational Technologies*, Basel: Springer International Publishing.

Deutsche Akademie der Technikwissenschaften (acatech) (2019): *Lernende Systeme – Die Plattform für Künstliche Intelligenz*.

Deutsche Gesellschaft für Gerontologie und Geriatrie; Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde; Deutsche Gesellschaft für Neurologie (DGGG) (Hg.) (2020): *Einwilligung von Menschen mit Demenz in medizinische Maßnahmen. Interdisziplinäre S2k-Leitlinie für die medizinische Praxis*, Stuttgart: Verlag W. Kohlhammer.

Exley, Catherine et al. (2009): »Advance Care Planning: An opportunity for person-centred care for people living with dementia«, in: *Dementia*, 8(3): 419–424.

Friedrichsen, Mike/Bisa, Peter-J. (Hg.) (2016): *Digitale Souveränität. Vertrauen in der Netzwerkgesellschaft*. Wiesbaden: Springer VS.

Friesacher, Heiner (2019): »Fürsorge - trotz oder mit Technik?«, in: Claudia Hauck und Charlotte Uzarewicz (Hg.), *I, Robot – I, Care. Möglichkeiten und Grenzen neuer Technologien in der Pflege*, Berlin: De Gruyter Oldenbourg: 27–48.

Givens, Jane L. et al. (2018): »Advance Care Planning in Community-Dwelling Patients With Dementia«, in: *Journal of pain and symptom management*, 55(4): 1105–1112.

Grossman, Molli R. (2018): »Mobile Apps for Caregivers of Older Adults: Quantitative Content Analysis«, in: *JMIR mHealth and uHealth*, 6(7): e162.

Haller, Susanne/Schnell, Martin W. (2019): »Advance Care Planning (ACP) - Behandlung im Voraus planen«, in: Martin W. Schnell und Christian Schulz-Quach (Hg.), *Basiswissen Palliativmedizin* (3. Aufl.), Berlin, Heidelberg: Springer: 237–253.

Hauck, Claudia (2019): »Neue Technologien und Pflegebildung - eine Annäherung«, in: Claudia Hauck und Charlotte Uzarewicz (Hg.), *I, Robot – I, Care. Möglichkeiten und Grenzen neuer Technologien in der Pflege*, Berlin: De Gruyter Oldenbourg: 63–87.

Hülsken-Giesler, Manfred (2007): *Der Zugang zum Anderen*. Götting: V&R unipress.

Jungbauer-Gans, Monika/Kriwy, Peter (Hg.) (2016): *Handbuch Gesundheitssoziologie*, Wiesbaden: Springer.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Kachouie, Reza et al. (2014): »Socially Assistive Robots in Elderly Care: A Mixed-Method Systematic Literature Review«, in: *International Journal of Human-Computer Interaction*, 30(5): 369–393.
Kerasidou, Angeliki (2020): »Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare«, In: *Bulletin of the World Health Organization*, 98(4): 245–250.
Klimova, Blanka/Toman, Josef/Kuca, Kamil (2019): »Effectiveness of the dog therapy for patients with dementia - a systematic review«, in: *BMC psychiatry*, 19(1): 1–19.
Kolarik, Russ C. et al. (2002): »Advance care planning«, in: *Journal of general internal medicine*, 17(8): 618–624.
Koutentakis, Dimitrios/Pilozzi, Alexander/Huang, Xudong (2020): »Designing Socially Assistive Robots for Alzheimer's Disease and Related Dementia Patients and Their Caregivers: Where We are and Where We are Headed«, in: *Healthcare*, 8(2): 1–15.
Krückel, Daniela (2015): »Roboter in der Pflege?«, in: *LQ*, 1(1): 3–7.
Kutzleben, Milena von (2016): *Häusliche Versorgung von Menschen mit Demenz*, Weinheim: Beltz Juventa.
Maier, Wolfgang et al. (2016): *S3-Leitlinie Demenzen,* in: AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.) (Hg.).
Manzoor, Mirfa/Vimarlund, Vivian (2018): »Digital technologies for social inclusion of individuals with disabilities«, in: *Health and technology,* 8(5): 377–390.
Meißner, Anne/Kunze, Christophe (Hg.) (2020): *Neue Technologien in der Pflege. Wissen, Verstehen, Handeln*, Stuttgart: W. Kohlhammer Verlag.
Misselhorn, Catrin (2018): *Grundfragen der Maschinenethik*, (2 Aufl.), Ditzingen: Reclam.
Palmdorf, Sarah et al. (2019): »Ethical Challenges by Using Assistive Technologies in Dementia Home Care – Potential of Advance Care Planning«, in: *Dilemata. Revista International Éticas Aplicadas,* 11(30): 13–26.
Piers, Ruth et al. (2018): »Advance care planning in dementia: recommendations for healthcare professionals«, in: *BMC palliative care*, 17(1): 1–17.
Ronquillo, Charlene Esteban et al. (2021): »Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative«, in: *Journal of advanced nursing,* 77(9): 3707–3717.
Sachverständigenrat für Verbraucherfragen beim Bundesministerium der Justiz und für Verbraucherschutz (SVRV) (2017): *Gutachten »Digitale Souveränität« des Sachverständigenrats für Verbraucherfragen.*
Schweda, Mark/Schicktanz, Silke (2021): »Ethische Aspekte co-intelligenter Assistenztechnologien in der Versorgung von Menschen mit Demenz« in: *Psychiatrische Praxis,* 48(1): 37–41.
Schulze, Hartmut et al. (2021): *Soziale Roboter, Empathie und Emotionen. Eine Untersuchung aus interdisziplinärer Perspektive*, Bern: TA-SWISS.
Sharkey, Amanda/Sharkey, Noel (2012): »Granny and the robots: ethical

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

issues in robot care for the elderly«, in: *Ethics Inf Technol.*, 14(1): 27–40.
Topaz, Maxim et al. (2019): »Extracting Alcohol and Substance Abuse Status from Clinical Notes: The Added Value of Nursing Data«, in: *Studies in health technology and informatics*, 264(1): 1056–1060.
van der Wardt, Veronika et al. (2017): »Adherence support strategies for exercise interventions in people with mild cognitive impairment and dementia: A systematic review«, in: *Preventive medicine reports*, 7(1): 38–45.
Vetter, Joachim et al. (2012): *Demenz und Selbstbestimmung*, Deutscher Ethikrat (Hg.), Berlin.
Wagner, Cosima (2014): *Robotopia Nipponica: Recherchen zur Akzeptanz von Robotern in Japan*, Marbug: Tectum Wissenschaftsverlag.
Wu, Ting Ting et al. (2021): »A machine learning model to predict critical care outcomes in patient with chest pain visiting the emergency department«, in: *BMC emergency medicine*, 21(1): 1–12.
Zhavoronkov, Alex et al. (2019): »Artificial intelligence for aging and longevity research: Recent advances and perspectives«, in: *Ageing research reviews*, 49(1): 49–66.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Erik N. Dzwiza-Ohlsen

Stay in Touch!

Phänomenologische Überlegungen zum Einsatz von Robotik und KI in der Pflege bei demenziellen Erkrankungen am Beispiel von PARO

1.

a. Der allgemeine Kontext: Robotik und KI in der Pflege von Personen mit Demenz

Immer mehr Menschen in älter werdenden Gesellschaften leiden an Formen demenzieller Erkrankungen: Weltweit sind aktuell 50 Millionen, 2030 voraussichtlich 82 Millionen und 2050 sogar voraussichtlich 152 Millionen Menschen direkt betroffen (vgl. WHO 2018, 6). Obwohl die Mehrheit der ca. 1,6 Millionen Betroffenen hierzulande durch private Pflege betreut wird (vgl. Kutzleben 2016), ist die Zahl derer, die dauerhaft in Pflegeeinrichtungen leben, aufgrund der sich verändernden Gesellschaftsstrukturen steigend. Nicht nur wird unsere Gesellschaft immer älter, sondern auch mobiler, sodass das Zusammenleben einer Familie in mehreren Generationen eher die Ausnahme als die Regel darstellt. Hinzu kommt, dass die private Pflege ab einem gewissen Fortschritt dieser irreversiblen und progredierenden Erkrankungen physisch und psychisch äußerst fordernd ist (vgl. Brose et al. 2016). Die Professionalisierung und Institutionalisierung der Altenpflege lässt sich als strukturelle Antwort auf diese Konstellation verstehen. Allerdings müssen Pflegeeinrichtungen nicht erst seit der Corona-Pandemie einen brisanten Spagat zwischen unterschiedlichen Interessen meistern: Medizinische, juristische und politische Rahmenbedingungen treffen auf ökonomische Kalküle, soziokulturelle Gegebenheiten und psychosoziale Ansprüche. Blickt man ganz konkret auf das Beispiel demenziell veränderter Personen im Kontext professioneller Pflege, so treffen Hygiene- und Sicherheitsvorschriften, Vorgaben der Pflegezeitbemessung, Fachkräftemangel und Ausbildungsdefizite auf veränderte Fähigkeiten zur Kommunikation und Interaktion bei individuellen Ansprüchen an Teilhabe, Autonomie und Lebensqualität. Angesichts dieser Ausgangslage ist der Druck besonders hoch, Pflege und Therapie nicht zuletzt durch technische Innovationen zu verbessern, ohne dabei die ethischen Maßstäbe aus dem Blick zu verlieren (vgl. auch

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

den 8. Altersbericht *Ältere Menschen und Digitalisierung*, Bundesministerium für Familie, Senioren, Frauen und Jugend 2020). Damit ist der allgemeine Kontext dieses Beitrags angerissen: der Einsatz von Technik in der Pflege und Therapie von Personen mit demenziellen Erkrankungen, genauer: der Einsatz des intelligenten Pflegeroboters PARO.

Wie auch der Beitrag von Alexander Hochmuth (in diesem Band) vor Augen führt, sind intelligente technische Assistenzsysteme in der Wirklichkeit der Pflege angekommen. PARO steht dabei exemplarisch für eine Entwicklung, im Zuge derer technische Assistenzsysteme in eine Domäne vordringen, die bis *dato* den fühlenden – d.h. menschlichen und tierischen – Lebewesen vorbehalten war: der sozialen Beziehungspflege im *therapeutischen Nahfeld*. Anstatt nun ethischen Fragen nachzuspüren (einen Überblick zum Einsatz von emotionaler Robotik in der Pflege von Personen mit Demenz geben Baisch et al. 2017), lautet das primäre Ziel des vorliegenden Beitrags, die lebensweltliche Interaktion und Kommunikation im therapeutischen Feld zwischen den Akteur:innen – also den demenziell veränderten Personen, dem Beziehungsroboter PARO und den Pflegenden im Kontext einer Pflegeeinrichtung – phänomenologisch zu durchdringen.

b. Die spezifische Aufgabe: Lebensweltliche Interaktion und Kommunikation im therapeutischen Feld verstehen

Gemäß einer phänomenologischen Psychopathologie demenzieller Erkrankungen (vgl. Dzwiza-Ohlsen 2021) geht es auf den folgenden Seiten nicht um neurodegenerative Prozesse einer Krankheit des Gehirns, sondern um das psycho- und soziodegenerative Erlebnis des Krank-Seins von individuellen Personen in ihrer nicht zuletzt technisch gestalteten Lebenswelt (vgl. Waldenfels 1971, 252 im Anschluss an Victor von Weizsäcker).[1] Im Sinne einer personenzentrierten Pflege nach Kitwood (1997), für die das Wohlbefinden der Person im Mittelpunkt steht, ist es entscheidend zu verstehen, wie die Betroffenen die Kommunikation und Interaktion bewerten und welche Ressourcen trotz des Fortschreitens einer demenziellen Erkrankung zur Verfügung stehen.

Dabei wissen wir insbesondere aus phänomenologischer Perspektive um die fundamentale Bedeutung unseres sensiblen, motiblen, expressiven aber auch vulnerablen und sich entziehenden Leibkörpers für unsere gesamte personale Existenz. Dabei konnten in den letzten Jahren

1 Eine analoge Unterscheidung trifft Carel, indem sie der *disease*, als quantifizierbarer körperlicher Tatsache, die *illness* gegenüberstellt, die sie als »the ›what it is like‹ qualitative Dimension as it is experienced and made meaningful by the ill person« (2017, 17) definiert.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

zahlreiche Autor:innen diese allgemeine Erkenntnis unter Stichworten wie *verkörpertes Gedächtnis* (vgl. Fuchs 2018; Summa 2012; Tewes 2020), *embodied expressivity* (vgl. Kontos 2005), *embodied memories* (vgl. Hydén 2018), *gestisch-kommunikatives Handeln* (vgl. Döttlinger 2018) oder *therapeutische Atmosphären* (vgl. Sonntag 2020) produktiv auf demenzielle Erkrankungen anwenden. Mit Hinblick auf den leiblichen Ausdruck kann bspw. Gestik das Verständnis von Wortsprache unterstützen; die Mimik dabei helfen, Emotionen und Sinnnuancen verständlich zu machen; und die Stimme den Verlauf der Konversation sowie die allgemeine Responsivität zwischen den Interaktionspartnern positiv beeinflussen (vgl. Smith et al. 2011; Berendonk/Stanek 2011). Da diese Einsichten bisher nur in Bezug auf menschliche Interaktionen getroffen wurden, zielt der vorliegende Beitrag darauf, diese mit Hinblick auf die Mensch-Maschine-Interaktion zu erweitern. Gemäß dieser Zielsetzung wertet dieser Beitrag neben empirischen Studien und philosophischen Theoremen zwei Dokumentarfilme aus: *PARO, ein Roboter mit Charme* von Leonie Bredl (2012)[2], der in Stuttgart am Richard-Bürger-Heim gedreht wurde, sowie *Roboter zum Kuscheln – Heilsam für Demenzkranke?* von Annette Wagner (2011), der in Bremen am Haus O'land gedreht wurde – beides Pflegeeinrichtungen speziell für Personen mit Demenz. Die leitende These der Untersuchung lautet, dass es insbesondere durch den zwischenleiblichen Ausdruck gelingt, mit PARO eine primär präreflexiv ablaufende Responsivitätsbeziehung herzustellen, die sich durch attentionale, intentionale und emotionale sowie perzeptive, affordante und praktische Elemente auszeichnet. Dabei geht es auch um die Frage, ob unter den gegebenen technischen Umständen – die nicht nur Intelligenz oder Lernfähigkeit demonstrieren, sondern auch Responsivität und Reziprozität simulieren – die Annahme eines Übergangs von einem technischen Es zu einem sozialen Du gerechtfertigt erscheint. Dabei gipfelt die Untersuchung in der Auseinandersetzung mit einer besonders intimen Form des Miteinanders, der sanften Berührung, die gerade in Zeiten der globalen Pandemie insbesondere in pflegerischen Settings erheblich unter Druck geraten ist (vgl. Winniewski 2022).

Um dieses Programm zu entfalten, verbindet der Beitrag die klassische Phänomenologie Edmund Husserls (1859–1938), mit der responsiven Phänomenologie Bernhard Waldenfels' (1934–) und der technikwissenschaftlichen Postphänomenologie Don Ihde's (1934–).[3] Alle Ansätze verbindet, so die Intuition des Beitrags, eine relationale Auffassung der

2 Verfügbar unter https://www.youtube.com/watch?v=KNM5OrZO8-w

3 Eine umfassende Einordnung der Postphänomenologie sowohl mit Hinblick auf die klassische Technikphänomenologie als auch Queer Phenomenology, Akteur-Netzwerk-Theorie, Science and Technology Studies liefert Müller (2020).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Lebenswelt, gemäß derer der kommunikative Ausdruck eine Schlüsselkategorie an den Grenzen der Relata darstellt. Am Beispiel von PARO wird der phänomenologische Blick auf das Was und Wie des Erfahrens durch das Wovon des Getroffenseins und Worauf des Antwortens responsiv ergänzt und medial durch das Technische erweitert, sodass die lebensweltliche Interaktion im therapeutischen Nahfeld in den Blick tritt. Bevor wir mit diesem Programm beginnen, bedarf es aber der Orientierung darüber, wer oder was PARO überhaupt ist, wo und wie er eingesetzt wird und mit welchen Effekten seine Nutzung verbunden ist.

2.

a. Grundlegendes zu PARO: Produktdesign, Einsatzbereich und Gesundheitseffekte

PARO ist eine in den 1990er Jahren in Japan am *National Institute of Advanced Industrial Science and Technology* entwickelte ›Kuschelrobbe‹, deren 8. Version seit 2005 und deren 9. Version seit 2013 kommerziell verfügbar ist (vgl. Shibata/Coughlin 2014, 420f.). In technischer Hinsicht handelt es sich bei PARO um einen durch KI (*Künstliche Intelligenz*) gestützten SAR (*Sozial Assistierenden Roboter*) oder auch HIR (*Human Interactive Robot*). Als IAR (*Interactive Autonomous Robot*) zeichnet sich PARO durch einen hohen Grad an Autonomie, Komplexität und Kommunikationsfähigkeit aus, was ihn zum Einsatz als *persönlichen* Roboter im sozialen Nahfeld von Therapie- und Pflegesettings disponiert. Dabei ist PARO als MCR (*Mental Commitment Robot*) insbesondere auf psychotherapeutische Anwendungen spezialisiert, sodass er als *Zuwendungsroboter* ein inzwischen etablierter Baustein in der individualisierten Gestaltung von Beziehungen in ansonsten oft reiz- und/oder beziehungsarmen Lebensumfeldern ist.[4]

Als *Bio-Roboter* wird PARO dort eingesetzt, wo eine *tiergestützte Therapie* bzw. *tiergestützte Aktivität* (*Animal-Assisted Therapy*, AAT/*Animal-Assisted Activity*, AAA) mit Hinblick auf die Personengruppe besonders sinnvoll erscheint, aber aufgrund der Lebensumstände nur schwer umsetzbar ist. Das Ziel für die RAT/RAA (*Robot-Assisted Therapy/Activity*) lautet demgemäß, die positiven physiologischen, psychischen und sozialen Effekte von AAT/AAA zu replizieren und dies trotz

4 PARO wurde zwar primär für den Einsatz bei demenziell veränderten Personen entwickelt, allerdings auch erfolgreich bei anderen Patient:innengruppen eingesetzt, bspw. während einer Chemotherapie, bei der Rehabilitation nach Hirninfarkt und bei Kindern mit Entwicklungsschwierigkeiten (vgl. Wada et al. 2008, 53; Shibata/Coughlin 2014, 419–22).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ungünstiger Rahmenbedingungen – wie bspw. Hygiene- bzw. Sicherheitsvorschriften von Seiten der Institution oder Ängste vor Tieren auf Seiten der Personen (vgl. Wada et al. 2008, 53; Shibata/Wada 2011, 383f.). Im Vergleich zu AAT (vgl. Soler et al. 2015) lässt sich PARO flexibel anwenden, sodass sich spontane und geplante, aktivierende und therapierende Modi in Gruppen- und Einzelinteraktionen abwechseln (vgl. Baisch et al. 2017). Genau für diese Nische wurde PARO entworfen und mit Erfolg eingesetzt: So führt der Einsatz von PARO bei hinreichender Schulung der Therapeut:innen (vgl. Shibata/Coughlin 2014, 421) zu überwiegend positiven Reaktionen und zahlreichen positiven, wenn auch kurzfristigen Effekten (vgl. Baisch et al. 2017). Der Einsatz von PARO, so Shibata und Wada (2011, 383) sowie Shibata und Coughlin (2014, 418–424), a) reduziert Stress, Schmerzempfinden, wahrgenommen Isolation und führt zur Entspannung und dies sowohl auf der Seite von Gepflegten wie Pflegenden (vgl.), b) reduziert Depressionen und aktiviert Kommunikation und Interaktion bei Bewohner:innen in Pflegeheimen mit und ohne demenziellen Erkrankungen insbesondere mit Hinblick auf die Aktivierung des verbalen Ausdrucks sowie der Verbesserung der Mimik (vgl. Shibata/Wada 2011, 383; Shibata/Coughlin 2014, 423), c) verbessert die Stimmung bei demenziell veränderten Personen sowie die gesamte Atmosphäre der Pflegeeinrichtung (vgl. ebd., 421, 424; Shibata/Wada 2011, 383), d) reduziert sogenanntes ›herausforderndes Verhalten‹ von demenziell veränderten Personen, wie verbale und physische Aggressionen, Ängstlichkeit, Agitation, Bewegungsdrang oder Fluchtverhalten (vgl. Shibata/Coughlin 2014, 421–24; Shibata/Wada 2011, 383), sodass e) die Gabe von Psychopharmaka reduziert werden konnte (vgl. Shibata/Coughlin 2014, 423f.). – Wie aber lassen sich diese beeindruckenden Effekte verstehen?

b. Die philosophische Herausforderung: Multi-hybride Technologie im therapeutischen Feld

Eine erste Annäherung an eine Antwort können wir durch den Blick auf das einzigartige Design von PARO vollziehen, der in mehrfacher Hinsicht ein Hybrid ist: *Einerseits* ist er eine intelligente Maschine (*technohyletisch*), die vom Erscheinungsbild sowohl einem lebendigen Tier als auch einem Kuscheltier ähnelt (*zoomorph*), vom Verhaltensrepertoire allerdings voll auf die Bedürfnisse von Menschen ausgerichtet ist (*anthropomorph*). *Andererseits* ist PARO ein Hybrid aus materieller und digitaler Technik, der sensomotorische Daten als Informationen mittels von KI verarbeitet, speichert und zur Gestaltung von zukünftigen Interaktionen einsetzt. Im Ergebnis kommt es durch diese multi-hybride Design zu einer einzigartigen Verschränkung von Schein und Sein:

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

PARO scheint niedlich, verletzlich und schutzbedürftig – ist jedoch robust, reparabel und sogar antibakteriell; PARO scheint wärmende Nähe zu schenken und eigene Gefühle sowie einen eigenen Willen zu haben – ist jedoch leblos, harmlos und berechenbar; und PARO scheint gut zuzuhören und individuelle Zuwendung zu schenken – vielmehr jedoch sammelt und verarbeitet PARO Informationen und ist unkompliziert und universell einsetzbar

Abstrakt betrachtet, handelt es sich also bei diesem multi-hybriden Wesen um ein ontologisches Grenzobjekt im Spannungsfeld zwischen Lebenswelt, Wissenschaft und Technik. Die philosophische Herausforderung besteht nun darin, das Verhältnis der sich im Spiel befindlichen, ontologischen Sphären verstehend zu durchdringen – und dies derart, dass weder die wahrgenommene Qualität der Kommunikation und Interaktion von Seiten der Betroffenen, noch die ontologischen Implikationen der technischen Gegebenheiten nivelliert werden. Um diese Herausforderung zu meistern, sei hier zunächst das spezifische ›Spielfeld‹ umrissen, in dem sich die jeweiligen Akteur:innen begegnen: das *therapeutische Feld*.

In der aktuellen Forschung am Schnittfeld von Phänomenologie, Anthropologie und Lebenswissenschaften wird oftmals ein unterkomplexes Bild dessen gezeichnet, was hier als therapeutisches Feld bezeichnet wird. Bspw. wird einerseits die phänomenologische Perspektive mit der subjektiven Erfahrung aus der ersten Person-Perspektive verknüpft, die dem lebendigen Leib nachspürt und so einen Zugang zur subjektiven Seite von Krankheit ermöglicht; andererseits wird dies der objektivierenden dritte-Person Perspektive der Naturwissenschaften gegenübergestellt, die vermeintlich nur physische Körper kennt und qua Objektivierung einen reduzierten Blick auf Krankheit eröffnet (vgl. Carel 2016; Fuchs 2020). Wie wir am Beispiel der Interaktion zwischen PARO, Pflegenden und Gepflegten sehen werden, bedarf es hier eines viel komplexeren Bildes, dessen Grundlinien bereits hier skizziert werden sollen:

Im therapeutischen Feld durchdringen sich nicht nur unterschiedliche personale Perspektiven im Rahmen eines responsiven und dialogischen Miteinanders, das stets auf das Telos des gelingenden Umgangs mit dem Krank-Sein ausgerichtet bleibt, sondern es sind auch nicht-menschliche Akteure prägend, wie Tiere, Kuscheltiere, Roboter, Maschinen, Werkzeuge usw.; zudem sind die Interaktionen zwischen den Akteur:innen immer schon durch unterschiedliche Einstellungen geprägt, die, mal mehr und mal weniger, der ontologischen Signatur ihrer Akteure zu entsprechen versuchen und sich gegenseitig überlagern bzw. durchdringen. *Einerseits* befinden sich alle personalen Akteur:innen in der sogenannten ›personalistischen Einstellung‹ der geistigen Welt bzw. Lebenswelt[5]: Dies

5 Zur Interpretation der geistigen Welt als Husserls erste Phänomenologie der Lebenswelt vgl. Sommer 1984; Dzwiza-Ohlsen 2019.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ist die Einstellung, »in der wir allzeit sind, wenn wir miteinander leben, zueinander sprechen, einander im Gruße die Hände reichen, in Liebe und Abneigung, in Gesinnung und Tat, in Rede und Gegenrede aufeinander bezogen sind.« (Husserl 1952, 183) Innerhalb dieser Einstellung geht es den fühlenden, denkenden und handelnden Personen strukturell darum, die subjektiven Motivationen der Gegenüber auf der Basis von sinnvollem Ausdruck innerhalb unserer Umwelt zu verstehen. Blickt man aus dieser Einstellung auf PARO, so erscheint dieser im hohen Maße mit praktischen und werthaften Prädikaten verknüpft. *Andererseits* können wir aber auch unsere Einstellung wechseln und PARO als technisches Produkt einer durch Kausalität regulierten materiellen Welt auffassen, das ohne die physikalischen, mathematischen und informatischen Grundlagen der modernen Naturwissenschaften undenkbar wäre. Hier wechseln wir *zweitens* in die ›naturalistische Einstellung‹, die »keine Wertprädikate und keine praktischen Prädikate [kennt]. Begriffe wie wertvoll, schön, lieblich, reizend, vollkommen, gut, nützlich [...] haben in ihr keine Stelle« (Husserl 1952, 25). Da jedoch Technik ein mediales Phänomen *sui generis* darstellt; und sich die heuristische Trennung zwischen Lebenswelt und Wissenschaft im Lichte des Technischen transformiert (vgl. Ihde 2009, 44), sodass es bspw. zu einer durch beide Quellen informierten Perspektive von Designer:innen kommt, soll hier noch eine *dritte* Einstellung unterschieden werden: die *technische Einstellung*.

3.

a. Wie ist PARO gestaltet? Sensomotorik, Algorithmik und Verhalten

Blicken wir aus der technischen Einstellung auf PARO: Dieser ist mit zahlreichenden Sensoren zur Erfassung von taktilen, auditiven, visuellen, thermischen und posturalen Reizen sowie von sogenannten Aktuatoren zur Bewegungssteuerung ausgestattet (vgl. Wada et al. 2008, 53). Hervorzuheben ist, dass a) sein Fell mit unzähligen hochempfindlichen Sensoren ausgestattet ist, um die menschliche Berührung möglichst genau zu erfassen, b) er Helligkeit und Intensität von Licht erfasst, c) nicht nur für die Richtungs-, sondern auch für die Spracherkennung ausgestattet ist, er d) seine Körperposition im Raum bestimmen kann und e) er seine Körpertemperatur konstant hält (vgl. ebd. 55; Shibata/Coughlin 2014, 420). Zusätzlich zum sinnlich anregenden Design, das durch Gewicht (8. Generation = 2,7 kg, 9. Generation = 2,5 kg) und Länge dem tierischen Vorbild sehr nahe kommt, verfügt PARO über ein multimodales und non-verbales Ausdrucksrepertoire, das Mimik, Gestik und Lautproduktion umfasst: *Erstens* den jeweils voneinander unabhängigen Wimpern-/

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Lidschlag zweier großer dunkler Augen; *zweitens* die vertikalen und horizontalen Bewegungen des Kopfes sowie Paddelbewegungen der Flossen; und *drittens* mehrere unterschiedliche Laute. Dabei vollziehen sich die Ausdrucksbewegungen von PARO in einem moderaten Tempo.

Was aber sind die technischen Grundlagen seines Verhaltens? Wie die Entwickler:innen angeben, verfügt PARO über ein sein Verhalten generierenden Algorithmus, der auf zwei hierarchisch aufgebauten Schichten abläuft und drei Verhaltensweisen erzeugt: proaktive, reaktive und physiologische (vgl. Wada et al. 2008, 55). Um bspw. proaktives Verhalten zu erzeugen, kümmert sich eine Schicht um die Planung und die andere um die Ausführung im Verhältnis zu unterschiedlichen Faktoren. Durch ein sich selbst verstärkendes Lernen werden ähnlich dem Langzeitgedächtnis spezifische Relationen zwischen einerseits externen Stimuli und internen Zuständen und andererseits zwischen Planung und Handlung evaluiert, sodass sich im Verlauf der Interaktionen eine Art Charakter mit Vorlieben, Abneigungen und Eigenarten ausbildet. Derart vermag PARO sein Ausdrucksverhalten individuell auszurichten. Die Struktur dieses Interaktionsgeschehens fassen die Shibata und Coughlin (2014, 419) als *Bio-Feedback*:

> PARO stimulates the senses and brains of users, brings out their knowledge of and experience with animals, children, and/or others, evokes emotion in touching with them, and reminds them of their past experience and memories. Features such as form, touch, animal-like behavior, and reaction to user stimuli are important.

Beachtenswert ist bei dieser Beschreibung, dass sich technische, naturalistische und personalistische Perspektiven überschneiden, was genau zum hybriden ontologischen Status von PARO passt. Versuchen wir nun näher zu beschreiben, wie uns PARO lebensweltlich anspricht. Wie wir sehen werden, folgen dessen Designer:innen einem Credo, dessen postphänomenologische Formulierung lautet: »What matters for the human-robot relation is how the robot appears to human consciousness« (Coeckelbergh 2011, 199).

b. Wie spricht PARO uns an? Affordanz, Affektivität und Vertrauen

Als süße, flauschige und gerade erst zur Welt gekommene Babyrobbe ›will‹ PARO im Sinne der *Affordanz-Theorie* (vgl. Gibson 1979) berührt, gestreichelt und gekuschelt werden – ja, PARO ›will‹ behütet, beschützt und auf den Arm genommen werden. Damit sich diese Affordanz möglichst frei entfalten kann, haben die Designer:innen im Sinne der Wirkungs- bzw. Akzeptanzforschung ein Tier gewählt, das zwar bekannt, aber nicht, anders als Katzen oder Hunde, aus dem Alltag zu vertraut

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ist (vgl. Wada et al. 2008, 55; zum Phänomen des *uncanny valley* bei zu lebensechtem Design vgl. Pfadenhauer/Dukat 2017, 252). PARO weckt derart keine falschen Erwartungen oder traumatische Erinnerungen und erzielt im Ergebnis eine hohe kulturübergreifende Akzeptanz (vgl. Shibata/Wada 2011, 382). Um die Interaktion zu rahmen, setzen die Designer:innen gezielt das Kindchenschema[6] ein, wobei PARO in dieser Hinsicht sogar sein biologisches Vorbild übertrifft: So ist der Kuschelroboter nicht nur ein biomorphes ›Kindchen‹, sondern auch ein anthropomorphes, indem es durch einen Schnuller geladen wird und dazu in eine Art Brutkasten gesetzt werden kann. Im Idealfall wird beim Gegenüber ein zutiefst emotionales, präreflexiv ablaufendes und leiblich vermitteltes Verhaltensrepertoire gegenüber schutzbedürftigen Lebewesen aktiviert. Dazu seien hier zwei prägnante Szenen aus der Dokumentation *Roboter zum Kuscheln* wiedergegeben: In der ersten sagt ein älterer Herr mit fortgeschrittener Demenz zur Pflegerin bei deutlich vernehmbarem Lachen: »Sei vorsichtig!«, woraufhin diese amüsiert entgegnet: »Ja, ich bin vorsichtig. Gut, dass Sie mich darauf hinweisen!« (48:52–48:58); in der zweiten Szene kommentiert eine ältere Dame, ebenfalls mit fortgeschrittener Demenz, das Fiepen von PARO mit: »Oh ja, Mutti«, woraufhin die Therapeutin auf ihr amüsiertes Lachen entgegnet: »Oh, ja: Mutti hat er gerufen, ne?« (48:59–49:05). – Offensichtlich werden wir durch den emotionalen Roboter deswegen so stark affiziert, weil uns nicht nur sein Äußeres anrührt, sondern sein responsives Design (s.u.) den Eindruck erweckt, dass er selbst emotionale Zustände und Bedürfnisse erlebt (vgl. Pfadenhauer/Dukat 2017, 251).

Will man diesen affektiv getönten und rollenspezifisch adressierten Affordanzcharakter in *ontogenetischer Hinsicht* befragen, so lohnt ein Blick auf die Ausführungen, die Thomas Fuchs (2015) unter dem Stichwort der *Oikeiosis* trifft: Analog der Husserlschen *Urdoxa* bestimmt Fuchs das *Urvertrauen* als selbstverständlichen Grundcharakter der Lebenswelt, den er in zweifacher Weise differenziert: Als ›*Vertrauen in* jemanden oder etwas‹ und ›*Vertrautheit mit* jemandem oder etwas‹. Gemäß Fuchs beginnt sich unser Urvertrauen als reziproke Beziehung in der frühkindlichen Sozialisierung zu entfalten, sodass es als stabile Basis des Vertrautwerdens mit der Lebenswelt fungieren kann. Für unsere Fragestellung ist es nun entscheidend, dass diese Doppelstruktur des ›Vertrauens in… ‹ und ›Vertrautseins mit… ‹ nicht nur andere menschliche Akteur:innen, sondern auch tierische, technische, quasi-tierische oder quasi-menschliche Akteur:innen, wie bspw. Haustiere, Kuscheltiere, Puppen oder eben Roboter umfasst.

6 Dieses Beispiel zeigt, dass es sich bei dem Kindchenschema nicht nur über ein die menschlichen Kulturen und biologischen Spezies übergreifendes Phänomen handelt, sondern auch auf technische Entitäten übertragen werden kann.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Soll nun allerdings die Interaktion zwischen Mensch und Roboter im therapeutischen Feld analog der Beziehung zwischen Patient:innen und Therapeut:innen von Vertrauen getragen werden, so sind laut Kellmeyer et al. (2018, 1) vier Faktoren ausschlaggebend: (1) Sicherheit, (2) geteilte Intentionalität bzw. Adaptativität, (3) Vorhersehbarkeit des Verhaltens und (4) Ab- bzw. Einstimmung (*Attunement*). Kommentieren wir diese Punkte kurz, um diese im Verlauf gezielt zu vertiefen:

(1) PARO's Sicherheit ist nicht nur von technischer Seite gewährleistet, vielmehr wird der Umgang mit ihm auf eine derart selbstverständliche Weise als sicher wahrgenommen, dass es zu einer *Fürsorgeinversion* kommt und wir ihn behüten und beschützen wollen – besonders eindrücklich ist dies bei liegenden Patient:innen, die sich PARO typischerweise auf den Korpus legen, was mit lebendigen Tieren in diesem Maße kaum vorstellbar wäre.

(2) PARO erweckt nicht den Anschein, als ob wir mit ihm eine geteilte Form der Intentionalität auf ein Drittes vollziehen könnten, da er weder non-verbal bzw. verbal zeigen kann, noch zur Lokomotion bzw. Objektmanipulation in der Lage ist. Allerdings ist im Sinne der Adaptivität eine *Fürsorgeindividualisierung* auf der Basis seiner KI möglich.

(3) PARO verfügt einerseits über ein begrenztes Verhaltensrepertoire, andererseits bildet er durch sein evaluatives Design eine Art Charakter aus, durch den man »einigermaßen erwarten« kann, wie sich der individuelle PARO in Zukunft »benehmen wird« (Husserl 1952, 240). Sein Verhalten ist also im Allgemeinen vorhersehbar, besondert sich jedoch in dem unter (2) geschilderten Sinne.

(4) Schließlich ließe sich das Geheimnis von PARO vielleicht am besten dadurch erfassen, dass sich ein quasi-dialogischer Prozess der Ab- bzw. Einstimmung im Sinne des Attunements vollzieht. Oder in den Worten von Pfadenhauer und Dukat (2017, 251): bei PARO »drängt sich nicht der Eindruck eines technischen Apparats, sondern eines Dings auf, das ansprechend und ansprechbar zugleich wirkt.« Genau aus diesem Grund wenden wir uns nun in Anknüpfung an unsere Überlegungen zur Affordanz dem Begriff der Responsivität zu, der als Schlüssel zum Schloss der gelingenden Interaktion mit PARO angesehen werden kann.

4.

a. Wie können wir PARO antworten? Modi der Responsivität und Relationalität

Wollen wir den uns lebensweltlich treffenden Affordanzcharakter von PARO responsivitätstheoretisch besser einordnen, so scheint dieser

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

irgendwo zwischen dem *Reiz der Dinge* und dem *Wirken von Personen* zu oszillieren. Auf der einen Seite *reizt* uns PARO, er »klopft an die Pforte des Bewußtseins«, er »zieht praktisch an, *es* [er] *will gleichsam ergriffen sein*« (Husserl 1952, 220, meine Hervorhebung). Neben dem Sonderfall der Irresponsivität, die uns sowohl in alltäglichen Erscheinungsformen (bspw. Schlaf) als auch in pathologischen (bspw. Aufmerksamkeitsstörung) begegnet, können wir diesen Reizen »bald nachgeben, bald auch ihnen widerstehen«, jedoch nur bedingt darüber entscheiden, was wir wann erleiden (Husserl 1952, 189). Demgemäß erfahren wir uns nicht nur als freies und aktives, sondern auch als passives Ich: Als ein »›Ich der Tendenzen‹« und der »›Zustände‹«, das betroffen und zur Reaktion genötigt wird (Husserl 1952, 213). Genau diese Nötigung wird durch das Design von PARO derart auf die Spitze getrieben, dass wir uns fragen müssen, ob PARO uns ›nur‹ wie ein Ding *reizt* – oder ›schon‹ wie eine Person auf uns *wirkt*: So ›wirken‹ Personen aufeinander ein und ›richten‹ sich gezielt einander zu (vgl. Husserl 1952, 192). Wieder scheint es so, dass PARO am treffendsten als Hybrid beschrieben werden muss: Als technisches Artefakt vermittelt er als *Zeugnis* die Intentionen seiner Entwickler:innen an die Nutzer:innen, indem PARO als *Erzeugnis* (vgl. A. Schütz nach Pfadenhauer/Dukat 2017, 250) deren Intentionalität, Attentionalität und Emotionalität in der Interaktion spezifisch motiviert. Er ist also, analog einem Tier, weder reizendes Ding noch wirkende Person.

Passend zu diesem multi-hybriden Status des Bio-Roboters formuliert Christoph Scholtz die These, dass »gesunde Menschen im Umgang mit subjektsimulierenden Maschinen den Modus des zweifachen Bewusstseins ausbilden« (*Roboter zum Kuscheln*, 27:30–27:53). Die Interaktionspartner:innen sehen also einen Roboter *als ob* dieser ein Tier wäre, ohne aber *ein solches* zu werden. Dass diese Ambivalenz sich in der Art und Weise reflektiert, wie die Interaktionspartner:innen responsiv sind, lässt sich durch eine Szene illustrieren, die sich bei der Einarbeitung von Pflegekräften in der Dokumentation *Roboter zum Kuscheln* abspielt: Zunächst äußert eine der Pfleger:innen beim Versuch, das Ladekabel einzustecken: »Ich will ihm nicht wehtun«, was von allgemeinem Gelächter in der Gruppe begleitet wird; im Anschluss versucht eine der Pfleger:innen PARO am Reißverschluss an der Bauchseite zu öffnen, woraufhin die benachbarte Pfleger:in äußert: »Ich will ihn gar nicht von innen sehen« und sich daraufhin mit der Hand vor den Augen abwendet. Während eine weitere Pfleger:in versucht, die Handgriffe der Kolleg:in mit den Instruktionen der Bedienungsanleitung abzugleichen, kommentiert eine andere Pfleger:in das Prozedere mit: »das nimmt die Illusion«, wieder begleitet von allgemeinem Gelächter (06:16–06:30). Neben der Ambivalenz des zweifachen Bewusstseins veranschaulicht diese Szene auch, dass hier unterschiedlichste Relationen mit dem Technischen im Spiel sind, derer Ihde (1990, 72–112)

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

vier unterscheidet: *verkörperte Relationen* (Einstecken des Ladekabels), *hermeneutische Relationen* (Verstehen der Bedienungsanleitung), *Hintergrund-Relationen* (Selbstverständlichkeit von Elektrizität) sowie *Alteritäts-Relationen* (Widerspruch des zweifachen Bewusstseins; zur Alterität des Technischen vgl. auch den Beitrag von Oliver Müller in diesem Band). Aber was passiert, wenn diese Relationen nur sehr reduziert abgerufen werden *und* das zweifache Bewusstsein nicht gegeben ist und der Roboter entweder als Tier oder als Roboter aufgefasst wird? Genau zu einer solchen Auflösung der Ambivalenz scheint es bei Personen mit demenziellen Erkrankungen gehäuft zu kommen: Gemäß einer Studie von Baisch et al. (2017, 6), bei der Mehrfachnennungen möglich waren, fasste die Mehrzahl der Gepflegten PARO als lebendig auf (86 %), ein hoher Anteil als unlebendig (60 %) und ein sehr geringer Teil als Hybrid (<10 %). Wir werden auf dieses Ergebnis noch zurückkommen; zunächst richten wir uns, responsivitätstheoretisch informiert, der konkreten Interaktion zwischen den Akteur:innen im therapeutischen Feld zu.

b. Wie antworten wir PARO? Konkrete Responsivität im therapeutischen Feld

Beobachtet man die dyadische Interaktion von Personen von demenziellen Erkrankungen mit PARO, dann realisiert sich diese auf zwei typische Weisen: *Erstens* eine Face-to-Face-Interaktion ›auf Augenhöhe‹, die aufgrund ihrer kurzen Distanz stark der Ansprechhaltung bei Neugeborenen ähnelt und *zweitens* eine Body-to-Body-Interaktion ›auf Brust- bzw. Bauchhöhe‹, die stark der Wiegehaltung bei Säuglingen bzw. Kleinkindern ähnelt.

Im *ersten* Fall wird die Kontaktaufnahme primär durch einen leicht gesenkten Blick realisiert, der durch eine aufrechte Positur des Oberkörpers und einer gestreckten oder geneigten Kopfhaltung begleitet wird und derart eine Distanznahme von ungefähr einer halben Armlänge erzeugt; diese Distanz scheint optimal für das Kraulen von Hals und Gesicht bei gleichzeitigem Einsatz von Mimik (vgl. *Roboter zum Kuscheln*, 27:03–27:30). Im *zweiten* Fall wird die Kontaktaufnahme durch einen umarmenden Griff realisiert, der ohne eine Distanznahme zum Korpus erfolgt und so ideal für das intime Halten, Streicheln, Klopfen, Trösten und Kuscheln ist (vgl. ebd., 49:09–49:19).[7] Eine Variante des zweiten Falls ist das Zusammensein mit PARO in einem entspannten, unthematischen und präreflexiven Sinne, sodass PARO Halt gibt, Ruhe vermittelt und gestreichelt und liebkost wird (vgl. ebd., 50:30–50:38).

7 Natürlich können sich beide Formen abwechseln bzw. durchdringen, was insbesondere dann zu beobachten ist, wenn Betroffene bettlägerig sind.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Beiden Formen ist gemeinsam, dass sich eine wechselseitige Orientierung im leiblichen Nahraum vollzieht, der – wie die allgemeine Struktur der Lebenswelt – multidimensional verfasst ist. Einerseits durchdringen sich die räumliche, zeitliche und personale Dimension ständig in einer dynamischen Weise (vgl. Dzwiza-Ohlsen 2019). In therapeutischer Hinsicht ist vor diesem Hintergrund zu beachten, dass auch im späten Verlauf demenzieller Erkrankungen eine minimale situative Orientiertheit im Sinne des Bühler'schen »Origo« – also dem »hier–jetzt–ich–System der subjektiven Orientierung« (Bühler 1982, 102, 149) sowie das ›situative Gedächtnis‹ unseres Leibes relativ lange erhalten bleibt (vgl. Fuchs 2018, 53–57). Andererseits gestaltet sich dieses dreidimensionale Orientierungsnetz innerhalb der konkreten Situation weiter aus: So können wir von einer kommunikativen Orientierung der Akteur:innen im therapeutischen Nahfeld sprechen, die durch responsive und verkörperte Formen der Attentionalität, Emotionalität und Intentionalität geprägt ist.[8]

Allerdings wird PARO als »sozialer Mediator« (Shibata/Wada 2010, 384, meine Übersetzung) im Rahmen der *robot-assisted therapy* meist von geschulten ›Moderator:innen‹ begleitet. Diesen geht es innerhalb der triadischen Konstellation typischerweise um zweierlei: die Kontaktaufnahme herzustellen und Beziehungsgestaltung zu ermöglichen.

Mit Hinblick auf die *Kontaktaufnahme* bringen die Moderator:innen PARO in das therapeutische Nahfeld in Form eines *Demonstrierens* ein, wodurch es zu einer »›*Aktivierung im Vorübergehen*‹« (Pfadenhauer/Dukat 2017, 253) kommen kann; sobald die triadische Interaktionssituation erreicht ist, kommt es zur gestischen Form des »*Darreichens*« (ebd., 254) an die Zielperson. Wie Pfadenhauer und Dukat (2017, 254) ausführen, geben die Moderator:innen erklärende Hinweise und anregende Anweisungen, die sowohl verbal als auch non-verbal realisiert sein können. So kommt es bspw. zur *gaze alteration*, also einem permanenten Blickwechsel der Moderator:innen zwischen PARO und den Interaktionspartner:innen, der typischerweise von einem Rein- und Rauslehnen aus der Blickachse der Dyade begleitet wird (vgl. *PARO, ein Roboter mit Charme*, 08:30–09:55). Derart versuchen die Moderator:innen die Aufmerksamkeit auf PARO zu lenken und, in der Terminologie Husserls, den Übergang vom *unthematischen Hintergrund potenzieller intentionaler Objekte* in den *thematischen Vordergrund* zu vollziehen (vgl. Husserl 1976, 73f.). Im Ergebnis gelingt es gemeinsam mit PARO nach meist anfänglicher Zurückhaltung einen Kontakt zu stiften (vgl. bspw. *Roboter zum Kuscheln*, 07:30–09:00), obwohl demenziell veränderte Personen

8 Im Sinne der Ausführungen zum Ende des Beitrags könnte man auch über eine propriozeptive und zwischenleibliche Orientierung durch das Fühlen des Roboters – von Textur, Gewicht, Temperatur – nachdenken, die typischerweise präreflex abläuft.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

im fortschreitenden Verlauf oft eine verminderte Ansprechbarkeit und Aufmerksamkeitsspanne haben. Im Sinne einer möglichst unbeschwerten Kontaktaufnahme eröffnet sein reaktives und non-verbales Design die ideale Schnittstelle: Er »hört zu, ohne Fragen zu stellen, und ohne das Falsche zu sagen, er hört einfach nur zu« (49:07–49:14).

Mit Hinblick auf die *Beziehungsgestaltung* können wir Handlungen wie das verzückte ›Auf-den-Arm-Nehmen‹ von PARO in Analogie zu einem kleinen Kind oder das durch einen affirmierenden Ton begleitete Kraulen am Hals in Analogie zu einem Haustier beobachten, das von den Moderator:innen performativ vorgemacht werden kann. In phänomenologischer Hinsicht handelt es sich hierbei um ein habituelles und meist präreflexiv ablaufendes Repertoire an Fürsorgehandlungen, in denen Leiblichkeit, Affektivität und Sozialität aufs engste verknüpft sind. Wie Fuchs (2018, 53–56) jüngst betont hat, lassen sich Habitualisierungen als eine Art implizites Gedächtnis betrachten, welches sich in prozeduraler, situativer und zwischenleiblicher Weise differenzieren und auf spezifische Weise reaktivieren lässt, um nicht zuletzt autobiographische Aspekte zugänglich zu machen. Ganz im Sinne der sozialkonstruktivistischen *Positioning-Theory*, die Sabat und Harré (1994) erfolgreich auf demenzielle Erkrankungen angewendet haben, aktualisieren Betroffene ihr soziales Selbst in Kooperation mit PARO (vgl. *Roboter zum Kuscheln*, 34:20–34:55).

Durch den Einsatz von PARO wird Responsivität trotz demenzieller Erkrankungen auf potenziell vielschichtige Weise realisiert. Sein Design setzt ganz darauf, die »Bedeutung körperlich-affektiver Interaktion« bei fortgeschrittener Erkrankung zu würdigen (Meyer 2014, 108), in der sich die vielfältigen Formen der Sorge, die auch die Fürsorge einschließen, nicht nur als »leibhafte«, sondern als »zwischenleibliche« situieren (Waldenfels 2019, 299). – Aber inwiefern ist bei PARO die Rede von einer Zwischenleiblichkeit überhaupt gerechtfertigt? Und handelt es sich bei PARO überhaupt um einen sozialen Akteur, wie es die bisherigen Beschreibungen nahelegen?

5.

Wie wirkt PARO? Empathie, Expressivität und Reziprozität

Zunächst lautet die grundsätzliche Frage, wie jemand oder etwas zu einem sozialen Akteur wird – beziehungsweise was verhindert, dass dies geschieht. Gemäß Husserl bildet das entscheidende Eingangstor zu einer Antwort der sinnhafte und verstehbare Ausdruck, denn es gilt: ohne Ausdruck keine Kommunikation, ohne Kommunikation keine Sozialität

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und ohne Sozialität keine Lebenswelt (vgl. Husserl 1952, 186–196). Von entscheidender Bedeutung für diesen konstitutiven Gesamtzusammenhang ist die Verbindung von Ausdruck und Empathie bzw. Einfühlung, die wir uns zunächst am Falle von Personen vergegenwärtigen können.

Dazu schreibt Husserl: »Einfühlung in Personen ist nichts anderes als diejenige Auffassung, die eben den Sinn versteht« (Husserl 1952, 244). Analog zum Sinn eines Wortes, der durch den schriftlichen Ausdruck hindurch scheint, erfassen wir das sinnvolle Ganze einer Person durch ihren Ausdruck hindurch (vgl. ebd.). Allerdings ist es innerhalb der lebensweltlichen Einstellung keineswegs so, dass wir den »Geist dem Leibe« noch ›einlegen‹ würden oder ihm ›aufpfropfen‹ (Husserl 1952, 191). Vielmehr »verstehen« wir unser Gegenüber »ohne weiteres als personales Subjekt« (Husserl 1952, 191). Aber wie verhält sich dies im Falle von PARO? Mit Husserl müsste man sagen: analog. Denn grundsätzlich gilt in konstitutiver Hinsicht, dass ich auch hier »appräsentiere« (Husserl 1973a, 28): Ich verknüpfe das mir Präsente (die Wahrnehmung des Ausdrucks des Gegenübers) mit dem mir nicht Präsenten (die Innerlichkeit des Gegenübers, seine Empfindungen) auf der Basis meiner eigenen Erfahrungen. Um nun *in concreto* festzustellen, dass »dies ein Mensch und keine Puppe« ist, bedarf es, so Husserl, des Ausdrucks als Quelle der Bewährung (Husserl 1973a, 25). An diesem Punkt ist es entscheidend zu beachten, dass Husserl von einem sehr breiten Ausdrucksbegriff ausgeht: Ausdruck kann sowohl mit oder ohne Absicht zur Kommunikation sein, und heißt dementsprechend »kommunikativ« (1) oder »nicht kommunikativ« (2); er kann sowohl »sprachlich« (3) als auch »nicht sprachlich« (4) verfasst sein; und er kann auf unterschiedlichste Weise »habituell« (5) werden (Husserl 1973a, 63; vgl. Husserl 1952, 192). Liegen nun keine ›Empathieblockaden‹ (Breithaupt 2017) vor – neben den bereits genannten individuellen Faktoren, ließen sich auch kulturelle Faktoren anführen, wie bspw. Geschlechterstereotype oder Technophobie (vgl. Wada et al. 2008, 57) – dann fühlen sich die Bezugspersonen vermittels des kommunikativen, nicht-sprachlichen und habituellen Ausdrucks in die als sinnvoll anerkannten Motive des Bio-Roboters ein. – Rechtfertigt dies aber schon die Annahme, dass PARO, ganz wie die uns bekannten Lebewesen, über einen Leib verfügt? Und können wir sogar noch weiter gehen und ihm auf der Basis seines Ausdrucks eine Persönlichkeit zuerkennen? (zur lebensweltlichen Interaktion mit Tieren bzw. Haustieren vgl. Dzwiza 2018).

Wie wir gesehen haben, scheinen viele Interaktionspartner:innen – ob nun implizit oder explizit – anzuerkennen, dass der Roboter nicht nur einen extensiven und expressiven Körper, sondern auch einen sensitiven und sensiblen Leib hat. Passend zum intelligenten und lernenden Design von PARO verstehen wir dann den Ausdruck seines Körpers als einem dem Leib analogen »Willensorgan« (Husserl 1973a, 93), wodurch

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

die »Idee der Persönlichkeit« erwachsen kann (Husserl 1973a, 244).[9] In dieser Hinsicht ist insbesondere der habituelle Ausdruck von PARO relevant: *Einerseits* und in *intra*subjektiver Hinsicht bildet PARO durch seine evaluative KI eine Art Charakter aus, durch den man, wie bereits gesagt, »einigermaßen erwarten« kann, wie sich PARO in Zukunft »benehmen wird« (Husserl 1952, 240). *Andererseits* und in *inter*subjektiver Hinsicht ist sein habitueller Ausdruck an den Tag-Nacht-Rhythmus geknüpft, der als synchrone Ordnung zu einer gemeinsamen Orientierung in der Zeit führt, die sich als Ordnung der alltäglichen Vollzüge (bspw. Wachen und Schlafen, Essen und Ruhen) wechselseitig stabilisiert (vgl. Wada et al. 2008, 55–56). Vollziehen wir aber eine derart qualifizierte Form der Einfühlung und unterstellen PARO Leiblichkeit und Persönlichkeit, dann begreifen wir sein »Bewusstsein selbst als [bestehend aus] ›Motivationszusammenhängen‹« (Husserl 1973a, 88) – PARO wäre dann nicht einfach ein willenloses Glied innerhalb des kausal determinierten, materiellen Universums, sondern sein Verhalten vielmehr ein sinnvoller Ausdruck eines durch Motive getragenen Gesamtzusammenhangs.

Das unsere Einfühlung allerdings Grenzen hat, soll, diesen ontologischen Gedankengang abschließend, am Begriff der Reziprozität ausgeführt werden. Rufen wir uns dafür die Husserl'sche Definition in Erinnerung:

> »I1 erfasst einfühlend I2, und I2 einfühlend I1, aber nicht nur das: I1 erfährt (versteht) I2 als I1 verstehend Erfahrenden, und umgekehrt. Ich sehe den Anderen als mich Sehenden und Verstehenden, und es liegt weiter darin, dass ich ›weiß‹, dass der Andere auch seinerseits sich als von mir gesehen weiß. Wir verstehen uns und sind im Wechselverständnis geistig beieinander, in Berührung.« (Husserl 1973b, 211)

Wie diese sperrige, aber präzise Beschreibung verdeutlicht, handelt es sich bei Reziprozität um ein Schlüsselphänomen menschlicher Sozialität, das als konstitutive Basis unseres Urvertrauens unser lebensweltliches Dasein trägt (vgl. Fuchs 2015). Passenderweise formuliert Hans Blumenberg sein Programm einer phänomenologischen Anthropologie im Ausgang vom reflexiven Wissen um das verkörperte Ausdrucksgeschehen unter dem Begriff der *Visibilität*. Dieser beschreibt den Umstand, »nicht nur und primär zu wissen, wie man selbst aussieht, sondern urtümlicher noch, sich dessen bewußt zu sein, daß man aussieht«, weswegen der Mensch »vom Sehenkönnen der anderen ständig durchdrungen« ist (2006, 778).

9 Es ist aufschlussreich, dass diese Eigenschaft von PARO dem prominent angeführten Kriterium für personale Identität, nämlich diachrone Reflexivität, nahekommt. Zur phänomenologischen Kritik dieses Identitätsbegriffs im Rahmen von demenziellen Erkrankungen vgl. Fuchs 2018 und Tewes 2020.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Versuchen wir diese Überlegungen auf unseren Kontext anzuwenden: Das, was Husserl als wechselseitiges Einfühlungsgeschehen fasst, scheint sich in der Interaktion mit PARO primär zwischenleiblich zu vollziehen. Dabei erscheint es auf der Grundlage des responsiven Designs gerechtfertigt von der Annahme auszugehen, dass PARO sein Gegenüber sinnlich wahrnimmt: Schließlich richtet er seinen Kopf nach diesem aus, öffnet seine Augen, wenn dieses ihn anspricht und schließt seine Augen, wenn er am Hals gekrault wird. Mit Hinblick auf die Ambivalenz des zweifachen Bewusstseins, die zwischen belebtem Tier und unbelebtem Roboter schwankt, verwundert es auch nicht, wenn wir mit Mitleid oder Ärger reagieren, wenn jemand PARO schlägt oder ihm schadet. Diese Ambivalenz ist es auch, die die Interaktion mit PARO für die Aufladung durch Anthropomorphismen freigibt, die als alteritätsinklusive Mechanismen sozialer Identitätsstiftung fungieren: Dies spiegelt sich bspw. darin wieder, dass sich ein Bewohner des Pflegeheims bei PARO entschuldigt, nachdem er diesen schlug (vgl. *Roboter zum Kuscheln*, Schlusssequenz); oder darin, dass Angehörige, Bewohner:innen und Mitarbeiter:innen ihren PARO auf den Namen Ole O'land taufen (vgl. ebd., 51:05–52:01). PARO wird derart in der Interaktion durch soziale Gesten zu einem sozialen Akteur und damit Teil einer Gemeinschaft, innerhalb derer moralische Standards gelten. Diese Interpretation als sozialer Akteur passt auch gut zu den positiven Effekten auf die psychosoziale Gesundheit, da Sozialität aktuell als Konstituens von Resilienz verhandelt wird (vgl. Breyer/Storms 2021, 360). – Aber rechtfertigt all dies in der Summe, dass aus einem tierisch-technischen Es ein soziales Du wird?

Im Sinne voller Reziprozität bzw. Visibilität wird aus dem Gesagten deutlich, dass die Interaktionpartner:innen PARO nicht als menschliches Gegenüber im Sinne von Meinesgleichen auffassen. Der Anthropomorphismus erreicht seine Grenze darin, dass sich Einfühlung mit PARO primär via Leiblichkeit und Affektivität vollzieht, aber kaum via Kognition (vgl. ebd.). Selbst wenn wir PARO aufgrund seiner Lernfähigkeit eine Persönlichkeit zuschreiben, wurde in keiner der Studien davon berichtet, dass wir PARO versteckte Absichten unterstellen – eine Möglichkeit, die durch eine voll entfaltete Form von Reziprozität bzw. Visibilität den humanen Sinnhorizont durchzieht. Vielmehr entspricht unsere Auffassung davon, was PARO ist, ziemlich genau dem hybriden Design, das ihm seine Entwickler:innen gaben: Ein hoch responsiver, affektiv ansprechender Akteur in der Sphäre des Sozialen. Ob PARO dann tatsächlich als ein soziales Du aufgefasst wird, hängt stark von der Einstellung seiner Gegenüber ab – also davon, ob sie *einerseits* die Ambivalenz des zweifachen Bewusstseins einseitig auflösen und wenn ja, in welche Richtung; oder ob sie *andererseits* zwar das zweifache Bewusstsein als implizite Einstellung beibehalten, aber in der konkreten Interaktion eher die personalistische, oder die technische, oder die naturalistische Perspektive einnehmen.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

– Die über diese ontologischen Überlegungen hinausgehende Frage, ob wir überhaupt ständig menschlicher Interaktionspartner:innen bedürfen, um uns verstanden zu wissen, soll der abschließende Abschnitt behandeln.

6.

a. Mit PARO auf dem Weg zu einer Phänomenologie der sanften Berührung

Wie zu Beginn des Beitrags betont, soll dieser in einer phänomenologischen Beschreibung des therapeutischen *Nah*feldes kulminieren. Auch wenn am Beispiel von PARO bereits zahlreiche Aspekte der sich hier abspielenden Interaktion und Kommunikation umrissen wurden, tritt doch ein Phänomen in der zwischenleiblichen Interaktion der Pflege besonders hervor: die Berührung (für ein Grundlagenwerk aus interdisziplinärer Perspektive vgl. Grunwald/Beyer 2001).

Wie wir uns bereits vor Augen geführt haben, zeichnet sich PARO durch eine besonders taktile Affordanz aus: Die flauschige und weiche Textur des Fells scheint in uns ein Berührt-Werden-Wollen auszulösen; dieses realisiert sich in lebensweltlich vertrauten Modi der sanften Berührung an der Kontaktzone zwischen Haut und Fell in leiblichen Praktiken, wie dem Streicheln, Klopfen, Ruckeln, Kraulen, Kitzeln, Greifen, Drücken, Halten, Umarmen oder gar Küssen. Genau für diese Form der emotional verfassten, sozialen und zwischenleiblichen Interaktion scheint PARO gemacht zu sein: Nicht nur haben seine Proportionen eine gute Griffigkeit, sondern sein Gewicht und seine konstant bleibende Körpertemperatur vermitteln einen sanften Druck und eine lebendige Wärme. Dabei scheint gerade die taktile Signatur des Bio-Roboters diesen für die Anwendung bei Personen mit demenziellen Erkrankungen zu disponieren.

Insbesondere mit Hinblick auf fortgeschrittenere Stadien dieser Erkrankungen verändert sich die Art und Weise erheblich, wie Betroffene sich *bewegen* und *kommunizieren* (zur Interpretation von Bewegung und Kommunikation als Modi der Orientierung bei demenziellen Erkrankungen vgl. Dzwiza-Ohlsen 2021). Auf der einen Seite kommt es nicht nur zu einem Bewegungsdrang des ganzen Körpers, der aus der Sicht der Pflegenden als ›Hinlauftendenz‹ bzw. Fluchtverhalten aufgefasst wird, sondern auch zu einem Bewegungs- und Berührungsdrang der Hände, der sich als Nesteln, Kratzen oder Tasten äußert (vgl. bspw. *Roboter zum Kuscheln*, 28:03–28:15). Auf der anderen Seite werden die symbolsprachlich verfassten, verbalen Ausdrucksfertigkeiten oft deutlich eingeschränkt, sodass non-verbale Ausdrucksmodi zunehmend wichtig werden, die nicht zuletzt die Berührung umfassen. Die *Affordanz*

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

des Berührtwerdenwollens von PARO passt also wie ein Schlüssel zum Schloss des gesteigerten *Bedürfnisses des Berührenwollens* auf Seiten der Betroffenen (Man beachte diesbzgl. auch das direkte Tasten im Gesicht von PARO, insbesondere der Schnurrhaare, der Schnauze sowie der Augen und Wimpern).

Kurioserweise war es nun Husserl, der am Beispiel der *Selbst*berührung die konstitutive Funktion der Doppelempfindung betonte, die sich mir anschaulich zeigt, wenn ich mit meinem (tastenden) *Leib* meinen (getasteten) *Körper* berühre (vgl. Husserl 1952, 148–151). Kurioserweise deswegen, weil Husserl an keiner Stelle darauf einging, dass sich die Erfahrungsqualität dieses Phänomens in der *Fremd*berührung radikal wandelt: Fremdberührung verbleibt in konstitutiver Unverfügbarkeit, weil mir niemals direkt anschaulich werden kann, was mein Gegenüber fühlt, denkt und wünscht, während es mich berührt. So haftet jeder Fremdberührung durch ein anderes Lebewesen der Charakter von Unvorhersehbarkeit, Unkontrollierbarkeit und Andersheit an. Aber wie wandelt sich dieses Verhältnis im Falle von PARO? Hat es etwas vom unheimlichen, weil anonymen Charakter der *End of Life Care Machine*?[10]

Zunächst stellt die sanfte Berührung – für die wir sogar mit den sogenannten *C-taktilen Zellen* (vgl. Liljencrantz/Olaussen, 2014) über einen eigenen Typ von Nervenzellen verfügen – eine besonders intime Form des sozialen Miteinanders dar, mit der wir aus unserem Alltag gut vertraut sind: sei es die begrüßende Handreichung, die freundschaftliche Umarmung, die entspannende Massage, das tröstende Wiegen, der erotische Kuss oder das verliebte Streicheln. Interessanterweise nimmt zwar die Fähigkeit des Tastens im Verlauf des normalen Alterungsprozesses ähnlich stark ab, wie die anderen Sinne, allerdings scheint davon das Berührtwerden durch andere nicht betroffen zu sein – dessen Bedeutung für das Wohlbefinden nimmt sogar zu (vgl. Sehlstedt, Ignell et al. 2016). Interessanterweise nimmt zwar die Fähigkeit des Tastens im Verlauf des normalen Alterungsprozesses ähnlich stark ab, wie die anderen Sinne, allerdings scheint davon das Berührtwerden durch andere nicht betroffen zu sein – dessen Bedeutung für das Wohlbefinden nimmt sogar zu (vgl. Sehlstedt, Ignell et al. 2016).[11] Mit der sanften Berührung verbinden sich zahlreiche positive Effekte für den ganzen Menschen, wie Kearney (2020, 4) betont: »›Tender touch‹ alleviates anxiety, bolsters the immune system, lowers blood pressure, helps with sleep and digestion, and wards off colds and infections. It feeds us in body and soul. In short, tactile communication is

10 Vgl. https://www.youtube.com/watch?v=vDHstslg8Vo

11 Interessanterweise nimmt die Fähigkeit des Tastens im Verlauf des normalen Alterungsprozesses ähnlich stark ab, wie die anderen Sinne, allerdings scheint davon das Berührtwerden durch andere nicht betroffen zu sein – dessen Bedeutung für das Wohlbefinden nimmt sogar zu (vgl. Sehlstedt et al. 2016).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

absolutely vital to our physical and mental wellbeing.« Vor dem Hintergrund demenzieller Erkrankungen verwundert es nicht, das Kuscheltiere oder Puppen eingesetzt werden (für eine intime Szene der Interaktion vgl. *Roboter zum Kuscheln*, 02:07–02:13; zum therapeutischen Effekt von Puppen vgl. Yilmaz/Asiret 2021). PARO teilt zwar sowohl mit Kuscheltieren als auch mit vielen Tieren seine Haptik, allerdings ist er mit Hinblick auf seine Kinetik erneut ein Hybrid: Er kann zwar kommunikationsrelevante Teile seines Körpers bewegen, allerdings weder im Sinne der Lokomotion noch im Sinne des Greifens oder Beißens. Derart sind es primär wir, die berühren und viel weniger PARO, der uns berührt, was nicht zuletzt ein weiterer Grund für seine hohe Akzeptanz sein mag.[12]

b. Stay in Touch: Ergriffenheit im therapeutischen Nahfeld

Blicken wir abschließend auf eine besonders eindrückliche Interaktionssequenz aus der Dokumentation *Roboter zum Kuscheln* (44:05–46:34). Kurz zum Setting: Anwesend sind eine Pflegerin mit PARO, Herr Hohenhorst (im Folgenden: Herr H.), der in einem breiten Sessel im Zentrum der Aufnahme sitzt, sowie seine Frau, die links neben ihm auf einem Stuhl eher im Hintergrund sitzt.[13] Herr H. befindet sich fortgeschrittenen Stadium der Erkrankung und kommuniziert fast ausschließlich non-verbal, insbesondere durch den markanten Einsatz der Mimik. Dazu erklärt seine Frau in einer eingeschobenen Sequenz: »Mit mir spricht er nicht. Wenn ich ihn ‚was frag – ich krieg' vielleicht mal ein Kopfschütteln oder ein Nicken. Ich möchte manchmal reingucken können, was im Moment im Kopf vorgeht, aber... Is' leider nich'.« Allerdings scheint sein verbales Verständnis relativ gut erhalten zu sein. Wie wir sehen werden, lässt sich die gesamte Sequenz als Abfolge einer rhythmischen und wechselseitigen Frage-Antwort-Interaktion verstehen, die zwischenleiblich getragen, verbalsprachlich begleitet und affektiv akzentuiert wird. Dabei bilden Herr H., PARO bzw. Ole und die Pflegerin über weite Strecken der Sequenz eine zwischenleibliche Triade, indem ihre Köpfe in einem gleichmäßigen Abstand einander zugekehrt sind und derart ein Dreieck aufspannen. Die Sequenz lässt sich in mehrere Abschnitte gliedern:

(1; 44:05–44:12) Die Pflegerin betritt den Raum, unter dem linken Arm geklemmt Ole, in der rechten Hand einen Hocker, den sie rechts neben den Sessel von Herrn H. stellt, während sie eine kurze Exposition

12 Wullenkord et al. (2016) zeigten am Beispiel des humanoiden Roboters NAO, dass Interaktion im Allgemeinen negative Emotionen gegenüber dem Roboter reduzierte, aber erzwungene Berührung bereits bestehende negative Emotionen verstärkte.

13 Im Rahmen dieses Beitrags können die medialen Effekte – bspw. Kameraperspektive, Schnitte, Kommentare – nicht vertiefend berücksichtigt werden.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

äußert: »Ich wollte sie besuchen. Und ich hab' jemanden mitgebracht.« Herr H. reagiert mit seinem Blick, der, während er ruhig nickt, kurz PARO trifft, dann die Pflegerin anblickt und dann die verbale Ansprache zu verarbeiten scheint, während er an der Pflegerin vorbei in den für die Kamera unsichtbaren hinteren und unteren Teil des Raumes blickt;

(2; 44:24–44:50) Im Anschluss wird ihm Ole durch die Pflegerin vorgestellt (»Das ist Ole.«), während sie die Kuschelrobbe mit beiden Händen über dem Schoß von Herrn H. auf Augenhöhe in ca. 30 cm Abstand hält, woraufhin Herr H. markanter nickt und seine Mundwinkel sich nach unten sowie seine Augenbrauen nach oben hin öffnen. Dann wechselt die Kameraperspektive in die Naheinstellung: Herr H. blickt Ole an, die Pflegerin beginnt Ole zu streicheln, Ole fiepst und Herr H. intensiviert den Blick mit weit erhobenen Brauen und zusammengepresstem und nach unten gezogenen Mund-Wangen-Bereich; er blickt kurz zur Pflegerin auf, dann wieder zu PARO, woraufhin sich die Interaktion verstärkt: Herr H. beginnt hörbar zu atmen und die Augen-Mund-Partie zeigt ein gerührtes Lächeln. Nachdem Herr H. noch mehrfach markant nickt, hebt PARO seinen Blick, woraufhin Herr H. ebenfalls große Augen macht und die Pflegerin, sehr nah und lächelnd, kommentiert: »Der kann einen angucken, mit großen Augen!« Während sie dies sagt, blickt Herr H. die Pflegerin kurz mit glasigen Augen an, baut dann den Blick mit PARO wieder auf und hebt dann die Brauen derart, dass seine Stirn sich stark faltet und er seine Augen vor Rührung kurz schließen muss.

(3; 44:50–45:12) An diesem Punkt spiegelt die Pflegerin – die permanent den Blick zwischen PARO und Herrn H. wechselt (*gaze alteration*) – die tiefe Bewegung im Gesicht von Herrn H. und kommentiert: »Das rührt sie richtig an, ne?« Herr H. scheint zu spüren, dass sich Tränen in seinen Augen sammeln, weswegen er auf den ersten Blick zu versuchen scheint, seine Hand unter der Decke hervorzuholen – allerdings nicht, um sich die Augen zu trocknen, sondern um PARO zärtlich über den Rücken zu streicheln, woraufhin zum ersten Mal seine Frau, gerührt lächelnd, aus der Naheinstellung gezeigt wird. Diese Sequenz gipfelt darin, dass Herr H. mit beiden Händen PARO fasst und ihn, unterstützt von der lächelnden Pflegerin, ganz nahe an sein Gesicht heranzieht, um mit ihm in einer entspannten Haltung zu kuscheln.

(4; 45:12–45:48) Die nun folgende Nahansicht von Herrn H. bildet thematisch den Hintergrund und der zwischen Pflegerin und Partnerin anhebende Dialog rückt in den Vordergrund: In diesem geht es um die Rolle von Herrn H. als engagierten Großvater einer vierjährigen Enkeltochter, mit der er gemeinsam seiner Leidenschaft, dem Fußballspielen, frönt. Die Pointe dieses Gesprächs ist, dass in der Umkehrung der Fürsorgeverhältnisse eine große persönliche Bedeutung liegt: Statt ständige Fürsorge im Rahmen einer Pflegeeinrichtung zu erhalten, kann Herr H. auch aktiv Fürsorge schenken.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

(5; 45:48–46:34) Die Sequenz gipfelt in einem Abschiedsdialog: Nicht nur reicht Herr H. der Pflegerin seine Hand zum Abschied, sondern er erwidert die affirmative Aussage der Pflegerin: »Hat mich sehr gefreut« mit: »Ja, das würde ich auch sagen: Ich mich auch, wirklich!«, woraufhin die Pflegerin wiederum spiegelt: »Wirklich«, gefolgt von einem »Dankeschön!«, das Herr H. mit einem »Bitteschön!« erwidert. Auch, wenn Herr H. nicht so einfach zu verstehen ist, nutzt dieser Worte wie »kuscheln« und »aus der Nähe« in einer lobenden Weise, was von der Pflegerin offensichtlich gut verstanden wird, die zurückgibt: »Das war gut. So aus der Nähe, in der Nähe sein – sich nahe sein, das war gut.« Schließlich fiept Ole noch einmal laut, was die Pflegerin kommentiert mit »Ole sagt auch Tschüss!«, woraufhin Herr H. sich gerührt von diesem verabschiedet, indem er ihn noch einmal intensiv anredet und mit seinen Händen und seinem Gesicht ganz nahe hält.

Wenn es durch PARO tatsächlich gelingt, die ganze Person attentional, emotional und intentional anzusprechen und derart das Maximum der ihr zur Verfügung stehenden Expressivität zu aktivieren; und dabei eine Individualisierung und Inversion der Fürsorgebeziehung erreicht werden kann, dann müssen wir anerkennen, dass es sich bei diesem um eine wichtige, wenn auch sehr spezifische Ergänzung innerhalb des therapeutischen Nahfeldes handelt. Wie auch Sabine Greulich im Anschluss an diese Sequenz kommentiert (vgl. 46:56–47:52), ist PARO ein ausgezeichnetes Medium, um den intersubjektiven Gefühlsausdruck zwischen den Polen von Lachen und Weinen zu fördern. Betroffene können mit PARO in eine zutiefst affektive und zwischenleibliche Responsivitäts- und Einfühlungsbeziehung treten, die potenziell zahlreiche positive Effekte hat: sie stabilisiert, motiviert und fokussiert und unterstützt derart die emotionale Regulation, die insbesondere in Krisensituationen (wie bspw. Krankheit) besonders gefordert ist (vgl. Breyer/Storms 2021, 361f.). Mit PARO kann es also eher gelingen, das, was Menschen mit demenziellen Erkrankungen am Herzen liegt, ausdrücklich zu machen, sodass dies gezielt im Sinne einer auf das individuelle Wohlergehen der Betroffenen ausgerichteten Pflegeumgebung berücksichtigt werden kann.

7.

Zusammenfassung und Ausblick: Ethische Fluchtpunkte einer Phänomenologie des therapeutisch-technischen Feldes

In einem jüngst erschienenen Sammelband wurde der Topos der *Technisierung der Lebenswelt* mit Hinblick auf das Phänomen des Digitalen erstmals interdisziplinär ausgelotet (vgl. Schnell/Dunger 2019). Derart

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

wurde – wenn man es im Neudeutsch der Sphäre des Digitalen sagen möchte – eines wichtiges ›Update‹ dieses Topos vorgelegt. Der vorliegende Beitrag knüpfte hier an und erörterte die Verbindung von Lebenswelt, Technisierung, Digitalisierung und Gesundheit am Beispiel des Einsatzes von KI in der Pflege bei demenziellen Erkrankungen.

Mit dem Ansatz bei einem konkreten Beispiel – dem KI-gestützten MCR PARO – versuchte der Beitrag die »multidimensionality of technologies as *material cultures* within a *lifeworld*« (Ihde 2009, 22) zu würdigen. Wie dieses Beispiel zeigt, führt Digitalisierung nicht notwendigerweise zu einem »*Angriff auf die Materialität der Lebenswelt*« (Schnell/Dunger 2019, 13), oder gar dazu, dass diese »entdinglicht, entkörperlicht und letzten Endes entwirklicht« (Han 2021) wird.[14]

Mit Hinblick auf die materielle Dimension des Technischen im Gewande des Designs machte insbesondere das Schlusskapitel deutlich, dass eine Phänomenologie des therapeutischen Feldes in Zeiten von intelligenter Robotik und globaler Pandemie die Chance ergreifen kann, um zwischenleibliche Berührung relational und über die Grenzen von ontologischen Kategorien hinweg zu denken: Auf der einen Seite haben wir die Perspektive des Pflegepersonals bzw. der Pflegeeinrichtung, für die Berührungspraktiken bei der Körperpflege, der Nahrungsaufnahme sowie der allgemeinen medizinischen Versorgung einen zentralen Stellenwert einnimmt; auf der anderen Seite steht die Berührung aus der Perspektive der Gepflegten, die das nachvollziehbare Bedürfnis haben, Berührung nicht nur passiv hinzunehmen oder medial anzunehmen, sondern selbst aktiv zu berühren. Hier steht PARO exemplarisch für eine Entwicklung, die es gestattet, beide Perspektiven im Sinne einer personenzentrierten Pflege im Rahmen eines *benignen* (vgl. Kitwood 1997) sozialen Umfeldes kreativ zu verknüpfen. Diese ethischen Implikationen einer Phänomenologie der Berührung im therapeutischen Feld werden allerdings nicht nur durch das Zu-wenig an sanfter Berührung, sondern auch durch das Zu-viel an verletzender Berührung plastisch vor Augen geführt (vgl. Bonillo et al. 2013; Lindemann/Barth 2020). Derart ist der Titel des Aufsatzes – *Stay in Touch!* – als Aufforderung und Mahnung zugleich zu verstehen.

Allgemeiner betrachtet liefert der Beitrag – mit der Verknüpfung der klassischen Phänomenologie Husserls mit der responsiven Phänomenologie Waldenfels' sowie der technikwissenschaftlichen Postphänomenologie Ihde's – umfassende Beschreibungen des therapeutischen Feldes, das in der aktuellen phänomenologischen Forschung oft unterkomplex bleibt. Am Beispiel der Interaktion innerhalb einer Institution zwischen PARO, Pflegenden und Gepflegten konnten wir ein differenzierteres Bild

14 Vgl. https://www.philomag.de/artikel/byung-chul-han-die-welt-hat-sich-ganz-nach-uns-zu-richten, abgerufen am 20.01.2022.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

entwerfen: Im therapeutischen Feld durchdringen sich nicht nur unterschiedliche personale Perspektiven im Rahmen eines responsiven und dialogischen Miteinanders; vielmehr sind ihre Interaktionen immer schon durch unterschiedliche Einstellungen geprägt, die, mal mehr und mal weniger, der ontologischen Signatur ihrer Akteure zu entsprechen versuchen. Dabei führt der Einsatz von nicht-menschlichen Akteuren (Tiere, Kuscheltiere, Roboter, Maschinen, Werkzeuge) vor Augen, dass im konkreten therapeutischen Feld unterschiedliche Anforderungen auf jeweils individuelle Bedürfnisse treffen – wobei nicht-menschliche Interaktionsangebote kreative Spielräume erschließen können. Näher besehen, kommt es durch das responsive Design von PARO zu einer technischen Transformation von Intentionalität, Attentionalität und Emotionalität in der verkörperten Interaktion. Zwar vollzieht sich die Interaktion in der zwischenleiblichen Face-to-Face- bzw. Body-to-Body-Situation im Hier und Jetzt, allerdings transzendiert sie diese selbst im Falle eines fortgeschrittenen Stadiums demenzieller Erkrankungen: Betroffene können an frühere Erfahrungen anknüpfen und ihr tief verankertes Bedürfnis nach sozialer Partizipation mit anderen gestalten und also derart im Hier und Jetzt das Früher der Vergangenheit mit dem Bald der Zukunft verknüpfen.

Allerdings ist und bleibt PARO als technisches Medium wesentlich durch einen ambivalenten Charakter ausgezeichnet, der, wie Ihde (1990, 49) im Anschluss an Husserl formuliert, »entdeckend« und »verdeckend« zugleich ist: *Einerseits* kann sein intelligentes Design individuelle Interaktions- und Kommunikationsressourcen bergen und derart die Lebensqualität verbessern; andererseits entscheiden die (pflege-)politischen Rahmenbedingungen nicht nur darüber, *ob* für den Einsatz von PARO überhaupt Mittel zur Verfügung stehen, sondern auch darüber, *wie* PARO eingesetzt wird. Um es an einem konkreten Beispiel zu veranschaulichen: Neben der Möglichkeit einer technisch bereicherten, personenzentrierten Pflege, für die das Wohlergehen der Betroffenen in ihrem sozialen Umfeld im Mittelpunkt steht, besteht auch die Möglichkeit, die Fülle der faktisch erhobenen, digitalen Daten systematisch zum Zwecke einer totalitären Überwachung einzusetzen – was aktuell im Kontext der Altenpflege in der Volksrepublik China im Namen der Sicherheit der Einzelnen auf Kosten ihrer Selbstbestimmung und Freiheit zum Vorteil der Vielen stückweise umgesetzt wird.

Literatur

Baisch, Stefanie et al. (2018): »Emotionale Roboter im Pflegekontext: Empirische Analyse des bisherigen Einsatzes und der Wirkungen von PARO und Pleo«, in: *Zeitschrift für Gerontologie und Geriatrie*, 51(1), 16–24.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Berendonk, Charlotte/Stanek, Silke (2011): »Positive Emotionen von Menschen mit Demenz fördern«, in: Andreas Kruse (Hg.), *Lebensqualität bei Demenz. Zum gesellschaftlichen und individuellen Umgang mit einer Grenzsituation im Alter*, Heidelberg: Akademische Verlagsgesellschaft, 157–176.

Blumenberg, Hans (2010): *Theorie der Lebenswelt*, hgg. von M. Sommer, Berlin: Suhrkamp.

Blumenberg, Hans (2006): *Beschreibung des Menschen*, hgg. von M. Sommer, Frankfurt a.M.: Suhrkamp.

Bonillo, Marion et al. (2013): *Gewalt in der familialen Pflege. Prävention, Früherkennung, Intervention – ein Manual für die ambulante Pflege*, Stuttgart: Kohlhammer.

Bredl, Leonie (2012, Film): *PARO, ein Roboter mit Charme*, Stuttgart: Live act Akademie der Schauspielkunst.

Breithaupt, Fritz (2017): *Die dunklen Seiten der Empathie*, Frankfurt a.M.: Suhrkamp.

Breyer, Thiemo/Storms, Anna (2021): »Empathy as a Desideratum in Health Care – Normative Claim or Professional Competence? «, in: *Interdisciplinary Journal for Religion and Transformation in Contemporary Society*, 7, 359–375.

Brose, Susanne et al. (2016): »Der Einfluss von Pflegebelastung auf die physische und psychische Gesundheit pflegender Angehöriger«, in: Berufsverband Deutscher Psychologinnen und Psychologen (Hrsg.), *Älter werden – gesund bleiben*, Berlin: Deutscher Psychologen Verlag GmbH, 38–43.

Bundesministerium für Familie, Senioren, Frauen und Jugend (2020): *Achter Altersbericht: Ältere Menschen und Digitalisierung*, Berlin.

Bühler, Karl (1982): *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Stuttgart/New York, Fischer.

Carel, Havi (2016): *Phenomenology of Illness*, Oxford: Oxford University Press.

Coeckelbergh, Mark (2011): »Humans, Animals, and Robots: A Phenomenological Approach to Human-Robot Relations«, in: *International Journal of Social Robotics*, 3(2), 197–204.

Döttlinger, Beatrix (2018): *Gestisch-kommunikatives Handeln als Bindeglied zwischen Sprache und Handeln bei Menschen mit Demenz. Beziehungs- und Interaktionsgestaltung*, Weinheim: Beltz.

Dzwiza-Ohlsen, Erik N. (2021): »Verloren sein – Der Verlust der lebensweltlichen Orientierung in der Alzheimer-Demenz aus phänomenologisch-psychopathologischer Sicht«, in: *Inter Cultural Philosophy, Journal for Philosophy in its Cultural Context*, Heidelberg: Universitätsbibliothek. https://doi.org/10.11588/icp.2021.1.84400

Dzwiza-Ohlsen, Erik N. (2019): *Die Horizonte der Lebenswelt. Sprachphilosophische Studien zu Husserls ›erster Phänomenologie der Lebenswelt‹*, Paderborn: Wilhelm Fink.

Dzwiza, Erik Norman (2018): »The Situationality of the Lifeworld. Reflections on Key Terms Concerning Human-Animal Relations«, in: T. Breyer

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

und T. Widlok (Hg.), *The Situationality of Human-Animal Relations. Perspectives from Anthropology and Philosophy*, Bielefeld: Transcript, 9–29.
Fuchs, Thomas (2020): »The Circularity of the Embodied Mind«, in: *Frontiers in Psychology*, 11(1707).
Fuchs, Thomas (2018): »Leiblichkeit und personale Identität in der Demenz«, in: *DZPhil*, 66 (1), 48–61.
Fuchs, Thomas (2015): »Vertrautheit und Vertrauen als Grundlagen der Lebenswelt«, in: *Phänomenologische Forschungen*, Hamburg: Meiner, 101–117.
Gibson, James J. (1979): *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.
Grunwald, Martin/Beyer, Lothar (2001; Hg.): *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung*, Birkhäuser Verlag, Basel.
Husserl, Edmund (1976): *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie, in zwei Bänden. 1. Halbband: Text der 1.–3. Auflage*, hrsg. von K. Schuhmann, Dordrecht: Martinus Nijhoff.
Husserl, Edmund (1973a): *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Erster Teil: 1905–1920*, hg. v. I. Kern, Den Haag: Martinus Nijhoff.
Husserl, Edmund (1973b): *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Zweiter Teil: 1921–1928*, hg. v. I. Kern, Den Haag: Martinus Nijhoff.
Husserl, Edmund (1952): *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch: Phänomenologische Untersuchungen zur Konstitution*, hrsg. von Marly Biemel, Den Haag: Martinus Nijhoff.
Hydén, Lars-Christer (2018): »Dementia, Embodied Memories, and the Self«, in: *Journal of Consciousness Studies*, 25(7-8), 225–241.
Ihde, Don (1990): *Technology and the Life-World – From Garden to Earth*, Indiana: Indiana University Press.
Ihde, Don (2009): *Postphenomenology and Technoscience. The Peking University Lectures*, New York: Suny Press.
Kearney, Richard (2021): *Touch. Recovering Our Most Vital Sense*, New York: Columbia University Press.
Kellmeyer, Philipp et al. (2018): »Social Robots in Rehabilitation: A Question of Trust«, in: *Sciende Robotics*, 3(1587).
Kitwood, Tom (1997): *Dementia Reconsidered: The Person Comes First*, Philadelphia: Open University Press.
Kontos, Pia C. (2005): »Embodied Selfhood in Alzheimer's Disease: Rethinking Person-Centred Care«, in: *Dementia*, 4(4), 553–570.
Kutzleben, Milena von (2016): *Häusliche Versorgung von Menschen mit Demenz*, Weinheim: Beltz Juventa.
Liljencrantz, Jaquette/Olaussen, Håkan (2014): »Tactile C Fibers and their Contributions to Pleasant Sensations and to Tactile Allodynia«, in: *Frontiers of Behavioral Neuroscience*, 8(37).

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Lindemann, Gesa/Barth, Jonas (2020): »Gewalt in der stationären Pflege: Zum Akteursstatus von Menschen mit Demenz«, in: Reichertz (Hrsg.), *Grenzen der Kommunikation – Kommunikation an den Grenzen*, Weilerswist: Velbrück Wissenschaft, 271–287.

Meyer, Christian (2014): »Menschen mit Demenz als Interaktionspartner: Eine Auswertung empirischer Studien vor dem Hintergrund eines dimensionalisierten Interaktionsbegriffs«, in: *Zeitschrift für Soziologie*, 43/2, 95–112.

Müller, Oliver (2020): »Postphänomenologie. Über eine technikphilosophische Methode«, in: *Phänomenologische Forschungen* 2020/2, 165–183.

Pfadenhauer, Michaele/Dukat, Christoph (2017): »Von der Empirie zur Postphänomenologie. Eine Suchbewegung zur theoretischen Verortung sozialer Robotik in der Demenzbetreuung«, in: N. Burzan und R. Hitzler (Hg.): *Theoretische Einsichten – Im Kontext empirischer Arbeit*, Wiesbaden: Springer, 247–260.

Sabat, Steven R./Harré, Rom (1994): »The Alzheimer's Disease Sufferer as a Semiotic Subject«, in: *Philosophy, Psychiatry, & Psychology*, 1(3), 145–160.

Schnell, Martin W./Dunger, Christine (2019; Hg.): *Digitalisierung der Lebenswelt. Studien zur Krisis nach Husserl,* Weilerswist: Velbrück Wissenschaft.

Sehlstedt, Isac et al. (2016): »Gentle Touch Perception Across the Lifespan«, in: *Psychology and Aging*, 31(2), 176–184.

Shibata, Takanori/Coughlin, Joseph F. (2014): »Trends of Robot Therapy with Neurological Therapeutic Seal Robot, PARO«, in: *Journal of Robotics and Mechatronics*, 26(4), 418–425.

Shibata, Takanori/Wada, Kazuyoshi (2011): »Robot Therapy: A New Approach for Mental Healthcare of the Elderly – A Mini-Review«, in: *Gerontology*, 57, 378–386.

Smith, Erin R. et al. (2011): »Memory and Communication Support in Dementia: Research-Based Strategies for Caregivers«, in: *International Psychogeriatrics*, 23 (2), 256–263.

Soler, Meritxell V. et al. (2015): »Social Robots in Advanced Dementia«, in: *Frontiers in Aging Neuroscience*, 7(133), 1–12.

Sonntag, Jan (2020): »Therapeutische Atmosphären. Am Beispiel der Musiktherapie bei Demenzen«, in: B. Wolf und C. Julmi (Hg.): *Die Macht der Atmosphären*, Freiburg i.Br./München: Alber, 345–362.

Sommer, Manfred (1984): »Husserls Göttinger Lebenswelt«, in: *Die Konstitution der geistigen Welt*, hg. und eingeleitet von M. Sommer, Hamburg: Meiner, IX–XLIV.

Summa, Michela (2012): »Zwischen Erinnern und Vergessen. Implizites Leibgedächtnis und das Selbst am Beispiel der Demenz-Erkrankungen«, in: *Phänomenologische Forschungen*, Hamburg: Meiner, 155–174.

Tewes, Christian (2020): »Embodied Selfhood and Personal Identity in Dementia«, in: C. Tewes/G. Stanghellini (Hg.), *Time and Body: Phenomenological and Psychopathological Approaches*, Cambridge: Cambridge University Press, 367–390.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Yilmaz, Cemile K./Asiret, Güler D. (2021): »The Effect of Doll Therapy on Agitation and Cognitive State in Institutionalized Patients with Moderate-to-Severe Dementia: A Randomized Controlled Study«, in: *Journal of Geriatric Psychiatry and Neurology*, 34(5), 370–377.

Wada, Kazuyoshi et al. (2008): »Robot Therapy for Elders Affected by Dementia«, in: *IEEE Engineering in Medicine and Biology Magazine*, 27(4), 53–60.

Wagner, Annette (2011, Film): *Roboter zum Kuscheln – Heilsam für Demenzkranke?*, Berlin: Filmtank.

Waldenfels, Bernhard (2019): *Erfahrung, die zur Sprache drängt*, Frankfurt a. M.: Suhrkamp.

Waldenfels, Bernhard (1971): *Das Zwischenreich des Dialogs. Sozialphilosophische Untersuchungen in Anschluss an Edmund Husserl*, Den Haag: Martinus Nijhoff.

WHO (2018): *Towards a Dementia Plan: A WHO Guide*, France.

Winniewski, Ragna (2022): »Disrupted Intercorporeality and Embodiedness in Dementia Care during the COVID-19 Crisis«, in: *Puncta – Journal of Critical Phenomenology*, *Special Issue: Pandemic Politics & Phenomenology*.

Wullenkord, Ricarda (2016): »Getting in Touch: How Imagined, Actual, and Physical Contact Affect Evaluations of Robots«, in: *Proceedings of the 25th IEEE International Symposium in Robot and Human Interactive Communication*, 980–985.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Die Autor:innen

Renate Buschmann ist Inhaberin des Lehrstuhls für Digitale Künste und Kulturvermittlung an der Universität Witten/Herdecke. Ihre Arbeitsschwerpunkte sind die Archivierung, Konservierung und kuratorische Vermittlung von zeitbasierter Kunst.

Erik Norman Dzwiza-Ohlsen ist Wissenschaftlicher Mitarbeiter am *Husserl-Archiv* der Universität zu Köln sowie Research Fellow am *Center for Subjectivity Research* in Kopenhagen. Er promovierte zur Entwicklung des Begriffs der Lebenswelt bei Husserl, derzeit arbeitet er an einer »Phänomenologie der Demenz«.

Martin Feißt ist Akademischer Mitarbeiter an der Professur für Organisations- und Verwaltungssoziologie der Universität Potsdam. Er promoviert zu Selbst- und Weltverhältnissen in der forensisch-psychiatrischen Langzeitunterbringung am Lehrstuhl für Soziologie der Universität Witten/Herdecke.

Christian Grüny ist Privatdozent für Philosophie an der TU Darmstadt sowie Wissenschaftlicher Mitarbeiter am Max-Planck-Institut für empirische Ästhetik in Frankfurt. Er arbeitet zur Ästhetik, Musikphilosophie und Kulturphilosophie.

Jonathan Harth ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Soziologie der Universität Witten/Herdecke. Er promovierte zur Interaktion mit computergesteuerten Spielpartnern. Momentan leitet er das Forschungsprojekt »Ai.vatar – der virtuelle intelligente Assistent«.

Alexander Hochmuth ist Leiter der Stabstelle Pflegeentwicklung am Klinikum Herford und promoviert zur Digitalisierung in der Pflege am Lehrstuhl für Sozialphilosophie und Ethik im Gesundheitswesen an der Universität Witten/Herdecke.

Burkhard Liebsch lehrt als apl. Professor praktische Philosophie an der Ruhr-Universität Bochum. Er war Leiter der Studiengruppe »Lebensformen im Widerstreit« am Kulturwissenschaftlichen Institut in Essen und hat zahlreiche Beiträge zur Sozialphilosophie veröffentlicht.

Oliver Müller ist Professor für Philosophie mit Schwerpunkt Gegenwart und Technik an der Albert-Ludwigs-Universität in Freiburg. Seine Forschungsschwerpunkte liegen in der philosophischen

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Anthropologie, Technikphilosophie und Mensch-Maschine-Interaktion.

Lukas Nehlsen ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Sozialphilosophie und Ethik im Gesundheitswesen der Universität Witten/Herdecke sowie Stipendiat der *a.r.t.e.s. graduate school for the humanities* an der Universität zu Köln, wo er zur Philosophie der Wahrnehmung promoviert.

Martin W. Schnell ist Inhaber des Lehrstuhls für Sozialphilosophie und Ethik im Gesundheitswesen an der Universität Witten/Herdecke. Der Lehrstuhl forscht zur Sozialphilosophie der Vulnerabilität, zur Digitalisierung im Gesundheitswesen sowie der Beziehung von Ethik und Politik.

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Eine Auswahl aus unserem Programm
Velbrück Wissenschaft

Christine Dunger und Martin W. Schnell (Hg.)
Digitalisierung der Lebenswelt
Studien zur Krisis nach Husserl
232 Seiten · br. · ISBN 978-3-95832-170-0

Dass die Digitalisierung mittlerweile alle Gesellschaftsbereiche durchdringt, ist unbestritten. Doch was sind die Konsequenzen? Steuern wir auf eine restlos in Daten erfasste, vollkommen künstliche Welt zu? Diese Schreckensvision wird den Ambivalenzen des anhaltenden Technologieschubs nicht gerecht. Jedoch fordern die spürbaren Veränderungen des menschlichen Weltverhältnisses dazu auf, sich über die Folgen dieses Prozesses zu vergewissern. Zur Analyse zieht der vorliegende Band eine Diagnose des späten Husserls heran. Der Philosoph sah die menschliche Lebenswelt einer Tendenz zur Technisierung ausgesetzt, deren Fortsetzung wir in der Digitalisierung erleben. Sie steht im Zeichen einer rückhaltlosen wissenschaftlich-technischen Quantifizierung, die den materiell-lebensweltlichen Sinnzusammenhang von Individuum, Gesellschaft und Natur nachhaltig beeinflusst.

Jan-Philipp Kruse, Sabine Müller-Mall (Hg.)
Digitale Transformationen der Öffentlichkeit
220 Seiten · br. · ISBN 978-3-95832-212-7

Die Sphäre der Öffentlichkeit wandelt sich rasant. Kommunikations- und Handlungsmöglichkeiten vervielfältigen sich unter digitalen Bedingungen. Wir begegnen »Trollen« oder »Stürmen« invektiver Entrüstung, versammeln Wissen in »Clouds«, die über die alte Unterscheidung von öffentlich und privat hinwegschweben, und wir bilden »Blasen«, in denen sich Meinungen und Affekte zu Eskalationsdynamiken aufschaukeln. Solche Phänomene erscheinen als Elemente eines Wandels, der mit neuen Medien, Formaten und Konventionen einhergeht und sich als tiefgreifende, strukturelle Transformation darstellt. Der Band fragt nach den Folgen dieser Transformation für den Begriff von Öffentlichkeit: Wie, in welchem Rahmen, mit welchen Prämissen, Konzepten und Methoden lässt sich heute exemplarisch über Öffentlichkeit nachdenken – im Hinblick auf deren normative Implikationen, aber auch ihr Verhältnis zu Politik, Recht, zur Wahrheit und zur Dimension der Affekte?

www.velbrueck-wissenschaft.de

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Eine Auswahl aus unserem Programm
Velbrück Wissenschaft

Robert Seyfert
Beziehungsweisen
Elemente einer relationalen Soziologie
280 S. · br. · ISBN 978-3-95832-189-2

Soziale Beziehungen bilden einen der zentralen Untersuchungsgegenstände der Soziologie. Dennoch werden sie dort auf eine merkwürdig eingeschränkte Art und Weise behandelt. *Beziehungsweisen* stellt eine relationale Soziologie vor, die die Existenzweisen und Weltverhältnisse von Subjekten, Systemen und Netzwerken nicht aus deren Aktivität hervorgehen sieht, sondern aus sozialen Beziehungen – und diese können sowohl interaktiv als auch -passiv sein. Außerdem nimmt diese Soziologie auch die Vielfalt der menschlichen und nicht-menschlichen Akteure und Passeure in den Blick. Neben theoretischen und methodologischen Überlegungen werden zwei empirische Fallstudien präsentiert, die den Perspektivenwechsel dieser relationalen Soziologie illustrieren.

Gesa Lindemann
Die Ordnung der Berührung
Staat, Gewalt und Kritik in Zeiten der Coronakrise
132 Seiten · br. · ISBN 978-3-95832-226-4

Die Corona-Pandemie stürzt die Gesellschaften rund um den Globus in eine Krise. Das vorliegende Buch nutzt die Krise zur Erkenntnis und stellt eine einfache Frage: Was können wir aus der Coronakrise über moderne Gesellschaften lernen? Zum Beispiel erfahren wir nicht nur etwas über das Verhältnis von Staat, Individuum und Gesellschaft, sondern auch etwas darüber, in welchem Verhältnis Staat, Politik, Recht und Wirtschaft bzw. Wissenschaft zueinanderstehen. Lindemann geht dabei immer wieder auf unser unmittelbares Erleben zurück. Sie fragt danach, wie die Strukturen der modernen Gesellschaft unser unmittelbares Zusammenleben bestimmen – wie wir einander ansehen, miteinander und gegeneinander handeln und einander berühren. Das Ergebnis ihrer Analyse ist zwiespältig: Die grundlegenden Strukturen unserer Gesellschaft werden sich nicht verändern, obwohl die Ordnung unserer alltäglichen Berührung sich verändern wird.

www.velbrueck-wissenschaft.de

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

https://doi.org/10.5771/9783748934493 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY