WAHRNEHMUNG – ERINNERUNG – IDENTITÄT

Birte Platow

Selbstwahrnehmung und Ich-Konstruktion im Angesicht von Künstlicher Intelligenz

1. Einleitung

Der Diskurs zur fortschreitenden Digitalisierung der Gesellschaft, zum damit einhergehenden Potenzial, aber auch zu erwartenden Umbrüchen im individuellen wie sozialen Leben, wird derzeit in nahezu allen wissenschaftlichen Disziplinen vor dem jeweils spezifischen Fachhintergrund sowie in diversen Institutionen der Gesellschaft geführt. Besonders hitzig, weil polarisierend, scheint dabei die Diskussion um positives Potenzial bzw. potenzielle Gefahr von Künstlicher Intelligenz (nachfolgend ‚KI') zu sein. Während die Einen damit technizistische Erlösungshoffnungen verbinden (bspw. Ray Kurzweil), sehen die Anderen die Menschheit in apokalyptischen Untergangsszenarien gefangen (u. a. Stephen Hawking). Dazwischen scheint es wenig zu geben. Die theologisch konnotierten Begrifflichkeiten ‚Erlösung' bzw. ‚Apokalypse' sind dabei keineswegs unbedacht gewählt, verweisen sie doch auf ein Phänomen, das im Zentrum des nachfolgenden Beitrags steht: Der Diskurs um KI ist in hohem Maße religiös motiviert, was den daran Teilnehmenden freilich überwiegend nicht bewusst ist. Der vorliegende Text verfolgt daher im Rahmen einer qualitativen empirischen Studie das Anliegen, verborgene religiöse Motive der menschlichen Selbstwahrnehmung im Gegenüber zu KI sichtbar zu machen, zu analysieren und zu diskutieren. Im Hintergrund steht eine theologisch orientierte, qualitativ-empirische Studie zu individuellen Selbst- und Wirklichkeitskonstruktionen im Angesicht von KI, die die Autorin derzeit bei der *Digital Society Initiative* der Universität Zürich als Fellow durchführt.[1] Nachfolgend genannte Teilhypothesen bilden einen ersten Ertrag der Studie und stellen zugleich übergeordnete Gliederungspunkte für den Text dar:

1 Als der vorliegende Beitrag verfasst wurde (04/2019), war die Studie noch nicht abgeschlossen. Bis dato waren alle Interviews geführt, transkribiert, codiert und zu zwei Dritteln ausgewertet. Insofern stehen die zitierten Hypothesen unter einem gewissen, dem Forschungsstand angemessenen Vorbehalt.

1) KI ist eine relevante Größe für die Konstruktion von individuellen Selbst- und Weltbildern.
2) Die unter dem Eindruck von KI spezifisch konturierten Konstruktionen der menschlichen Selbst- und Weltbilder erweisen sich als konstitutiv für den Diskurs um KI sowie – in langfristiger Perspektive – für das zukünftige ‚Miteinander' von KI und Menschen in einer zunehmend digitalisierten Gesellschaft.
3) Die systemimmanente Prägekraft genuin menschlicher Vorstellungen und Verhaltensweisen wird – so der Eindruck der Autorin – derzeit im Diskurs nicht hinreichend wahrgenommen oder aber unterschätzt. Es gilt daher, den spezifisch theologisch-religionspädagogischen Beitrag zum Diskurs themenspezifisch stärker zu konturieren.
4) Religionspsychologische und theologische Perspektiven auf eben genannte Phänomene erweisen sich für die Deutung in zweifacher Hinsicht als produktiv: zum einen als Erklärungsansatz für spezifische Muster in den Selbstwahrnehmungen und Wirklichkeitskonstruktionen von Individuen im Gegenüber zu KI, zum anderen für die Klärung erkenntnistheoretischer Positionierungen, wie sie im gegenwärtigen Diskurs um KI vorfindlich sind.
5) Im Kontext fachspezifischer, religionspädagogischer Reflexion erweist eine theologisch fundierte Bildung ihren spezifischen Beitrag zu einem konstruktiven Umgang mit KI.

2. Induktives Präludium

Auch wenn es für einen wissenschaftlichen Beitrag stilistisch etwas unkonventionell erscheint, möchte ich die Lesenden im Zuge eines induktiven Vorgehens zu einem kleinen Gedankenexperiment einladen, indem sie das folgende persönliche Erlebnis nachvollziehen.

Vor geraumer Zeit ließ mir der Bordcomputer meines Autos auf einer längeren Fahrt den folgenden Hinweis zukommen: „Zeichen von Müdigkeit erkannt", dazu das Symbol einer Kaffeetasse, das mir wohl nahelegen sollte, eine Pause zu machen. Meine erste, spontane Reaktion war die Einschätzung, dass ich nicht müde sei und den Hinweis daher ignorieren würde. Aber schon unmittelbar danach ertappte ich mich, wie ich nachrechnete, wie lang ich bereits wach war, wie gut ich geschlafen hatte, was ich bereits geleistet hatte etc., um validen Aufschluss über meinen physischen Zustand zu erhalten. Im Ergebnis kam ich zu dem Schluss, dass sich für etwaige Müdigkeit reliable Argumente finden ließen. In der Folge wollte ich fest-

stellen, welche Kriterien und Parameter beim Bordsystem meines Autos die Hypothese ‚Müdigkeit beim Fahrer' generiert hatten. Ich blinzelte, wackelte (kontrolliert) mit dem Lenkrad, fuhr (ebenfalls verantwortungsvoll kontrolliert) über Fahrbahnbegrenzungen, jedoch konnte ich das genannte Signal nicht gezielt initiieren. Stattdessen meldete es immer wieder, für mich ohne erkennbaren Auslöser, dass es feststelle, ich sei müde. Bereits etwas zermürbt von den ständigen Warnungen meinte ich mich daran zu erinnern, dass im Auto vorhandene Minikameras Pupillenbewegungen der Augen wahrnehmen und schnelle Zuckungen der Pupille als physiologischen Indikator für Ermüdung deuten und einen entsprechenden Hinweis an den Fahrer geben. Dessen eingedenk war ich sehr verunsichert. Auf einmal formte sich nämlich in mir der Gedanke aus: ‚Was, wenn mein Auto besser weiß als ich selbst, dass ich gefährlich müde bin?' Immerhin kann es mir unzugängliche Signale meines Körpers wahrnehmen und auswerten – und womöglich generiert dieser Umstand besseres Urteilsvermögen auf Seiten des Autos als das meine es ist.

An dieser Stelle muss offenbleiben, wie dieser Gedanke ggf. mein Verhalten beeinflusst hat. Stattdessen ist der Frage nach zugrunde liegenden Motiven und handlungsauslösenden Impulsen nachzugehen. Augenscheinlich wurde hier nämlich kurzfristig die individuelle Wahrnehmungs-Deute- und Urteilshoheit (hier über einen individuellen physischen Status), also die individuelle Autonomie an ein externes Steuerungssystem abgegeben. Aber warum?

3. Individuelle Selbst- und Wirklichkeitskonstruktionen im Spiegel Künstlicher Intelligenz

Der Schlüsselmoment für die eben skizzierten Ereignisse war die Erkenntnis, dass Steuerungssysteme (hier eines Autos) zunehmend nicht länger nur stupide programmierte, von der Person isolierte Programme darstellen, sondern dass sie in spezifischer Art und Weise mit dem Individuum interagieren, indem sie bspw. physiologische Regungen erfassen, deuten und zurückmelden. Entscheidend war nun die Annahme der betreffenden Person, dass dies auf Basis sogenannter „künstlicher Intelligenz" erfolge. Über fundiertes Wissen zu den Funktionsweisen und Differenzierungen von KI verfügt die Protagonistin zu diesem Zeitpunkt zunächst nicht. Vielmehr ist für sie die Vorstellung leitend, dass KI sich durch hohe Verarbeitungskapazität von Daten und einen spezifischen Grad an Autonomie auszeichne. Diese Eigenschaften verleihen KI den Status eines mehr oder

minder unabhängigen Gegenübers in beratender Funktion. Von besonderem Interesse ist nun die Frage, was die Wahrnehmung der skizzierten Art in Kombination mit der Chiffre ‚künstliche Intelligenz' an persönlichen Zuschreibungen auslöst, und welche Reaktionen sie in der Folge generiert. Unaufgefordert und weitgehend unbewusst stellen sich Individuen in Situationen der genannten Art nämlich die Frage, wie sich *menschliche* Intelligenz zu *künstlicher* Intelligenz verhält. Dabei werden diese beiden Formen von Intelligenz zu Größen in *ein- und demselben,* von Funktionen bestimmten Referenzsystem gemacht.

Am oben skizzierten Beispiel veranschaulicht offenbart sich in derartigen interaktiven Situationen der folgende Mechanismus: Individuen definieren eine Aufgabe, hier lautet sie, herauszufinden, ob der Fahrer müde sei. Im Anschluss stellt sich die Frage, wie gut und mit welchen Mitteln, weiter wie schnell, umfassend und fehlerfrei menschliche Intelligenz diese Aufgabe bewältigen kann, und wie sich die Fähigkeiten künstlicher Intelligenz dazu verhalten. Der in der Diskussion um KI oft und völlig zu Recht eingebrachte Einwand, dass Systeme künstlicher Intelligenz immer (bzw. bislang) nur in sehr eng begrenzten Funktionsbereichen leistungsfähig sind und daher menschlicher Intelligenz bei weitem unterlegen sind, ist dem Individuum dabei situationsbedingt nicht bewusst. In der Folge werden die menschlichen Fähigkeiten nicht selten in (rational) nicht zu rechtfertigender Weise in der individuellen Wahrnehmung abgewertet. Manch eine(r) mag das belächeln, aber die skizzierte Taktik ist allgegenwärtig und beschränkt sich keineswegs auf banale Alltagssituationen wie die einer Autofahrt. Im Gegenteil: Ähnliche Verhaltensweisen sind in den unterschiedlichsten gesellschaftlichen Funktionsbereichen vorfindlich, nämlich überall dort, wo Menschen mehrheitlich in Daten abgebildet werden. Dies ist in besonderem Maße in der Ökonomie und Medizin (u. a. im Versicherungswesen, am Kapitalmarkt sowie in big data affinen medizinischen Disziplinen) der Fall. Nicht wenige Mediziner verlassen sich bei der Diagnose bildgebender Verfahren zunehmend auf die Verarbeitungskapazität künstlicher Intelligenzen. Mitunter entstehen dabei Entscheidungssituationen wie die oben skizzierte. Im Zuge der nachfolgend skizzierten Interviewstudie gab etwa ein Radiologe zu bedenken: „Das menschliche Auge kann nur 15 Grauschattierungen auf Bildern von Gewebe entscheiden, eine KI 450. Wie sollte ich der KI (...) nicht glauben?" In Zukunft treffen wir vielleicht auf vergleichbare Wahrnehmungen in wachsendem Maße auch in anderen Fachbereichen. So ist etwa an die sogenannten *digital humanities*, die auch digitale Abbildungen des Individuums umfassen können, in den Geistes-

und Sozialwissenschaften zu denken, bspw. an die Sprachwissenschaft, Soziologie oder Psychologie.

Betrachtet man die Diskurse zum Thema KI im öffentlichen Raum, so stellt man fest: Auch hier werden in vielen Diskussionen menschliche und KI unreflektiert zu funktionalen Teilen *eines* gemeinsamen Referenzsystems erklärt. So wird etwa im Kontext von zukünftig zu erwartenden Risiken Künstlicher Intelligenz definiert und vergleichend festgelegt, zu welchen Anteilen und in welcher Qualität relevante Aufgaben von Systemen künstlicher Intelligenz übernommen werden können, und wie sich die Kapazitäten humaner Intelligenz dazu verhalten. Man denke nur an die Diskurse zum autonomen Fahren oder den Einsatz von Assistenz- oder Operationsrobotern in der Pflege bzw. im OP. Manche Funktionen gelten dabei freilich (vorerst noch?) als exklusiv von menschlicher Intelligenz erfüllbar. Gleichwohl – und das mag in Anbetracht des eben Beobachteten überraschen – bleiben Aspekte menschlichen Verhaltens im Diskurs um die Zukunft von KI in unserer Gesellschaft seltsam unterbeleuchtet. Die Frage, wie Menschen sich selbst wahrnehmen und in der Folge verhalten, wenn sie mit KI konfrontiert sind, ist jedoch mitnichten nur ein philosophisch-theologischer oder ethischer Luxus in einer Frage, die letztlich durch die Bewertung technologischer Parameter zu klären wäre.

4. Technikbezogene Selbst- und Weltbildkonstruktionen als Konstitutivum zukünftiger Verhältnisbestimmung von Mensch und Künstlicher Intelligenz

Die individuelle Selbstwahrnehmung von Menschen im Angesicht von KI wird, wie gesehen, durch menschliche Projektionen und Suggestionen beeinflusst. Menschen begeben sich in besagtes gemeinsames Referenzsystem und konstruieren über den Vergleich ihrer Fähigkeiten mit denjenigen künstlicher Intelligenz Selbstbilder, die nicht selten defizitorientiert sind, weil das Individuum die im jeweiligen Referenzsystem herrschenden Funktionen nicht in der Perfektion zu erfüllen vermag, wie Systeme künstlicher Intelligenz dies tun. In dieser Perspektive wird KI zum Beitrag für die „Definition der Natur des Menschen“ (Battaglia 2016, 83). Außerdem wird sie „somit als Gesamtheit der Fähigkeiten, mit denen ein Mensch die Natur in seinen Dienst setzt, verstanden. Sie ist daher Teil des Wesens des Menschen“ und „wahrer Spiegel des Menschen“ (Battaglia 2016, 83). Individuelle, den Blick in diesen Spiegel rahmende Motive sind einerseits die Hoffnung, den menschlichen Mangelzustand zu überwinden und gleich-

zeitig die Angst, qua Technik bzw. im Fall von künstlicher Intelligenz von der Technik selbst unwiderruflich überwunden zu werden. Vor diesem Hintergrund erschließt sich auch, warum die Bewertung von KI in so starkem Maße polarisiert (s. o.). „Da Hoffnung und Angst letztlich Projektionen unserer inneren Zustände sind, ist die Beschäftigung mit Robotern [Einfügung Autorin: und die mit KI] wohl auch und vor allem eine tiefgehende Auseinandersetzung mit uns selbst" (Dittmann 2016, 20).

Nun könnte man annehmen, dass einer von spezifischen Motiven getragenen Nabelschau des Individuums keine konstitutive Bedeutung für gesamtgesellschaftliche Zusammenhänge zukommt. Insofern wäre es unbedeutend, welche Selbstbilder die einzelne Person im Gegenüber zu KI vertritt. Dagegen ist allerdings einzuwenden, dass der individuellen Selbstwahrnehmung und dem korrespondierenden Verhalten über die normierende Kraft des Faktischen durchaus konstitutive Bedeutung zukommen, nämlich dann, wenn viele Menschen zu einem Urteil der beschriebenen Art kommen und sich entsprechend verhalten. Irgendwann wird derart nämlich schleichend aus faktischem Verhalten eine Norm. Gerade in Bereichen, die sich der empirischen Erfahrung (noch) entziehen, für die es ergo auch noch keine gesetzlichen Regelungen gibt, und die in besonderer Weise menschliche Identitätsfragen berühren, ist anzunehmen, dass sich Werte und Normen tendenziell in bottom-up Prozessen entwickeln.

Die beiden skizzierten Prozesse – polarisierende Selbstbilder sowie die verstärkende Weiterentwicklung korrespondierender Verhaltensmuster zu Werten und Normen qua normativer Kraft des Faktischen – erfährt im besonderen Falle von KI nun eine weitere Verstärkung durch der Technik immanente Mechanismen. So stellen „Deep Learning" oder „Machine Learning", eine spezifische, die *eigentliche* Form KI dar. Deep Learning, auch als „starke künstliche Intelligenz" bezeichnet, löst Aufgaben bekanntlich durch die Analyse von unvorstellbaren Mengen an Daten und beständige schematische Trial-Error Lernprozesse. Es handelt sich – in pädagogischen Kategorien gesprochen – also um Lernen durch Verstärkung statt durch Induktion, wie dies bislang mit der Programmierung durch einen spezifischen Algorithmus der Fall war. Mit Hilfe von Deep Learning kann KI immer neue, spektakulär anmutende Erfolge erzielen – etwa den Sieg gegen einen Menschen in dem komplexen Spiel „Go". Unter lernfort-

schrittsrelevante Datenmengen fallen inzwischen alle Arten von Daten[2] und zunehmend auch solche, die Verhaltensmuster abbilden, wie Menschen sie gegenüber KI an den Tag legen. In Form digitaler Spuren manifestieren wir nun möglicherweise das Gefühl menschlicher Unterlegenheit (oder genau gegenteilig das Gefühl von Überlegenheit) – und hinterlegen dieses Muster als Lerngegenstand für selbstlernende Systeme starker Künstlicher Intelligenz, die jene digitale Spuren für das Lösen ihnen gestellter Aufgaben nutzen. Eine sich gegenseitig bedingende, doppelte Verstärkung von Unterlegenheit bzw. Überlegenheit ist die Folge und ein Novum, das uns aufhorchen lassen sollte und die Frage dringlich werden lässt, welche Selbstbilder, wie und unter welchen Bedingungen Menschen entwickeln, wenn sie sich in Systemen Künstlicher Intelligenz spiegeln.

Ein Exkurs in die Vergangenheit zeigt übrigens, dass das Phänomen einer komplementären Bezogenheit des Menschenbildes auf die technische Innovation der jeweiligen Gegenwart kein Novum darstellt, sondern uns aus der Geschichte durchaus bekannt ist. „Blickt man auf die vergangenen Jahrhunderte zurück, so zeigt sich, dass das Verhältnis von Menschen und Maschinen durchaus dynamisch ist und der Mensch in den von ihm erschaffenen Maschinen stets in einen Spiegel blickt“ (Manzeschke/Karsch 2016, 8). So wurde der Mensch im 18. Jahrhundert in Analogie zur mechanischen Maschine interpretiert, im 19. Jahrhundert als hydraulisches System, im 20. Jahrhundert als kybernetische Maschine (Blum/Haas 2003). Heute sieht man im Menschen ein Zusammenspiel komplexer Faktoren bestehend aus Genetik, Epigenetik, chemischen Prozesse uvm. – durchaus in Analogie zu technologischen Netzwerksystemen. Eine interessante Beobachtung ist in diesem Kontext, dass die Anpassung von Menschenbild und Maschine reziprok ist. So ist die technologisch nicht notwendige leibliche Verkörperung von KI in Form von humanoiden Robotern oder Maschinen mit menschenähnlichen Extremitäten, Gesichtern, Mimik oder Gestik bzw. mit Namen und eigenen nationalen Identitäten als Ausdruck einer anthropomorphen Ideologie zu interpretieren. Und weiter ist diese als ein Hinweis darauf zu deuten, dass sich in diesem Paradigma wechselseitiger, komplementärer Bezogenheit spezifische identitätsbildende Prozesse ereignen.

2 Für die Entwicklung von Sprache oder Emotionen simulierenden Systemen oder aber für die automatisierte Textvereinfachung werden etwa dezidiert Texte zugrunde gelegt, die den Alltag und Interaktion von Menschen abbilden, um genau dort dann später auch mitzuwirken.

5. Religionspsychologie und christliche Theologie als paradigmatische Deutungskategorien

Der Sozialpsychologe Heiner Keupp (2002) versteht Identität als stets offenen, lebenslangen ‚Passungsprozess', bei dem vergangene, gegenwärtige und zukunftsbezogene Selbsterfahrungen unter verschiedenen Identitätsperspektiven reflektiert und zu Teilidentitäten zusammengefasst sind. Dieser Prozess hält durchaus Widersprüche und Spannungen aus und ereignet sich im dynamischen Gleichgewicht von Identität und Sozialität. Diese Grundannahme ist unter religionspädagogischer Perspektive anschlussfähig an die Ausführungen von Rudolf Englert (2014) oder Henning Luther (1992).

Keupp identifiziert insgesamt drei für diesen Passungsprozess relevante Größen: die situativen wie überdauernden Selbsterfahrungen des Individuums, die Perspektive anderer sowie das individuelle Identitätsgefühl, das im weitesten Sinne als Bewertungsinstanz interpretiert werden kann. Die jeweiligen Konstruktionsbeiträge der genannten Größen vereint das Individuum schließlich in einer Kernnarration, die Kohärenz stiftet und das Individuum nach innen sowie außen repräsentiert und verständlich macht und dabei nach Anerkennung in Balance mit Autonomie strebt. Kohärenz ist hier übrigens nicht statisch oder synthetisch zu verstehen, sondern durchaus als spannungsvoll und dynamisch – und in dieser Hinsicht auch als fragmentarisch, wie Henning Luther (Luther 1992, 281) dieses Phänomen bezeichnet.

Aus religionspsychologischer Sicht und im Kontext obenstehender Fragestellung ist nun die folgende Beobachtung bedeutsam: Die Selbsterfahrung eines Individuums weist zwei Aspekte auf, nämlich zum einen situative bzw. überdauernde Selbsterfahrungen und zum anderen *hypothetische* Selbsterfahrungen. Das Oxymoron ‚hypothetische Selbsterfahrung' ist hier nicht als stilistisches Mittel eingesetzt, sondern stellt eine inhaltlich relevante Kategorisierung dar. Hypothetische Selbsterfahrungen manifestieren sich nämlich in Form von Projektionen und werden gerade dann virulent, wenn das Ziel dieser Projektionen kein menschliches Gegenüber ist, sondern ein Ultimates, in christlicher Prägung ‚Gott'.

In religionskritischer Perspektive fühlen wir uns hier sofort an Ludwig Feuerbach erinnert, der in den Projektionen menschlicher Wünsche auf ein ultimates Gegenüber einen reinen Selbstausdruck des endlichen Menschen sah, noch dazu mit Negativwirkung für diesen, weil er sich in Wunschprojektionen und Sehnsüchten verlöre (Feuerbach, 1984). Feuerbach ist insofern zuzustimmen, als dass in der Vorstellung von Gott es der Mensch

immer auch mit sich selbst zu tun hat. Allerdings – so lautet ein religionspädagogischer Einspruch – nicht in dem Sinne, dass der Gottesgedanke nur ein Abziehbild des Menschen sei. Im Gegenteil, bildungstheologisch gesehen liegen in der Vorstellung von Gott und seinen klassischen Attributen Anregungen zu einer neuen Selbstwahrnehmung und Selbstinterpretation, die sich eben gerade nicht in Wunschprojektionen erschöpft. So unterstützt der Gottesgedanke das Individuum dabei, abstrakte Selbstbezüglichkeit aufzugeben und das eigene verengte Referenzsystem zu revidieren. Man könnte dies vielleicht als kritisch konstruktive Selbstentfremdung bezeichnen oder in einer spezifisch religionspädagogischen Perspektive als genuin religiöse Bildung.

Nun ist die hypothetische Annahme eines transzendenten, ultimaten Gegenübers mit den Eigenschaften der Omnipotenz, Omnipräsenz und Omniszienz allerdings mitnichten exklusiv der Vorstellung eines christlich konnotierten Gottesbildes vorbehalten. Gerade eine unkritische Technikgläubigkeit reproduziert pseudoreligiöse Phantasien. Unter anderem können Systeme künstlicher Intelligenz in der individuellen Wahrnehmung zum funktionalen Äquivalent des Gottesgedankens werden. Offen ist hingegen noch, worin sich ein christliches Gottesbild und sein funktionales Äquivalent, wie es sich in der KI findet, unterscheidet, und was dies für die Selbstwahrnehmung und Identitätskonstruktion in religionspädagogischer Perspektive bedeutet. Die jüdisch-christliche Tradition ist jedenfalls voll von Erzählungen, die vor menschlicher Hybris und Omnipotenzfantasien warnen. Man denke nur an die Erzählung vom legendären Turmbau zu Babel oder an die jüdischen Fabeln von Golem, den man vielleicht als Vorläufer eines aus Algorithmen geborenen Hominiden verstehen könnte.

Die im Rahmen der oben genannten Studie geführten (Leitfaden-)Interviews erlauben erste – dezidiert unter der Prämisse der Vorläufigkeit und noch nicht vorhandener Validität stehende – Hinweise zur Frage, wie sich individuelle Selbstkonstruktionen gestalten, wenn KI in der Funktion eines quasi göttlichen Gegenübers diesen Prozess prägt, und inwiefern theologische Lehren ein ggf. produktives Interpretament darstellen.

Bei den Interviewten ist eine dezidiert teleologisch ausgerichtete Grundrichtung der Selbstwahrnehmung und -beschreibung augenfällig. Teleologisches Denken ist in diversen christlichen Vorstellungen angelegt, etwa in der christlich anthropologischen Bestimmung des Individuums zur Gottebenbildlichkeit, in der Eschatologie (unabhängig davon, ob man sie präsentisch oder futurisch interpretiert), in der Soteriologie. Weiter ist damit nachweislich die Etablierung einer linearen Zeitvorstellung und entsprechenden Ausrichtung des Menschen verbunden. Diese Eigenschaften

der Denkrichtung und Vorstellungswelt werden nun augenscheinlich auch adaptiert und spezifisch zugespitzt, wenn funktionale Äquivalente an die Stelle christlicher Vorstellungen treten. Wenn KI zum ultimaten Gegenüber wird, wird die Perfektionierung des*der Einzelnen und der Welt als ganzer zur Motivation und zum Programm, das am Ende des Prozesses Letztgültigkeit und Unübertreffbarkeit anstrebt. Teleologien können also augenscheinlich nicht nur religiös, natural, sozial (Mieth 2018, 20), sondern eben auch technologisch motiviert sein. Die digitalen Abbildungen und Vermessungen des Menschen, wie ich sie eingangs kurz zitiert habe, zielen in letzter Konsequenz auf die grundlegende Veränderung der *conditio humana* ab. Wenn KI zum ultimaten Gegenüber wird, wird das Individuum in seiner Selbstwahrnehmung radikal objektiviert und entfremdet. Die Selbsttranszendierung des Individuums ist dabei von einem quantifizierbaren Ziel her bestimmt, nicht – wie etwa in der christlichen Theologie – von ihrer genuinen Bestimmung zur Gottebenbildlichkeit, die u. a. das Zentrum christlicher Anthropologie ausmacht. Entsprechend stehen auch eher defizitäre Selbstwahrnehmungen („Ich kann mich halt auch mal verrechnen", „müde sein", „ich bin nicht so schnell") im Zentrum des Bewusstseins anstelle von Potenzial.

Hier ist auch an die transhumanistische Idee zu denken, die durch die Vereinigung von KI basierter Technik und Mensch die Überwindung „menschlicher Makel", wie Sterblichkeit, Krankheit überhaupt jede Form von Begrenzung anstrebt. Diese Denkrichtung stellt sicherlich ein Extrem dar. Aber eine analoge, teleologisch ausgerichtete Selbstwahrnehmung habe ich in jedem bislang geführten Interview beim Nachdenken über KI vorgefunden. Meist manifestiert sich diese als ein individuell ausgeprägtes Programm zur Perfektionierung, quasi als Aufgabe für den Einzelnen oder die Einzelne. Dabei geht es weniger um das Erreichen eines ultimativen Endpunktes, sondern vielmehr um das Programm einer kontinuierlichen Verbesserung, als „Optimierung" bezeichnet oder aber zumindest um den Erhalt vorhandener Fähigkeiten, die Individuen im Angesicht von KI oft als gefährdet sehen. Manche Studienteilnehmer*innen vermeiden bspw. bewusst, Navigationssysteme zu benutzen oder Taschenrechner, weil sie ihre kognitiven Fähigkeiten trainieren wollen, um „besser zu werden" oder zumindest nicht ihren Status quo zu verschlechtern. Oder aber es gilt, Schwächen auszugleichen. Eine Probandin gab an, Nahrungsmittel (Linsen) auf Empfehlung ihres Fitnesstrackers hin zu essen, obwohl sie diese nicht mag, um nachteilige Körperwerte „zu verbessern". Befragt nach dem Motiv gibt sie an, potenzielle Schwächen (Alterungsprozesse) ausgleichen zu wollen. Der Körper als Einfallstor potenzieller Unterlegenheit ist übrigens in besonderer Weise bewusst, wenn Menschen mit Künstlicher Intelligenz

umgehen. Interpretament ist auch hier stets das teleologisch ausgerichtete Ideal, nämlich die perfekte fehlerfreie Erfüllung von Aufgaben und die Hoffnung, angesichts der zunehmenden Präsenz und Bedeutung von KI als „nur Mensch" nicht ins Hintertreffen zu geraten.

Dieser primär an Defiziten ausgerichteten Selbstwahrnehmung steht allerdings eine gegenläufige Tendenz gegenüber, die zunächst überraschen mag. So provoziert die bewusste Selbstwahrnehmung, die Konturierung von menschlichen Schwächen im Angesicht von KI nämlich zugleich eine durchaus selbstbewusste Profilierung des ‚menschlichen Menschen'. Mehrfach verbindet sich in den Ausführungen der Proband*innen der Hinweis auf individuelle Fehlbarkeit, Schwäche etc. mit der selbstbewussten Deutung: „Das macht mich halt als Mensch aus." Wenn Dittmann (2016, 39) für Roboter feststellt, dass „die Entwicklung und Konstruktion von Robotern kein allein originär wirtschaftlich getriebener technischer Prozess, sondern nicht zuletzt ein sozialer Akt der Selbstreflexion des Menschen im Technikzeitalter" ist, so scheint dies in besonderer Weise auch für die Entwicklung von KI zu gelten. Sie trägt zur Profilierung des Menschlichen bei, auch und insbesondere von dessen Schwächen. Im Zuge individueller Selbstkonstruktionen zeichnet sich die Deutung menschlicher Defizite durch eine neue, fast trotzige Akzeptanz der Substanz aus – selbst dort, bzw. *gerade dort,* wo diese im Abgleich mit der ultimat gesetzten KI nicht perfekt erscheint. Allerdings findet sich in der Deutung eine weitere Differenzierung: So wird die mitunter als defizitär identifizierte Substanz zwar als solche positiv gewertet, die motivationalen Veranlagungen des Menschen erfahren jedoch hingegen eine ungleich stärkere Stigmatisierung. Eigenschaften, wie Bequemlichkeit, Faulheit, mangelndes Wissen oder Bewusstsein, das Delegieren von Verantwortung werden – so zeigen diverse Codes und Subcodes in den Interviews – als negativ, punktuell sogar als bedrohlich empfunden, weil sie dazu verleiten, die vermeintliche Unterlegenheit des Menschen gegenüber KI zu vergrößern und zu verstetigen. Während das Menschsein an sich also besondere, wenn man so will rechtfertigungstheologische, Akzentuierung Akzeptanz im Angesicht von KI findet, treten andere vermeintliche Mängel, wie etwas mangelnde Motivation zum eigenständigen Handeln und Entscheiden, ungleich mehr in den Vordergrund. Eine Versöhnung ist – zumindest nach dem bisherigen Stand der ausgewerteten Interviews – indes noch nicht in Sicht.

Von besonderem Interesse ist nun die Frage, inwiefern sich eine solch spezifisch ausgeprägte Selbstkonstruktion auf das Urteilen und Handeln von Individuum auswirkt. Bislang wurde deutlich, dass „die saubere von Kant eingeführte Trennung zwischen demjenigen, was die Natur aus uns

macht, und demjenigen, was wir aus uns selbst machen können und sollen, angesichts solcher wissenschaftlichen und technischen Entwicklung nicht weiterhin Geltung in Anspruch nehmen [kann]“ (Battaglia 2016, 67). Mithilfe von KI wird das Unverfügbare – Wissensgrenzen, Begrenzungen durch genuin menschliche Eigenschaften wie Alterung, Müdigkeit, kognitive Begrenztheit uvm. – zunehmend verfügbar und entzaubert (Rosa, 2019). In der Folge wird auch der Raum des ehemals Unverfügbaren einer, für den wir potenziell in stärkerem Maße verantwortlich sind. Vor diesem Hintergrund erschließt sich möglicherweise auch die besondere Bedeutung, die motivationalen Veranlagungen zugeschrieben wird (s. o.), denn im Angesicht schier unbegrenzter Möglichkeiten, wird Motivation zum entscheidenden Impuls. KI erweitert in der individuellen Wahrnehmung die potenzielle Selbstwirksamkeit und Reichweite des/der Einzelnen in scheinbar nie da gewesenem Maße – es bedarf einzig des notwendigen Willens, diese Option in konstruktiver Weise zu nutzen.

Tatsächlich weisen eine frappierende Anzahl an Interviews Parallelen in Bezug auf eine auffällige Terminologie auf, indem die Befragten quasi unisono das „Instrument“ oder „Werkzeug“ im Kontext von KI erwähnten. Gefragt nach ihren Gefühlen im Umgang mit KI bzw. nach ihren Vorstellungen zur Zukunft wird variantenreich davon berichtet, dass „alles gut sei, solange KI ein Werkzeug in der Hand des Menschen“ bliebe und nicht vice versa. Eine solche „gute Verhältnisbestimmung“ war oft verbunden mit einer utopischen Version der erwarteten Zukunft – dem Ende der niederen (körperlichen) Arbeit, der Wende in der ökologischen Krise, der Chance, sich „dem Wichtigen zu widmen“ uvm. Das Modell, das hier im Hintergrund für derartige Visionen Pate steht, ist vereinfachend zu bezeichnen als ‚neue Technik, neues Zeitalter, neue Welt‘. Ausnahmslos alle Proband*innen vertreten sehr ausgeprägt die Überzeugung, dass eine neue Ära anbreche. Die damit einhergehende apokalyptische Grundstimmung wird von den Studienteilnehmer*innen indes unterschiedlich ausgestaltet. Auffällig viele plausibilisieren jedoch vor dem Hintergrund umweltethischer Fragestellungen das notwendige Ende einer Ära und den Bedarf eines markanten Umbruchs. Vor dem Hintergrund von Klimawandel, Energiekrise, Überbevölkerung und Ressourcenknappheit erhält das Potenzial von KI für sie eine ganz andere Plausibilität und Wertigkeit, nämlich als Problemlöser und Regulativ des nicht vernünftigen, in kapitalistisch geprägter Unersättlichkeit nach immer mehr strebenden Menschen. In diesem Kontext erzählen sie von einer Revolution der Mobilität (Auto fahren als „Social Entrepreneurship“), vom Ende vertrauter Berufsbilder, von Umbrüchen in der Kommunikation und in Beziehungen. Was die individuelle Bewertung

dieser Prozesse angeht, polarisiert diese und ist auch im Einzelfall nicht in sich konsistent. Eine präzise Auswertung steht hier zu diesem Zeitpunkt noch aus, allerdings ist schon jetzt deutlich, dass mit der Endzeitstimmung ein sehr explizit wahrgenommener Gestaltungs-/Schöpfungsauftrag für den Einzelnen oder die Einzelne einhergeht. In theologischer Perspektive ist dies an die Idee einer *creatio continua* anschlussfähig. KI fungiert dabei als „Werkzeug", mit dem Laien, Profis, aber auch die Forschenden und Entwickelnden versuchen wollen, ihrer Verantwortung gerecht zu werden. Ihr Bezugsrahmen ist dabei oft utilitaristisch oder libertär-individualistisch geprägt, so ein erster Eindruck.

Neben handlungsleitenden Impulsen geht mit einer Selbstwahrnehmung der beschriebenen Art als *cooperator dei* jedoch implizit ein weiterer Effekt einher, der sich unter anderem unter erkenntnistheoretischen Aspekten als relevant erweist und zugleich auf die besondere Rolle theologisch konturierter Deutungen der skizzierten Phänomene verweist. So dienen „Vorstellungen über die Natur des Menschen als unreflektierte Stellungnahmen, um eine normative Position zu verteidigen. Angesichts des Regelungsbedarfs dienen solche Positionen dazu, die Normativität bestimmter Überzeugungen als unabdingbar zu erklären, und das lediglich aufgrund ihrer tiefen Verwurzelung und ihrer praktischen Unaufhebbarkeit." (Battaglia 2016, 71–72). Hier ist etwa an die rationalen Argumentationslinien eines Nick Bostrom (2016) oder die von Ray Kurzweil (2005) zu denken, die im Widerspruch zu Deutungen des Menschen in christlich-anthropologischer Perspektive stehen. Derzeit kann im interdisziplinären Diskurs um Künstliche Intelligenz, wie ich ihn wahrnehme, weder eine rein rationale noch eine theologische Bestimmung des Menschen (etwa im Sinne christlicher Anthropologie) überzeugen. Mehr noch: oft bricht der Diskurs ab, bevor er beginnt, weil bereits kategoriale Unterschiede in den diversen Disziplinen und Positionen zu den grundlegenden Vorannahmen zum Wesen des Menschen diesen verhindern (vgl. Beitrag von Markus Huppenbauer in diesem Band).

Nun verweist die Studie, die im Zentrum dieses Beitrags steht, jedoch darauf, dass sowohl die Wahrnehmung von Künstlicher Intelligenz als auch die sich im Gegenüber dazu ereignenden individuellen Selbstkonstruktionen in hohem Maße als religiös zu qualifizieren sind. Dieser Eindruck ist gleichermaßen zutreffend für Laien, wie reflektierte Anwender*innen und Entwickler*innen. Unter dieser Prämisse ist dann auch nicht von zwei parallelen, sich einander ausschließenden erkenntnistheoretischen Horizonten zu sprechen. Wenn nämlich auch rationale Betrachtungsweisen funktional religiöse Motive aufweisen, können sich diese nicht kategorisch gegen

eine theologische Interpretation und Argumentation verschließen. Vielmehr sollte *religiöse* wie *technische* Selbstaufklärung das genuine Anliegen *aller* am Diskurs Beteiligten sein, denn „die Beschreibung der Natur des Menschen liefert keinen normativen Standard, aus welchem Festlegungen darüber möglich wären, was ethisch richtig oder falsch ist" (Battaglia 2016, 74). Vielmehr wird der Mensch im Kant'schen Verständnis der menschlichen Natur weder essentialistisch noch naturalistisch wahrgenommen und auch nicht als Objekt. Stattdessen konstituiert er sich über sein Denken, Verhalten und Handeln. Menschsein ist in diesem Sinne als Projekt unter bestimmten, gegebenen Bedingungen und als offene Option zu verstehen. Vor diesem Hintergrund erschließt sich auch die besondere Bedeutung einer (theologisch fundierten) Bildung des Individuums für eine konstruktive Begegnung mit künstlicher Intelligenz.

Dabei bleibt bislang jedoch offen, wie das Phänomen KI, wie es sich oben vielseitig und komplex darstellt, in spezifisch religionspädagogisch-bildungstheoretischer Perspektive zu antizipieren wäre. Wie muss Bildung in unserer Gesellschaft fundiert und ausgestaltet sein, um das Individuum auf eine Zukunft vorzubereiten, in der KI eine zunehmend größere Rolle spielen wird und ggf. noch weitreichendere Prägekraft für Selbstbilder haben wird als dies bislang schon der Fall ist? Wie ist mit bereits vorhandenen Selbstbildern (s. o.) in konstruktiver Weise im Zuge von Bildung umzugehen? Und welche Rolle kommt bei all dem ggf. der Theologie, respektive der Religionspädagogik zu?

6. Theologisch fundierte Bildung für eine konstruktive Begegnung von Mensch und Künstlicher Intelligenz

Natürlich können Antworten auf Fragen, wie die oben formulierten, im Rahmen des vorliegenden Beitrags nur skizzenhaft beantwortet werden. Der Fokus liegt dabei auf der bildungstheoretischen Fundierung – etwa der Frage, welches Bild vom Menschen die Grundlage und Ausgangspunkt von Bildung sein soll. Denn gerade in Zeiten des Umbruchs und einer Neufindung von Bildung als „Digitale Bildung" wird die Frage nach dem im Hintergrund von Bildung stehendem Menschenbild virulent. Denn unter anderem von dort aus ergibt sich dann die Struktur von Bildung, u. a. wie Ziele zu bestimmen sind, ob sie primär nach formalen oder materialen Kriterien definiert sind, welche Inhalte sich daraus ergeben uvm. Eine theologisch fundierte Reflexion eben dieser Fragen ist als Schlüsselaufgabe der Religionspädagogik zu sehen, die somit weit mehr ist als eine rein

fachdidaktische Disziplin. Die Legitimation einer theologischen Fundierung von Bildung ergibt sich jedoch nicht nur vor dem Hintergrund der Fachdisziplin, sondern im Fall von KI auch aus der Sache selbst, da bei diesem Thema der religiösen Veranlagung des Menschen – wie oben gesehen – eine konstitutive Funktion zukommt.

Für eine theologisch orientierte bildungstheoretische Grundlegung eignen sich diverse Lehren christlicher Theologie; exemplarisch konzentrieren sich die nachfolgenden Ausführungen auf die der Gottebenbildlichkeit als zentrale Bestimmung des Menschen in der christlichen Anthropologie, die zugleich die Grundlage für eine Ethik der christlichen Lebensführung bildet. Aus dem Gedanken der Gottebenbildlichkeit aller Menschen unabhängig von ihren diversen Orientierungen ergibt sich etwa, dass jedem Menschen eine unabdingbare Würde zukommt, auch wenn er beispielsweise akut würdelos oder unwürdig erscheint oder etwa unvollendet und anfällig in einem funktional bestimmten Paradigma seiner Selbst- und Fremdwahrnehmung. Die Rede von der Würde eines Menschen meint allerdings keine an-sich-seiende Qualität, sondern kennzeichnet einen Anspruch und eine Aufgabe – ähnlich wie dies für die Selbstkonstruktion im Gegenüber zu künstlicher Intelligenz festgehalten wurde (s. o.). Allerdings erfährt Mensch-Sein in christlich anthropologischer Perspektive eine spezifische Konturierung. Hier ist die Wahrnehmung als Ebenbild Gottes nämlich eine Motivationsressource für kritische Selbstfindung und Akzeptanz. Unter der Perspektive von Gottesebenbildlichkeit gilt: der Mensch ist *niemals* das, was er sein könnte und sein will. Die bleibende Unterschiedenheit von Gott und Mensch ist Konstitutivum ihrer Beziehung in christlicher Überzeugung. Aus dieser Differenz ergibt sich allerdings auch gerade eine Bildungsaufgabe. In lutherisch-protestantischer Perspektive ist der gerechtfertigte Mensch nämlich dazu befreit, seine Verantwortung um ihrer selbst willen zu übernehmen als Freiheit in Verantwortung. Unter dieser Prämisse steht auch ein protestantisch fundiertes Bildungsverständnis, das die Differenz zwischen Faktizität und Potentialität auch als die Unterscheidung von Subjekt und Subjektivität oder von ungebildeter und gebildeter Identität interpretiert.

Durch den Gedanken der Gottesebenbildlichkeit wird zudem gesagt, dass der oder die Einzelne auf eine transzendente Größe außerhalb seiner oder ihrer Selbst bezogen ist. Der Mensch geht eben gerade nicht in dem auf, was ihm sozial, ökonomisch, politisch und technologisch zugemutet und abgefordert wird. Der Gedanke der Gottebenbildlichkeit stellt so auch einen Gegenentwurf – oder in gewisser Weise einen Protest – zu funktionsbezogenen Abbildungen des Menschen dar, wenn KI als funktionales Äqui-

valent den Gottesgedanken repräsentiert. Zugleich schützt der Gedanke vor einer Vergottung – oder besser vor einer ‚Vergötzung' – technischer Intelligenzen.

Traditionell besagt der Gedanke der Gottesebenbildlichkeit, dass der Mensch zum Mitschöpfer, zum *cooperator dei*, werden soll. Durch ihn soll – durchaus im Sinn einer *creatio continua* – das Werk des Schöpfergottes bewahrt und weiterentwickelt werden. Daraus ergibt sich eine hohe ethische Verantwortung für alle, die technische Intelligenzen entwickeln oder anwenden. Auf der anderen Seite ist damit eine Verteufelung des technologischen Fortschritts oder eine religiös motivierte Maschinenstürmerei ausgeschlossen. Technische Intelligenz darf durchaus auch den Anspruch erheben, Element einer Weiterschöpfung zu sein. Aber dann muss sie sich auch den Kriterien unterstellen, die im Gedanken der Gottesebenbildlichkeit und der Schöpfungsverantwortung liegen. Das ‚Woraus' ist also nicht auf Kosten eines ‚Woraufhin' zu eliminieren.

Um bei der Frage nach einer theologisch fundierten Bildung nicht im Sumpf fundamentalistischer, nach hinten orientierter Handlungsanweisung versinken, muss man allerdings althergebrachte Theologumena vor dem Hintergrund gegenwärtiger und zukünftiger Herausforderungen dekonstruieren und rekonstruieren. Man muss den Kerngedanke einer theologischen Aussage herausarbeiten und dabei z. B. exklusivistische Verengungen oder historisch bedingte einseitige Interpretationen überwinden – man denke etwa Bultmanns existentiale Interpretation oder Pannenbergs Forderung nach einer Theologie, die sich nicht auf einen Sonderbereich der Wirklichkeit beschränken darf. In diesem Sinne ist die christliche Anthropologie als regulatives Prinzip und zugleich hermeneutische Kraft in Bildungsprozessen zu interpretieren. Implizit ist mit der theologischen Bestimmung von Bildung nämlich ein normatives Prinzip implementiert: zum einen die bleibende Differenz zwischen „dem Selbst" und „dem Anderen". Das Fremde kann und soll nicht erreicht oder vereinnahmt werden, sondern in seiner Andersartigkeit und Unverfügbarkeit Impulse für die Subjektwerdung als kontinuierliche Aufgabe bieten. Andersherum darf aber auch das Subjekt von keinem Gegenüber vereinnahmt werden weder explizit und augenscheinlich noch implizit und geheim.

Subjektorientierte religiöse Bildung ist demnach eine kontinuierliche Aufgabe in der dynamischen Spannung von Zuspruch und Anspruch. Dieses Kriterium einer gebildeten Subjektivität könnte sich auch für die Selbstbestimmung im Angesicht von KI als wichtig erweisen. So wären Fixierungen des Verhältnisses ausgeschlossen: Weder ein Verständnis von Künstlicher Intelligenz als einfacher Handlanger für niedere Aufgaben

(„Werkzeug“ oder „Arbeitssklave“) noch als mehr oder minder baugleicher Kooperationspartner und erst recht als überlegene Macht wäre festzuschreiben. Stattdessen bleibt der beständige Auftrag, die theologische Bestimmung des Menschseins immanent-hermeneutisch als Anlass zur Selbst- und Weltreflexion zu nehmen.

Literaturverzeichnis

Battaglia, Fiorella. 2016. „Vorstellungen über ‚die Natur des Menschen‘.“ In *Roboter, Computer und Hybride. Was ereignet sich zwischen Menschen und Maschinen?,* edited by Arne Manzeschke und Fabian Karsch, 67–88. Baden-Baden: Nomos.

Blum, Hubert E. und Rudolf Haas. 2003. *Über das Menschenbild in der Medizin.* Stuttgart: Thieme.

Bostrom, Nick. 2016. *Superintelligenz. Szenarien einer kommenden Revolution.* Berlin: Suhrkamp.

Dittmann, Frank. 2016. „Mensch und Roboter – ein ungleiches Paar.“ In *Roboter, Computer und Hybride. Was ereignet sich zwischen Menschen und Maschinen?,* edited by Arne Manzeschke und Fabian Karsch, 17–46. Baden-Baden: Nomos.

Englert, Rudolf. 2014. „Komposition des Differenten. Inwieweit ist so etwas wie eine ‚religiöse Identität‘ heute noch möglich?“ In *Mensch – Religion – Bildung. Religionspädagogik in anthropologischen Spannungsfeldern,* edited by Thomas Schlag und Henrik Simojoki, 128–139. Gütersloh: deGruyter.

Feuerbach, Ludwig. 1984. *Das Wesen des Christentums.* Leipzig: Reclam.

Keupp, Heiner, und Thomas Ahbe, und Wolfgang Gmür, und Renate Höfer, und Beate Mitzscherlich, und Wolfgang Kraus, und Florian Straus. 2002[2]. *Identitätskonstruktionen. Das Patchwork der Identitäten in der Spätmoderne.* Reinbek bei Hamburg: Rowohlt.

Kurzweil, Ray. 2005. *Menschheit 2.0.* Berlin: Lola Books.

Luther, Henning. 1992. *Religion und Alltag. Bausteine zu einer Praktische Theologie des Subjekts,* Stuttgart: Radius Verlag.

Manzeschke, Arne, und Fabian Karsch. 2016. „Einleitung.“ In *Roboter, Computer und Hybride. Was ereignet sich zwischen Menschen und Maschinen?,* edited by Arne Manzeschke und Fabian Karsch, 7–14. Baden-Baden: Nomos.

Mieth, Dietmar. 2018. „Perfektionierung und Meliorisierung (‚Enhancement‘) – ein Versuch über Menschenbilder?“ In *Die Perfektionierung des*

Menschen? Religiöse und ethische Perspektiven, edited by Thomas Bahne und Katharina Waldner, 19–42. Münster: Aschendorff-Verlag.
Rosa, Hartmut. 2019[2]. *Unverfügbarkeit.* Wien-Salzburg: Residenz Verlag.