

## 11. Künstliche Intelligenz und Handeln I

### Ist die Entscheidungsfreiheit des Menschen bedroht?

---

Menschen haben von jeher versucht, ihre Aktivitäten durch geeignete Hilfsmittel zu erleichtern oder zu unterstützen. Die Entwicklung erstreckt sich von ersten Werkzeugen aus Stein und vielfältigen Hilfsmitteln für die Handarbeit und für einfache körperliche Tätigkeiten über Kraftmaschinen, Fortbewegungsmittel und herkömmliche Informations- und Kommunikationstechnologien bis zu den heutigen digitalen Möglichkeiten. So entstanden nicht nur bedeutsame Erweiterungen für die manuelle Tätigkeit und die mechanische Arbeit, sondern auch neue technische Möglichkeiten für die bis dahin nur dem Menschen vorbehaltenen »geistige Arbeit«. Mit den Entwicklungen im Bereich der Mikroprozessortechnologie in den 1980er-Jahren konnten computerbasierte technische Systeme bereits schematisierbare »geistige Arbeit« sowie die Steuerung und Regelung von technischen Prozessen mit relativ hoher Flexibilität durchführen.<sup>1</sup> Das Verhältnis zwischen Mensch und Maschine war dabei durch klare »Schnittstellen« gekennzeichnet, bei denen der Mensch eingreifen bzw. bestimmen konnte, was die Maschine oder das Informatiksystem im Einzelnen zu tun hatte. Für menschliches Handeln bedeutete dies, dass der Mensch sich der informatischen Möglichkeiten gemäß seinen Intentionen bedienen konnte. Insofern war auch das Verhältnis zwischen Mensch und Maschine klar definiert: Der Mensch entwirft Algorithmen und setzt sie in einem Prozess der Programmierung in maschinell lesbare Programme um,

welche dann von der Maschine ausgeführt werden. Etwas vereinfacht und kürzer: Mensch sagt Maschine, was zu tun ist.

Durch die rasanten Weiterentwicklungen im Bereich der künstlichen Intelligenz (im Folgenden kurz: KI), z.B. beim maschinellen Lernen und bei der Robotik, ergeben sich in zunehmendem Maße Entgrenzungen zwischen Mensch und Maschine. Beispiele sind das Erkennen und die Analyse sowie die Produktion von Sprache oder von Bildern und Musik, die Ausführung hochkomplexer kognitiver Prozesse und das Treffen moralisch relevanter Entscheidungen unter Einbezug riesiger Datenmengen. Dabei überschreitet die Leistungsfähigkeit von KI-Technologien in manchen Bereichen die Leistungsmöglichkeit des Menschen. So geben Informatiksysteme heute – unter Zugrundelegung umfangreicherer Datenmengen als sie ein Mensch verarbeiten kann – Empfehlungen zu der Frage, wer einen Kredit bekommen und wem man ihn besser verweigern sollte; sie diagnostizieren zum Teil besser als ein Arzt, ob und welche Veränderungen an Zellen vorliegen; sie zeigen auf, ob und wie man sein Ess- oder Bewegungsverhalten zum Erhalt der Gesundheit verändern kann. Für solche und viele weitere Fälle scheint sich die Situation von »Mensch sagt Maschine, was zu tun ist« in Richtung »Maschine sagt Mensch, was zu tun ist« zu verändern.<sup>2</sup> Hinzu kommt, dass Informatiksysteme in manchen Bereichen bereits mehr oder weniger »autonom« agieren, z.B. bei der Regelung der Heizungen in einem Haushalt mit einer Optimierung des Energieverbrauchs oder beim sogenannten autonomen Fahren.

Mit solchen Entwicklungen ergeben sich Problemstellungen, die sich zwar schon im Zusammenhang mit der Digitalisierung und Mediatisierung angedeutet haben, nun allerdings mit den Fragen weitergeführt werden sollen, wie sich Entscheidungsfreiheit unter den Bedingungen künstlicher Intelligenz darstellt und ob – die mit dem Menschenbild verbundene – Leitidee eines reflexiv eingestellten und gesellschaftlich handlungsfähigen Subjekts für die Zukunft einer Revision unterzogen werden muss und welche Konsequenzen sich aus den Überlegungen für menschliches Handeln ergeben. Dazu wird zunächst das Feld künstlicher Intelligenz näher in den Blick genommen.

## 11.1 Künstliche Intelligenz (KI) als Forschungs- zweig und Technologie

Die erste Verwendung des KI-Begriffs wird dem amerikanischen Informatiker John McCarthy (1927– 2011) zugeschrieben, der ihn 1956 in einem Antrag für eine wissenschaftliche Konferenz benutzte, bei der unter anderem Programme für das Dame- und das Schachspiel vorgestellt wurden.<sup>3</sup> Seitdem wird der Begriff in unterschiedlichen Zusammenhängen gebraucht. Dabei kommen ihm hauptsächlich zwei Bedeutungen zu. Zum einen steht er für das Teilgebiet der Informatik, in dem es um Forschungen und Entwicklungen geht, die auf eine Ausführung von Aufgaben durch einen Computer bzw. eine Maschine zielen, für die beim Menschen Intelligenz vorausgesetzt wird;<sup>4</sup> zum anderen gilt KI als Oberbegriff für alle Technologien, mit denen menschliche Denk- und Handlungsvollzüge nachgebildet werden.<sup>5</sup> Hierbei signalisiert die Formulierung »menschliche Denk- und Handlungsvollzüge«, dass KI nicht auf einen fest umrissenen Intelligenzbegriff – etwa im Sinne psychologischer Konzepte – festgelegt ist. Vielmehr können in die KI-Forschung alle Fähigkeiten einbezogen werden, bei denen aufseiten des Menschen psychische oder psychomotorische Prozesse vorliegen.

Im Zusammenhang konkreter Software- und Hardwarelösungen spricht man auch von KI-gestützten Systemen oder kurz: KI-Systemen.<sup>6</sup> Häufig sind KI-Komponenten in andere Zusammenhänge eingebettet. Dabei können sie in einer virtuellen Umgebung zur Geltung kommen, z.B. als Sprachassistent, oder in größeren technischen Systemen, z.B. in Autos. Um einen Eindruck von Leistungen der KI zu vermitteln, werden im Folgenden verschiedene Anwendungsbereiche mit Beispielen angesprochen.

## 11.2 Anwendungsbereiche von KI-Technologien<sup>7</sup>

Bereits in den 1950er-Jahren gelang es dem Elektroingenieur Arthur Samuel (1901–1990), ein Programm zum Brettspiel »Dame« zu schreiben. So konnte der Computer gegen einen menschlichen Spieler antreten,

wobei er zunächst verlor. Mit der Zeit wurde das Programm so weiterentwickelt, dass der Computer auch gegen sich selbst spielen konnte, sich hierbei immer mehr verbesserte und schließlich ein so guter »Dame-Spieler« war, dass ein Mensch keine Chance mehr gegen ihn hatte. Ähnliche Entwicklungen ergaben sich beim Schachspiel: 1997 besiegte ein Rechner den damaligen Weltmeister Garri K. Kasparow (geb. 1963). Für das noch komplexere Spiel Go wurde 2017 ein Programm entworfen, das innerhalb von drei Tagen die Spielstärke eines Profis erreichte und dabei eine Programmversion übertraf, die ein Jahr zuvor den Ersten der Weltrangliste, Lee Sedol (geb. 1983), besiegt hatte. Hierbei war es dem Programm – wegen der Fülle an Spielvarianten – nicht mehr möglich, alle denkbaren Spielzüge durchzurechnen, es musste vielmehr auf der Grundlage von »Spielerfahrungen« Entscheidungen treffen, die man in Analogie zu menschlichem Verhalten als »intuitiv« bezeichnen kann. Weitere KI-Anwendungen im Bereich des *Spielens* bestehen unter anderem darin, dass in Computerspielen computergesteuerte »Mitspielerinnen« oder »Mitspieler« analog zu menschlichem Verhalten agieren.

In ähnlicher Weise beeindruckend sind die Leistungen von KI-Technologien, wenn es um den Rückgriff auf *vorhandenes Wissen* geht. So trat beispielsweise das wissensbasierte System Watson 2011 in der Quizshow »Jeopardy!« gegen frühere Champions an und entschied die Duelle für sich (wobei es bei »Jeopardy!« in »Umkehrung« der herkömmlichen Quizshows darum geht, zu vorgegebenen Antworten die richtigen Fragen zu finden). Überhaupt liefern Suchmaschinen immer bessere Ergebnisse, z.B. bei Anfragen zu bestimmten Ereignissen, Sachverhalten oder Vorgehensmöglichkeiten in verschiedenen Wissensgebieten – von der Geschichte und Meteorologie bis zum Finanz- und Gesundheitswesen. Darüber hinaus ist es möglich, vorhandenes Wissen nicht nur wiederzugeben, sondern auch maschinell *auszuwerten*, z.B. in der Form von Zusammenfassungen oder als Bündelung unter spezifischen Fragestellungen. So mag ein KI-System auf der Basis ihm zur Verfügung stehender Wissens- und Datenbestände z.B. die Frage nach den vier häufigsten Argumenten für oder gegen den Klimaschutz beantworten. Insgesamt lassen sich bei Vorliegen entsprechender Datenbestände immer komplexere Fragen stellen – beispielsweise: Wel-

ches ist die günstigste Zeit, um an einem gewünschten Ort bei guten Wetterbedingungen zu möglichst niedrigen Preisen und optimalen Verkehrsmöglichkeiten Urlaub zu machen?

Für Auskünfte solcher und anderer Art spielen unter anderem *Big Data-Analysen* eine wichtige Rolle. Diese sind von besonderer Bedeutung, wenn es nicht nur um die Aufbereitung vorhandenen Wissens, sondern auch um die Generierung *neuer Informationen* durch Rückgriff auf unterschiedliche Datenbestände und deren Verknüpfung geht.<sup>8</sup> Hierbei kann es sich sowohl um unmittelbar praktisch relevante Informationen als auch um grundlegende wissenschaftliche Erkenntnisse handeln. So ist es z.B. für den Handel wichtig, noch vor dem Erscheinen eines neuen Produktes zu wissen, in welchem Umfang es in welchen Regionen vermutlich nachgefragt werden wird. Oder: Für die Verteilung der Polizeipräsenz in einer Stadt ist es bedeutsam, die Wahrscheinlichkeit zu kennen, mit der man in einzelnen Stadtvierteln zu welcher Zeit mit bestimmten Straftaten rechnen muss. Besonders relevant ist es darüber hinaus, die Folgen bestimmter medizinischer Therapien und ihre Erfolgswahrscheinlichkeit bei unterschiedlichen gesundheitlichen Bedingungen zu bestimmen. Des Weiteren: Wenn in der Kommunikationswissenschaft mit Blick auf Genderfragen herausgefunden werden soll, wodurch sich weibliches und männliches Kommunikationsverhalten in sozialen Netzen – z.B. hinsichtlich Wahrhaftigkeit, Täuschung, Mobbing, Aggression – unterscheiden, können entsprechende Datenströme einer Analyse unterzogen werden. Hinsichtlich des Erkenntniswertes ist allerdings zu beachten, dass Big Data-Analysen letztlich (nur) korrelative Zusammenhänge und keine kausalen Beziehungen ausweisen.<sup>9</sup>

Für wissens- und datenbasierte Systeme hat zudem die Tatsache an Bedeutung gewonnen, dass die *Kommunikation* nicht nur schriftlich, sondern auch mündlich sowie in unterschiedlichen Sprachen erfolgen kann. Dies hängt mit Fortschritten im Bereich der automatischen Spracherkennung und Sprachproduktion zusammen, welche auch dazu führen, dass die Kommunikation nicht nur nach dem Motto »Mensch fragt, Computer antwortet« ablaufen kann, sondern auch als mündlicher und schriftlicher Dialog zwischen Mensch und Maschi-

ne. Dabei haben sich die entsprechenden Dialogsysteme – bei allen Grenzen die ihnen im Vergleich zu direkter personaler Kommunikation (noch) bleiben – gegenüber früheren Entwürfen deutlich verbessert, z.B. gegenüber der Simulation eines psychotherapeutischen »Dialogs« durch das 1966 von Joseph Weizenbaum (1923–2008) entwickelte Programm »Eliza«. Des Weiteren ist es mittlerweile möglich, während eines Gesprächs zwischen zwei Personen unterschiedlicher Muttersprache mittels Smartphone sofort eine automatische Übersetzung in der jeweils anderen Sprache zu liefern.

Spracheingabe und Sprachausgabe vereinfachen zugleich die Verwendung sogenannter *virtueller Assistenten*. So lässt sich ein virtueller Assistent beispielsweise in mündlicher Form beauftragen, präzise Antworten auf komplexe Fragen zu geben, Kontakte für Terminabstimmungen mit mehreren Personen aufzunehmen, im Bedarfsfall den Kauf von Druckerpatronen oder Nahrungsmitteln einzuleiten, auf Zahlungsfristen zu achten, im Kontext von *smart home*-Entwicklungen anzuzeigen, dass im Kühlschrank die Vorräte ausgehen, für eine wettergerechte Beschattung durch vorhandene Gardinen oder Rollläden zu sorgen oder eine angemessene Bewässerung verschiedener Pflanzen im Garten zu gewährleisten. Demgemäß können virtuelle Assistenten auch Aufgaben an Roboter übertragen, z.B. an einen Mäh- oder Poolroboter.

Für Aufgaben dieser und weiterer Art ergeben sich durch die Verbindung von KI mit der Sensorik und Aktorik bzw. mit der *Robotik* vielfältige Möglichkeiten – von der Entwicklung robotischer Mobilitätshilfen bzw. neurotechnologischer Hilfsmittel zum Ausgleich körperlicher Defizite bis zum Bau tierähnlicher oder humanoider Roboter, über deren Einsatz beispielsweise in der Pflege diskutiert wird. So muss KI in Verbindung mit der Robotik gegebenenfalls auch in hochkomplexen Umgebungen »agieren«, z.B. beim autonomen Fahren im Straßenverkehr. Robotische Prozesse spielen aber nicht nur eine Rolle, wenn KI-Technologien in augenscheinlichen technischen Geräten eingesetzt werden, sondern auch, wenn sie im Rahmen anderer Systeme verwendet werden – etwa als *Social Bots* oder *Chatbots* in sozialen Netzwerken. Solche Bots kommen unter anderem zum Einsatz, um Nutzende in ihrem Kauf- oder Wahlverhalten zu beeinflussen (siehe dazu auch Kapitel 3). Sie kön-

nen des Weiteren als juristische Unterstützung agieren – wie z.B. der Legal-Bot »DoNotPay«, der hilft, bei einem Bußgeldbescheid einen auf den Einzelfall bezogenen und juristischen einwandfreien Widerspruch einzulegen, nachdem er vorher in einem Chat alle relevanten Informationen erfragt hat.

Spätestens mit der Diskussion um *autonomes* Fahren wird zudem offensichtlich, dass in KI-gesteuerten Systemen unter anderem auch *moralisch relevante* Entscheidungen getroffen werden müssen. Dabei ist allerdings zu beachten, dass hier ein *technischer* Autonomiebegriff zugrunde liegt. In der Technik verwendet man das Adjektiv »autonom« in der Regel dann, wenn ein technisches System auf der Grundlage von KI-Technologie Erfahrungen verarbeiten bzw. lernen kann und Entscheidungen zwischen zwei oder mehr Möglichkeiten selbstständig, d.h. ohne menschliches Eingreifen, trifft. Im Hinblick auf *moralisch relevanten* Entscheidungen könnte ein autonomes Fahrzeug z.B. in die Situation kommen, bei einem nicht zu vermeidenden Unfall entweder vier alte Menschen in einer Gruppe oder eine Mutter mit Kind in Lebensgefahr zu bringen. In diesem Zusammenhang hat beispielsweise die Ethik-Kommission des Bundesministeriums für Verkehr und digitale Infrastruktur betont, dass es unzulässig und mit dem Grundgesetz nicht vereinbar wäre, potenzielle Verkehrsoffer nach persönlichen Merkmalen, z.B. nach Alter oder Geschlecht, zu bemessen. Moralisch brisant ist auch die Frage, ob Kampfroboter zur Terrorbekämpfung oder gar zur Kriegsführung gebaut und eingesetzt werden sollten. Entsprechende Diskussionen werden zum Teil von futuristischen Visionen einer neurotechnologischen Optimierung menschlicher Fähigkeiten oder der Entwicklung einer »Superintelligenz« überlagert. Dabei kann die Perspektive entweder auf einen – in seinen Fähigkeiten vielfältig erweiterten – Menschen gerichtet sein oder auf ein neues Gebilde, bei dem Gehirn und Geist des Menschen losgelöst von seinem Körper in eine superintelligente »Maschine« übergegangen sind. Die erste Denkrichtung kann als kennzeichnend für den sogenannten Transhumanismus gelten und die zweite als charakteristisch für den Posthumanismus.<sup>10</sup> Visionen solcher Art werden z.B. als *Science-Fiction* in Filmen oder Romanen in utopischer oder dystopischer Weise ausgestaltet.<sup>11</sup>

### 11.3 Funktionen von KI-Systemen und grundlegende Ansätze

Um die obigen Anwendungen zu ermöglichen oder zu unterstützen, führen KI-Systemen einzelne oder mehrere der folgenden Funktionen aus: (a) Wahrnehmung der Umwelt über Sensoren, (b) Umwandlung der aus der Umwelt aufgenommenen Signale in Daten und deren Interpretation, (c) Ziehen von Schlüssen und Treffen von Entscheidungen, (d) Ausführung gewählter Vorgehensweisen sowie (e) Kontrolle von Ausführungen und Optimierung. Bei entsprechenden Prozessen erfolgt gegebenenfalls eine Abstimmung mit menschlichen Tätigkeiten. Beispielsweise könnte bei der Flugsteuerung einer Drohne die Wahrnehmung eines Hindernisses an einen menschlichen Operateur gemeldet werden, der dann das notwendige Ausweichmanöver in Gang setzt. Allgemein gilt, dass die technische Autonomie eines KI-Systems umso größer ist, je mehr Funktionen von ihm wahrgenommen werden.<sup>12</sup>

Für die Realisierung der Funktionen von KI-Systemen lassen sich zwei grundlegende Ansätze unterscheiden:<sup>13</sup>

- *Symbolverarbeitende KI*: Bei dieser geht man davon aus, dass menschliches Denken wesentlich auf logisch-begrifflichen Zusammenhängen beruht und entsprechend rekonstruiert bzw. in funktionaler Hinsicht »nachgebildet« werden kann. Demgemäß ist das Wissen bei der symbolverarbeitenden KI in der Form logisch-begrifflicher Zusammenhänge repräsentiert. Ein logisch-begrifflicher Zusammenhang liegt z.B. bei folgender Aussage vor: Wenn gilt, dass die Krankheit X mit den Symptome A und B und C verbunden ist und bei einem Menschen die Symptome A, B und C zu beobachten sind, dann verweist dies darauf, dass dieser Mensch die Krankheit X hat. In dieser Formulierung stehen die Symbole A, B, C und X für bestimmte Begriffe: A, B und C beispielsweise für Grippe-symptome und X für Grippe. Dabei ist die begriffliche Beschreibung eines Symptoms, z.B. hohes Fieber, selbst wieder ein (sprachliches) Symbol für den entsprechenden Sachverhalt. Auf dieser Grundlage kann mithilfe eines entsprechenden Algorithmus eine Krankheit diagnostiziert werden.

Falls in einem Expertensystem weitere Wenn-Dann-Beziehungen gespeichert sind, z. B. »Wenn die Krankheit X vorliegt ist, dann ist unter der Bedingung D das Medikament Y geeignet«, kann von dem System außerdem eine Empfehlung für die Therapie abgegeben werden. Damit ist die symbolverarbeitende KI an das Vorhandensein begrifflichen Wissens gebunden, das ein Programmierer als Grundlage entsprechender Algorithmen für maschinelle Diagnosen und Entscheidungen nutzen kann.

- *Neuronale KI*: Diese ist nicht mehr an vorheriges begriffliches Wissen gebunden, sondern entwickelt mithilfe künstlicher Neuronen und deren Verknüpfung von sich aus Antworten auf bestimmte Fragen. So besteht beispielsweise die Möglichkeit, unterschiedliche Sachverhalte im Sinne der Mustererkennung zu erfassen. Dabei wird versucht, das Funktionsprinzip des Gehirns mit seinen Neuronen als kleinsten Funktionseinheiten in künstlichen neuronalen Netzwerken nachzubilden. Technisch gesehen bestehen künstliche neuronale Netze aus einer Eingabeschicht, einer oder mehreren Zwischenschichten und einer Ausgabeschicht. Die Eingabeschicht nimmt Daten auf, die ein Sensor liefert, z. B. Daten von schriftlichen Texten, Bildern, Tönen oder Bewegungen. Von dort wird der Input je nach der Stärke der von ihm ausgehenden Signale an Zwischenschichten weitergeleitet, ehe er in der Ausgabeschicht als Output erscheint. Wenn beispielsweise auf einem Bild ein Reh als Reh »erkannt« werden soll, wird das KI-System solange trainiert, bis es die von einer Kamera aufgenommenen optischen Werte mit hoher Wahrscheinlichkeit richtig benennt. Da in den verschiedenen Schichten unzählige Neuronen wirksam werden, lassen sich KI-Systeme auch so trainieren, dass sie hochkomplexe Sachverhalte »erkennen«, z. B. Gemälde, Musikstücke, Bewegungsabläufe oder menschliche Gesichter. Auf diese Weise können auch Sachverhalte erfasst werden, deren Merkmale aufgrund ihrer Komplexität nicht in umfassender sprachlicher Fassung vorliegen. Die Ergebnisse entsprechender Lernalgorithmen lassen sich dann als Basis nutzen, um z. B. ein Bild zu malen als wäre es von van Gogh, ein Musikstück zu komponieren als wäre es von Bach, eine Bewegung auszuführen

als wäre sie von einem menschlichen Akteur oder ein Lächeln auf dem »Gesicht« eines humanoiden Roboters zu erzeugen.

Die Ansätze einer symbolverarbeitenden und einer neuronalen KI können in KI-Systemen auch miteinander verbunden werden. Beispielsweise lässt sich bei einer Drohne die neuronale Technologie nutzen, um einen Baum auf dem Flugweg zu erkennen. Der Baum kann dann im Sinne der Symbolverarbeitung als Hindernis eingeordnet werden, sodass sich über einen symbolverarbeitenden Algorithmus eine Flugwegänderung in Gang setzen lässt.<sup>14</sup>

Angesichts der Entwicklungen im Bereich künstlicher neuronaler Netze verwundert es nicht, dass maschinelles Lernen im Rahmen der KI-Forschung immer bedeutsamer geworden ist. Dabei spielt das sogenannte tiefe Lernen (*deep learning*) eine zunehmend wichtige Rolle. In diesem Zusammenhang unterscheidet man auch zwischen einem überwachten Lernen (*supervised learning*) und einem unüberwachten Lernen (*unsupervised learning*). Beim überwachten Lernen sind die angestrebten Ergebnisse bekannt, z.B. als Klassifikationen oder Bewegungsabläufe, sodass sie als sogenannte Labels für die zu generierenden Prozesse mit eingegeben und »Lernprozesse« solange fortgesetzt werden, bis zufriedenstellende Ergebnisse erreicht sind. Bei unüberwachtem Lernen werden dagegen nur Beispieldaten eingegeben, für die ein KI-System Muster und Strukturen sucht. Wenn z.B. festgestellt werden soll, welche Themen in einem offenen politischen Chat angesprochen werden, kann das unüberwachte Lernen helfen, solche Themen herauszufinden. Unter Umständen lassen sich die Vorgehensweise überwachten und unüberwachten Lernens auch miteinander verbinden. So kann z.B. ein KI-System zur Analyse von Themen in einem Chat zunächst darauf trainiert werden, Beiträge zu erwarteten Themen zu identifizieren. Wenn Zuordnungen zu erwarteten Themen nicht möglich sind, können durch unüberwachtes Lernen weitere Themen entdeckt werden.

## 11.4 Problemlagen und Forschungsfelder, Chancen und Risiken

Derzeitige Probleme liegen beim maschinellen Lernen mittels künstlicher neuronaler Netze unter anderem darin, dass in der Regel eine sehr große Anzahl von Beispielen benötigt wird, um zu zufriedenstellenden Ergebnissen zu kommen. Während beispielsweise Kinder aufgrund weniger Erfahrungen mit Hunden lernen, solche in unterschiedlichen Kontexten zu identifizieren, brauchen Deep Learning-Systeme eine unvergleichlich höhere Anzahl von Beispielen. Zudem hängt die Qualität der Lernergebnisse sehr stark von der Qualität der jeweiligen Beispiele ab. So musste beispielsweise der Chatbot Tay, der 2016 lernen sollte, in sozialen Netzen zu kommunizieren, bereits nach einem Tag zurückgerufen werden, weil er so viele rassistische Beiträge zum »Lernen« genutzt hatte, dass er in kurzer Zeit selbst rassistisch (re-)agierte. Allgemein gilt, dass ein Lernalgorithmus mit einer zunehmenden Anzahl von (guten) Trainingsbeispielen seine Modellbildung verbessern und die Fehlerhäufigkeit verringern kann.

Insgesamt stellen sich für maschinelles Lernen verschiedene *Forschungsaufgaben*. Diese verweisen zugleich auf *Begrenzungen, Probleme* und *Desiderata* jetziger Möglichkeiten. Das gilt gemäß Überlegungen der Fraunhofer-Gesellschaft z. B. für folgende Aspekte:<sup>15</sup>

- *Nutzung großer Datenmengen*, insbesondere wenn kontinuierlich neue Daten in extrem großem Umfang anfallen, die möglichst in Echtzeit ausgewertet werden sollen, wie es z. B. im Finanzbereich der Fall sein kann,
- *Lernen mit geringen Datenmengen*, um auch mit wenigen verfügbaren Daten zu zufriedenstellenden Modellbildungen und Ergebnissen zu kommen, z. B. wenn in der Medizin bei seltenen Krankheiten nur wenige Daten bzw. Beispiele vorliegen.
- *Anpassungsfähigkeit und Flexibilität*, damit KI-erzeugte Modellbildungen auch auf neue Situationen und Kontexte übertragen werden können und sich dabei noch zufriedenstellende Ergebnisse einstellen,

- *Lernen mit zusätzlichem Wissen*, sodass beim maschinellen Lernen Wissen, das bei Fachleuten bereits vorhanden ist, mit verwertet werden kann,
- *flexible Kollaboration zwischen Mensch und Maschine*, wodurch zum einen die jeweilige Maschine bzw. der Lernalgorithmus aus der Interaktion mit dem Menschen lernen und zum anderen der Mensch von den maschinellen Leistungen profitieren kann,
- *Transparenz und Erklärbarkeit*, wobei Transparenz in Sinne einer vollständigen Nachvollziehbarkeit der Prozesse in künstlichen neuronalen Netzen wegen der hohen Komplexität bei vielen Anwendungen kaum zu erreichen sein dürfte, sodass man sich gegebenenfalls mit Hinweisen auf wesentliche Einflussfaktoren für Einzelentscheidungen zufriedengeben muss,
- *Fairness und Vermeidung von Diskriminierung*, womit verhindert werden soll, dass ethisch inakzeptable Ergebnisse entstehen, wenn z.B. bei einer automatischen Vorauswahl von Bewerbungen Geschlecht, Herkunft, Religion oder Alter in unzulässiger Form in die Entscheidungsfindung eingehen,
- *Sicherheit und Robustheit*, die garantieren sollen, dass einerseits keine Fehler passieren, z.B. bei der automatischen Erkennung von Verkehrsschildern beim autonomen Fahren, zugleich aber unwesentliche kleine Änderungen die Funktionsfähigkeit nicht beeinträchtigen.

Bei allen Detailaufgaben stellt die Vereinbarkeit von KI-Forschung und KI-Technologie mit unserem Menschen- und Gesellschaftsbild sowie mit unseren Rechts- und Wertvorstellungen eine unhintergehbare Anforderung dar.

Vor diesem Hintergrund lassen sich als allgemeine Chancen für KI-Anwendungen z.B. nennen:<sup>16</sup> Entlastungen bei beruflichen Arbeiten und mehr Freizeit für den Menschen, Anpassung von Angeboten an Voraussetzungen und Interessen von Nutzenden, höhere Produktivität und geringere Fehlerquoten, neue berufliche und wirtschaftliche Möglichkeiten, mehr Komfort im Alltag, weitergehende Fortschritte in vielen Bereichen der Wissenschaft und Technik.

Solchen Chancen stehen verschiedene Risiken gegenüber, z.B.:<sup>17</sup> zunehmende Abhängigkeit des Menschen von der Technik, unter Umständen bis zur Dominanz der Technik gegenüber dem Menschen, erweiterte Gefahren bezüglich Datenschutz und Machtmissbrauch bzw. Machtverteilung, gegebenenfalls gepaart mit unkontrolliertem Datenkapitalismus, Verlust von Arbeitsplätzen, Verringerung von sozialer Nähe durch Robotereinsatz, Verstärkung selektiver Wahrnehmung durch »Filterblasen«, manipulative Beeinflussung der Meinungsbildung, Diskriminierung durch kaum durchschaubare Algorithmen, Gefahren durch KI-gestützte Waffensysteme.

## 11.5 Fragen angesichts von KI-Entwicklungen

Nach den Überlegungen im vorherigen Kapitel kann und sollte für menschliches Handeln ein Menschenbild zugrunde liegen, das auf der Leitidee eines reflexiv eingestellten und gesellschaftlich handlungsfähigen Subjekts beruht. Bei einer solchen Leitidee wird – auch unter Berücksichtigung gegebenenfalls hinderlicher Bedingungen – unterstellt, dass der Mensch über Entscheidungsfreiheit verfügt sowie selbstverantwortlich und mindestens teilautonom zu handeln vermag. Dieser Gedanke beruht letztlich auf der Vorstellung, dass der Mensch einen »Sonderentwurf« darstellt<sup>18</sup> bzw. dass er sich im Laufe der Evolution zu einem Wesen entwickelt hat, das durch eine einzigartige Verbindung von Körper und Geist oder Leib und Seele gekennzeichnet ist und über Bewusstsein und Freiheit verfügt.<sup>19</sup>

Im neunten und zehnten Kapitel wurden bereits einzelne mögliche Kritikpunkte und Aspekte diskutiert, die mit der Entscheidungsfreiheit und dem hier vertretenen Subjektverständnis zusammenhängen. Auch wenn man solche Fragen und damit verbundene Probleme im Bewusstsein hat, erweist es sich mit Blick auf vernunftbezogene Argumentationen als sinnvoll, das oben noch einmal zusammengefasste Menschenbild im Sinne einer offenen Leitidee zu postulieren

Vor diesem Hintergrund stellt sich mit den KI-Entwicklungen zunächst die Frage, ob die subjektiv empfundene Entscheidungsfreiheit

nicht erneut zur Debatte steht. Erweist sie sich vielleicht doch als bloße Illusion, weil sich auch KI-Systeme so verhalten können, als ob sie Entscheidungsfreiheit hätten, obwohl sie letztlich nur auf physikalisch beschreibbaren Prozessen beruhen? Muss im Umkehrschluss nicht doch angenommen werden, dass sich auch das Phänomen des Freiheitsempfindens letztlich vollständig durch neurobiologische Vorgänge erklären lässt?

Grundsätzlich behält auch angesichts der KI-Entwicklungen das, was dazu im neunten Kapitel gesagt wurde, seine Gültigkeit: Die schon unabhängig von KI-Entwicklungen aufgrund von hirneurophysiologischen Untersuchungen gezogene Schlussfolgerung, dass Entscheidungsfreiheit eine Täuschung des Bewusstseins sei, erweist sich als unangemessen, weil selbst bei einem weitgehenden Nachweis neurophysiologischer Korrelate für Bewusstseinsprozesse noch offenbliebe, wo letztlich die Steuerung und Kontrolle dafür läge. Mit Blick auf die KI-Entwicklungen wäre zudem ein Rückschluss von KI-Prozessen auf menschliche Bewusstseinsprozesse deshalb irreführend, weil es sich bei den ersteren um algorithmische Vorgänge mit physikalisch darstellbaren Daten in technischen Artefakten handelt, während es bei letzteren um Prozesse geht, die sich im Laufe von Millionen von Jahren des Lebens auf der Erde herausgebildet haben. Eine Reduzierung solcher Prozesse auf den Aspekt, dass mit KI und Gehirn vergleichbare Funktionen realisiert werden können, würde der Komplexität menschlichen Denkens – einschließlich seiner emotionalen Komponenten – nicht gerecht.

Ergänzend ist noch auf ein wissenschaftsmethodisches Problem hinzuweisen: Der eigentliche Anspruch der neurobiologischen Forschung ist es, existierende Phänomene zu erklären. Demgemäß müsste es darum gehen, das Phänomen der introspektiven Erfahrung von Bewusstsein und Freiheit einer Erklärung zuzuführen. Der Hinweis auf korrelative Bezüge zwischen Denkprozessen und neurophysiologischen Prozessen stellt jedoch *keine* Erklärung für das Phänomen der introspektiven Erfahrung dar, sondern führt mit der Schlussfolgerung, Bewusstsein und Freiheit seien eine bloße Illusion, nur zu einer Leugnung des Phänomens.<sup>20</sup>

Über diese Überlegungen hinaus stellen sich mit den KI-Entwicklungen vor allem vier weitere Fragen:

- (1) Wie ist die menschliche Entscheidungsfreiheit angesichts der Tatsache zu sehen, dass der Mensch in seinen Entscheidungen immer stärker von Ergebnissen KI-gesteuerter Prozesse abhängig wird, wobei die Kriterien, die dabei zur Geltung kommen, zum Teil kaum noch nachvollziehbar sind?
- (2) Liegt es angesichts der zunehmenden Vernetzungen nicht nahe, den Menschen bei seinem Denken und Handeln weniger als eigenständig handelndes Subjekt, sondern vor allem als Teil einer Online-Community mit vielfältigen Wechselbeziehungen zu anderen sowie zu KI-gesteuerten Agenten aufzufassen?
- (3) Was bedeutet es für die Verantwortung des Menschen, wenn bestimmte Entscheidungen maßgeblich auf Ergebnissen von Datenverarbeitungsprozessen beruhen und andere Entscheidungen an sogenannte autonome Systeme abgegeben werden (müssen)?
- (4) Wäre es für die Zukunft angebracht, dem Leitgedanken eines – in seinen physischen und psychischen Möglichkeiten optimierten – transhumanen Menschen oder gar eines posthumanen Gebildes zu folgen, oder ist es letztlich doch angemessener, an einem Bild vom Menschen festzuhalten, bei dem sich dieser zwar technologischer Möglichkeiten bedienen kann, dabei aber nicht mit ihnen verschmilzt oder gar in ihnen aufgeht?<sup>21</sup>

Wegen der Komplexität dieser Fragen und der umfangreichen Diskussion dazu, kann es hier nicht darum gehen, alle Facetten der Diskussion zu referieren oder gar zu abschließenden Antworten zu kommen. Im Folgenden sollen vor allem mögliche Denkrichtungen skizziert sowie erste Konsequenzen angesprochen und begründet werden. Dabei erfolgt in diesem Kapitel eine Konzentration auf die Fragen (1) und (2), während die Fragen (3) und (4) im nächsten Kapitel diskutiert werden.

## 11.6 Zur Entscheidungsfreiheit bei KI-basierten Informationen, Diagnosen oder Empfehlungen

Wie die obigen Anwendungsbeispiele zur KI zeigen, sind heute bei vielen menschlichen Entscheidungen Informatiksysteme einbezogen. Dabei kann es sich z.B. um die Nutzung digital bereitgestellter Informationen, um die Abfrage von spezifischen Diagnosen oder um das Einholen expliziter Empfehlungen handeln. Die Einbeziehung von computerbasierten Informationen, Diagnosen oder Empfehlungen ist sowohl für private als auch für berufliche und andere Entscheidungen von Bedeutung: So kann im privaten Alltag unter Umständen die Entscheidung, ob jemand eine Kurzreise antritt und wohin sie geht, davon abhängig gemacht werden, was die computergenerierten Wettervorhersagen für bevorzugte Regionen erwarten lassen; in beruflichen Zusammenhängen entscheidet sich ein Investor möglicherweise für ein Bauprojekt aufgrund einer simulierten Architekturanimation; in der Politik sind Entscheidungen – etwa zu Familien-, Renten-, Umwelt-, Gesundheits-, Verteidigungs- oder Wirtschaftsfragen – gegebenenfalls von umfangreichen Datenanalysen und Simulationen abhängig. Allerdings bleibt der Mensch in solchen Fällen letztlich der Herr der Entscheidungen, weil sie im Prinzip nach dem Muster verlaufen »Mensch fragt, Maschine antwortet«. Dies gilt auch dann, wenn es im Einzelfall schwerfallen mag, eine computerbasierte Diagnose abzulehnen oder von einer Empfehlung abzuweichen, weil man nicht sicher sein kann, ob das jeweilige KI-System etwas entdeckt hat, was man selbst übersieht. Grundsätzlich besteht jedoch die Möglichkeit, etwaigen Zweifeln nachzugehen und zu anderen als zu den vom Informatiksystem nahegelegten Informationen, Diagnosen oder Empfehlungen zu kommen. Unter Umständen werden entsprechende Zweifel auch dadurch provoziert, dass computerbasierte Auskünfte Unklarheiten enthalten oder sich die Informationen, Diagnosen oder Empfehlungen aus unterschiedlichen Systemen widersprechen oder »quer« zu intuitiven Einschätzungen liegen. In jedem Fall kann der Mensch als Nutzer von KI-Technologien möglicherweise entstehende Abhängigkeiten reflektieren und darauf reagieren bzw. entsprechend agieren. Insofern

wird seine prinzipielle Entscheidungsfreiheit durch die Nutzung von KI-Technologien in den oben genannten Fällen nicht aufgehoben.

Allerdings – auch wenn eine kritische Auseinandersetzung mit computergenerierten Ergebnissen notwendig ist, sollte in solchen Zusammenhängen *nicht* die Furcht vor möglichen Abhängigkeiten oder Einschränkungen von Freiheit im Zentrum stehen, sondern die Frage, wie das Zusammenwirken von Mensch und Maschine so gestaltet werden kann, dass es letztlich einem sachgerechten, selbstbestimmten, kreativen und sozial verantwortlichen Handeln dient. Aus dieser Sicht entstehen vor allem zwei bedeutsame Anforderungen: Erstens sollten Informatiksysteme (einschließlich ihrer Algorithmen) so konstruiert werden, dass ihre Outputs für den Nutzer hinsichtlich der Kriterien, die bei Informationen, Diagnosen oder Empfehlungen wirksam waren, nachvollziehbar bleiben (wenn dies, wie oben bezüglich des maschinellen Lernens angedeutet, auch nicht immer erreichbar sein wird, wobei zugleich noch einmal die Bedeutung der oben genannten Forschungsaufgaben zum maschinellen Lernen zu betonen ist). Zweitens stellen sich an den Nutzer erweiterte Kompetenzanforderungen:<sup>22</sup> Bloße Anwendungen von Programmen im Sinne der Nutzung möglicher Befehle und einer funktionalen Handhabung der Mensch-Maschine-Schnittstelle reichen nicht mehr aus; es müssen – mindestens bezogen auf die benutzten Systeme – Kenntnisse über interne Abläufe hinzukommen. Wenn es auch nicht einfach sein wird, beiden Anforderungen gerecht zu werden, bleiben sie doch unabdingbar, wenn ein optimales Zusammenwirken zwischen Mensch und Maschine erreicht werden soll.<sup>23</sup>

Werden beide Anforderungen erfüllt, stellen KI-Technologien auch kein Problem für die menschliche Freiheit dar, sie können den Menschen vielmehr bei seinen Entscheidungen im Sinne eines humanen Handelns maßgeblich zum Wohle derer unterstützen, die von den Entscheidungen und einem entsprechenden Handeln betroffen sind.

## 11.7 Zum Menschen als Subjekt seines Handelns in Online-Communitys

Bei den obigen Überlegungen war vor allem der einzelne Mensch – gegebenenfalls in Partnerschaft mit einem oder mehreren menschlichen Mitentscheidern – im Fokus. Der Mensch kann aber mit seinem Handeln auch von vornherein als Mitglied einer Community betrachtet werden, wobei diese zunächst allgemein als Gruppe von Menschen mit Zusammengehörigkeitsgefühl oder gleichen Interessen zu verstehen ist. Bei dieser Betrachtungsweise geht es nicht mehr vorrangig um individuelles Handeln, sondern um das Handeln von Gruppen. Entsprechende Sichtweisen gibt es schon lange. Sie spielen unter anderem in der Soziologie und Sozialpsychologie eine große Rolle. Mit der zunehmenden Vernetzung und den KI-Entwicklungen wächst allerdings die Wahrscheinlichkeit, dass der Einzelne nicht mehr als alleiniger Akteur agiert oder reagiert, sondern in einem Zusammenhang mit mehreren menschlichen Akteuren und/oder Informatiksystemen steht. Wenn sich dabei Vernetzungen ergeben, die insgesamt durch ein Zusammengehörigkeitsgefühl oder gleiche Interessen gekennzeichnet sind, kann man ebenfalls von Communitys bzw. Online-Communitys sprechen. Unterstellt man einen weiten Begriff von Online-Communitys, so unterscheiden sich diese von herkömmlichen Gemeinschaften dadurch, (a) dass sich in ihnen Zusammengehörigkeitsgefühle und gleiche Interessen entwickeln können, ohne dass die Möglichkeit gegeben sein muss, sich persönlich kennenzulernen, wobei dies auch damit zusammenhängt, dass sie nicht an räumliche Nähe gebunden sind und unter Umständen sogar global agieren, (b) dass die Anzahl derer, die sich ihnen anschließen können, praktisch unbegrenzt ist, (c) dass man sich selbst gegebenenfalls beliebig vielen Online-Communitys zugehörig fühlen kann, (d) dass mit dem Anschluss an die jeweilige Online-Community in der Regel keine spezifischen Verpflichtungen verbunden sind, vieles unverbindlich bleibt und man üblicherweise davon ausgeht, dass die Zustimmung zu oder die Ablehnung von einzelnen Beiträgen sowie das Einbringen eigener Beiträge – auch wenn sie gegebenenfalls brisant sind – risikolos ist (wenn dabei unter Umständen auch die

Bedeutung eigener »Datenspuren« unterschätzt wird), (e) dass sie auch virtuelle Agenten oder Informatiksysteme umfassen können – unabhängig davon, ob sie für andere Mitglieder erkennbar oder aufgrund der Simulation von Zugehörigkeit und Interessen nicht identifizierbar sind.

Für den einzelnen Menschen ist außer dem vermittelten Zugehörigkeitsgefühl und anderen Merkmalen von Online-Communitys gegebenenfalls bedeutsam, dass er der Meinung sein mag, dass er durch die Teilhabe an einer Online-Community etwas bewirken kann, was er allein nicht könnte.<sup>24</sup> So wird er sich möglicherweise an einem Aufruf gegen Rassismus oder für Faktenchecks bei politisch relevanten Aussagen beteiligen – auch wenn man solche Aufrufe oder Beteiligungen nicht mit allzu großen Hoffnungen hinsichtlich ihrer Wirksamkeit verbinden sollte. Letzteres gilt mindestens solange, wie sich – in einem postdemokratischen Sinne – zwar alle äußern können, Entscheidungen jedoch nur von einigen wenigen getroffen werden.<sup>25</sup>

Es wäre nun zwar interessant, die Vorzüge und Probleme von Online-Communitys – auch vor dem Hintergrund von soziologischen oder sozialpsychologischen Untersuchungen – zu diskutieren. Dies würde jedoch den Rahmen dieses Bandes sprengen (wenn dazu im vierten Kapitel auch einzelne Gedanken angesprochen werden konnten). Im Folgenden soll es ergänzend nur darum gehen, die Frage aufzunehmen, inwieweit die Sichtweise vom Menschen als Teil von Online-Communitys zu einer Korrektur des zuvor vertretenen Subjektbegriffs führt – zumal es einzelne Ansätze gibt, in denen dem Menschen in digitalen Netzen der Subjektstatus abgesprochen wird. Dies geschieht unter anderem in bildungstheoretischen Studien zu Praktiken und Bildungsvorgängen in digitalen Netzen.<sup>26</sup> Dabei erscheint nicht – wie bisher – das Individuum als Subjekt von Bildungsprozessen. An seine Stelle treten vielmehr Formationen vernetzter Akteure. Mündigkeit wird nicht mehr als Eigenschaft oder Fähigkeit von individuellen Subjekten verstanden, sondern der Community zugeschrieben.

Gegen Ansätze solcher Art lässt sich allerdings einiges einwenden: *Erstens* setzt auch eine Formation vernetzter Akteure letztlich Akteure voraus, die – nimmt man den Begriff des Akteurs ernst – aufgrund

von bedachten Argumenten handeln. Insofern muss auch den Akteuren in Online-Communities die Möglichkeit zugestanden werden, dass ihr Tun auf der Basis eigener Abwägungen erfolgt oder zumindest erfolgen kann. *Zweitens*: Wenn man den Menschen vor allem unter dem Aspekt sieht, dass er als Teil einer Online-Community agiert oder reagiert, wird die Wahrnehmung insbesondere auf die Beeinflussung bzw. auf Wechselwirkungen des Handelns mit dem Handeln anderer gelenkt. Das Individuum erscheint unter Umständen bereits durch die eingenommene Blickrichtung nicht mehr als frei, sondern als abhängig. So ließe sich z.B. darauf verweisen, dass jemand gegebenenfalls eine Stellungnahme für den Umweltschutz allein aus Gründen der Zugehörigkeit unterzeichnet, obwohl er sein eigenes Umweltverhalten noch nie ernsthaft reflektiert hat. Mit einer Änderung der Perspektive könnte man jedoch genauso gut die subjektbasierten Leistungen betonen, die Einzelne im Rahmen einer Online-Community erbringen. Bezogen auf das Beispiel der »AVAAZ-Gemeinde« ließe sich so z.B. feststellen, dass auch deren Aktivitäten davon abhängig sind, dass einzelne Akteure aktiv werden und als freie Subjekte Missstände anprangern oder Petitionen entwerfen.<sup>27</sup> *Drittens* ist es zwar möglich, dass sich einzelne Menschen vor allem über ihre Online-Communities definieren und auch einen großen Teil ihrer Zeit in Netzwerken verbringen, im Regelfall bleiben aber immer noch viele online-unabhängige Aktivitäten, sodass es nicht gerechtfertigt erscheint, das Bild vom Menschen allein aus der Community-Perspektive bestimmen wollen. Dies gilt auch angesichts des Eindrucks, dass der Zustand des »*permanently online, permanently connected*« bei vielen Menschen zunimmt.<sup>28</sup>

Insgesamt ergibt sich, dass durch die zunehmende Eingebundenheit in Online-Communities der in diesem Band postulierte Subjektbegriff in seinen Grundpositionen nicht hinfällig wird. Im Gegenteil – angesichts möglicher Manipulationen oder Abhängigkeiten in Netzzusammenhängen erscheint es umso wichtiger, die Leitidee eines reflexiv eingestellten und gesellschaftlich handlungsfähigen Subjekts mit dem Vermögen eines sachgerechten, selbstbestimmten, kreativen und sozial verantwortlichen Handelns zu betonen. Zugleich sollten allerdings

mögliche Gefährdungen und potenzielle Chancen durch das Agieren in Online-Communitys im Blick bleiben.

## 11.8 Zusammenfassende Bemerkung

In der Zukunft werden für das Handeln KI-basierte Informatiksysteme eine zunehmende Rolle spielen – sei es in responsiven Formen (Mensch fragt oder gibt Anweisungen und Maschine antwortet oder führt aus), sei es in adaptiven Formen (Maschine versucht, auf der Grundlage von Datenanalysen mediale Botschaften unter Berücksichtigung von kognitiven, emotionalen oder wertbezogenen Voraussetzungen der Nutzenden zu erzeugen und diese zielgerecht zu vermitteln), sei es in aktiven Formen (Maschine führt Aktionen mit grundsätzlicher oder spezifischer Auftragserteilung durch den Menschen in technisch autonomer Form aus).

Dabei können KI-basierte Systeme in unterschiedlichen Handlungsbereichen zur Geltung kommen z.B. in den Bereichen Information und Lernen, Analyse und Simulation, Unterhaltung und Spiel, Dienstleistung und Produktion sowie Steuerung und Kontrolle. Solche Handlungsbereiche sind für verschiedene gesellschaftliche Handlungsfelder relevant, z.B. für Beruf und Freizeit, für Wirtschaft und Verwaltung, für Bildung und Erziehung, für Gesundheit und Umwelt, für Kunst und Kultur, für öffentliche Meinungsbildung und Politik. Dabei vergrößern sich – wie die obigen Beispiele zur KI andeuten – die Möglichkeiten und Risiken in allen Handlungsfeldern.

Vor diesem Hintergrund ging es in diesem Kapitel darum zu diskutieren, wie sich angesichts der KI-Entwicklungen die *Entscheidungsfreiheit* und die *Subjektidee* darstellen. Dazu lässt sich resümierend feststellen:

Hinsichtlich der Frage der *Entscheidungsfreiheit* ist zum einen wichtig, dass die Nutzung von KI-Technologien einer kritischen Auseinandersetzung mit diesen bedarf. Damit die Auseinandersetzung gelingen kann, sind bei den Nutzenden Kompetenzen erforderlich, die eine reflektierte Verwendung der KI-basierten Technologien ermög-

lichen. Zugleich müssen Entwickler bereit und in der Lage sein sowie gegebenenfalls verpflichtet werden, entsprechende Anwendungen so zu gestalten, dass KI-basierte Informationen, Diagnosen und Empfehlungen bezüglich der zugrunde liegenden Kriterien oder mindestens bezüglich entscheidungsrelevanter Aspekte nachvollziehbar sind.<sup>29</sup> Wenn beide Bedingungen erfüllt werden, bleibt einerseits die Entscheidungsfreiheit des Menschen erhalten und andererseits kann die Frage in den Mittelpunkt rücken, wie die jeweiligen Mensch-Maschine-Schnittstellen gestaltet werden sollten, damit Mensch und KI-Technologien in optimaler Weise bei einem humanen Handeln zum Wohle aller zusammenwirken können.

Im Hinblick auf die *Subjektidee* ist der Leitgedanke eines reflexiv eingestellten und gesellschaftlich handlungsfähigen Subjekts – auch bei zunehmender Eingebundenheit menschlicher Akteure in Online-Communitys – keineswegs überholt. Im Gegenteil: Angesichts vielfältiger Beeinflussungsversuche oder Abhängigkeiten in Netzzusammenhängen erscheint es umso wichtiger, diesen Leitgedanken zusammen mit der Forderung nach einem sachgerechten, selbstbestimmten, kreativen und sozial verantwortlichen Handeln zu betonen. Dabei sollte ein entsprechendes Handeln in zeitgemäßer Form ausgelegt und hinsichtlich sowohl möglicher Gefährdungen als auch potenzieller Chancen bedacht werden.

Im folgenden Kapitel werden die Überlegungen zu KI-Entwicklungen mit den Fragen fortgesetzt, wie ihre Bedeutung für die Verantwortung beim Handeln und für das Menschenbild der Zukunft – angesichts trans- oder posthumanistischer Vorstellungen – einzuschätzen ist.