

### 3 Kann ethisches Urteilsvermögen an Technologien delegiert werden?<sup>184</sup>

#### 3.1 Einführung

Angesichts der hohen Komplexität des Umgangs mit künstlicher Intelligenz und der digitalen Transformation aus ethischer Sicht, wie oben dargestellt, wäre es für den Menschen verlockend, die ethische Verantwortung an die künstliche Intelligenz zu delegieren und auf «moralische Technologien» und eine «ethische künstliche Intelligenz» zu vertrauen.<sup>185</sup> Darüber hinaus «werden KI-Systeme aufgrund ihrer zunehmenden Intelligenz, Autonomie und Interaktionsmöglichkeiten zunehmend als moralische Agenten wahrgenommen und erwartet.»<sup>186</sup> Begriffe wie «moralische Technologien», die im aktuellen Diskurs über Digitalisierung, Automatisierung, Maschinisierung, Robotisierung und den Einsatz von künstlicher Intelligenz verwendet werden, legen eine solche Option nahe. Sie drücken die Erwartung aus, dass es möglich wäre, moralische Technologien zu schaffen. Das Hauptziel wäre es, zu verhindern, dass Roboter beispielsweise Menschen Schaden zufügen. Der Begriff «moralische Technologien» bezieht sich auf deren Fähigkeiten, ethische Regeln zu befolgen, moralische Entscheidungen zu treffen und auf dieser Grundlage Handlungen auszuführen. Der Begriff «moralische Technologien» bringt den Wunsch zum Ausdruck, Maschinen mit ethischen Prinzipien und Normen auszustatten. Dies würde nicht durch Programmierung, sondern durch Lernen erreicht werden.

Die Rede von «moralische Technologien» kann zu Irritationen führen. Können technische Erfindungen wirklich moralisch sein? Kann künstliche Intelligenz ethisch sein? Kann man technologischen Systemen eine Moral zutrauen? Kann man ihnen eine moralische Handlungsfähigkeit zuschreiben?<sup>187</sup> Oder verfügen sie über eine begrenzte, aber nicht vollständige Moral, wie eine «funktionale Moral»<sup>188</sup>, die es ihnen erlaubt, die ethischen

---

184 Vgl. Kirchschräger 2017c.

185 Vgl. Anderson / Anderson 2011: 1-4.

186 Dignum 2019: 36.

187 Vgl. Sullins 2006.

188 Wallach / Allen 2009: 39.

Folgen ihrer Handlungen abzuschätzen, oder eine geistlose Moral, ohne die Eigenschaften zu erlangen, die Menschen als Grundlage für ihre Moral besitzen?<sup>189</sup> Oder ist es unmöglich, an Technologien mit Moral zu denken? Im Folgenden wird die Charakterisierung als «moralische Technologien», die über Moralfähigkeit verfügen, aus ethischer Sicht kritisch untersucht.

### 3.2 Verletzbarkeit

«Moralische Technologien» und «ethische künstliche Intelligenz» sind zunächst mit einer der Besonderheiten der moralischen Fähigkeit des Menschen konfrontiert, die den Menschen von Maschinen und künstlicher Intelligenz unterscheidet: die Verletzbarkeit des Menschen in Verbindung mit seiner «Erste-Person-Perspektive» und seinem «Selbstverhältnis». Der Begriff «Verletzbarkeit»<sup>190</sup> umfasst die Möglichkeit, angegriffen oder verletzt zu werden, und gleichzeitig die fehlende Fähigkeit oder die fehlenden Mittel, sich aus dieser Situation zu befreien und sich vor Verletzungen zu schützen.<sup>191</sup> Die Verletzbarkeit hat ihren Ursprung in der physischen und psychischen Hilflosigkeit des Menschen gegenüber sich selbst, seinen Mitmenschen, seinem Umfeld und seiner Umwelt<sup>192</sup> sowie in seiner Abhängigkeit von der Welt. Der Mensch ist insofern von sich selbst, den Mitmenschen, dem Kontext und der Umwelt abhängig, als er einerseits von ihnen verletzt werden kann und andererseits von ihnen vor Verletzung geschützt werden kann.<sup>193</sup> Hier zeigt sich ein wesentlicher Unterschied zwischen dem Verständnis von «Verletzbarkeit» und dem Begriff «Demütigung»<sup>194</sup>. «Demütigung» bezieht sich nur auf menschliche Handlungen oder Unterlassungen und nicht auf die natürliche Umwelt.<sup>195</sup> Die Umwelt (verstanden als menschliche und natürliche Umwelt) ist hier aber auch in ihrem zweiten Sinn gemeint, nämlich als Quelle der Verletzbarkeit des Menschen, weil sie wesentliche Elemente und Bereiche der menschlichen Existenz berühren kann und weil der Mensch verschiedene Möglichkeiten hat, mit ihr umzugehen und auf sie zu reagieren. So sollten Menschen,

---

189 Vgl. Floridi / Sanders 2004a.

190 Vgl. Kirchschräger 2013a: 241-267.

191 Vgl. Schröder / Gefenas 2009: 113-121; Kottow 2004.

192 Vgl. Ong-Vang-Cung 2010: 119.

193 Vgl. auch Butler 2004: 77.

194 Vgl. Margalit 1998.

195 Vgl. Margalit 1998: 9-10.

die Opfer einer Naturkatastrophe geworden sind, nicht einfach sich selbst überlassen werden, sondern Unterstützung erfahren.

Man kann zwischen zwei Arten der Verletzbarkeit unterscheiden, wobei die zweite Art in drei Herleitungen unterteilt werden kann:

- *Grundlegende Verletzbarkeit (A)*, die eine Eigenschaft aller Menschen ist und vom Menschen nicht beeinflusst werden kann (z. B. menschliche Vergänglichkeit);
- *Selektive und variable Verletzbarkeit (B)*,
  - die auf Ungerechtigkeit zurückgeführt werden kann;
  - die auf Unglücksfälle zurückgeführt werden kann;
  - die auf eigenes Verschulden zurückgeführt werden kann.

Alle Arten der Verletzbarkeit können verschiedene Bereiche bzw. Aspekte der menschlichen Existenz umfassen. Zum einen kann es sich um physische oder psychische Verletzbarkeit handeln. Physische und psychische Verletzbarkeit gehören gleichsam zum Individuum und können daher als „*innere Bereiche und Aspekte der Verletzbarkeit*“ (C) bezeichnet werden. Die Verletzbarkeit wird zu Verletzungen bei Folter, Gewalt, ... Beide – sowohl physische als auch psychische Verletzbarkeit – werden durch die menschliche Heterogenität nicht eingegrenzt. Z. B. sind Menschen mit Behinderungen subjektiv und objektiv in gleichem Masse der Verletzbarkeit ausgesetzt wie Menschen ohne Behinderungen.

Die «inneren Bereiche und Aspekte der Verletzbarkeit» sind von den «*äußeren Bereichen und Aspekten der Verletzbarkeit*» (D) zu unterscheiden, wie z. B. Glaube, Religiosität und Weltanschauung,<sup>196</sup> Recht, Nahrung, medizinische Versorgung, Bildung, finanzielle Zuwendungen, Infrastruktur usw. Die Verletzbarkeit verwandelt sich in Verletzungen, Entbehrungen, Diskriminierung und Ausgrenzung, wenn sie missachtet wird.<sup>197</sup> Im Gegensatz zu den «inneren Bereichen und Aspekten der Verletzbarkeit» (C) findet die Umwandlung von Verletzbarkeit in Verletzung in den «*äußeren Bereichen und Aspekten der Verletzbarkeit*» (D) außerhalb der Person statt. Das bedeutet natürlich nicht, dass diese Umwandlung keine Auswirkungen auf die Person oder ihre inneren Sphären hat – im Gegenteil.

Von den inneren und äusseren Bereichen und Aspekten der Verletzbarkeit zu unterscheiden sind die oben eingeführte *äussere Form (E)* (Mög-

196 Vgl. Barnes 2002: 3.

197 Vgl. Ferrarese 2009.

lichkeit des Angriffs und der Verletzung von aussen) und *innere Form* (*F*) (Unfähigkeit und Mangel an Mitteln für einen korrespondierenden Schutz) der Verletzbarkeit, die ebenfalls bei beiden Arten der Verletzbarkeit vorkommen können.

Was die zweite Art von selektiver und variabler Verletzbarkeit und ihre Unterformen betrifft, so besteht ein Interesse daran, die verschiedenen disziplinären Positionen zu messen und operationell zu machen, um auf die Verletzbarkeit mit geeigneten Maßnahmen zu reagieren.<sup>198</sup>

Der Mensch teilt die Verletzbarkeit mit anderen Lebewesen, was die moralische Relevanz der Letzteren begründet.<sup>199</sup> Wenn die eigenen Entscheidungen und Handlungen einem anderen Lebewesen Schmerz zufügen würden, ist dies moralisch relevant. Denn wenn die Moral nach dem guten Leben und dem richtigen Handeln strebt, dann ist der Schmerz eines anderen Lebewesens aus zwei Perspektiven wichtig: Die eine Perspektive ist die Sicht des Moralfähigen. Moralfähig bedeutet, dass der moralfähige Mensch in der Lage ist, selbstbestimmt über das gute Leben und das rechte Handeln zu reflektieren, moralische Regeln zu definieren und nach ihnen zu urteilen und zu handeln. Die andere Perspektive vertritt die Sichtweise des moralbedürftigen Wesens oder des Wesens mit moralischer Geduld. Moralische Bedürftigkeit wird hier nicht in dem kantischen Sinne verstanden, dass es eine Eigenart des Menschen ist, dass er auf Moralisierung angewiesen ist.<sup>200</sup> «Ein Tier ist schon alles durch seinen Instinkt; eine fremde Vernunft hat schon alles dasselbe getan. Aber der Mensch braucht seine eigene Vernunft. Er hat keinen Instinkt und muss sein Verhalten selbst planen»<sup>201</sup>. Dieses Verständnis des Bedürfnisses nach Moralität zeigt eine große Nähe zum sittlichen Vermögen, da Ersteres auch als Teil des Letzteren gedacht werden kann, nämlich als Hinweis darauf, dass das sittliche Vermögen durchaus ein Entwicklungs- und Aktivierungspotential enthält («Der Mensch soll erst seine Anlagen zum Guten entwickeln; die Vorsehung hat sie noch nicht fertig in ihn gelegt; sie sind bloße Anlagen und ohne den Unterschied der Sittlichkeit. Der Mensch muss sie selbst machen, sich selbst kultivieren und (...) die Sittlichkeit in sich selbst erzeugen»<sup>202</sup>). Moralisches Bedürfnis oder moralische Geduld bedeutet hier

---

198 Vgl. Alwang et al. 2002.

199 Vgl. Ladwig 2007.

200 Vgl. Ladwig 2007.

201 Kant 1995b:697.

202 Kant 1995b:702.

vielmehr, dass die Objekte von Entscheidungen und Handlungen sittlichen Charakters (z. B. Tiere), die von den Moralfähigen getroffen oder vollzogen werden, auf die sittliche Qualität dieser Entscheidungen und Handlungen angewiesen sind; d.h. in diesem Sinne, dass sie einer Sittlichkeit bedürfen, welche die Entscheidungen und Handlungen der Moralfähigen prägt. Umgekehrt sind diese Objekte nicht moralisch gleichgültig gegenüber den Entscheidungen und Handlungen moralfähiger Menschen, d.h. sie machen einen Unterschied, wenn die moralfähigen Menschen nach dem guten Leben und dem richtigen Handeln streben, was im Beispiel der Tiere durch ihre Verletzbarkeit gerechtfertigt ist. Moralisches Bedürfnis oder moralische Geduld hängt nicht von der Moralfähigkeit ab. Um moralische Relevanz aufzuzeigen, kann man sowohl vom Moralbedürftigen als auch vom Moralfähigen ausgehen. Vom Standpunkt des Moralbedürftigen oder Moralgeduldigen aus ist es zum Beispiel moralisch falsch, wenn jemand es schlägt. Gleichzeitig ist es aus der Sicht des Moralfähigen moralisch falsch, einen Moralbedürftigen oder einen Moralgeduldigen zu schlagen, d.h. Schmerzen zuzufügen (schlechtes Gewissen). Mit der Verletzbarkeit ist es möglich, die Beziehung zwischen Moralbedürfnis (oder Moralgeduld) und Moralfähigkeit neu zu überdenken und die moralische Relevanz von Moralbedürfnis oder Moralgeduld sowie Moralfähigkeit aufzuzeigen. In der moralischen Relevanz der Verletzbarkeit liegt auch ein Anspruch auf einen Schutz derselben, der in besonderer Weise durch das «Recht auf Leben» geltend gemacht wird. Weil der Mensch als moralisch relevantes Wesen aufgrund seiner Verletzbarkeit des Schutzes bedarf, ist er Subjekt des «Rechts auf Leben». Die Verletzbarkeit des Menschen dient also als Hinweis auf die Notwendigkeit des «Rechts auf Leben». Diese Verletzbarkeit gilt sowohl für moralfähige als auch für moralbedürftige Wesen bzw. für moralgeduldige Wesen. Dementsprechend ist der Schutz sowohl für moralfähige als auch für moralbedürftige bzw. moralgeduldige Wesen moralisch notwendig. Daraus ergibt sich die Forderung, die Subjektivität des «Rechts auf Leben» des Menschen auch auf Tiere auszudehnen.<sup>203</sup>

Im Hinblick auf das, was weiter unten in Unterkapitel 6.4 Menschenrechte als ethische Referenzpunkte beschrieben wird, um die Menschenrechte auf der Grundlage des Prinzips der Verletzbarkeit zu begründen, wird eine Unterscheidung zwischen Menschen und Tieren in Bezug auf das «Recht auf Leben» in folgender Weise vorgeschlagen: Menschen teilen

---

203 Vgl. Ladwig 2007.

die Verletzbarkeit mit Tieren. Bedeutet dies, dass die Subjektivität der Menschenrechte auch auf die Tiere ausgedehnt wird? Weist der Unterschied zwischen Moralbedürftigkeit bzw. Moralgeduld und Moralfähigkeit auch auf einen Unterschied in den Schutzansprüchen hin, der nur im Falle der letzteren den besonderen Schutz des «Rechts auf Leben» einschließt? Oder verlangt das neu konzipierte Verhältnis von Moralbedürftigkeit bzw. Moralgeduld und Moralfähigkeit zugleich den besonderen Schutz des «Rechts auf Leben» für Moralbedürftige bzw. Moralgeduldige und Moralfähige? Zunächst lässt sich feststellen, dass dieser Ansatz, der auch Moralbedürftigkeit oder Moralgeduld als moralisch relevant ansieht, den für die Definition des moralischen Status von Maschinen und künstlicher Intelligenz relevanten Grenzbereich zwischen Moralfähigkeit und Moralbedürftigkeit oder Moralgeduld berücksichtigt.<sup>204</sup>

Wenn die erste Frage bejaht wird, dient Moralfähigkeit dazu, das «Recht auf Leben» zu rechtfertigen.

Eine entsprechende Verneinung würde zu einer Ausweitung von moralisch fähigen auf moralisch bedürftige Wesen oder Wesen mit moralischer Geduld führen, was einer Bejahung der zweiten Frage gleichkäme. Dann stellt sich die Frage, ob wirklich alle moralisch bedürftigen oder moralisch geduldigen Wesen den Schutz des «Rechts auf Leben» beanspruchen können (z.B. Tiere). Dies würde dem menschenrechtlichen Konzept des «Rechts auf Leben» widersprechen, so dass innerhalb der moralisch bedürftigen bzw. moralisch geduldigen Wesen zwischen menschlichen moralisch bedürftigen bzw. moralisch geduldigen Wesen und nicht-menschlichen moralisch bedürftigen Wesen zu differenzieren ist.

In dieser Frage sollte also eine Abgrenzung zu Tieren vorgenommen werden. Es ist jedoch zu beachten, dass diese Abgrenzungsdiskussion hier nicht abschließend geführt wird, sondern nur insoweit und ausschließlich, als sie für eine Begründung der Menschenrechte unmittelbar relevant ist. Das Ausschlusskriterium ist, ob ein direkter Beitrag zur Beantwortung der Frage zu erwarten ist, warum Menschen sich gegenseitig Menschenrechte gewähren, oder eine Antwort auf eine Frage, die sich in der Diskussion dieser Grundfrage stellt. Mit diesem kurzen Einbezug der Frage nach der Abgrenzung von Mensch und Tier werden daher – in Anerkennung dieser Einschränkungen – keine Lösungsvorschläge für die vielen in diesem Zusammenhang auftretenden Fragen gemacht.

---

204 Vgl. LaBossiere 2017.

Es liegt auf der Hand, darauf hinzuweisen, dass Menschenrechte trotz aller möglichen Überschneidungen mit Tieren nicht zwangsläufig als Tierrechte anzusehen sind, zumal der Begriff «Menschenrechte» das Wort «Mensch» und nicht «Tier» enthält und somit keine Tierrechte darstellt. Dieser Hinweis auf die Terminologie reicht jedoch nicht aus, um die Frage zu beantworten, warum nur der Mensch Träger:in von Menschenrechten ist. Der Hinweis hilft ein wenig weiter: «Die Menschenrechte haben keineswegs den diskriminierenden Sinn, den Menschen vom Tier zu unterscheiden und ihm etwaige höherwertige Rechte gegenüber ‚niederen‘ Lebensformen zuzuerkennen. Sie zielen lediglich auf ein *zwischenmenschliches* Rechtsverhältnis, das auch dann der Regelung bedürfte, *wenn* Tiere ähnliche Rechte haben sollen.»<sup>205</sup> Definitionsgemäß erweist sich also der Mensch, nicht das Tier, als Träger:in von Menschenrechten. Zugleich ist darauf hinzuweisen, dass dies nicht automatisch bedeutet, dass die Möglichkeit kategorisch verneint wird, dass theoretisch ähnliche oder im Wesentlichen gleiche moralische Rechte gewährt werden könnten.<sup>206</sup>

Man kann sich dem Umstand der gemeinsamen Verletzbarkeit von Mensch und Tier auch auf andere Weise nähern, nämlich indem man feststellt, dass es einige Anhaltspunkte für die Annahme gibt – die hier nur kurz angedeutet werden soll, da sie nicht im Zentrum des Versuchs steht, einen begründenden Ansatz für die Menschenrechte zu formulieren –, dass der Mensch wegen der Verletzbarkeit der Tiere Tierschutzgesetze formuliert. Die unmittelbare Schlussfolgerung, dass Menschen und Tiere die gleichen Rechte haben sollten, weil sie beide verletzlich sind, ist insofern nicht zulässig, als drei Punkte (und insbesondere der dritte Punkt) zu beachten sind: *Erstens* verständigen sich Menschen – aufgrund ihrer Verletzbarkeit – auf Rechte, die für sie selbst als Menschen gelten. Darüber hinaus entwerfen Menschen auch für andere Lebewesen, wie z.B. Tiere, mit denen sie die Verletzbarkeit teilen, entsprechende Tierschutzgesetze. Diese beruhen auf ihrer Selbstwahrnehmung und der Wahrnehmung ihrer eigenen Verletzbarkeit und, getrennt davon, auf der Wahrnehmung von Tieren und deren Verletzbarkeit.

*Zweitens* ist es denkbar, dass Menschen sich so vorstellen oder wahrnehmen, dass die Verletzbarkeit von Menschen und die Verletzbarkeit von Tieren phänomenologisch unterschiedlich ist. Dieser Zusammenhang beinhaltet nicht den Versuch zu behaupten, dass dies empirisch tatsächlich

205 Menke / Pollmann 2007: 140, Hervorhebung im Text.

206 Vgl. Menke / Pollmann 2007: 140-141.

der Fall ist, weder in der Unterscheidung zwischen menschlicher Selbstwahrnehmung und der davon abweichenden Wahrnehmung von Tieren, noch in der Feststellung von Unterschieden im Zuge der phänomenologischen Betrachtung der Verletzbarkeit von Menschen und Tieren. Dies ist für diesen Ansatz irrelevant. Vielmehr ist es wahrscheinlich, dass Menschen – im Bewusstsein und in der Anerkennung der Verletzbarkeit von Tieren – sich selbst als Menschen anders denken als Tiere, oder dass Menschen ihre eigene Verletzbarkeit, die sie mit anderen Menschen teilen, anders denken als die Verletzbarkeit von Tieren.

*Drittens* unterscheidet die sich anschließende Bewusstwerdung der eigenen Verletzbarkeit, die sich der «Erste-Person-Perspektive»<sup>207</sup> und dem «Selbstverhältnis» öffnet und in das Prinzip der Verletzbarkeit mündet, den Menschen von anderen Lebewesen. Beim Prinzip der Verletzbarkeit – weiter unten im Unterkapitel 6.4 Menschenrechte als ethische Referenzpunkte – geht es nicht nur um die rein empirisch wahrnehmbare Verletzbarkeit, sondern vor allem darum, wie der Mensch seine eigene Verletzbarkeit wahrnimmt und über sie denkt. Im Mittelpunkt stehen die Untersuchung und Reflexion von Verletzbarkeit und die moralischen Konsequenzen, die das Prinzip der Verletzbarkeit im Menschen auslöst. Verletzbarkeit kann zum Beispiel bedeuten, dass ein Mensch, der heute gesund ist, weiß, dass er morgen krank werden könnte. Oder – während er heute glücklich lebt – dass er morgen von anderen getötet werden könnte. In diesem Denkprozess durchläuft diese Person einen Prozess der Verunsicherung. Denn sie wird sich ihrer eigenen Verletzbarkeit und in letzter Konsequenz auch ihrer Vergänglichkeit bewusst.<sup>208</sup> Diese Möglichkeit der Selbstwahrnehmung gilt für alle Menschen.

Die Bewusstwerdung der eigenen Verletzbarkeit ist ein Selbstwahrnehmungsprozess des Menschen, dessen empirische Korrektheit nicht relevant ist. Während dieses Bewusstwerdungsprozesses, wenn sich ein Mensch seiner eigenen Verletzbarkeit bewusst wird, erkennt er *ex negativo* die «Erste-Person-Perspektive»<sup>209</sup>. Dies umfasst das Bewusstsein der Menschen, dass sie als singuläre Personen Subjekte der Selbsterfahrung sind, durch das sie Zugang zu ihrer eigenen Verletzbarkeit haben. Andererseits erfahren sie diese anthropologische Grundsituation der Verletzbarkeit als Subjekt (d.h. in der ersten Person Singular). Die Handlungen, Entscheidungen, Leiden

---

207 Vgl. Runggaldier 2003.

208 Vgl. Hoffmaster 2006: 42.

209 Vgl. Runggaldier 2003.



und das Leben eines Menschen gehen von ihnen als Subjekte aus. Darüber hinaus interpretieren sie diese anthropologische Grundsituation der Verletzbarkeit als Subjekt: «Denn handelnd und erleidend erfährt er sich als das Lebewesen, das nicht einfach lebt wie alle anderen Lebewesen, sondern das nur lebt, indem es *sein Leben* führt. Sich zu sich zu verhalten, weder naturnotwendig noch beliebig zu handeln, sondern sich an Gründen zu orientieren und frei gewählte Zwecke zu verfolgen, macht die Lebensform aus, die ihn mit allen Menschen als *seinesgleichen* verbindet. Sie macht ihn zugleich verletzbar, ist doch das zu seiner Lebensform gehörende Selbstverhältnis auf fundamentale Realisierungsbedingungen angewiesen.»<sup>210</sup> Der Mensch ist in der Lage, mit sich selbst in ein «Selbstverhältnis» zu treten.

Da die Menschen sich ihrer Verletzbarkeit bewusst sind, aber gleichzeitig nicht wissen, ob und wann sich diese Verletzbarkeit manifestiert und in eine konkrete Verletzung oder Übertretung umschlägt, sind sie bereit, allen Menschen die «Erste-Person-Perspektive» und das «Selbstverhältnis» auf der Grundlage der Gleichheit aller Menschen zuzugestehen, weil dies für sie die rationalste, klügste und vorteilhafteste Lösung darstellt. Das heißt, allen Menschen Rechte – also Menschenrechte – zuzugestehen, um sich selbst und alle anderen zu schützen, denn die Verletzbarkeit beinhaltet auch die «Erste-Person-Perspektive» und das «Selbstverhältnis». (Dies wird weiter unten in Unterkapitel 6.4 Menschenrechte als ethische Referenzpunkte näher erläutert).

Der Mensch ist also nicht deshalb Träger:in von Menschenrechten, weil er verletzbar ist, sondern wegen des Prinzips der Verletzbarkeit, das zum Genuss der Menschenrechte für alle Menschen führt, den Unterschied zur Verletzbarkeit von Tieren deutlich macht und als *differentia specifica* dient, die ein unterschiedliches Verständnis der Verletzbarkeit von Menschen und der Verletzbarkeit von Tieren begründet. Die Tatsache, dass Verletzbarkeit Menschen und Tiere betrifft, widerspricht daher nicht der Relevanz des Prinzips der Verletzbarkeit für die Begründung der Menschenrechte.

Verletzbarkeit, «Erste-Person-Perspektive» und «Selbstverhältnis»: Das Prinzip der Verletzbarkeit deckt wahrscheinlich das ab, was «Bewusstsein», «Willensfreiheit», «Selbstreflexivität»<sup>211</sup> und «Subjektivität»<sup>212</sup> als Unterscheidungsmerkmale zwischen Menschen und Maschinen bzw. künstlicher Intelligenz vorsehen. Gleichzeitig gehen Verletzbarkeit, «Erste-

210 Honnefelder 2012: 171-172.

211 Misselhorn 2018: 214.

212 Ohly 2019a: 49-72.

Person-Perspektive» und «Selbstverhältnis» in ihrer argumentativen Wirkung auf die Unterscheidung zwischen Menschen und Maschinen (oder künstlicher Intelligenz) über diese Unterscheidungsmerkmale hinaus, weil sie *erstens* nicht von der Bedingung eines angenommenen Dualismus zwischen Geist (oder Seele) und Körper abhängen und es daher erlauben, diese kontroverse Frage zu vermeiden. *Zweitens* – und das ist noch wichtiger – schaffen sie kein Potenzial für Diskriminierung. Wenn man «Bewusstsein», «Willensfreiheit», «Selbstreflexivität» und «Subjektivität» zu *den* Merkmalen oder Fähigkeiten unter den vielen menschlichen Merkmalen oder Fähigkeiten macht, die den Menschen von Maschinen und künstlicher Intelligenz unterscheiden, läuft man Gefahr, diejenigen Menschen zu diskriminieren, die sich in diesen Merkmalen oder Fähigkeiten unterscheiden, wie Menschen mit Behinderungen, Komapatient:innen oder Embryonen. Die Gefahr der Diskriminierung bleibt auch dann bestehen, wenn sie potenziell zur Rechenschaft gezogen wird. Martha C. Nussbaum zeigt dieses Diskriminierungspotential besonders deutlich, wenn sie – dieser Argumentationslinie folgend – so weit geht, Menschen mit Behinderungen ihre Menschlichkeit abzuspochen, weil ihnen bestimmte Fähigkeiten fehlen, ihnen aber dennoch moralische Relevanz zubilligt, um ihre Argumentation abzuschwächen: «[...] dass bestimmte schwer behinderte Säuglinge niemals Menschen sind, selbst wenn sie von zwei menschlichen Eltern geboren werden: wiederum solche mit einer globalen und totalen Sinnesunfähigkeit und/oder ohne Bewusstsein oder Denken; auch, so denke ich, solche, die überhaupt nicht in der Lage sind, andere zu erkennen oder eine Beziehung zu ihnen aufzubauen. (Das sagt natürlich nichts darüber aus, was wir ihnen moralisch schulden, es trennt diese Frage nur von moralischen Fragen über Menschen.)»<sup>213</sup>

Das Fehlen des Prinzips der Verletzbarkeit ist ein erstes Argument gegen die Moralfähigkeit von Maschinen und künstlicher Intelligenz. Das Fehlen der Verletzbarkeit ist ein Argument dagegen, Maschinen und künstliche Intelligenz als moralbedürftig oder als Wesen mit Moralgeduld zu betrachten. Als Gegenargument könnte angeführt werden, dass der Status der Moralbedürftigkeit oder der Moralgeduld von den Menschen in Abhängigkeit von ihren Beziehungen zu den Maschinen und zur künstlichen Intelligenz zugewiesen werden sollte.<sup>214</sup> Die implizite Willkür, die dieser

---

213 Nussbaum 1995: 82.

214 Vgl. Coeckelbergh 2012.

Sichtweise entspricht, erfüllt nicht die Anforderungen einer kritischen rationalen Ethik. Darüber hinaus könnte das Gegenargument angeführt werden, dass es ethisch besser ist, ein moralisch wertvolles Wesen besser zu behandeln, als es sein Status verdient, und es nicht als bloßes Objekt zu behandeln.<sup>215</sup> Die Grenze dieses Arguments ist jedoch, dass dies voraussetzen würde, dass Maschinen und künstliche Intelligenz Wesen und mehr als Objekte sind. Beides konnte bisher nicht nachgewiesen werden.<sup>216</sup> Im Gegenteil, die fehlende Verletzbarkeit und die fehlende Identität<sup>217</sup>, die oben dargelegt wurden, widersprechen der Vorstellung, dass Maschinen und künstliche Intelligenz Wesen sind, und sprechen dafür, sie als Objekte zu verstehen. Diese Wahrnehmung als Objekte bedeutet nicht automatisch, dass der Mensch ihnen gegenüber keine ethischen Verpflichtungen hat,<sup>218</sup> weil eine Misshandlung eines Objekts das menschliche Subjekt in ethisch negativer Weise verändern würde<sup>219</sup> und weil es nicht ausschließt, dass Objekte insofern moralische Relevanz haben, als ihr materieller Wert unter dem Gesichtspunkt der Nachhaltigkeit einschließlich der ökonomischen, ökologischen und sozialen Dimension betrachtet werden muss.

### 3.3 Gewissen<sup>220</sup>

Eine zweite Infragestellung des Begriffs «moralische Technologien» basiert auf dem Konzept des Gewissens, das für den Menschen und seine Moral von zentraler Bedeutung ist. Das Gewissen vereint das, was objektiv geboten ist, und das, was subjektiv in einer spezifischen und konkreten Situation, in einem spezifischen Kontext, bei einer einzigartigen Begegnung mit einzigartigen Menschen erlebt wurde. «Das Gewissen ist ein aktives Vermögen, das in der Komplexität einer jeden Situation das Gute entdeckt und erkennt»<sup>221</sup>. Das Gewissen schafft in einer Person eine Autorität, die sich *a priori*, aber auch *a posteriori* auf eine Handlung auswirkt. Das Gewissen ist ein Prozess, der einer Handlung vorausgeht, sich aber auch auf die

215 Vgl. LaBossiere 2017.

216 Vgl. zu dieser Frage Gunkel 2018.

217 Vgl. DiGiovanna 2017.

218 Vgl. Bryson 2010.

219 Vgl. Kant 1990.

220 Vgl. Kirchschräger 2017d.

221 Hogan 2004: 86-87.

kritische Befragung und Prüfung nach einer Handlung bezieht.<sup>222</sup> Aber das Gewissen handelt nicht selbst.<sup>223</sup> «Das Gewissen macht man nicht geltend, wenn es einem in den Kram passt. Das Gewissen macht sich selber geltend, oft wenn es uns gar nicht in den Kram passt. Was das Gewissen ist, ist uns nicht im letzten verfügbar. [...] Diese Erfahrung hört nicht auf. Daraus lernen wir, dass das Gewissen im Werden ist, einen Prozess darstellt, der uns zugleich vorgegeben und aufgegeben ist.»<sup>224</sup> Das Gewissen erschließt also den Unterschied zwischen der Seinsebene und der Sollensebene.<sup>225</sup>

Als Zeugnis der menschlichen Moralfähigkeit verbindet das Gewissen die Menschen über Kulturen, Traditionen, Religionen und Weltanschauungen hinweg. Das Gewissen kann daher auf einer konstituierenden Ebene des Menschseins verortet werden. Es umfasst das Potenzial, das moralisch Richtige und Gute zu erkennen und in den jeweils spezifischen Kontext einzuordnen.<sup>226</sup>

Darüber hinaus beschreibt das Gewissen eine Wechselwirkung zwischen dem normativen System, auf dem es beruht, und den inneren Aspekten des Individuums.<sup>227</sup> Ersteres drückt sich auch durch die Einordnung des Individuums in einen gesellschaftlichen Kontext mit den daraus resultierenden Pflichten und der entsprechenden Verantwortung aus.<sup>228</sup> Letzteres garantiert eine kritische Distanz des Individuums gegenüber dem normativen System einer Gesellschaft.<sup>229</sup>

Das Selbstverhältnis, die Beziehung zur Normativität und die Beziehung zu einem normativen System beruhen also auf dem Gewissen oder sind darin vereint.<sup>230</sup> Bei näherer Betrachtung kann das Gewissen verstanden werden als die «Fähigkeit des menschlichen Geistes zur Erkenntnis der sittlichen Werte, Gebote und Gesetze (*Synderesis*), im engeren Sinne aber deren Anwendung auf das eigene, unmittelbar zu vollziehende Handeln.»<sup>231</sup> Das Gewissen übernimmt die Rolle einer inneren Stimme des Menschen in moralischen

---

222 Vgl. auch Holzhey 1975: 7; Schüller 1980: 40-57.

223 Vgl. auch Wolbert 2008: 170.

224 Mieth 1992: 225.

225 Vgl. Reiter 1991: 11.

226 Vgl. Schuster / Kerber 1996: 144.

227 Vgl. Kranich-Strötzel 2008: 124.

228 Zum Diskurs über konkrete Gewissensfragen vgl. Schaupp 2014; Höfling 2014; Martinsen 2004.

229 Vgl. Reiter 1991: 15.

230 Vgl. Kranich-Strötzel 2008: 125.

231 Schuster / Kerber 1996: 144; zum philosophischen Diskurs über das «Gewissen» vgl. Hübsch 1995.

Fragen und Entscheidungen. Diese innere Stimme ist nicht immer eindeutig, sondern das Gewissen kämpft mit sich und seinen Entscheidungen, wie die Definition von Immanuel Kant zeigt: «Das Bewusstsein eines inneren Gerichtshofes im Menschen (vor welchem sich seine Gedanken einander verklagen oder entschuldigen<sup>232</sup>) ist das Gewissen.»<sup>232</sup> Anstelle eines heteronomen Gewissensverständnisses, bei dem das Gewissen durch etwas bestimmt wird, das der menschlichen Freiheit und Vernunft fremd ist, verwendet Kant die Autonomie zur Wahrung der Unabhängigkeit des moralischen Phänomens oder der Pflicht. Moral im engeren Sinne bedeutet, dass sie rein subjektbezogen ist. Sittlichkeit ist die Pflicht, die der Mensch als etwas anerkennt, was ihm seine praktische Vernunft als seine Pflicht vorgibt. Nicht nur pflichtgemäßes Handeln, sondern Handeln aus Pflichtgefühl ist moralisch gut und moralisch richtig, denn rein pflichtgemäßes Handeln kann aus verschiedenen Gründen geschehen, die nichts mit Moral oder gar Unmoral zu tun haben. Da das Gewissen sich nur an sich selbst misst, ist ein falsches Gewissen nach Kant unmöglich. Nach seinem Gewissen zu handeln, kann keine zusätzliche Pflicht sein, sonst bräuchte es ein Gewissen über dem Gewissen, das das gesuchte Gewissen wäre. Es ist auch Sache des Gewissens selbst, zu erkennen, was eine Pflicht ist. Die Pflicht zur Gewissensbildung ergibt sich aus dem Gewissen selbst. Nur dann ist sie eine moralische Pflicht. In jedem anderen Fall käme die Pflicht von außen, wäre fremdbestimmt und daher nicht moralisch, sondern legalistisch.

Die praktische Vernunft ist gewissermaßen der Gegenpol zum Gewissen. Die praktische Vernunft erinnert uns daran, ob eine Handlung pflichtgemäß ist oder nicht und bestimmt die Objekte, die pflichtgemäß sind. Das heißt, ihr werden alle inhaltlich bestimmten Funktionen entzogen. Stattdessen ist das Gewissen die Pflicht an sich. Nicht das Gewissen, sondern die praktische Vernunft entscheidet über den Inhalt der Pflicht. Das Gewissen beurteilt nur das Verhältnis des Menschen zu seiner Pflicht wie «eines inneren Gerichtshofes im Menschen (vor welchem sich seine Gedanken einander verklagen oder entschuldigen<sup>233</sup>)»<sup>233</sup>. Auf der Grundlage der Autonomie der Moral ist das Gewissen mit einer Selbstverpflichtung des Menschen gleichzusetzen.

Hilfreich ist auch die von Thomas von Aquin eingeführte Unterscheidung zwischen dem «Urgewissen» (synteresis) – dem natürlichen Habitus,

---

232 Kant 1997: 572.

233 Kant 1997: 572.

in dem der Mensch an der ewigen Wahrheit der praktischen Vernunft teilhat, die auch den obersten Grundsatz des moralischen Gewissens «Du sollst das Gute tun und das Böse unterlassen» enthält – als Grundlage des Handelns und dem darauf aufbauenden konkreten Gewissen bzw. Funktionsgewissen, das in seiner Anwendung auf das konkrete Handeln (*conscientia*) besteht.<sup>234</sup> «Das Gewissen hilft, in der Gestaltung der Freiheit realistisch und vernünftig zu sein – ja, auch seine moralischen Grenzen und Abhängigkeiten zu erkennen. Aber es macht zugleich immer wieder Mut, nicht zu resignieren und an die Möglichkeit zu glauben, dass man sein Leben nach moralischen Überzeugungen und Idealen tatsächlich verändern kann.»<sup>235</sup>

Darauf aufbauend geht die Unterscheidung des Gewissens in ein «Urgewissen», ein «Wertegewissen» und ein «Situationsgewissen» noch weiter.<sup>236</sup> Während sich das «Urgewissen» auf die grundsätzliche Forderung konzentriert, das Gute dem Bösen vorzuziehen und entsprechend zu handeln, entscheidet das «Situationsgewissen» anhand der konkreten Situation, welche konkreten Handlungen von moralischem Wert zu setzen sind, und berücksichtigt dabei persönliche Bedürfnisse, natürliche Neigungen, persönliche Werte, Pflichten und Verantwortlichkeiten. Das «Wertegewissen» vermittelt zwischen dem «Urgewissen» und dem «Situationsgewissen», indem es die persönlichen Eindrücke und Einstellungen berücksichtigt. Die Vermittlung des Gewissens beinhaltet also die verschiedenen Dimensionen der Moral, die sich auch in der Reifung des Gewissens manifestieren kann. Auf dieser Grundlage lässt sich über das Gewissen Folgendes sagen: «In seinem Vollzug als Urteil ist es wie das sittliche Urteil ein mannigfach bedingtes, nämlich mehrfach gestuftes und inhaltlich von höchst vielfältigen Grössen bestimmtes Geschehen. In dem Anspruch dagegen, den es habituell festhält und im Urteil zur Geltung bringt, ist es gänzlich unbedingte, kommt doch in ihm im Modus des Selbstbezugs nichts anderes zum Ausdruck als das, was Grund sittlicher Verbindlichkeit überhaupt ist, nämlich das Selbstsein des Menschen als Vernunft- und Freiheitswesen.»<sup>237</sup>

Das Gewissen verbindet also die Moral oder die Pflicht mit den verschiedenen Ebenen des Menschen und seiner Existenz. Letztere sind von

---

234 Thomas von Aquin *Summa theologiae* I q 79 a 12-13; vgl. Anzenbacher 2015; Noichl 1993: 264-274.

235 Römel 2011: 58.

236 Vgl. Teichtweiter 1976.

237 Honnefelder 1993: 121.

unterschiedlicher Qualität und Intensität und werden durch individuelle Entwicklung oder gesellschaftliche Einflüsse geprägt.<sup>238</sup> Der Mensch bildet auch sein bzw. ihr Gegenstück. «Jede Beanspruchung von Gewissen greift auf Selbstbejahung zurück und wiederholt sie. Wie das Selbst der Bejahung nichts Einsames ist, so handelt auch das Gewissen nicht lautlos-verschwiegen, verallgemeinerungsunfähig, selbstisch. [...] Wiederholt Gewissen im Rückbezug auf Selbstbejahung den entscheidenden Augenblick der Einleitung von Individuation und Sozialisation, dann unternimmt es dabei nichts zur 'einsamen' und 'nackten' Selbstfindung, Selbstwahl und Selbstverwirklichung.»<sup>239</sup>

Darüber hinaus geht es beim Gewissen um Entscheidungen, Handlungen und Unterlassungen des Menschen in einer einzigen Lebenssituation, die nicht übersehen werden kann. Alle diese Situationen erheben Anspruch auf Moral. Der Mensch betrachtet diesen Anspruch auf unterschiedlichste Art und Weise. Der Gegenstand des moralischen Handelns ist die Gestaltung der eigenen Lebensgeschichte. In die jeweiligen bewussten Entscheidungen werden auch emotionale Momente einbezogen. Weil «nur die Vernunft das Vermögen sein [kann], das (im Akt des Gewissens) die Vielheit der Handlungen auf eine letzte Einheit, nämlich auf das Gelingen der menschlichen Existenz als ganzer bezieht.»<sup>240</sup> Diese umfassende Besorgnis lässt auch viele Wege zur Erkenntnis und Wahrnehmung des Gewissens zu. «Das Gewissen missrät nur dann nicht zur autoritären Kontrollinstanz, wenn es sich öffnet, kontextuell agiert und aus eigener – kritisch begleiteter – Anschauung der Wirklichkeit seine sozialpraktische Sensibilität und Urteilskraft aufbaut.»<sup>241</sup>

Zugleich wird der Mensch nachdrücklich dazu angehalten, die objektiven und subjektiven Bedingungen zu überprüfen, indem er sich seiner eigenen Identität, seines Selbstverständnisses und seiner Selbsteinschätzung im Gewissen durch die Reflexion über Normen und die Gestaltung und Setzung von Normen bewusst wird. Das Gewissen stellt sich also in den Dienst der Verwirklichung des Willens im Menschen, damit dieser sich dem ersten so weit wie möglich annähert.

Darüber hinaus wirkt sich der Horizont des Gewissens aus, der nicht bei einer einzigen Entscheidung oder Handlung endet, sondern die gesamte

238 Vgl. Schmitt 2008.

239 Marten 1975: 124.

240 Honnefelder 1982: 32-33; vgl. Honnefelder 1993.

241 Schmitt 2008: 162.

Existenz des Menschen umfasst. Das Gewissen erweist sich als «der geduldige Weg der lernenden Freiheit.»<sup>242</sup>

Die im Gewissen erfahrene moralische Vernunft des Menschen und sein moralisches Vermögen, zwischen Gut und Böse zu unterscheiden, macht das Gewissen schließlich zu einer absoluten Herausforderung für den Menschen – immer wieder, wenn er in einem letzten praktischen Urteil, das zum Handeln führt, eine Handlung plant und erwägt, und wenn er nach einer Handlung die Infragestellung dieses Urteils hinsichtlich seiner Stellung zur moralischen Vernunft, seines Beitrags zum Gelingen des Menschseins und seines Verhältnisses zu den eigenen Handlungsprinzipien kritisch überprüft.

Schließlich drückt das Gewissen ein Vertrauen in den einzelnen Menschen aus. Vom einzelnen Menschen wird erwartet, dass er diese innere Stimme in moralischen Fragen hat, dass er sie erkennt, auf sie hört und dann verantwortungsvoll handelt. Es wird respektiert und hochgehalten, dass die Würde des Gewissens dem einzelnen Menschen zukommt.

Auch wenn im Sinne der epistemischen Bescheidenheit einerseits darauf hingewiesen werden muss, dass die Bestimmung des Gewissens äußerst komplex ist, und andererseits daran erinnert werden muss, dass man versucht, auf der Grundlage des gegenwärtigen Forschungsstandes denkbare und plausible Aussagen über die Zukunft zu treffen, so lässt sich doch Folgendes festhalten: Man kann nicht sagen, dass Technologien ein Gewissen haben. Die Möglichkeiten, die Technologien in Bezug auf ethische Entscheidungen und Handlungen besitzen, reichen bei weitem nicht an das menschliche Gewissen heran. Es fehlt ihnen an verschiedenen Ebenen der Moral oder der Pflicht sowie an einer im Gewissen verankerten Existenz in unterschiedlicher Qualität, Intensität und geprägt von individueller Entwicklung oder gesellschaftlichem Einfluss.<sup>243</sup> Daher kann man nicht sagen, dass Technologien ein Gewissen haben. Wenn man das Gewissen als wesentlich für die Moral versteht, ist das Fehlen des Gewissens ein zweites Argument gegen die Moralfähigkeit der Technologien.

Es ist nicht das stärkste Argument, denn im gegenwärtigen Diskurs gerät das Gewissen sowohl aus psychologischer und soziologischer als auch aus neurowissenschaftlicher Perspektive unter Druck. In der ersten belassen sie es bei tiefenpsychologischen Ansätzen, die die Selbstbestimmung zwar

---

242 Römelt 2011: 58-61.

243 Vgl. Schmitt 2008.



noch hochhalten, ihr aber einige Einschränkungen auferlegen. Letztere beruft sich auf Hinweise auf neuronale Prozesse und chemisch-biologische Vorgänge, die es nur zulassen, dass die Autonomie des Gewissens tatsächlich nur vorgetäuscht wird und das Gewissen in empirischer Wahrheit nichts anderes als das Ergebnis dieser Vorgänge ist.<sup>244</sup> Diesen beiden Argumentationslinien zur Infragestellung des Gewissens kann man im Sinne einer Analogie zu den Argumenten Immanuel Kants bezüglich der geistigen Freiheit des Menschen entgegentreten:<sup>245</sup> Wenn dies zuträfe, d.h. wenn der Mensch in seiner Selbstbestimmung eingeschränkt wäre und nichts als chemische und biologische Prozesse wäre, wie wäre es dann zu erklären, dass sich der Mensch gegen «sinnliche Reize» entscheiden und entsprechend handeln kann oder dass sich der Mensch von «Bewegursachen, welche nur von der Vernunft vorgestellt werden»<sup>246</sup> leiten lassen?

### 3.4 Freiheit

Ein drittes Fragezeichen hinsichtlich der Moralfähigkeit von Maschinen ergibt sich aus der Freiheit. Freiheit ist eine *conditio sine qua non* für die Moral, denn erst die Freiheit eröffnet die Möglichkeit, sich für oder gegen das Gute bzw. das Richtige zu entscheiden. Freiheit ist vieldeutig. Als formale Beziehung kann Freiheit als «Freiheit von...» und «Freiheit zu...» beschrieben werden. Freiheit bedeutet, nach den eigenen Wünschen und Plänen zu handeln. Sie kann die Freiheit umfassen, zu wollen, was man will, und die Freiheit, zu wollen, was man nicht will. Letzteres bedeutet, dass Freiheit auch bedeuten kann, das «Gesollte», d.h. das ethisch Gesollte zu wollen, auch wenn dies nicht den eigenen Wünschen, Bedürfnissen, Vorlieben oder Interessen entsprechen mag. Damit öffnet sich der soziale Horizont der Freiheit.

Darüber hinaus ist die Freiheit der Ursprung von Wissenschaft, Forschung und Technologie. Dieser Aspekt muss in einer Zeit hervorgehoben werden, in der manche Stimmen die Existenz der Freiheit gänzlich leugnen.<sup>247</sup> «Freiheit, die nun verneint wird, hat die Entwicklungen der Wissenschaft, im Namen derer sie nun verneint wird, erst möglich gemacht. In

244 Vgl. Roth 2003.

245 Vgl. auch Nida-Rümelin 2005.

246 Kant 1995a: 675.

247 Zu diesem Diskurs vgl. Holderegger et al. 2007; Fink / Rosenzweig 2006; Fleischer 2012; Bloch 2011; Bauer 2007; Achtner 2010; Guckes 2003.

der Tat hätte es nie eine Wissenschaft gegeben ohne die dem menschlichen Geist eigene Fähigkeit, zwischen Falsch und Wahr zu unterscheiden und das Wahre dem Falschen gegenüber vorzuziehen. Falsch und Wahr ergeben keinen Sinn, wenn nicht für einen freien Geist, der fähig ist, das eine anzustreben und das andere zu verwerfen. Ohne diese wesentliche Voraussetzung bleibt jede Erklärung bloss ein lautstarker, sinnloser Akt. Aus diesem Grund kann mit Fug und Recht gesagt werden, dass die Wissenschaft das glanzvollste Denkmal ist, das die Freiheit sich selbst errichtet hat, und dass wissenschaftliche Forschung ohne Freiheit völlig undenkbar ist.»<sup>248</sup>

Maschinen haben keine Freiheit. Technologien werden von Menschen entworfen, entwickelt und gebaut, das heißt, sie werden heteronom produziert. Daher wird auch das Erlernen von ethischen Prinzipien und Normen von Menschen geleitet. In letzter Konsequenz würden Maschinen immer von außen gesteuert werden. Bildlich gesprochen: Maschinen – auch selbstlernende Maschinen – werden auf eine erste Codezeile zurückgehen, die immer vom Menschen stammt. Freiheit ist ein drittes, *wichtiges* Argument gegen die Moralfähigkeit von Maschinen.

### 3.5 Verantwortung

Die Freiheit zu wollen, was man nicht will, zeichnet Verantwortung aus.<sup>249</sup> Verantwortung gelingt es, die eigene Freiheit mit der Freiheit aller anderen Menschen zu verbinden und die Menschenwürde aller Menschen zu achten. Verantwortung ermöglicht die Freiheit, über die eigenen Bedürfnisse und Interessen hinaus den Horizont für die Freiheit aller anderen Menschen und für gesellschaftliche Aufgaben und Ziele zu entdecken. «Die Verantwortung bricht die individualistische und auf eigene Bedürfnisse konzentrierte Freiheit auf und bindet sie ein in soziale Gefüge, in gemeinsame Aufgaben und Ziele.»<sup>250</sup> Verantwortung ist auch eine *conditio sine qua non* für Moral. Um Verantwortung zu tragen oder übertragen zu bekommen – um ein Subjekt der Verantwortung sein zu können – sind Freiheit und Rationalität notwendig.

Es stellt sich die Frage, ob Verantwortung von Maschinen übernommen werden kann. Die Antwort muss negativ ausfallen, denn Maschinen kön-

---

248 Hersch 1992: 60-61.

249 Vgl. Kirchschräger 2014a.

250 Holderegger 2006: 401.

nen kein Verantwortungssubjekt darstellen, weil ihnen die Freiheit fehlt – ein viertes, *wichtiges* Argument gegen die Moralfähigkeit von Maschinen.

### 3.6 Autonomie

Eine fünfte grundsätzliche Frage bezüglich der Zuschreibung von Moral an Maschinen ergibt sich aus der vom Menschen für sich proklamierten Autonomie. Immanuel Kant verknüpft die Würde des Menschen mit seiner Autonomie.<sup>251</sup> Der Mensch ist Träger:in der Würde und darf deshalb nicht instrumentalisiert werden, weil er als vernünftiges Wesen allgemeine moralische Regeln und Prinzipien für sich selbst erkennt, sie selbst bestimmt und seinem Handeln zugrunde legt.<sup>252</sup> Die Würde des Menschen «als vernünftiges Wesen, das keinen anderen Gesetzen gehorcht als denen, die er sich selbst gegeben hat»<sup>253</sup>, beruht auf der Fähigkeit des Menschen, sich selbst Vernunftregeln zu setzen. Das bedeutet, dass moralische Regeln und Prinzipien, die der Mensch in seiner Autonomie formuliert, den folgenden Anforderungen einer kritischen, rationalen Moral genügen müssen, was ihre Universalität garantiert: Universalität setzt die Erfüllung des Prinzips der Verallgemeinerbarkeit durch die Vorlage rationaler und plausibler Argumente – «guter Gründe» – voraus. «Gute Gründe» bedeutet, dass es denkbar sein muss, dass alle Menschen in ihrer effektiven Freiheit und Autonomie sowie in ihrer vollen Gleichheit diesen Gründen – innerhalb eines Denkmodells und nicht innerhalb eines realen globalen Referendums – aus ethischen Gründen zustimmen würden.

Entspricht die Beschreibung der menschlichen Autonomie, die von Menschen zum Ausdruck gebracht werden kann, dem Potenzial von Technologien,<sup>254</sup> moralischen Regeln zu folgen, dementsprechend moralische Entscheidungen zu treffen und entsprechende Handlungen auszuführen? Hinsichtlich des Begriffs «Autonomie» besteht eine Kluft zwischen Technologie und Ethik.<sup>255</sup> Während der Mensch allgemeine moralische Regeln und Prinzipien für sich selbst erkennt, sie für sich selbst festlegt und sein Handeln daran ausrichtet, ist dies bei Technologien nicht möglich. Techno-

251 Vgl. Kant 1974: 69.

252 Vgl. Kant 1974: 74.

253 Kant 1974: 67.

254 Vgl. z. B. Decker 2019a; Decker 2019b; Thimm/Bächle 2018.

255 Vgl. Kirchschräger 2017c.

logien sind in erster Linie auf ihre Tauglichkeit hin gemacht und können als selbstlernendes System Regeln aufstellen, zum Beispiel um ihre Effizienz zu steigern. Aber diese Regeln enthalten keine ethischen Qualität. Maschinen scheitern an dem oben erwähnten Prinzip der Verallgemeinerbarkeit. Diese Negation ist ein fünftes, *wichtiges* Argument gegen den Begriff «moralische Technologien». Verstärkt wird diese Verneinung noch dadurch, dass Technologien ohne Freiheit keine Autonomie haben können.<sup>256</sup>

Was ist mit «selbstlernenden Systemen» und ihrer Moralfähigkeit? Unter selbstlernenden Systemen versteht man Maschinen, die danach streben, keinen oder so wenig wie möglich menschlichen Input zu benötigen, um ein Ziel zu erreichen. «Der Unterschied ist, dass Maschinen nun denken können, wenn auch limitiert. Sie lösen Probleme, treffen Entscheidungen und – am wichtigsten: Sie lernen.»<sup>257</sup> Wenn sich selbstlernende Systeme auch ohne menschliches Zutun verbessern könnten, bestünde die Möglichkeit, dass sie ethisch besser werden, was letztlich zu ihrer Autonomie führen könnte. Selbst im Fall von «selbstlernenden Systemen» würde das Fehlen von Verletzbarkeit, Gewissen, Freiheit, Verantwortung und Autonomie, wie oben gezeigt wurde, ihrer Moralfähigkeit widersprechen. (Dies würde auch für die künstliche allgemeine Intelligenz gelten.) Da ihr Selbstlernen auf praktischen Fehlern beruht, kann die Möglichkeit eines solchen moralischen Lernprozesses wegen des Vorrangs des Prinzips der Nichtschädigung des Menschen grundsätzlich bezweifelt werden.<sup>258</sup>

Ändert sich diese Einschätzung hinsichtlich der Moralfähigkeit von Maschinen im Falle der «Superintelligenz»? Unter «Superintelligenz»<sup>259</sup> versteht man Systeme, die generell intelligenter sind als der Mensch. Da Maschinen den Menschen bereits heute in verschiedenen Bereichen der Intelligenz massiv übertreffen (z.B. Gedächtnis, Umgang mit großen Datenmengen etc.), ist zu erwarten, dass weitere Bereiche der Intelligenz hinzukommen werden.

Es lassen sich drei Formen von Superintelligenz unterscheiden: Superintelligenz in Bezug auf die Geschwindigkeit (wie menschliche Intelligenz,

---

256 Dieser Aspekt wurde z.B. auch in der Stellungnahme der European Group on Ethics in Science and New Technologies to the European Commission berücksichtigt, vgl. European Group on Ethics in Science and New Technologies to the European Commission 2018a.

257 Metzler 2016.

258 Vgl. Neuhäuser 2012.

259 Vgl. Bostrom 2014.

aber viel schneller)<sup>260</sup>, kollektive Superintelligenz («überlegene Leistung durch die Zusammenfassung einer großen Zahl kleinerer Intelligenzen»<sup>261</sup>) und qualitative Superintelligenz (welche die menschliche Intelligenz bei einer Geschwindigkeit, die mindestens so hoch ist wie die menschliche Intelligenz, deutlich übertrifft)<sup>262</sup>.

Die Frage nach der Moralfähigkeit der «Superintelligenz» erweist sich als äußerst relevant. «Die Herausforderung, die die Aussicht auf Superintelligenz darstellt, und die Frage, wie wir am besten darauf reagieren können, ist möglicherweise die wichtigste und entmutigendste Herausforderung, der sich die Menschheit je gestellt hat. Und – ob wir Erfolg haben oder nicht –, es ist wahrscheinlich die letzte Herausforderung, der wir uns jemals stellen werden.»<sup>263</sup> Unter anderem stellt sich die ethische Frage, was mit den Menschen geschieht, wenn superintelligente technologische Systeme, die selbstlernend sind, beschließen, sich ihre eigenen Ziele zu setzen.

Aufgrund der obigen Ausführungen kann auch nicht gesagt werden, dass eine Superintelligenz Verletzbarkeit, Gewissen, Freiheit, Verantwortung und Autonomie hätte. Daher wäre es auch nicht möglich, bei ihr eine Moralfähigkeit zu denken.

Dieses technologische Potenzial schafft jedoch die ethische Chance, dass superintelligente technologische Systeme ethisch bessere Entscheidungen treffen und ethisch bessere Handlungen vollziehen als Menschen, obwohl sie beides tun, weil sie die langfristigen Vorteile für sich selbst sehen, die sich aus ethischen Entscheidungen und ethischem Verhalten ergeben, also letztlich auf pragmatischen Gründen beruhen. (Dies würde natürlich streng genommen die ethische Natur ihrer Entscheidungen und ihres Verhaltens in Frage stellen). Dennoch könnten superintelligente technische Systeme diese positiven Auswirkungen aus ethischer Sicht haben.

Um einem möglichen Missverständnis vorzubeugen, wenn die Moralfähigkeit von Maschinen negiert wird: Maschinen können mit ethischen Prinzipien und Regeln programmiert oder trainiert werden, um ethisch legitime Entscheidungen und Handlungen einer Maschine zu erreichen,<sup>264</sup> auch wenn sie selbst diese nicht erkennen und bestimmen können und auch wenn sie die ethische Qualität von Prinzipien und Regeln mangels Moralfähigkeit nicht erkennen können.

260 Vgl. Bostrom 2014: 53-54.

261 Bostrom 2014: 54.

262 Vgl. Bostrom 2014: 55-57.

263 Bostrom 2014: vii.

264 Vgl. Wallach / Allen 2009; Moor 2006.

Auf dieser Grundlage ist es schließlich notwendig, die bereits vorhandenen Möglichkeiten und Fähigkeiten von Technologien genau zu ordnen und begrifflich zu fassen. Deshalb sollte zum Beispiel im Bereich der Automatisierung der Mobilität der Begriff «autonomes Fahrzeug» vermieden und durch «automatisiertes Fahrzeug» oder «selbstfahrendes Fahrzeug» ersetzt werden.<sup>265</sup> Es ist beispielsweise zweifelhaft, ob die Behauptung, dass Roboter, die es geschafft haben, beim Go-Spiel – einem Spiel, das nicht nur auf logischer Schlussfolgerung, sondern auch auf Intuition<sup>266</sup> beruht – gegen Menschen zu gewinnen, über «Intuition» oder «strategisches Denken» verfügen, angemessen wäre, oder ob die Phänomene, die bei Robotern als «Intuition» oder «strategisches Denken» angesehen werden, nicht in letzter Konsequenz auf Wahrscheinlichkeitsberechnungen beruhen. Es erscheint auch unangemessen, einem Roboter ein «Pokerface» zuzutrauen, wenn er in einem 20-tägigen Pokermarathon gegen menschliche Spieler gewinnt.<sup>267</sup> Realistischerweise könnte dies auf die Fähigkeit zurückgeführt werden, mit unvollkommenen Informationen strategisch zu denken. Ähnlich verhält es sich mit dem Begriff «moralische Technologien». Roboter werden in erster Linie wegen ihres Nutzens gebaut. Sie können aber auch in ethischen Regeln geschult oder mit ethischen Regeln programmiert werden, um ein aktives Robotersystem in die Lage zu versetzen, auf ethisch vertretbare Weise zu handeln.<sup>268</sup> Dieses Potenzial führt zu dem Begriff «moralische Technologien». Dabei darf aber Folgendes nicht außer Acht gelassen werden: «Wenn wir über Roboterethik sprechen, sollten wir über normative Ethik für den Einsatz von Robotern sprechen, d.h. das richtige und falsche Verhalten von Robotern liegt in der Verantwortung der Roboterbenutzer:innen und nicht der Roboter selbst. (...) ein Roboter sollte nicht von sich aus ethisch sein; er sollte ethisch genutzt werden.»<sup>269</sup>

Daher wären andere Begriffe als «moralische Technologien» angemessener, wie «technische Werkzeuge»<sup>270</sup>, «intelligente Computerschnittstellen»<sup>271</sup>, «soziotechnische Systeme»<sup>272</sup> oder «impliziter ethischer Agent»<sup>273</sup>.

---

265 Vgl. Kirchschräger 2017b.

266 Vgl. Natur 2016.

267 Vgl. Spice 2017.

268 Vgl. Wallach/Allen 2009.

269 Krenn 2016: 17.

270 Vgl. Engineering and Physical Sciences Research Council 2011.

271 Vgl. Van Est / Stermerding 2012.

272 Vgl. Manzeschke et al. 2013.

273 Vgl. Krenn 2016.

Der letzte Begriff ist allerdings auch problematisch, da er Roboter als «Agenten» oder «ethische Agenten» bezeichnen würde, was moralische Autonomie als Voraussetzung voraussetzen würde.<sup>274</sup> Denn «Agenten» «überlegen, bewerten, wählen und handeln, um das zu erreichen, was wir als ein gutes Leben für sie ansehen»<sup>275</sup>.

Eine andere Lösung könnte in der Einführung mehrerer Ebenen der «Autonomie» (oder der Moral) bestehen: «Autonomie bei Maschinen und Robotern sollte in einem engeren Sinne als beim Menschen (d.h. metaphorisch) verwendet werden. Insbesondere kann die Autonomie von Maschinen und Robotern nicht absolut definiert werden, sondern nur relativ zu den erforderlichen Zielen und Aufgaben. Natürlich kann es häufig vorkommen, dass die Ergebnisse der Handlungen einer Maschine/eines Roboters der menschlichen Designer:in und Operator:in nicht im Voraus bekannt sind. Das bedeutet jedoch nicht, dass die Maschine/der Roboter ein (völlig) autonomer und unabhängiger Akteur ist, der selbst entscheidet, was zu tun ist. Tatsächlich können Maschinen und Roboter als teilautonome Agenten betrachtet werden, in diesem Fall können wir mehrere Ebenen der 'Autonomie' (...) [bzw.] mehrere Ebenen der 'Moral'«<sup>276</sup> haben. Diese Ebenen umfassen die «operationelle Moral» («die moralische Bedeutung und Verantwortung liegt vollständig bei den Menschen, die an ihrer Entwicklung und Nutzung beteiligt sind, weit entfernt von einer vollständigen moralischen Handlungsfähigkeit. Die Computer- und Software-Wissenschaftler:innen und -Ingenieur:innen, die heutige Roboter und Software entwerfen, können in der Regel alle möglichen Situationen vorhersagen, mit denen der Roboter konfrontiert wird»<sup>277</sup>), «funktionelle Moral» («die Fähigkeit des ethischen Roboters, moralische Urteile zu fällen, wenn er ohne direkte Anweisungen von oben nach unten eine Handlungsweise beschließt. In diesem Fall können die Designer:innen die Handlungen des Roboters und ihre Folgen nicht mehr vorhersagen»<sup>278</sup>) und «vollständige Moral» («ein Roboter, der so intelligent ist, dass er seine Handlungen völlig autonom auswählt und somit voll verantwortlich dafür ist. Tatsächlich kann die moralische Entscheidungsfindung als natürliche Erweiterung der technischen Sicherheit für Systeme mit mehr Intelligenz und Autonomie

274 Vgl. Griffin 2001: 311-312.

275 Griffin 2008: 32.

276 Tzafestas 2016: 2.

277 Tzafestas 2016: 73.

278 Tzafestas 2016: 73.

betrachtet werden»<sup>279</sup>). Selbst die Einführung von Autonomiestufen oder von Moral kann die oben dargelegten grundlegenden Kritikpunkte nicht ausräumen. Es bleibt problematisch, «Moral» oder «Autonomie» für ein technologisches System zu verwenden, das heteronom programmiert ist, um heteronom definierten Prinzipien und Regeln zu folgen, anstatt in der Lage zu sein, diese autonom für sich selbst zu definieren, autonom zu entscheiden, sie zu befolgen oder nicht, und sich deren ethischen Qualität bewusst zu sein. Es bleibt problematisch, «Moral» oder «Autonomie» für ein technologisches System zu verwenden, dem es an Verletzbarkeit, Gewissen, Freiheit, Verantwortung und Moralfähigkeit fehlt.

Ein mögliches Gegenargument könnte darin bestehen, auf die Stadien der moralischen Entwicklung des Menschen – z. B. im Modell von Kohlberg<sup>280</sup> – zu verweisen, eine Ähnlichkeit der Stufen mit diesen Stadien zu behaupten und den Grund abzuleiten, warum, wenn sogar Menschen auf niedrigeren Stufen der moralischen Entwicklung als «moralisch» gelten, dies nicht auf die Stufen der «Moral» von Robotern zutreffen sollte. Dieser Herausforderung könnte man begegnen, indem man auf das Potenzial verweist, das Menschen ergreifen können, um ihre moralischen Fähigkeiten weiterzuentwickeln (was realisiert werden kann oder auch nicht), als Grundlage für die Angemessenheit des Verständnisses von Menschen auf niedrigeren Stufen der moralischen Entwicklung als «moralisch».

Während die Definition dieser Ebenen als Ebenen der «Autonomie» und «Moral» kritisiert wird, ohne sie mit «Autonomie» und «Moral» zu verbinden und ohne sie als «autonom» und «moralisch» zu verstehen, kann die Unterscheidung dieser Ebenen hilfreich sein, um die verschiedenen Ebenen der Fähigkeit technischer Systeme zu kategorisieren, rechtliche und ethische Prinzipien und Regeln zu befolgen und – im Falle der letzten Ebene – zu unterscheiden, was für technische Systeme aufgrund ihrer fehlenden «Autonomie» und ihrer fehlenden Moralfähigkeit unmöglich ist.

---

279 Tzafestas 2016: 73.

280 Vgl. Kohlberg 1981; Kohlberg 1984.



### 3.7 Ethisches Urteilen von Menschen

Die obige Kritik führt zu der Hauptkonsequenz, dass der Mensch dafür verantwortlich ist, ethische Entscheidungen zu treffen,<sup>281</sup> ethische Prinzipien sowie ethische und rechtliche Normen festzulegen, den Rahmen, die Ziele und die Grenzen der digitalen Transformation und des Einsatzes von künstlicher Intelligenz zu definieren und technologiebasierte Innovationen aus ethischer Sicht zu untersuchen, zu analysieren, zu bewerten und zu beurteilen. Aufgabe der Ethik ist es, «die moralischen Instrumente bereitzustellen, die die Gesellschaft und den Einzelnen dazu ermutigen, den Missbrauch der technologischen Errungenschaften der Robotik gegen die Menschheit zu verhindern. Die Gesetzgebung sollte wirksame und gerechte rechtliche Instrumente zur Verfügung stellen, um von einem solchen Missbrauch abzuschrecken und ihn zu verhindern, und die Haftung in Fällen von Schäden durch Robotermissbrauch und menschliches Fehlverhalten zu regeln.»<sup>282</sup> Es ist notwendig, dass die Menschen einen ständigen ethischen und rechtlichen Diskurs über den Einsatz, die Prioritäten, die Zusammenhänge und die Grenzen des technologischen Fortschritts führen.<sup>283</sup>

### 3.8 Überwindung der Dichotomie «Mensch – Maschine» mit Hilfe des interaktionistischen Modells

Wenn man diese Verantwortung des Menschen akzeptiert und sie mit der oben in Kapitel 2 Das Verhältnis zwischen Ethik und Technologie dargelegten Interaktion zwischen Technologie und Ethik verbindet, sollte man dennoch eine Unterscheidung zwischen Menschen und Maschinen aufgrund ihrer wesentlichen Unterschiede treffen, von denen einige eine unmittelbare ethische Relevanz haben, wie sie gerade in diesem Kapitel erörtert wurde. Gleichzeitig sollte die Dichotomie «Mensch – Maschine» mit dem oben in Kapitel 2 Das Verhältnis zwischen Ethik und Technologie vorgestellten interaktionistischen Modell dahingehend aufgelockert werden, dass der Mensch seiner ausschließlichen Verantwortung gerecht wird, indem er sich mit den notwendigen Aspekten auseinandersetzt: Der Mensch muss ethische Prinzipien, ethische und rechtliche Normen sowie einen Rahmen,

---

281 Vgl. Johnson 2006a; Yampolski 2013.

282 Tzafestas 2016: 2.

283 Vgl. Kirchschräger 2013b.

Ziele und Grenzen für Maschinen, aber vor allem für die Interaktion zwischen Mensch und Maschine – für die «technologische Begleitung»<sup>284</sup> – und für das «hybride sozio-technische System aus Menschen, Maschinen und Programmen»<sup>285</sup> definieren, implementieren und realisieren.

Dies könnte die Kritik hervorrufen, dass die Lockerung der Dichotomie «Mensch – Maschine» nicht weit genug geht. «Menschen und technische Artefakte sind in unserem Alltag so eng miteinander verbunden, dass sogar unsere moralischen Wahrnehmungen und Entscheidungen technologisch vermittelt sind. Nur wenn wir diese Verflechtung von Menschen und Technologien anerkennen, können wir Verantwortung für die Art und Weise übernehmen, in der sich Technologien auf die Gesellschaft und die menschliche Existenz auswirken – in den Praktiken des Technolgie-designs, der Implementierung und der Nutzung.»<sup>286</sup> Das Verständnis von Artefakten als «moralische Einflussfaktoren»<sup>287</sup> oder als «moralische Faktoren» hebt den positiven oder negativen Beitrag eines Artefakts zur Verwirklichung eines ethischen Grundsatzes in einer Handlung oder einem Ergebnis<sup>288</sup> hervor. Während über die Verflechtung von Menschen und Technologien völlige Einigkeit besteht, muss *zunächst* die Frage gestellt werden, ob die Berufung auf den «Beitrag» nicht zu dem Ergebnis führt, dass jedes materielle Objekt zu einer positiven oder negativen Handlung oder zu einem positiven oder negativen Ergebnis gegenüber einem ethischen Prinzip beitragen kann. So kann z.B. ein Stein die Handlung, jemanden zu schlagen, negativ beeinflussen, indem er die Gewalt dieser Handlung verstärkt, und er kann das Ergebnis des Schlagens negativ beeinflussen, da der Zustand des Opfers dieser Gewalttat schwer verletzt werden kann. Da also alles ein moralischer Faktor sein kann, kann man nicht sagen, dass dies nur für Artefakte gilt. Darüber hinaus stellt sich die Frage, ob die Bezeichnung «moralischer Faktor» überhaupt eine erkenntnistheoretische Bedeutung hat.

*Zweitens* kann die Lockerung der Dichotomie «Mensch – Maschine» nicht weiter gehen, weil die «Forderung nach der Bewahrung der Einheit des Subjektes gegen die drohende Gefahr seiner funktionalen Aufspaltung [besteht]. Es ist wichtig zu sehen, dass diese Forderung nicht selber

---

284 Verbeek 2011b: 43.

285 Rammert 2003: 15.

286 Verbeek 2014: 76.

287 Moor 2006: 18.

288 Vgl. Brey 2014.

als moralische missverstanden werden darf. Sie ist – und hinter dieses ‚Paradigma‘ der Kantischen Moralphilosophie kann nicht zurückgegangen werden – Bedingung der Möglichkeit von Ethik.»<sup>289</sup> Es bleibt ein grundlegender Unterschied zwischen Menschen und Technologien in Bezug auf ethische Subjektivität und moralische Handlungsfähigkeit, denn eine Wirkung oder einen Einfluss zu haben – was der Fall ist: Technologien können einen Einfluss auf das menschliche Leben haben, einschließlich der ethischen Dimension – ist nicht gleichbedeutend damit, ethische Subjektivität oder moralische Handlungsfähigkeit zu besitzen. So kann beispielsweise das Wetter einen Einfluss auf die Stimmung des Menschen haben, aber niemand würde behaupten, dass das Wetter moralische Handlungsfähigkeit hinsichtlich der Stimmung des Menschen hat. Wie oben in diesem Kapitel gezeigt wurde, gibt es einige Voraussetzungen für ethische Subjektivität und Moralfähigkeit, die den Technologien fehlen. Daher besitzen Technologien – obwohl sie eng mit dem Menschen verwoben sind – keine ethische Subjektivität oder Moralfähigkeit. Die Unterscheidung zwischen der «kausalen Wirksamkeit von Artefakten bei der Erzeugung von Ereignissen und Zuständen», dem «Handeln für oder im Namen einer anderen Entität» und der «moralischen Handlungsfähigkeit»<sup>290</sup> bekräftigt diese Position, zumal es begrifflich nicht notwendig ist, alle drei unter «Handlungsfähigkeit» zusammenzufassen, während sich die ersten beiden von der Handlungsfähigkeit unterscheiden. Betrachtet man schließlich die Verflechtung zwischen Menschen und Technologien selbst genauer, so führt eine noch intensivere Form der Verflechtung nicht zur Auflösung des Unterschieds zwischen den beiden. Wenn zum Beispiel zwei Menschen ineinander verliebt sind, und zwar so tief wie möglich, und daher die denkbar intensivste Verflechtung erleben, würde man dennoch nicht für die Auflösung der beiden Individuen in einen Menschen plädieren.

Eine Möglichkeit, dieses Problem zu überwinden und dennoch Technologien in ethische Subjektivität und moralisches Handeln hineinzufordern, besteht darin, das Verständnis von Moral zu ändern. «Moral ist weder in den Objekten selbst, noch in autonomen Subjekten zu finden. Sie kommt nur in den Beziehungen zwischen Subjekten und Objekten vor, wo die Objekte moralische Bedeutung haben und die Subjekte in vermittelte Beziehungen mit der Welt treten.»<sup>291</sup> Das würde z.B. bedeuten,

289 Mathwig 2000: 288.

290 Vgl. Johnson / Noorman 2014.

291 Verbeek 2014: 87.

das Wetter für die Stimmung der Menschen verantwortlich machen zu können, obwohl die Stimmung unabhängig vom Wetter sein kann, wie sie ist. Mit anderen Worten, und das ist ein erstes Gegenargument gegen diese Anpassung des Moralverständnisses, gibt es auch eine Moral ohne Technologien. Zweitens: Auch wenn Technologien als Objekte ethisch bedeutsam sein können, so gehört die Fähigkeit, das ethisch Legitime zu erkennen, sich entsprechend zu entscheiden und danach zu handeln, nicht ihnen, sondern dem Menschen als ethischem Subjekt und moralischem Agenten, und zwar aus dem oben erwähnten Mangel an Verletzbarkeit, Gewissen, Freiheit, Verantwortung und Autonomie. *Drittens* können Technologien als Objekte zwar «unabhängig von den Menschen, die sie geschaffen haben, moralisch relevante Handlungen ausführen und 'künstlich Gutes' und 'künstlich Böses' bewirken»<sup>292</sup>, doch ist dies kein Grund, das Konzept der moralischen Handlungsfähigkeit so weit auszudehnen, dass Technologien in dieses Konzept passen.<sup>293</sup> Technologien können Gutes und Böses bewirken, aber sie tun dies, ohne die ethische Qualität dieser Handlungen zu erkennen, zu kennen oder sich ihrer bewusst zu sein. Es stimmt, dass «ein Computer Emotionen darstellen kann, ohne selbst Emotionen zu haben, und dass Computersysteme in der Lage sein können, so zu tun, als ob sie die Bedeutung von Symbolen verstehen, ohne tatsächlich über das zu verfügen, was man als menschliches Verständnis bezeichnen würde.»<sup>294</sup> In beiden Fällen mangelt es der Darstellung und dem Funktionieren an Authentizität, die als wesentlich für Emotionen und Moralfähigkeit angesehen wird. Mit anderen Worten: Die Darstellung von Emotionen hängt für die Erfüllung ihrer Aufgabe davon ab, dass die Darstellung emotional ist, und das Funktionieren hängt für die Erfüllung seiner Aufgabe davon ab, dass es Moralfähigkeit als Grundlage für dieses Funktionieren besitzt. In beiden Fällen würden Emotionalität und Moralfähigkeit für Maschinen verneint.

*Viertens* besitzt nur der Mensch – wie oben ausgeführt – Verletzbarkeit, Gewissen, Freiheit, Verantwortung, Autonomie und Moralfähigkeit. Deshalb sollte man vermeiden, «alles Seiende – sich selbst eingeschlossen – aus dem Blickwinkel des technisch Machbaren zu sehen. Beispielsweise sind viele bereit, sich selbst als hochperfektierte Roboter zu sehen – oder aber, gemäss der anfänglich gewählten wissenschaftlichen Sehweise, als Ergebnis von biologisch vererbten Merkmalen. Eine solche Auffassung des

---

292 Floridi 2014: 187.

293 Vgl. Floridi 2014; zu dieser Frage vgl. auch Torrance 2008; Torrance 2011.

294 Wallach / Allen 2009: 9.

Selbst hat einen tiefen Einfluss auf das ‚Ich‘. Diejenigen, die sich selbst so sehen, tendieren dazu, das ‚Ich‘ auf eine Summe von ‚Wirkungen‘ zu reduzieren, deren ‚Ursachen‘ analysiert werden können, damit der Anspruch auf Freiheit und auf ihre Unbeugsamkeit in der Verantwortung an der Wurzel zersetzt werden kann.»<sup>295</sup>

*Fünftens* ist es ausschließlich Sache des Menschen, Technologien ethische Werte, Prinzipien und Normen einzuprogrammieren und/oder ihnen diese anzutrainieren, die sie zu respektieren haben – auch wenn die Technologien deren ethische Qualität nicht erkennen. Dies führt zu der Frage, welche ethischen Werte, Prinzipien und Normen den Technologien beigebracht werden sollten<sup>296</sup> – ein Thema, das weiter unten in Kapitel 6 Instrumente für die ethische Bewertung näher erläutert wird.

*Sechstens* umfasst die ausschließliche Fähigkeit des Menschen, das ethisch Legitime zu erkennen, sich entsprechend zu entscheiden und danach zu handeln, auch die Möglichkeit, sich zu entscheiden, Technologien gar nicht erst zu schaffen, zu entwerfen, zu produzieren, zu verbreiten oder als Objekte zu nutzen, oder die Möglichkeit, sich zu entscheiden, Technologien abzuschaffen oder zu zerstören.

Gleichzeitig könnte diese Überwindung der Dichotomie «Mensch – Maschine» durch das interaktionistische Modell die Kritik provozieren, dass die Aufweichung der Dichotomie «Mensch – Maschine» zu weit geht. Das interaktionistische Modell könnte in der Weise wahrgenommen werden, dass es den Vorschlag aufgreift, den Ausgangspunkt des Subjekt-Objekt-Dualismus zu beseitigen, der dann durch die Beendigung des Maschinen Denkens überwunden werden soll, indem «Nicht-Maschinen» und «Post-Maschinen» sowie entsprechende «neue, nicht-moderne Formen der Subjektivität» postuliert werden.<sup>297</sup> Dies ist nicht der Fall, denn das interaktionistische Modell unterscheidet zwischen Menschen als Subjekten und Maschinen als Objekten und betont die wesentliche gegenseitige Beeinflussung. «Menschen und andere hoch entwickelte Lebewesen sind keine Maschinen. Die Natur als Ganzes ist keine Maschine. Viele Mythen seit der Antike bis ins heutige Hollywood kreisen um dieses Verhältnis Mensch/Maschine. In manchen Mythen ist der Mensch nur eine Maschine, in anderen ist die Natur als Ganzes eine Maschine, in anderen unterjochen Maschinen Menschen, und manche Utopisten glauben, dass das endgültige

295 Hersch 1992: 60.

296 Vgl. Bhargava/Kim 2017.

297 Vgl. Coeckelbergh 2017.

Reich der Freiheit darin bestehen wird, ausschliesslich Maschinen arbeiten zu lassen. Ein digitaler Humanismus transformiert den Menschen nicht in eine Maschine und interpretiert Maschinen nicht als Menschen. Er hält an der Besonderheit des Menschen und seiner Fähigkeiten fest und bedient sich der digitalen Technologien, um diese zu erweitern, nicht um diese zu beschränken.»<sup>298</sup>

Diese Unterscheidung von Mensch und Technologie aus ethischer Sicht liefert die Grundlage dafür, dass es aus ethischer Sicht ausgeschlossen ist, Maschinen und künstlicher Intelligenz Rechte und Rechtspersönlichkeit zuzusprechen. «Einem Computer Menschenrechte zuzugestehen, würde die Menschenwürde herabsetzen. Als Saudi-Arabien beispielsweise einem Roboter namens Sophia die Staatsbürgerschaft verlieh, erhoben menschliche Frauen, darunter auch feministische Wissenschaftler:innen, Einspruch und stellten fest, dass der Roboter mehr Rechte erhielt, als viele saudische Frauen haben.»<sup>299</sup>

Andere Risiken unterstreichen, warum Maschinen und künstliche Intelligenz keine Träger von Rechten sein sollten. «Zwei weitere Aspekte von Unternehmen machen die Menschen noch anfälliger für KI-Systeme mit menschlichen Rechten: Sie sterben nicht, und sie können politischen Kandidaten und Gruppen unbegrenzte Geldbeträge zukommen lassen. Künstliche Intelligenzen könnten Geld verdienen, indem sie Arbeitnehmer ausbeuten, Algorithmen zur Preisgestaltung von Waren und zur Verwaltung von Investitionen einsetzen und neue Wege zur Automatisierung wichtiger Geschäftsprozesse finden. Über lange Zeiträume hinweg könnte sich dies zu enormen Gewinnen summieren – die niemals unter den Nachkommen aufgeteilt würden. Dieser Reichtum könnte leicht in politische Macht umgewandelt werden. Politiker:innen, die von algorithmischen Einheiten finanziell unterstützt werden, wären in der Lage, gesetzgebende Gremien anzugreifen, Präsidenten anzuklagen und die Ernennung von Aushängeschildern am Obersten Gerichtshof zu unterstützen. Diese menschlichen Galionsfiguren könnten dazu benutzt werden, die Rechte von Unternehmen zu erweitern oder sogar neue Rechte speziell für Systeme der künstlichen Intelligenz einzuführen – was die Bedrohung der Menschheit noch weiter ausdehnen würde.»<sup>300</sup>

---

298 Nida-Rümelin 2018: 10-11; Zum Humanismus vgl. Mayer-Tasch 2006; Holderegger et al. 2011.

299 Yampolskiy 2018.

300 Yampolskiy 2018.

Angesichts der ethischen Chancen und Risiken der digitalen Transformation und der künstlichen Intelligenz ist aus ethischer Perspektive begriffliche Klarheit über den Bereich möglicher Verantwortungssubjekte notwendig. Innerhalb des interaktionistischen Modells macht diese Unterscheidung von Mensch und Technologie aus ethischer Perspektive einen wesentlichen Unterschied, um der ausschließlichen Verantwortung des Menschen für die Technologie gerecht zu werden, denn sie macht glasklar, wer das Sagen hat und wer das Subjekt der Verantwortung für die Technologie ist: der Mensch.

